# Towards Dialogue Strategies for Cognitive Workload Management

Jessica Villing

GSLT

The Graduate School of Language Technology

UNIVERSITY OF GOTHENBURG

Towards Dialogue Strategies
for Cognitive Workload Management

# Towards Dialogue Strategies for Cognitive Workload Management

Jessica Villing

GSLT
The Graduate School of Language Technology

UNIVERSITY OF GOTHENBURG
2015

For Björn, Max and Tess

Doctoral dissertation in linguistics, University of Gothenburg
October 15, 2015

© 2015 Jessica Villing
Printed by Ale Tryckteam, Bohus, 2015
Dissertation edition, October 2015

Distribution:
University of Gothenburg
Department of Philosophy, Linguistics and Theory of Science
P.O box 200, SE-405 30 GOTHENBURG, SWEDEN

**Abstract**

**Title:** Towards Dialogue Strategies for Cognitive Workload Management
**Author:** Jessica Villing
**Language:** English
**Department:** Graduate School of Language Technology and Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, Box 200, SE-405-30 Göteborg
**Year:** 2015

Although it has been shown that drivers are less distracted when using speech interfaces compared to traditional interfaces, using voice control instead of manual controls does not completely solve the problem with distracted drivers. The interaction with the dialogue system may itself add to the driver's cognitive workload and may therefore be a safety issue. The main purpose of this thesis is to learn more about in-vehicle dialogue during various types of cognitive workload, to use this knowledge to enable safe and non-distracting dialogue system interaction in vehicles. We do this by analysing a corpus of human-human in-vehicle dialogue to learn more about the dialogue strategies used by drivers and passengers during various types of workload. We discuss the types of cognitive workload that we believe are most important to consider when studying the multitasking activity of driving and interacting with a dialogue system, and suggest a method for distinguishing different types of workload by using information about the driver's workload and driving behaviour. We found that dialogue strategies such as interruptions – in the form of silent pauses and domain switches – are used in response to the driver's cognitive workload, as well as resumption of unfinished discussions. These behaviours are analysed in order to find strategies for preventing, or shortening the duration time of, high cognitive workload. We also indicate how these strategies can be implemented in in-vehicle dialogue systems.

**Keywords:** cognitive workload, workload types, dialogue system, vehicles, dialogue strategies, interruption in dialogue, resumption in dialogue

# Acknowledgements

There are so many people who have helped me getting this thesis done. First of all, of course, my supervisor Staffan Larsson. Your brilliance, patience and ability to always manage to make me feel that I can actually do this have been invaluable. My second supervisor Robin Cooper, who helped my in the beginning of my PhD journey and got me on the right track, and then helped me finish it off, thank you! Chris Howes, without you helping me with your amazing knowledge about statistics I don't know what I would have done!

Thank you GSLT for providing an excellent setting when I started my studies, and for funding. Thank you also to my fellow PhD students at GSLT, for interesting discussions and support. Thank you to everybody at the Philosophy, Linguistics and Theory of Science department at Gothenburg University. I have missed you all so much after I moved, I missed the stimulating discussions at seminars as well as "fika" breaks! A special thank you to Simon Dobnik who served as opponent at my final seminar, and did a great job.

A big thank you to all the members of the DICO project - Cecilia Holtelius, Johan Jarlengrip and Nina Åberg at Volvo, Anders Lindström at TeliaSonera and Alexander Seward at Veridict. I had a lot of fun working with you, and without you and the work we did in the project I would not have been able to write this thesis. To the people who participated in the DICO study I will be forever grateful, without you there would have been nothing to analyse in this thesis. And thank you Anders Lindström for reading the thesis at a very early stage and giving me valuable advice.

Thank you Robert Adesam for technical support over the years, and Pia Gårdmo, Helena Bjärnlind and Paula Wäne for all kinds of practical and administrative support.

I have been fortunate enough to be able to attend a number of conferences and workshops, and learned a lot from the feedback I got and from discussing with fellow researchers. Thank you Alexander Berman, Ellen Breitholtz, Fredrik Kronlid, Staffan Larsson, Peter Ljunglöf, Kristina Lundholm Fors, and the members of the DICO project for writing with me!

Thank you also to Vinnova and Volvo AB for funding the DICO project, and to Kungliga och Hvitfeldska stiftelsen for the scholarship I received.

Thank you to my friends, both in Sweden and all the new friends I have made in USA. You have been cheering me on all the time. Finally, a big thank you to my family, and especially my sister Nettan who has always been convinced that I will be able to make this even when I doubted myself. Thank you to my dad, my mom and my brother. I am so sad that you are no longer with us, you believed in me all the time but you never got the chance to see me finish. Last, but definitely not least – my children Max and Tess, thank you for your patience with a distracted mom, especially the last few months! And thanks for supporting me in my doctoral studies even though, as you like to point out Max, I will be a doctor but "not the useful kind"! And Björn, my husband. What would I have done without you? Your endless love and support will forever make me feel special!

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

The development of in-vehicle technologies and services, and the technology for controlling them, is moving quickly forward. In 1922 the first radio was installed in a car, something that was met with scepticism (DeMain, 2012). Tuning in a radio station while driving was considered a safety risk, since the driver had to take their attention away from the road and the task of manoeuvring the car. There was even a fear that listening to music could lull a driver to sleep, and in 1930 the state of Massachusetts, USA, proposed a law to ban radios in cars for these reasons. Nowadays, a radio is since long considered a standard equipment in a car, and the number of in-vehicle devices is constantly growing. To mention a few, we now have GPS systems in our vehicles to help us find the way, we want our smart phones connected to the car so that we can make phone calls, and we may even want to check text messages and catch up on social media (although many countries around the world have banned such activities while driving, for example, Australia, South Africa, China, France, Denmark and some states in USA[1]). In addition to the radio, we also want to stream the music of our own choice to the car, and in order to make a long ride smoother and less boring we may want our children or other passengers to be able to watch their

[1]Source: http://www.cellular-news.com/car_bans/

favourite movies on a screen in the backseat.

Controlling all these devices in a safe and non-distracting way is still an issue. The concern that arose in 1922, that managing in-vehicle devices would steal the driver's attention away from the driving task, is an ongoing concern. The growing number of vehicles on the road and the faster speed of vehicles are two factors that put heavy demands on the driver, and distraction from managing in-vehicle devices further adds to the cognitive workload of the driver. First and foremost, the driver needs to be able to concentrate on the driving task and react fast to sudden changes in the traffic situation, which makes the issue of potential distractions, and the amount of cognitive workload a driver is able to manage, a serious question.

It is therefore necessary to invent new techniques for controlling in-vehicle devices. Since the drivers already use their hands and eyes for driving the car, researchers began exploring voice control as an alternative to manual control, and in 1996, Mercedes Benz was the first vehicle brand to introduce voice control in cars. *Linguatronic* was a command-based system for controlling a cellphone and a navigation system using voice for input and voice and screen for output (Bühler et al., 2003).

Since then, the development of in-vehicle dialogue systems has continued to move forward. Nowadays, it is possible to speak more freely when using a dialogue system, and control several devices. For example, it is now possible to control in-vehicle devices such as the navigation system, radio or fan. It is also possible to connect your own smart phone to get access to the phone's applications as well as information from the Internet while driving. See, for example, CarPlay, developed by Apple[2].

## 1.1 Rationale

Nowadays, having the ability to use your voice to control at least some of the devices, technologies and services in a car is some-

[2] http://www.apple.com/ios/carplay/

thing we have started to take for granted (at least when buying a new car). However, even though voice control may solve the problem with drivers having to take their hands off the steering wheel and their eyes off the road, it does not solve the whole problem of distracted drivers. The reason is that the interaction with an in-vehicle dialogue system may be distracting in itself, since the driver must still allocate parts of their mental resources to the interaction task. In this thesis, we will therefore investigate cognitive workload in relation to the multitasking activity of driving and interacting with a dialogue system, and look at dialogue strategies for preventing - or shortening the duration time of - high cognitive workload. We will do this by analysing a corpus of in-vehicle driver-passenger dialogues. The corpus was developed in the DICO project and is described in more detail below.

## 1.2 Aim of this study

The aim of the thesis is to learn more about in-vehicle dialogue during various types of cognitive workload, and the long term goal is to improve in-vehicle dialogue system interaction. To accomplish this, we have analysed a corpus of human-human in-vehicle dialogue, where the driver's workload and driving manoeuvres have been measured. This data has been used when analysing the participants' dialogue behaviour. The corpus is rather small and we therefore take an explorative approach to the data. We will look for interesting behaviours in in-vehicle dialogue during various types of workoad that we believe are relevant. We will describe both statistically significant results and observable tendencies whcih have not (yet) been shown to be significant. We will allow ourselves some degree of speculation based on in depth-analyses of examples of dialogue behaviour. The purpose of this thesis is therefore not to formulate a complete theory about in-vehicle dialogue strategies. Instead, this thesis should be seen as a first step towards such a theory, and towards developing cognitive workload-aware dialogue systems. We hope that the results put forward here will be useful not only as evidence of relevant correlations between cognitive workload and dialogue behaviours, but also as a starting point for future work. For example, non-significant results can be tested for again in a larger dataset or in an experimental setting. Finally, to make clear how our results are relevant to cognitive workload-aware in-vehicle dialogue systems, we will sketch how selected behaviours could be implemented in such a system.

## 1.3 Outline of this thesis

The outline of this thesis is as follows:

- **Chapter 2: Cognitive Workload.** In order to understand why the driver's cognitive workload level is an important factor when it comes to road safety, we give an overview of research on cognitive workload and show how high workload may affect the driving performance. We discuss the types of cognitive workload that we believe are most important to consider when studying simultaneous driving and interaction with a dialogue system, and present a novel theory of how to distinguish between workload types by analysing the driver's cognitive workload and driving behaviour.

- **Chapter 3: Interruption in dialogue.** We found that passengers frequently switch domain in response to high driver workload. For example, the passengers often switch to the Navigation domain to give a navigation instruction or to clarify an earlier instruction. We also found that the participants often pause the dialogue when the driving task is demanding. We present an analysis of both behaviours in relation to the driver's workload type, and suggest how to implement interruption strategies into an in-vehicle dialogue system. We give examples of how an interaction with an in-vehicle dialogue system would look using these strategies. We conclude by discussing future research.

- **Chapter 4: Resumption in dialogue.** When studying interruption behaviour, we also noted that the participants strive to resume an interrupted topic if the task had not yet been solved. We found that resumption behaviour differed depending on the driver's workload level and whether the topic is within a domain of discussion that is related to the driving task or not. We suggest how to implement resumption strategies into an in-vehicle dialogue system, and conclude by discussing future research issues.

- **Chapter 5: Conclusion.** We conclude by summarizing the previous chapters and the findings we have gained.

## 1.4 The DICO corpus

DICO (Villing and Larsson, 2006) was a project that took place 2006-2008, funded by the Swedish governmental agency for innovations, Vinnova[3]. The members of the project were the University of Gothenburg[4], Volvo Technology (now Volvo Group Advanced Technology and Research[5]), Volvo Cars[6], Volvo Trucks[7], TeliaSonera[8] and Veridict[9]. The aim of the project was to build a proof-of-concept demo of an in-vehicle dialogue system – with focus on cognitive load management, multimodal menu-based dialogue and speech recognition in noisy environments – showing how a spoken dialogue system can be an aid for drivers. The dialogue system used in the DICO project is built on top of the GoDiS dialogue system (Larsson, 2002), which in turn is implemented using TrindiKit (Traum and Larsson, 2003). This system has since been commercialised by Talkamatic AB[10].

In order to adapt the GoDiS dialogue system to an in-vehicle environment, a data collection was carried out within the DICO project to find out how people talk while driving. The challenge was to elicit a natural dialogue (as opposed to giving the driver

[3]www.vinnova.se
[4]www.gu.se
[5]www.volvogroup.com
[6]wwww.volvocars.com
[7]www.volvotrucks
[8]www.teliasonera.com
[9]www.veridict.com
[10]www.talkamatic.se

a constructed task such as for example a math task) and make the participants engage in the conversation since we wanted to study how additional distraction or increase in the cognitive load would affect the dialogue behaviour.

### 1.4.1 Participants and test environment

Eight particpants (six male and two female) between the ages of 25 and 36 who had no experience of using speech technology or dialogue systems were recruited. The participants drove a car in (driver and passenger) pairs while performing two tasks, one navigation task and one interview task (see below).

The test car was a Volvo XC 90 (model year 2004). It was equipped with a dual headset microphone which enabled recording of driver and passenger on separate channels. Two digital video cameras were used, one camera directed towards the driver's face to record head- and eye-movement of the driver, and the other directed towards the road ahead to record the traffic situation and the surroundings.

### 1.4.2 Tasks

The participants were given two tasks to solve together, an interview task and a navigation task. The tasks were chosen to be similar to so called in-vehicle infotainment (IVI) applications. IVI applications are applications that deliver either information or entertainment to the driver, for example, navigation, information about road and weather conditions, radio, cell phone etc. (Pretschner et al., 2007). The Navigation domain can be compared to a navigation application in the way that the passenger provides information on where to go and when to make a manoeuvre in a similar way as a navigation system would do. The Interview domain may be compared to an entertainment system, even though the fit is not perfect. To perform each task within this domain (i.e., to receive enough information regarding each interview question to be able to answer questions about it afterwards), the passenger asks questions that the driver needs to answer.

Similarly, a dialogue system typically asks questions until the driver has provided enough information so that the system can perform the task of, for example, calling someone, changing radio channel or giving information about sport scores. Historically, most in-vehicle dialogue systems have been primarily command- and control-based, while the Interview domain is more similar to the question and answer-interaction that follows a user command in a form-based dialogue system. However, we will probably see a more natural and human-like dialogue in future dialogue systems, where both the user and the system are able to ask questions and negotiate to be able to perform a task in the best way. One example of development in this direction is Siri, which is used in Apple's CarPlay system[11].

The interview task involved 52 interview questions about personal interests. The questions varied in difficulty. Some were easy to answer (for example, *"what is your name"*, *"where do you live"*) and some needed to be reflected upon (for example,*"name a country you visited that you never want to visit again"*, *"what is your dream job"*). They were also told to motivate their answer, if possible. In the navigation task the passenger guided the driver step by step along a given route. The route was set to be rather challenging, in relatively dense daytime traffic in central Gothenburg, Sweden. The interview questions as well as driving instructions were given only to the passenger, who was told to only give verbal instructions (as opposed to, for example, point in the right direction). The passenger was also told to give the instructions step by step, no more than one intersection ahead. Hence, the driver did not know what questions to discuss and none of them knew the whole route in advance. Therefore, the driver had to signal, implicitly or explicitly, the need for a driving instruction or the appropriate time for a new question to discuss. The passenger, too, had to have a strategy for when to read and give the

[11] www.apple.com. Retrieved on January 19, 2015

next driving instruction and when to discuss interview questions. The reason for this setup was to

1. elicit a natural and fairly intense dialogue
2. force the participants to interrupt the current topic when necessary to change topic and/or domain (for example to interrupt the discussion of an interview question to get driving instructions)

To further encourage them to perform their best, the participants were told that they would be tested after the drive to see how much they remembered from the interview task. The driver was told to focus on the driving for safety reasons, but was told also to perform the best they could in the interview task. Each passenger / driver pair was free to solve the interview and navigation tasks in any way that was suitable for them, except from taking notes or use any other memory aids. The participants were told to switch roles after 30 minutes, hence each test session lasted for 60 minutes and during that time each participant acted as both driver and passenger.

### 1.4.3 Workload measurements

Apart from the interview and navigation tasks, the driver was also asked to perform a Tactile Detection Task (TDT). A TDT is used in test environments in order to measure the driver's cognitive workload level. When using a TDT, a buzzer is attached to the driver's forearm and a response button is attached to the index finger. At random intervals (between 3 and 5 seconds) the TDT is activated and the driver must then press the response button as quickly as possible. Workload is determined based on user hit-rate and reaction latency (van Winsum et al., 1999). The TDT also enables the measurement of cognitive load that is not related to the driving task, but instead caused e.g. by the dialogue itself or by memory processing, even when the car is not moving, e.g.

at stoplights. Studies have shown that the TDT is able to detect the mental demand of a control task, and there is no significant increase in reaction time when performing the control task (see, for example Krause et al. (2015)).

Driver activity was measured using an IDIS system (Broström et al., 2006). Having been developed by a company, the exact functioning of IDIS is proprietary, but IDIS determines activity based on the driver's behaviour. For example, steering wheel movements, sudden changes in speed or the driver manoeuvering certain devices in the car such as the radio or the fan, are monitored in order to determine if the driver is performing a manoeuvre that is assumed to be cognitively demanding. The output from IDIS was shown as a red light (high activity) or a green light (low activity), which was captured by a camera pointing towards the road. Neither the driver nor the passenger were able to see the IDIS output signal.

The measurement tools, TDT and IDIS, are explained in more detail in Chapter 2.6.

### Transcription and coding

Video recordings were ortographically transcribed by the members of the DICO project, using the ELAN[12] transcription tool. ELAN is able to handle both audio and video resources, and it allows annotation along multiple tiers (i.e., an utterance can be annotated with several independent annotation schemas). The DICO annotation schemas were designed to enable analysis of utterances and cognitive workload.

As mentioned above, eight 30-minute interactions were recorded. Due to technical problems, one driver/passenger pair had to be removed and therefore the corpus consists of almost 3 hours (164,9 minutes) of dialogue. All in all, 3590 driver utterances and 4382 passenger utterances were transcribed and coded.

[12] http://www.lat-mpi.eu/tools/elan/

Cognitive load according to TDT is annotated as:

- **workload**: an annotation on this tier means high workload, no annotation means that the workload is considered low
- **reliability**: indicates whether the measured workload level is reliable or not (reliability was low if response button was pressed more than 2 times after the event)

High workload utterances could then be found by searching for annotations where workload and reliability are overlapping, and low workload utterances could similarly be found by searching for annotations where reliability is annotated, but workload is not.

Driver activity detected by IDIS is annotated as:

- **high**: an annotiation on this tier means that the IDIS sensor is showing red, indicating increased driver activity.
- **low**: an annotiation on this tier means that the IDIS sensor is showing green, indicating low driver activity.

In addition to the annotations of workload and demanding driving manoeuvres, the DICO corpus has been annotated with the domain of discussion. When transcribing the dialogues, we found that the participants, apart from discussing the tasks that were given to them (i.e. interviewing each other and navigating through a pre-defind route), also discussed topics related to the traffic situation and the driving task as well as topics not related to either the given tasks nor the driving task. When annotating the corpus we therefore distinguished between four domains:

- *Interview* domain: discussions regarding interview questions

- *Navigation* domain: discussions regarding the navigation task
- *Traffic* domain: discussions about the driving task, traffic situation and fellow road-users
- *Other* domain: anything else

For descriptions of additional linguistic annotations, see Chapters 3 and 4.

## 1.5 Summary

In this chapter we presented the aim of this study and an initial motivation for exploring cognitive workload in relation to in-vehicle dialogue system interaction. We then gave a brief overview of the thesis. Finally, we presented the DICO project and described the corpus, developed within the project, that is used in this thesis for analysing in-vehicle dialogue. We also described the annotations that have been applied to the DICO corpus, and that have been used as a basis for the work presented in the coming chapters.

# Chapter 2

# Cognitive Workload

## 2.1 Introduction

In this chapter, we will give an overview of research regarding cognitive workload and discuss why we think the driver's cognitive workload is an important factor when it comes to driving and interacting with an in-vehicle dialogue system. We will present a taxonomy of workload types and show how workload type can be determined from TDT and IDIS signals.

The advantages of using speech technology for controlling in-vehicle devices are intuitively obvious - the driver is able to keep their hands on the steering wheel and their eyes on the road, and therefore the driver's attention can be kept on the driving task. However, even though tests have shown better driving ability when using speech interfaces compared to manual interfaces, it has also been shown that in-vehicle dialogue systems add to the driver's cognitive workload and may therefore be a safety issue.

## 2.2 Defining cognitive workload

The term *cognitive workload* is frequently used in traffic safety research but is often not defined, or described in general terms as a demand placed upon humans (De Waard, 1996).The concept is hard to define, since the term can have different meanings depending on the situation. Sweller et al. (1998) describe the amount of workload imposed on a person when learning to solve a problem, and talks about *intrinsic, extraneous* and *germane* cognitive load, by which they mean cognitive load related to the task, related to the nature of the instructions given to the person who will perform the task, and related to a person's ability to learn to perform that task, respectively.

However, workload may not only be task-specific but also person-specific, meaning that it is not only the complexity of the task (and each person's individual capability to manage to perform it) that matters but also the person's motivation to perform the task, their choice of strategies to perform it and the person's mood while performing it (De Waard, 1996). The workload described in this thesis is both task-specific and person-specific, since it originates from the act of performing two or more tasks simultaneously.

O'Donnell and Eggemeier (1986) define cognitive workload as *"that portion of the operator's limited capacity that is actually required to perform a particular task"* which is close to how we would define it. However, this definition implies that a certain task requires a certain amount of workload in all situations. In our context, with multiple tasks performed simultaneously, we believe that it depends on the situation how much of the mental resources has to be used to perform the tasks. Therefore, our definition is *that portion of a person's limited capacity that is actually used to perform a number of tasks*.

## 2.3 Cognitive workload awareness

Each of the tasks of driving and interacting with a dialogue system may be mentally demanding on their own. For example, driving in heavy traffic, making a left turn or parking the car in a tiny parking space is more demanding than driving straight ahead on a country road in light traffic. Using a dialogue system you have never used before or answering a question from the system that you are not prepared for is more mentally demanding than interacting in a way that you have done many times before and know by heart. Factors like experience (driving experience as well as experience of using a dialogue system), age and the number of tasks performed simultaneously influence if, and how much, the workload increases.

In this thesis, we want to investigate if, when the driver's workload is high, a temporary change of dialogue strategies can help decrease the driver's high workload and make the dialogue task less distracting. By analysing human-human in-vehicle conversations we have found clues about how to further develop in-vehicle dialogue systems and make them *cognitive workload-aware*. Our definition of a cognitive workload-aware dialogue system is one that is able to adapt its behaviour depending on the user's cognitive workload level. This minimally requires being able to estimate if the driver's workload is high or low and reacting to that. In this thesis, we propose to take cognitive workload-awareness one step further and also include being able to determine if the high cognitive workload is related to the driving task or not, and use different strategies for lowering the workload depending on type of workload. An additional step would be to also be able to determine if a behaviour that correlates with high workload is a behaviour that *causes* workload, or if it is a sign of the user being *affected* by high workload caused by some other task. However, this is something which we will leave for future work.

Later in this chapter, we will therefore define the types of workload we are mainly interested in, and propose a novel method for distingushing between them. First, we will give a brief survey of the history of in-vehicle dialogue systems and the previous research that has been made within this field.

## 2.4 Previous research

In this section we will give an overview over previous research regarding voice control of in-vehicle applications.

### 2.4.1 Early in-vehicle speech systems

Early in-vehicle speech recognition systems were command-and-control based, i.e. the user speaks a command and the system responds to that (Pieraccini et al., 2004). The reason for taking this approach was to limit the number of words the system needs to recognize, since it allows the system to listen for a few possible commands instead of the broader variety of possible user input a more conversational dialogue system offers. A smaller input vocabulary gives a better recognition rate and lowers the risk of system misinterpretation (McTear, 2004). In the beginning, this was what in-vehicle voice control looked like. Later the interaction was improved so that the interaction included not only one single user command but the user could navigate through a menu system, but the lengthy and formal type of interaction unfortunately gave voice controlled interaction a bad reputation. People thought of voice recognition systems as tedious to use since it was time-consuming to navigate step-by-step through the menu structure, and despite the small vocabulary the recognition rate was not good enough. This made the users not want to use it because of the many errors and the fact that it was so much faster to use manual controls. The public did not believe in the technology and the possibility to interact with a computer in a more natural and human-like way was more or less looked upon as something that belonged in science fiction movies (Sangani, 2013).

From a safety perspective, the major disadvantage of command-and-control based systems is the effort it takes to use them. The user is forced to learn a lot of specific commands by heart instead of using words that feel natural for them. They also have to know the menu structure since they can only give one piece of information within each utterance, and are thereby forced to navigate through the menu system step-by-step. This makes interaction with the system unnecessarily complicated and risks increasing the driver's cognitive workload.

### 2.4.2 Conversational dialogue systems

A better way of interaction is offerred by conversational dialogue systems. A conversational system allows the user to speak more freely; the user does not have to remember either a list of commands or the exact menu structure. Instead it is possible to express what you want to do using a wider range of words and include as much information as you want within each utterance, which makes the interaction both faster and less cognitively demanding (Pieraccini et al., 2004). Impovements of voice recognition techniques allowed for larger vocabularies and helped preparing for further research in conversational dialogue systems.

TRAINS was an early project aiming at studying human-human spoken dialogue and from that build a mixed-initiative multimodal dialogue system where the user is able to reason with the system about time, actions and events (Ferguson et al., 1996). In 1999, AT&T, IBM, Lucent and Motorola formed VoiceXML forum[1], aiming at creating a standard for spoken dialogue. The standard features "synthesized speech, digitized audio, recognition of spoken and DTMF key input, recording of spoken input, telephony and mixed initiative conversations" (McGlashan et al., 2004). VoiceXML is a now W3C standard.

### 2.4.3 In-vehicle dialogue systems

Research on conversational dialogue systems laid the ground for further research on free dialogue also in the in-vehicle environment. VICO (Bernsen and Dybkjaer, 2001), Project 54 (Kun et al., 2002), CHAT (Weng et al., 2006), and Commute UX (Tashev et al., 2009) are examples of research projects, both within the academic and the industrial, concerning in-vehicle dialogue systems. Larsson (2002) presented a way of modelling information-seeking dialogue using a set of not yet answered questions, where the user with each utterance provides answers and the system aims to

[1] http://www.voicexml.org

find a relevant question for each given answer. The question may have been already asked or not, and in the latter case the system tries to find a matching question which has not yet been asked. This makes it possible for each individual user to give as much information as desired in each utterance and in any order that feels natural which may ease the interaction, especially if the user is performing multiple tasks (such as driving and interacting with a dialogue system). Therefore, this platform was used in the DICO project (see Chapter 1.4 and Villing and Larsson, 2006).

Apple Inc.[2] and Google[3] are currently integrating their voice controlled smartphone systems into the in-vehicle environment, thereby taking one more step towards integrating the technology into the everyday life of car drivers.

So, does a spoken dialogue system, with a greater freedom in how the user can interact with the system, solve the problem with distracted drivers? Or, in other words, is the increased head-up time (i.e. the time the driver spends looking at the traffic instead of a secondary task) and the possibility to choose how to express oneself that comes with speech technology and a free dialogue the only things to consider when designing IVIS (In-vehicle Information Systems)?

Several studies have been made regarding driving performance when interacting with a device using manual vs. speech control. For example, Maciej and Vollrath (2009) carried out a user study where the participants performed four IVIS tasks. Driving performance was measured using LCT (Lane Change Task) and visual distraction by studying gaze behaviour to see how often the participants were looking at the LCT screen (i.e. the "road") and how often they were looking away from the screen towards the secondary task. The participants were also asked to estimate their perceived level of distraction. The authors found that the speech interface improved driving performance and was less distracting.

Villing and Larsson (2011) evaluated a dialogue system with three

[2]www.apple.com
[3]www.google.com

different modalities; speech control only, manual control only and a multimodal version where the user can choose to use both speech and manual controls in combination. It was found that even if task completion time was shorter when using manual controls, driving performance was better using speech and multimodal interfaces, and the participants themselves preferred the multimodal interface.

These results are promising and show that speech technology is an important key to improving safety during interaction with in-vehicle devices. But, even if the possibility of using voice instead of manual controls reduces the visual distraction, and even if a conversational dialogue system gives greater freedom for the user to express what they want to do and is thereby less mentally demanding, spoken dialogue systems can still be a safety risk if the dialogue task interferes with the driving task. If the driving task temporarily demands a higher degree of concentration from the driver and the dialogue system is not aware of that, the interaction with the system risks adding to the cognitive load. This could happen if the system insists on, for example, getting an answer to a question although the driver is not prepared to answer at the time.

### 2.4.4 In-vehicle interaction between humans

Pashler (1994) showed that when humans perform two tasks at the same time, where both of them involve reacting to stimuli (and we believe that the act of interacting with a dialogue system while manoeuvering a vehicle falls into that category) there is a ubiquitous slowing effect on the performance of both tasks which he called the *psychological refractory period* (PRP). So, even if using voice control makes the driving ability better compared to using manual controls, just exchanging buttons and screens with voice interfaces without consideration for the effect on the driver's workload level might not completely solve the problem of distracted drivers.

The issue of cognitive distraction in relation to conversation while driving has been studied in the context of in-vehicle mobile phone use, and studies of in-vehicle mobile phone conversations do seem to support the theory of PRP. Patten et al. (2003) found effects on the driver's sensory and perceptive abilities when talking on the phone while driving. The driver gets tunnel vision; each gaze at the road gets longer and the number of saccades decreases from about 90 per minute to about 80 per minute. Reaction time can increase from about 50 milliseconds to about 400 milliseconds. More complex conversations adds to the driver's cognitive workload, which has an effect on the ability to plan the route and manoeuver the vehicle. Another consequence of the increased workload is that steering wheel movements increase in frequency and size.

Hence, it seems like the mental effort of participating in an engaging and/or complicated conversation – even if it is another type of distraction than interacting with a dialogue system – also takes the driver's attention away from the driving task. Even if the distraction is not as bad as the physical distraction of using manual controls and a screen (for example, when looking away from the road while searching for the right button to push or when searching for information on a screen), it might still be a safety risk. Apparently, just adding a speech interface without special consideration of the fact that it is to be used as a secondary task might not be good enough from a safety perspective. Even if a speech interface decreases visual distraction, special concern has to be put on methods for also decreasing cognitive distraction.

Mobile phone conversation, as well as dialogue system interaction, involves two dialogue partners. One problem with many studies of cognitive workload while driving is that they are performed in a car simulator, and the test person is given a math task as a way of increasing their cognitive workload. This setup basically tests one-way communication and only gives us information about the driver's behaviour (which probably differs from the behaviour during two-way communication) and therefore does not tell us the whole story.

### 2.4.5 Signalling workload in interaction / cognitive workload awareness

So, what about looking at the problem from the aspect of both of the dialogue partners, to see if we can learn something from that? Esbjörnsson et al. (2006) studied drivers' dialogue behaviour when talking over the phone while driving in city traffic. They found that drivers, when having the ability to choose for themselves, plan their actions and only make phone calls in certain traffic situations. Analysis of the conversations revealed that drivers also give implicit information about increased cognitive load. For example, during intense driving they make comments about the traffic situation or give minimal contribution to the conversation (e.g. filled pauses or standard phrases like "what was I going to say") for as long as the cognitive workload is high. They also found that the remote caller is very adept at interpreting those signals and adapt their dialogue behaviour according to the driver's cognitive load.

This implicit way of communicating high workload and the way the remote caller is perceiving the signals (i.e. being cognitive workload-aware) are very important differences between the interaction between mobile phone users and interaction with a non-cognitive workload-aware dialogue system. Mimicking the human behaviour of perceiving signals indicating high workload and using dialogue strategies to lower the workload might be an effective way of making a dialogue system cognitive workload-aware and less distracting to use while driving.

Fors and Villing (2011) noted that in human-human dialogues, speakers tend to use more pauses both within and between utterances during high cognitive load. Even if only one of the speakers experience high workload (for example, when the driver's workload is high and the passenger's is not) the speaker with low workload also uses more pauses as a way of adjusting to the dialogue partner. The authors propose that a dialogue system should adjust its use of pauses within utterances when the driver shows signs of high cognitive workload as a way of making the utterances easier to interpret and thereby lowering the dialogue partner's workload.

Breitholtz and Villing (2008) studied whether *enthymemes* can decrease the cognitive load of a dialogue system user. Enthymemes can be seen as arguments pointing out the speaker's reason for making a certain statement, and thereby releasing the listener from the burden of trying to find an interpretation of the utterance. For example, by letting a navigation system say "Turn left, there are road works straight ahead", instead of just "Turn left", the driver does not have to try to figure out why they should turn left although it might seem like the proper way to go would be straight ahead.

Using dialogue strategies like these as a way of making a dialogue system cognitive workload-aware includes knowing *when* the workload is high. In recent years there has been a growing interest in how to estimate drivers' cognitive workload level. This research has mainly been focused on either analysing the driver's speech and/or behaviour, or using tools such as, for example, eye-tracking or other physical measurements (see, for example, Lindström et al. (2008); Palinko et al. (2010); Medenica and Kun (2012); Dozza (2013)).

Estimating workload level is the first step in making a dialogue system cognitive workload-aware. An idea that we will be investigating in the following is that it may be useful to distinguish between different *types* of workload, in order to be able to adjust the interaction depending on the cause of the increased workload and the effect the workload has on the driver's behaviour. Next, we will present a theory on how to distinguish between two types of cognitive workload.

## 2.5 Types of cognitive workload

The motivation for the research presented in this thesis is to find a way of making in-vehicle dialogue systems less distracting for the driver. As we could see in Section 2.4, research has shown that even if the use of spoken dialogue systems increases the driving ability compared to using manual systems, there is still a risk that the driver gets distracted by the interaction and thereby less attentive to the driving task.

The cause of the increased workload may be one or more of the following:

- the driving task
- the interaction with the dialogue system
- some third task

In a sense, all activity is mentally demanding in some way. Driving a car includes performing several tasks simultaneously, such as steering, accelerating/braking, keeping an eye on the traffic situation and surroundings to avoid accidents, etc. All these activities add to the driver's cognitive workload in various ways, but not all manoeuvres are sufficiently demanding to increase the driver's workload level enough to affect the driving performance in a negative way. Interacting with a dialogue system while driving further adds to the cognitive load, since the driver needs to split their attention between the driving task and the dialogue task.

So, what can we do in order to lower the driver's workload? We have no influence over the driving task, nor can we influence any third task the driver might be performing. We might be able to get information about some of the additional tasks the driver might be performing; there are, for example, vehicles equipped with cup holder sensors that can let us know when the driver lifts a cup from the cup holder to take a sip of coffee. However, we can not realistically prevent the driver from doing these tasks even if we believe that it might be a bad time for drinking coffee from a safety point of view. That leaves us with one option, namely the dialogue system task. That task *is* possible to influence. In fact, it is possible to let the system decide both when and how an in-vehicle interaction should be performed. Two tools available for doing that are:

- interrupting the interaction in order to let the driver concentrate on the driving task
- continuing the interaction but changing the system's dialogue behaviour in a way that is more suitable when the dialogue partner's workload is high

To be able to decrease the driver's workload level by changing dialogue behaviour, we need to know when to use which tool. As we will argue below, the choice of tool is dependent on what *type* of workload the driver is experiencing. We distinguish between workload related to the driving task and workload related to some other task. To be able to determine workload level and type, we need to be continously estimating the driver's workload level. There is still no way of measuring the exact cognitive workload level, all we can do is estimate if it is increased or not (Mahr et al., 2012). We also need to analyse the driving behaviour to see if the manoeuvres that are performed at the moment are (assumed to be) cognitively demanding to perform.

Next, we will define the two types of workload we are interested in; *manoeuvre-related* and *manoeuvre-unrelated* workload, and a third condition that we refer to as *state of alert*. During all types of workload, we assume that the driver is interacting with the dialogue system while driving the vehicle.

### 2.5.1 Manoeuvre-related workload

Manoeuvre-related workoad is our label for situations where the workload is high at the same time as the driving behaviour is changed in a way that correlates with high workload. The cause could be the driving task itself, the dialogue task or some third task. However, if it is caused by anything other than the driving task, it has to affect the driving task to count as manoeuvre-related workload.

Figure 2.1 shows manoeuvre-related workload induced by the driving task.

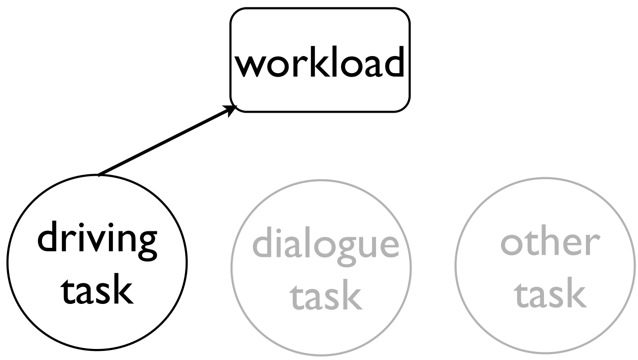



Figure 2.1: Manoeuvre-related high workload induced by the driving task.

This type of high workload is induced by the driving task itself when the driver is making a manoeuvre that causes high workload, which is illustrated by the arrow pointing from the driving task box to the workload box. For example, braking hard, presumably due to some sudden chance in the traffic environment, or making a left turn in right-hand traffic are manoeuvres that usually demand extra attention from the driver and hence are assumed to increase the workload (Harms, 1991). From the DICO corpus we have the following example of presumed manoeuvre-induced workload:

(1) [TDT = LOW, IDIS = LOW]

**Passenger (pointing left):** "let's see if you PAUSE change to" (Sw. "ska vi se om du PAUSE lägger dig i")

*The driver turns on the blinkers.*

[TDT = LOW, IDIS = HIGH]

**Passenger:** “the left lane” (Sw. “vänsterfilen”)

*driver looks in the rear-view mirror and turns left*
[TDT = HIGH, IDIS = HIGH]

**Passenger:** “here so that you can go straight ahead” (Sw. “här så att du kör rakt fram”)

In this example, IDIS’s continuous analysis of the driving behaviour reveals that the driver has turned on the blinkers to make a left turn. Since making a left turn in right-hand traffic is assumed to be cognitively demanding, IDIS starts signalling “increased driver activity”. The workload is still low at this time. The driver then looks in the rear-view mirror and turns the steering wheel to the left in order to shift lane. These actions require more resources from his working memory, which means that he has less resources left for other actions, and consequently the TDT notices an increased level of workload and starts signalling high workload. The increased workload in combination with the demanding driving manoeuvre (the left turn) make us draw the conclusion that it is the driving task that is causing the increased workload.

However, manoeuvre-related workload may also be induced by a secondary task. It can be either the interaction with the dialogue system or some third task that is causing the workload to increase, which in turn *affect* the driving behaviour.

In this case, the driving manoeuvre(s) performed correlate with high workload, although the driving task is not the cause. A possible cause could, for example, be that the driver finds it hard to interpret an utterance made by the dialogue system, or maybe is adjusting the fan or picking up something from the passengers seat. The participants in the DICO study were often engaged in intense discussions, and sometimes the dialogue task may have

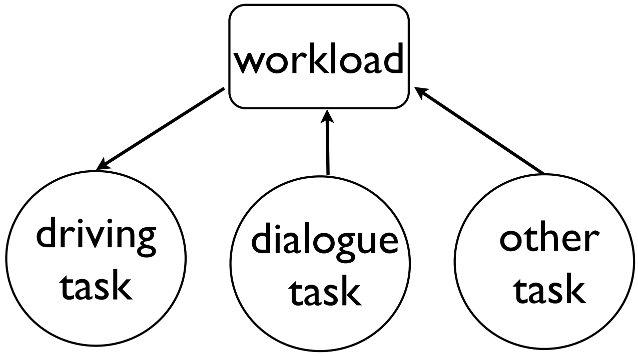



Figure 2.2: Manoeuvre-related high workload induced by a secondary task

taken focus away from the driving task. We see a possible example of this when the participants are discussing the driver's boat, and the driver suddenly gets unsure on where he is heading:

(2) [TDT = LOW, IDIS = LOW]

**Driver:** "yes it is a eeh motorboat a daycruiser it's nice to have when you" (Sw. "ja det är en eeh motorbåt en daycruiser så atte den är lite lagom att ta sej ut med när man")

*After crossing a street the driver starts looking around at both sides and the car wobbles a little bit back and forth.*

[TDT = HIGH, IDIS = HIGH]

**Driver:** "we are at Slottsskogs no" (Sw. "vi är in på slottsskogsg nej")

*The passenger starts reading the navigation instructions.*

[TDT = HIGH, IDIS = HIGH]

> Passenger: "er into Maria Street PAUSE is that here"
> (Sw. "eeh in på Mariagatan PAUSE fanns det här")

The driver is talking about his boat and seems to be rather occupied by the discussion. He drives straight through a crossing and after that he realises that he is not sure of the name of the street he just entered. He starts looking around to orientate himself, and wobbles a little bit back and forth while trying to figure out where he is. In this case, it is not the driving task itself that makes him wobble. The driver is driving straight ahead all the time and nothing on the street or in the traffic situation overall would have initiated such behaviour. Instead, the interview discussion may make him less concentrated on the navigation task to begin with. When the driver realises that he has lost track on where he is, the navigation task requires more resources from his working memory. This makes the driving task suffer in the sense that the driver is less able to keep the car straight in the lane. Instead of continuing the interview task (the boat discussion), the passenger interrupts that task and switches to the navigation task to try to help the driver figure out where to go next.

Distraction caused by a secondary task (the dialogue task or some third task) can also manifest itself in different ways. It may, for example, make the driver slowly decrease the speed without knowing it or check mirrors and instruments less frequently (Young and Regan, 2007).

However, when choosing what dialogue management tool to use, we do not *in our initial analysis* care about the cause of the increased workload. Regardless of the cause being the driving task or some secondary task, if the driving behaviour is changed in a way that correlates with high workload, we label it manoeuvre-related workload and take measures accordingly. This is motivated strictly from a safety point of view. Regardless of whether the driver is making a demanding manoeuvre or is distracted enough to change the driving behaviour, we need to change the dialogue behaviour so that the dialogue system interaction does not steal attention from the driving task and risks adding to the workload.

In both dialogue excerpts above, we could see that the passenger followed this principle. Frequently, it is done by switching to the navigation domain. In Example (1) the passenger helped the driver by giving him further instructions regarding which lane he should choose, and in Example (2) the passenger interrupted the interview task and instead started reading from the navigation intructions to get more clues on where to go. It is only at a later stage, when the driving behaviour is back to normal (i.e. is not showing signs of high workload), that it is time to take into account the cause of the increased workload. The system should probably adjust its behaviour to prevent the workload from increasing yet again, and there may be different strategies depending on the cause of the increased workload.

### 2.5.2 Manoeuvre-unrelated workload

Manoeuvre-unrelated workload is our label for situations where the workload is high although the driving behaviour is not changed in a way that correlates with high workload.

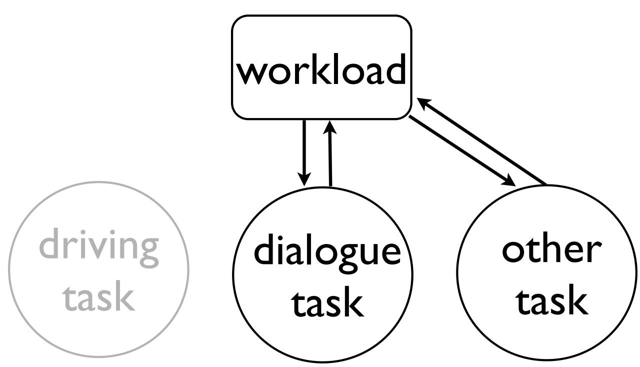



Figure 2.3: High workload related to a secondary task

Again, the workload may be induced by the interaction with the dialogue system or some third task, but in this scenario the driving behaviour shows no signs of high workload. Therefore, when the user is interacting with the dialogue system, we assume that we do not have to change the dialogue behaviour in a way that allows for the driver to concentrate more on the driving task, e.g. by switching to the Navigation domain. Instead, we need to simplify the dialogue task, e.g. by rephrasing an utterance or suggest alternative answers to a question.

For example, the dialogue task may temporarily be more difficult to perform due to a certain system utterance being hard to interpret, or it may be hard to figure out how to express what you want to do. In the DICO study, the participants were told that they would be tested after the driving session to see how much they could remember from interviewing each other. Sometimes the answers they got during the interview made them realise it would be hard to remember the answer correctly, for example, when memorizing the title of the most recently read book:

(3) [TDT = LOW, IDIS = LOW]

**Passenger:** "the most recently read book eeh years of warfare by peter englund" (Sw. "senast lästa bok eeh ofredsår av peter englund")

[TDT = HIGH, IDIS = LOW]

**Driver (whispers):** "i am never gonna remember that" (Sw. "det kommer jag aldrig komma ihåg")

IDIS is not signalling increased driver activity, which tells us that the high cognitive load probably does not arise from the driving task in this case, and neither is the workload affecting the driving behaviour.

User experience influences the amount of effort required to perform a task; a novice user has to mentally divide a task into the small steps that together form the entire task (Sweller et al., 1998). For example, the task of brewing coffee consists of a series of steps, such as opening the water container, pouring water in the container, opening the container for the coffee filter, putting in a new filter, and so on. The novice user performs one step at a time, while an experienced user has learned the steps by heart and does not look upon the task as a series of steps any longer, but instead sees it is one single task; "to brew coffee" (ibid). A conversational dialogue system offering greater freedom in expressing what you want to do makes it easier and less demanding to use even for a novice user, but there is still a learning process.

Since it is often hard to know the exact cause of high workload, we use a safety principle. If the workload increases while the driver is (at least part of the time) interacting with the dialogue system but there are no signs of the driving behaviour causing high workload, we can simply assume that the interaction is the cause. Even if the workload instead happens to be be induced by some third task, we believe that it is still useful, and at the very least not harmful, to try to decrease the workload caused by the interaction by changing the dialogue behaviour in the same way as if the dialogue task was the cause of the workload.

### 2.5.3 State of alert

There is a third scenario that is also relevant when adjusting dialogue system interaction to the driver's workload level, namely when the driving behaviour correlates with high workload although the estimated workload level is low. In this scenario, the driving analyser detects a driving manoeuvre that correlates with high workload, but the tool that measures the driver's workload estimates the workload level to low. In this case we probably do not have to change the dialogue behaviour, but we should be aware of the increased risk that the workload level might get high as the driver indeed is performing a task that is more demanding, and therefore, if possible, we should avoid adding to the workload.

## 2.6 Determining workload type

To be able to determine type of workload we need a tool for analysing workload level and a tool for analysing the driving behaviour.

### 2.6.1 Tools for estimating cognitive workload

Workload level can be estimated subjectively as well as objectively. One way of measuring workload subjectively is to simply ask the drivers to estimate how well they are driving at the moment compared to how they usually drive. This type of measurement is very useful during user tests to find out the participant's own comprehension of the situation, but for use in commercial systems one needs objective measurements that do not intrude or demand any additional actions from the driver.

There are several ways of objectively estimating cognitive workload level. For example, we can use physiological measurements such as heart rate, brain activity or eye-movements (see, for example, Brookhuis and de Waard (2010), Kun et al. (2013)), or various kinds of task performance measurements such as reaction time (see, for example, van Winsum et al. (1999)) or the lane change task (LCT) (Mattes, 2003). However, tools that require that you perform a task, such as tools for measuring reaction time, often interfere with the driving task to some extent. They might, for example, require the driver to push a button while holding the steering wheel and thereby force the driver to use their hands for both the driving and the performance tasks. When testing and evaluating interaction models this does not have to be a problem, but for vehicles in serial production the measurement tool must not interfere with the driving task.

When using an in-vehicle spoken dialogue system it could be convenient to measure workload level by analysing the driver's speech, since it does not interfere with the driving task and the user has to speak anyway to be able to use the dialogue system.

For example, studies have shown that an increased number of disfluencies such as deletions can indicate increased workload (Shriberg, 2001; Lindström et al., 2008). However, all humans do not produce the same disfluencies under high cognitive load, in the same way that they do not produce the same disfluencies under low cognitive load. Shriberg (2005) distinguishes, for example, between "repeaters" (people who tend to produce more repetitions than deletions or false starts) and "deleters" (people who show the opposite pattern). A problem with analysing user utterances to find a behaviour that correlates with high workload is that the utterances are mostly very short. Therefore, it can be hard to determine workload based only on clues like disfluencies since the material that is to be analysed is too small. There are currently no available commercial systems that determine workload based on speech analysis. Another problem with using a speech analyzer is that it is only able to measure workload when the user is speaking. Therefore, we do not know anything about the workload level when the user is silent, and as mentioned before, it is useful to be aware of the workload level also before the interaction starts, or resumes after a pause, to know if the driver's workload level already is high due to some third task.

What we want is a tool that measures the workload continuosly, without any interruptions. An eye-tracker (mounted in a non-intrusive place, for example, on the rear-view mirror) might be a good alternative, insofar as this is an effective and non-disturbing way of measuring the driver's workload level. An eye-tracker which measures pupil size and/or eye-movements does not interfere with the driving task and is considered to be a reliable tool for determining workload (Iqbal et al., 2005).

In the DICO study, a TDT (i.e. *Tactile Detection Task*) was used for measuring the driver's workload. As the stimuli is presented to the user by a buzzer attached to the wrist it is a "pure" way of measuring workload, since it is not affected by different lighting conditions or the driver looking away, as an eye-tracker or a PDT (i.e. *Peripheral Detection Task*, which is similar to a TDT except the user reacts to visual stimuli instead of tactile stimuli) potentially is (Engström et al., 2005). TDT measures task performance (hit-rate and reaction time) and is considered to be a reliable way of measuring workload (Merat and Jamson, 2005). A TDT requires the user to push a button at the same time as they are manoeuvering the steering wheel which could possibly make it more difficult to make a turn. However, since TDT is considered a reliable way of measuring workload while driving, it could still be preferable to use in a user study. Studies have shown that although it interferes with the driving task to some extent, it does not make a significant disturbance for the user (Mattes, 2005).

### 2.6.2 Tools for analysing driving behaviour

As mentioned, apart from measuring workload, we also need a tool that analyses driving behaviour. In the DICO study, IDIS was used to analyse the driving behaviour. IDIS (Broström et al., 2006) measures workload based on the assumption that some manoeuvres are more cognitively demanding than others. For example, in right-hand traffic, making a left turn is assumed to be more demanding than making a right turn. Therefore, it does not measure the individual driver's workload directly. Instead, it should be seen as a tool for measuring driving difficulty, as it reacts to driving manoeuvres that are complicated and *may* correlate with high workload. It is not clear from available documentation whether it separates between manoeuvre-induced and manoeuvre-affected behaviour, or simply categorizes all manoeuvres that correlate with high workload as manoeuvre-related workload.

IDIS was developed to decide when it is suitable to show alarms that are non-critical, such as the indicator for low level of screen washer fluid. The driver would possibly be distracted if the alarm signal is shown at a bad time (for example, when the driver is shifting lane), but since the driver is not expecting an alarm to appear at a certain moment, a few seconds delay of the signal is highly unlikely to be noticed.

The workload-inducing manoeuvres that IDIS reacts to are manoeuvres that are assumed to be demanding for the *average* driver.

Hence, depending on factors such as, for example, age and experience, not all drivers' workload is high even if the manoeuvre they are making indicates that it should be. Therefore, IDIS might over-generate and gives a false positive, i.e. indicates high workload although the driver's workload is actually low. If IDIS overgenerates and as a consequence does not show, for example, an alarm right away even though the driver's workload is in fact low at the moment, this will most likely not be noticed by the driver who does not expect the alarm to go off at a certain time. However, when using a dialogue system the user may be expecting a reaction from the system at a certain time (for example, an answer to a question), and therefore a system like IDIS cannot be used alone to determine when it is suitable to continue the interaction or not. Therefore, for our field of application (i.e. interaction with a dialogue system) it is probably better to look at IDIS as a tool for detecting potentially workload-correlated driver activity rather than determining the driver's workload level. If we make this assumption, we can use data from IDIS to understand how information about the driving activity can be used as an aid in determining type of workload. Also, having a cognitive workload detector which is independent of the driving activity detector allows adapting the interaction also to manoeuvre-unrelated workload.

### 2.6.3 Combining workload level and driver activity

Combining analyses of workload level and driver activity may be an effective way of distinguishing between workload types. By using tools for measuring workload and driver activity as proposed above, we believe that we can determine type of workload according to the following heuristic (Villing, 2009b):

**Manoeuvre-related workload** occurs when the workload analyser signals high workload and the tool for analysing driving behaviour signals driving behaviour correlating with high workload.

**Manoeuvre-unrelated workload** occurs when the workload analyser signals high workload but the driving behaviour analyser does not detect any manoeuvres that correlates with high workload.

**State of alert** occurs when the workload is estimated low, but the driver activity is increased.

These heuristics are summarized in Table 2.1, and are based on the presumption that the driver is interacting with the dialogue system when the workload is increasing.

| | + WL | – WL |
|---|---|---|
| + DA | manoeuvre-related | state of alert |
| – DA | manoeuvre-unrelated | low |

Table 2.1: Type of workload (WL) based on driver activity (DA).

## 2.7 Distribution of workload types and domains of discussion in the DICO corpus

Next, we will show descriptive data of the distribution of workload types and domains of discussion within the DICO corpus. Workload type is annotated according to the taxonomy described in Section 2.6.3.

### 2.7.1 Workload distribution

We start by calculating the distribution of workload types, to see which workload type is most common and which is least common. Table 2.2 shows the distribution of driver workload.

In total, the DICO corpus contains 164.9 minutes of in-vehicle dialogues. During most of that time, about 109 minutes, the workload is low and the driver is not performing any demanding driv-

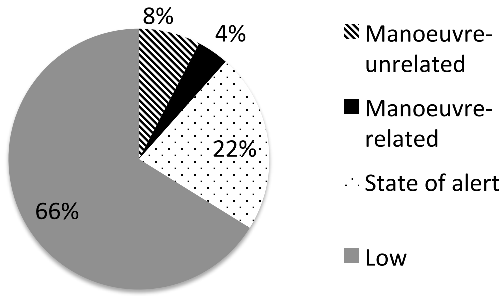



Figure 2.4: Distribution of workload types.

| Workload type | Duration | Average (seconds) | Range (seconds) | Median (seconds) | Occur-ances |
|---|---|---|---|---|---|
| **Manoeuvre-related** | 6 min 30 sec | 5.6 | .32 - 26.8 | 4.1 | 70 |
| **Manoeuvre-unrelated** | 12 min 36 sec | 6.3 | .26 - 31.3 | 4.7 | 120 |
| **State of alert** | 36 min 42 sec | 8.6 | .12 - 39.1 | 5.8 | 256 |
| **Low** | 109 min 6 sec | 21.2 | .27 - 121.2 | 13.8 | 308 |

Table 2.2: Duration time for each workload type.

ing manoeuvres. State of alert has the second longest duration time and number of occurances, and of the two workload types manoeuvre-unrelated workload is more common than manoeuvre-related.

As mentioned earlier (see Chapter 1.4), the study took place in the centre of Gothenburg, Sweden, with sometimes dense traffic and a lot of lane changes and turns. Therefore, it is not surprising that the manoeuvreing of the car is often demanding, causing state of alert to be the second most common workload type after low workload. However, since state of alert is more common than manoeuvre-related workoad (which is the least common workload type), it seems that even though the driving task is often demanding it is usually not demanding enough to cause high workload.

### 2.7.2 Domain distribution

As described in Chapter 1.4.3, we annotated the corpus with domain type, each annotation starts with the first utterance within the domain, and ends with the last utterance before a domain switch. The total duration time is then calculated for each domain. The duration times for the domains can be found in Figure 2.5 and Table 2.3.

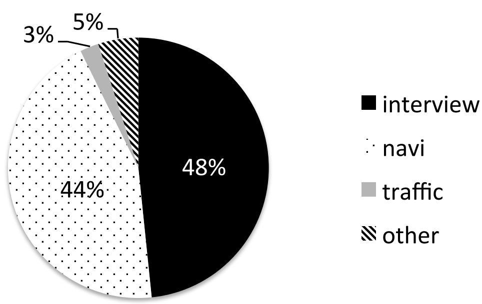



Figure 2.5: Distribution of domain duration.

| | Duration | Average (seconds) | Range (seconds) | Median (seconds) | Occurances |
|---|---|---|---|---|---|
| **Interview** | 79 min 48 sec | 32 | .2 - 211 | 23.2 | 150 |
| **Navigation** | 73 min 6 sec | 25.6 | .1 - 285.4 | 17 | 171 |
| **Traffic** | 3 min 48 sec | 6.7 | .1 - 20.6 | 5.3 | 34 |
| **Other** | 8 min 12 sec | 10.7 | 1.2 - 34.1 | 8.8 | 46 |

Table 2.3: Duration time for each domain.

As mentioned in Section 1.4, the participants could freely choose when and how they wanted to discuss each task they were given.

The participants were told that they should be tested afterwards to see how much they remembered from the interview questions they had discussed, and be given points for every right answer. From the recordings we can tell that they wanted to score as high as possible by discussing as many questions as they could, but also by motivating some of their answers since a motivation of the answer gave more points than a straight answer. The interview discussions are therefore often quite long, and the result shows that the Interview domain has the longest duration time overall.

The navigation instructions could sometimes be tricky to interpret, especially if the neighbourhood were unknown to the driver-passenger pair (for example, an instruction like "turn right into X Street" could sometimes start a longer discussion of which street to turn into if the street sign was missing or hard to see) or if the instruction was unclear (for example, some of the participants did not know how to interpret the instruction "turn left at the third exit in the roundabout", and had trouble figuring out which exit was referred to). However, many of the instructions were rather straightforward and easy to follow (for example, "go straight ahead at X Street for one kilometer") and did not need to be discussed or elaborated further. The reason for presenting the instructions in this way is that we wanted to study the participant's dialogue strategies and see if there was any effect on the cognitive workload level when trying to interpret the navigation instructions. We hoped that the tricky instructions would elicit problem-solving discussions, and the driving instructions are thus not necessarily comparable with instructions that are expected to be given by a navigation system (even though a navigation system can also be tricky to interpret in certain situations).

The duration times for the Traffic and Other domains are shorter than the other two. The shorter duration times are probably due to both the nature of these domains and the fact that the tasks that were actually given to the participants were the interview and navigation tasks and the participants wanted to perform as well as possible when solving them. As can be seen in Table 2.3, the comments within the Traffic domain are usually very short, average duration time is 6,7 seconds compared to 32 seconds for the Interview domain and 25,6 seconds for the Navigation domain. The comments within the Traffic domain are usually not aiming at starting a longer discussion, instead the speaker often utters only a single or at the most a few words to comment about something that just happened in the traffic. The same goes for the Other domain, the comments made here are usually short and are seldom elaborated further by the dialogue partner as both speakers are focused on the two tasks of interviewing each other and navigating along the route.

### 2.7.3 Distribution of workload between domains of discussion

We have measured the duration time for the workload types within each domain. Figure 2.6 and 2.7 shows how the workload types are distributed over the domains.

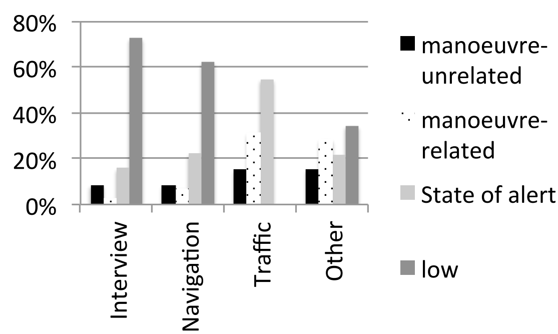



Figure 2.6: Workload per domain.

The Traffic domain and the Interview domain stand out as each others opposite. The total duration time within the Interview domain is almost 80 minutes (see Table 2.3). The workload is high about 9 minutes of the time (manoeuvre-unrelated workload about 7 minutes and manoeuvre-related about 2 minutes), and there is a state of alert during about 11 minutes of the time.

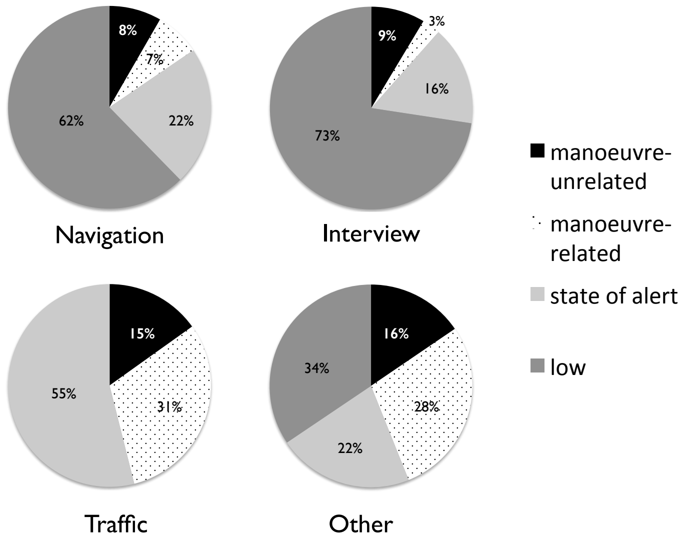



Figure 2.7: Workload and increased driving activity within each domain.

The workload is low about 59 minutes of the time. This means that the workload is low during most of the time, and if the workload increases, manoeuvre-unrelated workload is more common than manoeuvre-related. The total duration time for the Traffic domain is almost 4 minutes. The workload is high about 2 minutes of the time (manoeuvre-unrelated about 35 seconds and manoeuvre-related about 1,5 minutes) and there is a state of alert about 2 minutes of the time. The workload is never low within this domain.

This means that while the workload is usually low when the participants discuss interview-related issues, there is always either high workload or a state of alert during traffic-related discussions. Within the traffic domain, the workload is more often manoeuvre-related than manoeuvre-unrelated, while within the Interview domain we see the opposite pattern. State of alert is more than three times as frequent within the Traffic domain than within the Interview domain.

The Navigation domain is similar to the Interview domain. The workload is low about 46 minutes of the time, but unlike the Interview domain the two high workload types are almost equally common (about 5 minutes for manoeuvre-related and about 6 minutes for manoeuvre-unrelated workload). There is a state of alert about 16 minutes of the time.

The Other domain is similar to the Traffic domain in the way that the workload is either high or there is a state of alert during most of the time (about 2,5 minutes duration time for manoeuvre-related workload, about 1 minute for manoeuvre-unrelated workload and almost 2 minutes for state of alert). The workload is low almost 3 minutes of the time.

## 2.8 Summary

In this chapter, we have given a definition of the term *cognitive workload* in the context of the research presented in this thesis, and presented an overview of previous research regarding cognitive workload. We discussed the concept of cognitive workload-aware systems and how they may help lowering the cognitive workload of a driver who is using an in-vehicle dialogue system. We presented a taxonomy of workload types and showed how workload type can be determined from TDT and IDIS signals. Finally, we showed descriptive data of distribution of workload types and domains of discussion within the DICO corpus.

# Chapter 3

# Interruption in dialogue

## 3.1 Introduction

When the driver is experiencing manoeuvre-related high workload during interaction with a dialogue system, it may be necessary for the system to interrupt the dialogue in order to give information that can be useful, or to help the driver focusing on the driving task. There are many examples of this dialogue behaviour in the DICO corpus. See, for instance, Example (4), where the participants are discussing the driver's work assignments and the driver becomes unsure of what to do after beginning a left turn and then realizing there is a traffic jam and the car is blocking the street.

(4) [WORKLOAD = MANOEUVRE-RELATED]

**Driver:** "it is the technology itself when it comes to that that I am PAUSE interested in" (Sw. "det är själva tekniken där som jag är PAUSE road av") [*domain of discussion = interview*]

**Passenger:** "yes that's right PAUSE technology is fun" (Sw. "ja just det PAUSE nä teknik är roligt") [*domain of discussion = interview*]

**Driver:** "this feels a bit strange to be standing here right now but" (Sw. "det här känns lite konstit att stå här just nu men") [*domain of discussion = traffic*]

**Passenger:** "yes but I think you can stand" (Sw. "ja men du kan nog stå") [*domain of discussion = traffic*]

> *Pause for 1 second*

**Passenger**: "nobody else is getting anywhere either so" (Sw. "det är ingen annan som kommer nån vart heller så att") [*domain of discussion = traffic*]

**Driver**: "no it's the usual situation in this town" (Sw. "nä det är som vanligt i stan") [*domain of discussion = traffic*]

> *The car in front is starting to move.*

**Passenger**: "catch on to that" (Sw. "haka på här") [domain of discussion = traffic]

> *Pause for about 7 seconds while the driver is slowly making the left turn. When they have almost finished the left turn the passenger starts speaking again.*

**Passenger**: "have you ever been to a country you don't want to return to" (Sw. "har du varit i nåt land som du aldrig skulle vilja åka till igen") [*domain of discussion = interview*]

> [WORKLOAD = STATE OF ALERT]
>
> *While the passenger is making the last utterance the workload changes from manoeuvre-related to state of alert.*
>
> *Pause for about 3 seconds*
>
> [WORKLOAD = LOW]

**Driver**: "LAUGHS no no" (Sw. "nähä näe")

In this example, both participant's speech is disfluent (silent pauses) when they are discussing the interview question, which may be a sign of high workload (Lindström et al., 2008). When the driver is making the left turn the workload is manoeuvre-related, and both participants are silent. When the turn is almost finished the workload decreases according to the TDT, and the workload type is therefore changing from manoeuvre-related to state of alert. The passenger then starts speaking and asks the next interview question. The driver is quiet for another 3 seconds until the turn is finished and they are going straight on. Since the driver is no longer making a demanding manoeuvre the workload type changes to low, and the driver answers the question. The passenger made the decision that it was ok to start talking even though the left turn was not completely finished. Whether the driver hesitates before answering because of the manoeuvring of the car or because s/he needs to think about the answer, we do not know.

Another way of interrupting a conversation is to switch to another topic and/or domain of discussion although the ongoing topic is not finished yet. For example, by switching to the Traffic domain and by that indirectly indicating high workload. Example (5) is an example of this behaviour, when the participants are discussing their education while the car is approaching a crossing:

(5) *Driving straight ahead, approching a crossing street*
[WORKLOAD = LOW]

**Driver**: "and eeh well I guess that is basically PAUSE eeh" (Sw. "och eeh ja det är väl egentligen+ PAUSE eeh") [*domain of discussion = interview*]

*As they are approaching the intersection a pick-up truck is crossing the street and the driver brakes.*
[WORKLOAD = STATE OF ALERT]

**Driver:** "well that is basically PAUSE it PAUSE I studied it with focus on communications" (Sw. "ja det är väl egentligen PAUSE det PAUSE jag har läst en kommunikationsinriktning")

*Turns right, into the crossing street*
[WORKLOAD = STATE OF ALERT]

**Passenger:** "mm" (Sw. "mm") [*domain of discussion = interview*]

**Driver:** "so well" (Sw. "så atte+")

*The driver pauses for about 1 second. The car is entering the new street which is narrow, there are cars parked along both sides of the street*
[WORKLOAD = STATE OF ALERT]

**Driver:** "squeeze myself in here" (Sw. "knör mig emellan här") [*domain of discussion = traffic*]

*The driver pauses for about 4 seconds while driving in between the parked cars, then drives straight ahead on the street.*
[WORKLOAD = LOW]

**Driver:** "so that is basically it" (Sw. "så det är väl egentligen det") [*domain of discussion = interview*]

When the car is approaching the crossing street, the driver becomes more concentrated on the driving task, and less able to continue the interview task. He uses more disfluencies such as pauses and filled pauses, which could be a sign of high workload (Lindström et al., 2008; Shriberg, 2001). However, according to the TDT the workload is still low even though IDIS starts signalling that the driver is making a demanding manoeuvre when the driver brakes. When the driver turns into the crossing street, it is a bit tricky to navigate through the street which is rather narrow, especially since the street gets even narrower because of the parked cars along the sides. The driver comments upon what he is doing (*"squeeze myself in here"*), thereby interrupting the discussion within the Interview domain and switching to the Traffic domain.

This comment could be considered redundant information since the passenger is sitting beside him and is able to see what manoeuvres the driver is performing, and hence the driver should not have to inform the passenger (Breitholtz and Villing, 2008). The reason why the driver nevertheless chooses to make the comment is therefore probably not so much to inform the passenger of what he is doing. Instead, by switching to the Traffic domain, the driver is indicating that the driving task is more demanding at the moment and therefore he wishes to concentrate on that instead of continuing the interview task. In doing so, the driver uses dialogue strategies (domain switching) as a way of indicating high workload (Villing et al., 2008; Villing, 2009a). The passenger seems to interpret the driver's choice to switch domain in the same way as we do. He does not try to encourage the driver to continue the current discussion, nor does he ask a new interview question. Presumably the passenger is silent too because he wants to let the driver focus entirely on manoeuvering the car instead of simultaneously continue the interview task.

We can also find examples of how both speakers interrupt the ongoing discussion to ask for or give navigation instructions, if the information is needed immediately. This behaviour is illustrated in Example (6), where the participants are discussing which country they last visited for vacation:

(6) *going straight on, approaching a crossing*
[WORKLOAD = LOW]

**Passenger:** "a country that you would like to visit again" (Sw. "ett land som du gärna åker till igen") [*domain of discussion = interview*]

**Driver:** "that I would like to go to again eh" (Sw. "som jag åker till igen eh") [*domain of discussion = interview*]

> **Passenger:** "keep left after follow the sign that says tunnel" (Sw. "håll till vänster efter följ skylten mot tunneln") [*domain of discussion = navigation*]

The participants are driving straight ahead and the driver's workload is low. It seems like a good time to ask the next interview question, but while the driver is thinking about what to answer they are getting closer to where they need to decide which lane to choose. The passenger does not wait for the driver to answer the question, but interrupts to clarify where they should go next. A better strategy would probably have been to first give the navigation instructions before asking the interview question, but sometimes an interaction takes longer than expected and if the information that needs to be given is time-critical, it may be necessary to interrupt the ongoing conversation.

These examples show that during a conversation, there are (at least) two ways in which the participants can interrupt an ongoing topic – pausing the interaction, or switching topic. We believe that high workload during interaction with a dialogue system may be prevented, or its duration time shortened, if the dialogue system is able to use dialogue strategies like these to interrupt the ongoing topic when there is an increased likelihood of high workload.

We have discussed the disadvantages with using manual devices while driving (see Section 2.1). Instead of looking at the road and keeping both hands on the steering wheel, the driver is forced to look at a screen and push buttons to be able to control the device. One advantage of manual devices, however, is that they give the user freedom to choose when to pay more attention to the device (for example, look at the screen or press a button) and when to pay more attention to the driving task (for example, look at the road). Still, the advantage of allowing the user to keep their hands on the steering wheel and look at the road during the entire interaction makes spoken interaction a safer choice. As mentioned above, the interaction is less distracting and the driver's rate their driving ability as better when using speech (Gärtner et al., 2002; Vollrath and Maciej, 2008; Villing and Larsson, 2011). However, when using an in-vehicle dialogue system there is a chance that the system keeps the attention for too long. It is not possible to interrupt a spoken utterance in the same easy way as a written utterance on a screen. You can look away from the screen at any time you choose, it can be hard to avoid listening to a spoken utterance (unless you physically hit a "mute" button). Therefore, each system utterance should be planned so that the system does not speak at a bad time. Implementing dialogue strategies for managing cognitive workload may improve speech interfaces and make them safer and less distracting to use.

To achieve this, we also have to define places in the dialogue where it is (more) suitable to interrupt, since an interruption at an unsuitable place could possibly cause confusion and in the worst case increase the workload level.

## 3.2 Aims

In this chapter, we will look at interruptions in dialogue as a way of managing workload related to the driving task. We will investigate when and how humans interrupt in-vehicle dialogue, to learn more about the human-human behaviour as a basis for developing interruption strategies for human-machine interaction. We believe that interrupting an ongoing interaction may prevent, or shorten the duration of, high workload. We will explore two strategies concerning interruptions in dialogue. The first, and maybe most evident, strategy is to pause the dialogue to avoid any possible distractions from interacting with a dialogue system. However, it is possible that the driver in some situations would benefit from getting useful information from a dialogue system (for example, a navigation instruction). The second dialogue strategy therefore concerns interruption of an ongoing discussion with the purpose of switching topic and domain.

## 3.3 Background

Earlier research regarding interruption in human-computer interaction often aims at finding out how to interrupt an ongoing interaction in cases where the system needs to, for example, inform the user of something. Hence, it usually concerns the types of interruption that is initiated because the *system* needs to interrupt, and the interrupting task is usually a secondary task. The research is therefore concentrated on how to interrupt in a way that is least disruptive to make it as smooth and easy as possible to resume the interrupted (primary) task. What we are dealing with is another aspect of interruption, namely how to interrupt the dialogue because we believe the *user* needs the interruption, due to high cognitive load. Furthermore, the task that is interrupted (the interaction with an in-vehicle dialogue system) is the secondary task, and the task that calls for attention and is the cause of the interruption is the primary task (driving). This gives us a different angle on how and when to interrupt.

Kousidis et al. (2014) carried out a user study aiming at testing this behaviour. While using a driving simulator, the participants carried out a lane change task and a memory task (listening to a message and then being asked about facts in the message). The system giving the information either talked regardless of driving conditions, or paused when the user changed lane. When using the system that adapted to the driving conditions, there was no impact of the spoken interaction on either driving performance or information recall, but when using the system that spoke during lane changing, both driving performance and information recall was significantly worse. This result is encouraging, and indicates that a cognitive workload-aware dialogue system may help minimizing cognitive workload while driving.

Earlier research on pauses in dialogue shows that an increase in number of pauses and/or pause length may be a good indicator of cognitive load. A review of psycholinguistics and linguistics literature, carried out by Berthold (1998) regarding the effects of cognitive load on speech, revealed that both the number and duration of silent pauses tend to increase during high workload.

Pausing behaviour was investigated by Khawaja et al. (2008) in a study where the participants performed a reading comprehension task during low workload (where they were reading a text aloud and then answered questions) and during high workload (where they also had to listen to and remember the number of random series of digits while reading the text). When analysing the pausing behaviour, the authors found that there were significantly more silent pauses (meaning silent time out of total time including both number of pauses and duration time of the pauses) when the participants performed the dual-task (reading and listening) compared to when performing the single task (reading only).

As mentioned, if a dialogue system needs to interrupt to switch topic, we first need to know *when* to interrupt. It may not be suitable to interrupt the interaction immediately when the workload increases if we are at a point in the interaction where an interruption would be confusing or distracting in itself for the driver. We know that a user's cognitive workload level usually changes during task execution. The variations in workload level may be due to the complexity of the tasks, but researches have also found workload variations within single task execution (Iqbal et al., 2005). Interrupting during high vs. low cognitive load influences, for example, the time that is needed to resume the interrupted task (Bailey and Konstan, 2006; Iqbal and Bailey, 2006; Bailey and Iqbal, 2008). Salvucci and Bogunovich (2010) found that if users need to interrupt an ongoing task they wait, if possible, until the mental workload has been minimized before they interrupt to do the other task. When driving and interacting with a dialogue system it is not always possible to choose when to perform each task. Sometimes the traffic situation requires the driver's full attention, and there may also be situations when the system needs to interrupt to give time-critical information. However, these earlier studies show that if it is possible, the dialogue system should strive to interrupt when the driver's workload is low.

Looking at when we interrupt spoken dialogue, we find that humans normally strive to interrupt at the least disruptive place in the dialogue (Yang et al., 2011). For example, we usually avoid interruption in the middle of an adjacency pair. One reason for this may be social conventions. Levinson (1983) stated that adjacency pairs “are deeply inter-related with the turn-taking system as techniques for selecting a next speaker”. This is also emphasized by Schegloff and Sacks (1973); “having produced a first part of some pair, current speaker must stop speaking; and next speaker must produce at that point a second part to the same pair”. If an adjacency pair is being interrupted, the interrupting utterance is in most cases related to the first part of the adjacency pair, e.g. to clarify something in order to be able to answer a question. If the adjacency pair is aborted, i.e. the first part is not followed by the second or by a sub-dialogue, the speaker of the first part can may assume that the second speaker is, for example, not interested, is being deliberately rude or did not understand (Bridge, 2002). Therefore, dialogue partners strive to follow this turn-taking rule as far as they can and will not break it unless necessary.

When it comes to the cognitive workload aspect we can see another reason not to interrupt an adjacency pair, apart from social conventions. The workload usually decreases at subtask boundaries, i.e. during the time between consecutive subtasks (Iqbal et al., 2005), and that may be one reason why interruptions between adjacency pairs is less mentally demanding than interruptions within an adjacency pair. A finished pair makes for a simpler context to resume to, compared to an interrupted adjacency pair that might force the resuming dialogue partner to repeat (parts of) the discourse (Shyrokov et al., 2007). Yang et al. (2008) found that when the participants in a study were playing a card game and were interrupted in the middle of the game, they had to remind each other what cards they have at hand or even had to repeat parts of the dialogue before the interruption. When interrupted after a finished game they could start a new game without further ado, not having to remind each other of what happened before the game was interrupted.

When we are about to change topic, we usually signal this with certain *discourse markers* to prepare our dialogue partner(s) what we are about to do (Schriffrin, 1987). It has been found that discourse markers signal nearly the same set of conversational moves in task-oriented dialogue as they do in a more social dialogue (Byron and Heeman, 1997). We will take a closer look on this behaviour in the next chapter, to see if there is a difference in how we signal topic shifts depending on workload level and domain of discussion. Villing et al. (2008) distinguish between *standard phrases* (a discourse marker that is used to introduce a domain switch and can be used to switch to any domain, for example, *"let's see"*) and *domain specific phrases* (a phrase that is specific for a certain domain, for example, *"turn left now"*) as discourse markers, and found that standard phrases are used more often when the topic switch is planned, and that domain specific phrases are used more often when the topic shift is spontaneous. This result seems to correlate with what Yang et al. (2011) found in their card game study. The participants were interrupted by a real-time task, hence the interruptions were not planned ahead by the participants. Standard phrases like "oh", "wait", "alright" as well as a higher pitch (the higher the pitch, the more disruptive the interruption) were used for signalling task interruption, but the standard phrases were used in less than 40% of the topic switches.

## 3.4 Research questions

In Section 2.5 we describe a method for distinguishing between different types of cognitive workload. If the workload is determined to be manoeuvre-unrelated, we assume that the high workload is related to some second or third task. If the driver is interacting with the dialogue system when this type of workload occurs, we propose to adjust the dialogue system behaviour to ease the dialogue task and try to avoid adding to the cognitive workload. If the workload is considered to be manoeuvre-related (i.e. high workload in combination with a demanding driving manoeuvre), the dialogue may need to be interrupted to enable the driver to manoeuvre the car without being distracted by the conversation. We described the two strategies we believe are most useful in these situations (i.e. pause the dialogue or switch topic and/or domain), and we will now analyse the DICO corpus to see if the participants use these strategies and to what extent.

Regardless of whether we want to pause the dialogue or switch topic, we need to know *when* to do it. In spoken dialogue it is not always possible to plan ahead to choose when and how to interrupt the conversation. Although speakers strive to change topic at a less disruptive place, there could be situations (for example, due to time constraints) which may force the speaker to interrupt the ongoing conversation even if the timing is not optimal (Yang et al., 2008). We want to interrupt at a point in the dialogue where the interruption itself distracts the driver as little as possible. Pausing the dialogue at a bad time may cause confusion and make the driver wonder if s/he has done something wrong, or if there is something wrong with the system.

Finally, we want to learn more about how to switch topic. If switching topic during manoeuvre-related workload is a strategy that is used by the participants to aid the driver with the driving task, that may also be a good strategy to use for a dialogue system. By analyzing the passengers' behaviour when switching topic we can learn more about how to do this in a way that minimizes the risk that the driver gets confused by the topic switch, and by analyzing the drivers' behaviour we can learn more about how a user may want to express that they need to switch topic.

By doing this investigation, we want to find out which strategy to use; when should the interaction be paused and when should we switch to another topic or domain? To be able to make that decision, an agent needs to know the motivation behind the interruption. Would it be best for the driver to concentrate solely on the driving task at the moment and therefore pause the dialogue, or would the driver benefit more from getting help from the dialogue system in some way, for example, to get a navigation instruction?

## 3.5 Hypotheses

As just mentioned, the purpose of this investigation is to find out three things about interruptions in in-vehicle dialogue; whether to pause or switch topic, when to interrupt, and how to signal a topic switch.

Regarding whether to pause or switch topic, we will analyse the corpus with respect to silent pauses and topic switches in relation to type of workload, to see if workload is related to these types of dialogue strategies. Our hypothesis is that both types of interruptions occur more often during manoeuvre-related workload, since we believe that the speakers want to pause the dialogue if they believe it is necessary in order to let the driver concentrate on the driving task, but switch topic to the Navigation domain if the driver seems to need help with the navigation task.

To find out when it is suitable to interrupt an ongoing conversation, we will analyse the corpus with respect to interruptions within and outside an adjacency pair. Based on earlier research on interruption in dialogue (see Section 3.3), our hypothesis is that the participants will avoid interruption within an adjacency pair also in the in-vehicle environment. We will investigate interruption utterances in relation to adjacency pairs within the Interview domain, to see if there is a difference in the number of interruption utterances within and outside a pair.

Finally, we need to know more about how to switch topic, when that is the chosen strategy. Villing et al. (2008) propose that standard sequencing phrases are used more often when the topic switch is planned ahead, and domain-specific sequencing phrases are used more often when the topic shift is spontaneous (see Section 3.3). We will investigate this behaviour with respect to the driver's cognitive workload level, since we believe that topic switches made during high workload are more often spontaneous, and topic switches made during low workload are more often planned ahead. Following this reasoning, our hypothesis is that when switching topic during high workload, the participants will use domain specific sequencing moves more often than on average, and during low workload they will use standard phrases more often than on average.

## 3.6 Method

To learn more about how and when interruptions are being made, the DICO corpus has been annotated with *pauses*, *topic shifts*, *adjacency pairs* and *sequencing moves*, in addition to previous annotations (see Section 1.4).

### 3.6.1 Topic switches

We have annotated the corpus with *topic switches*, to be able to see how the participants switch between topics. We want to know how the participants start discussing a topic and how they end the topic discussion.

Topic switch has been coded as follows:

- the start of a topic:
  - **topic-begin**: whatever → topic A (new)
    - The beginning of a topic that has not been discussed earlier. For example:

      **Driver:** and you are a Gemini [*domain of discussion = interview*]
      **Passenger:** mm
      **Passenger:** do you have any pets? [*domain of discussion = interview, topic-begin*]
  - **topic-resume**: whatever → topic A (unfinished)
    - An utterance that resumes an earlier, interrupted, topic. The interrupted topic has therefore not been ended with an utterance annotated as *topic-end*. For example:

      **Driver:** what I am reading now is Sophie's World [*domain of discussion = interview*]

**Driver:** where do you want me to go [*domain of discussion = navigation*]
**Passenger:** straight ahead straight ahead [*domain of discussion = navigation*]
**Driver:** let's do that [*domain of discussion = navigation*]
**Passenger:** Enemy's Enemy was the last one I read [*domain of discussion = interview, topic-resume*]

- **topic-reraise**: whatever → topic A (finished)
  - An utterance that reraises an earlier topic that has been ended with an utterance annotated as *end-topic*. For example:

    **Passenger:** you said you played music instruments, piano and guitar [*domain of discussion = interview, topic-reraise*]

- the ending of a topic:
  - **topic-end**: topic A (finished) → whatever
    - An utterance that ends a topic, for example the answer to a question or feedback indicating uptake/acceptance. A *topic-end* annotation does not mean that the topic can not be discussed again, but if so a reraise of the topic is needed.

      **Passenger:** what instrument do you play? [*domain of discussion = interview*]
      **Driver:** mostly piano and guitar [*domain of discussion = interview*]
      **Passenger:** yes okay nice [*domain of discussion, topic-end*]

- In addition to the annotation of the start of a new topic, topic switches may also be marked as *interrupting* the preceding topic. In the cases where the last utterance of a topic is not annotated as being a *topic-end* utterance, the topic is assumed to be interrupted. The utterance marking the topic switch (i.e annotated as *topic-begin*, *topic-resume* or *topic-reraise*), is in these cases annotated as *topic-interrupt* on a separate tier:

- **topic-interrupt**: topic A (unfinished) → whatever
  - An utterance that starts a topic when the preceding topic is not ended with an utterance annotated as *topic-end*. For example:

    **Driver:** chocolate is unbelievably good [*domain of discussion = interview*]
    **Passenger:** this is Ekelund Street [*domain of discussion = navigation, topic-resume, topic-interrupt*]

Within the Interview domain, all types of topic switches are annotated. The first utterance regarding one of the interview questions is marked as starting the topic, and the topic is marked as ended if the last utterance regarding the topic is either the answer to a question or a confirmation of the answer.

Within the Navigation, Traffic and Other domains, *topic-end* utterances have not been annotated. The reason for this is that it is often hard to determine when these discussions are actually ended. The Interview domain has fairly well-defined topics (i.e. the interview questions), but this is not true for the other domains. One possible way of doing it would be to annotate confirming feedback as *topic-end*, as we do within the Interview domain. However, even though the speaker gives feedback to confirm, for example, a navigation instruction, we believe that it does not necessarily mean that the topic has ended. Giving confirmation feedback (for example, *"ok"*, *"alright"*) may, for example, mean that the speaker giving the feedback is listening and has received the instruction or the comment, but not necessarily that the instruction is understood or the speaker is done discussing the topic. The distinction between feedback meaning "I hear what you are saying" and feedback meaning "I understand and accept what you are saying" is hard to make within these domains. At least, it may be hard to tell whether positive feedback indicates acceptance or not and even if it does, acceptance-feedback may not end the topic.

Within the Interview domain, we assume that a confirmation of a relevant answer is the ending of the topic discussion, since we believe it is more clear in that context (when the speaker has asked a question) that a confirmation of a relevant answer to the question means that the speaker has accepted the answer (although in theory it is possible that this is not always true). Within the other domains we believe it is more common to confirm that, for example, the speaker has heard an instruction or a comment but not necessarily has understood it. Therefore it is not always possible to determine which utterance is ending a topic within these domains. We also have the opposite behaviour - that the instruction and/or comment is not confirmed at all, but they may still be both received, understood and accepted. So, because of the uncertainty factor, we have not annotated any utterances as *topic-end* within the Navigation, Traffic and Other domains.

This also means that it is not possible to make a distinction between *topic-resume* and *topic-reraise* within these domains. Since we do not annotate the ending of topics within these domains, we cannot determine whether the start of a topic that has been discussed earlier is a resumption or a reraising of the topic. Therefore, we annotate these topic switches as *topic-resume/reraise* to mark that the topic has been discussed earlier, as opposed to a topic switch annotated as *topic-begin* which marks a topic switch to a topic that has not been discussed earlier. Consequently, all topic switches annotated as *topic-interrupt* are topic switches from or within the Interview domain, since this is the only domain where it is possible to determine when a topic is ended. There are no utterances made from the Navigation, Traffic or Other domains where the speaker switches to another topic within the same domain.

### 3.6.2 Adjacency pairs

We want to find the best place to interrupt in-vehicle interaction without disturbing the driver. Previous dialogue research suggest that interruption should be avoided within an adjacency pair (Yang et al., 2008), and therefore we want to find out if this holds also for in-vehicle dialogue. We have annotated the corpus with adjacency pairs, but only within the Interview domain. This domain contains one of the tasks the participants were assigned (i.e. interviewing each other), and includes pairs of questions and answers. As mentioned before (see Section 1.4), the interaction is therefore comparable to an interaction with a (form-based) dialogue system where the passenger (who is interviewing) has the role of the system asking for information and the driver (who is being interviewed) has the role of the user providing that information.

Furthermore, only the adjacency pairs that contain the actual interview question and subsequent answer have been annotated. Often, the participants continue to talk about an interview topic after the answer has been given, but this conversation has here been considered to be "small talk" and not necessary for completing the interview task. For example, at one time the passenger asked which country the driver last visited for vacation. The driver answered that he drove to Norway, which completed the adjacency pair. Then the participants discussed the beautiful roads in the Norwegian mountains, but that was considered small talk and hence not a part of the adjacency pair. This type of small talk is not produced to inform the dialogue partner about something that may be essential for the performance of a task (as opposed to, for example, traffic-related comments). Instead, these discussions may aim to, for example, get to know each other better, and are therefore less likely to occur when interacting with a dialogue system which make them less relevant for this study. We want to know what dialogue strategies the participants use when performing the assigned task, and therefore the focus of this study is question-answer pairs rather than the small talk that may follow them.

Adjacency pairs have not been annotated within the Navigation, Traffic and Other domains. Unlike the discussions within the Interview domain, the discussions within the Navigation, Traffic and Other domains do not follow a strict structure. The interview questions are clearly specified which makes it easy to identify each question-answer pair. Within the Navigation domain, the participants do have a set of navigation instructions to follow, but since the instructions are sometimes hard to interpret, the participants discuss back and forth and uses gestures to complement what they are saying. This makes it harder to determine when an adjacency pair starts and ends, and therefore we have not included these discussions in this study. Discussions within the Traffic domain often contain only one dialogue turn, and hence these discussions are not as relevant to study for this purpose. Within the Other domain, the discussions are either similar to those within the Traffic domain, or are considered to be small talk (see above) and hence not as relevant from a dialogue system point of view. For these reasons, we have not included an adjacency pair analysis of the Navigation, Traffic and Other domains within this thesis.

Adjacency pairs have been annotated as:

- *Adjacency pair (question-answer)*: beginning with the utterance where an interview question is asked, ending with the first resolving answer to that question.

The annotations may contain only two turns, the pairs:

(7) **Passenger:** "what is your occupation?
" (Sw. "vad jobbar du med")
**Driver:** "hmi expert is what my business card says
" (Sw. "hmi-expert står det på mitt visitkort")

or several turns if the question is not immediately followed by the answer:

(8) **Passenger:** "star sign" (Sw. "stjärntecken")
**Driver** (interrupting to clarify a navigation instruction): "it's not this one" (Sw. "det är inte den här det")
**Passenger:** "nope" (Sw. "nä")
**Driver:** "nope" (Sw. "nä")
**Passenger:** "star sign" (Sw. "stjärntecken")
**Driver:** "Scorpio" (Sw. "skorpion")

As can be seen, the notion of "adjacency pair" has been stretched a bit. What we are interested in is to find out what happens from the moment where the question is being asked until a relevant answer is given. Therefore, there might be occasions such as in the previous example where the interview question, due to an interruption, is asked twice. In this case, one could argue that the first adjacency pair ("What star sign are you?") is aborted, and then a new adjacency pair is started ("Star sign?") which is completed. However, we have annotated the adjacency pair to start with the first question and end with a resolving answer. By doing so, we get all information we need, since the annotation includes how the pair is interrupted, the cognitive load level at the point of interruption and resumption, how the question is reraised and how it is answered.

### 3.6.3 Sequencing moves

We also want to investigate strategies concerning *how* to interrupt an ongoing conversation. To find out more about this behaviour we have annotated the corpus with sequencing moves.

Before switching topic, we often use discourse markers to signal that a topic switch is about to come (Schriffrin, 1987). In this thesis, we call these discourse markers *sequencing moves*, following Allwood et al. (2004) who use the term "sequencing" to refer to the organisation of a dialogue in meaningful sequences based on content rather than speaker's turn. Each sequence corresponds to a topic within a specific domain. Switching topic means that the speaker is switching from one sequence to another, and a sequencing move signals the topic shift that will follow. Utterances annotated with *sequencing move* are therefore utterances that indicate that the speaker is switching topic.

Sequencing-moves are annotated on two tiers, one for each speaker (*seq-driver* and *seq-passenger*) with three possible values:

- *std-phrase* (sequencing function, standard phrase): a standardized, domain-independent phrase

  - for example, “let’s see”, “where were we”

- *dom-spec* (sequencing function, domain-specific phrase): a phrase that can only be understood within a specific domain

  - for example, “turn right” (navigation instruction), “Wolfmother” (the answer to an interview question)

Domain-specific phrases are classified based on grammatical category according to the following schema:

- DEC: declarative sentence

  - for example, *“you are the guide now”*

- INT: interrogative sentence

  - for example, *“should I turn here”*

- IMP: imperative sentence

  - for example, *“drive towards Linnéplatsen”*

- ANS: *“yes”* or *“no”* answer

- NP: bare noun phrase

  - for example, *“lion”*

- ADVP: bare adverbial phrase

  - for example, *“further into Kaserntorget”*

- INC: inomplete phrase

  - for example, *“do you know when to”*

## 3.7 Results and discussion

In this section we will show the results from our analysis of the DICO corpus regarding interruptions in in-vehicle dialogue. We start with an analysis of interruption strategies during the various types of workload, followed by an analysis of topic switches in relation to adjacency pairs within the Interview domain, and finally we present an analysis of sequencing moves. When it is meaningful, we will indicate statistical significance of results, but in the cases where the data is too sparse we will discuss trends without drawing definite conclusions. As elsewhere in this thesis, we will also illustrate with example interactions from the DICO corpus. In addition, we will occasionally take the liberty of discussing anecdotal data in some detail, especially in cases where we believe this may give ideas for future investigations.

We start by analysing the corpus with regards to the two proposed strategies of interrupting the dialogue during high workload; either to pause the dialogue to let the driver concentrate on the driving task, or to switch topic if we believe it would benefit the driver to get information from another domain.

### 3.7.1 Pausing the dialogue

We want to know if the participants in the DICO study use the strategy of pausing the dialogue when the driving task is more demanding. To do that, we have analysed the corpus with respect to average number of pauses per second for each workload type. We have also looked at the average pause length during each workload type. The term "pause" is here used rather loosely, meaning silent intervals both within and between utterances. In the DICO corpus, very short pauses have been annotated simply as PAUSE (for example, *"take left into Engelbrekt Street and continue PAUSE zero point five kilometers"*) and are not included in this investigation. Included are instead longer pauses, which are annotated on a separate tier. Pauses within the DICO corpus are annotated manually. No threshold was set regarding the length of pauses annotated as PAUSE vs. the pauses annotated on a separate tier. This distinction is instead made subjectively by the annotator. An *utterance* in the DICO corpus is a chunk of talk ended with either a speaker change or a longer pause (annotated on a separate tier). The pause can be followed by a speaker change or a new utterance by the same speaker. Consequently, the longer pauses can be regarded as inter-utterance pauses.

In Example (9), we see two examples on pause annotations within the DICO corpus. As can be seen, we count both pauses between a speaker turn and pauses within one speaker's turn as pauses:

(9) **Passenger:** "it must be this roundabout and the third exit must be a" (Sw. "det måste vara den här rondellen och tredje avfarten är väl ett")

**PAUSE 450ms**

**Driver:** "it's right over there really" (Sw. "det är rakt över där egentligen")

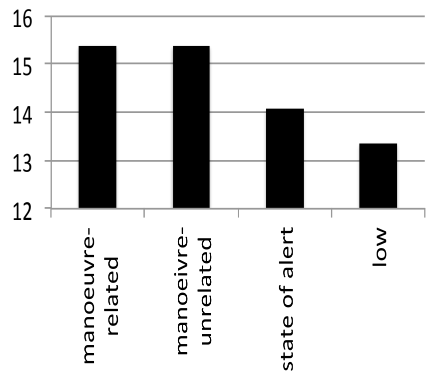
**Passenger:** "yes it must be" (Sw. "ja det måste det vara")

**PAUSE 4400ms**

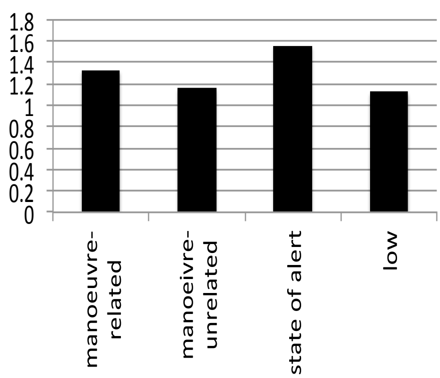
**Passenger:** "it should be Karlavagn Street later" (Sw. "Karlavagnsgatan ska det vara sen")

The actual purpose of the study was not to study pausing behaviour but domain and/or topic change. Because of the rather improvised way pauses are annotated, this investigation only aims to get a preliminary idea of whether the participants make more and/or longer pauses while making a demanding driving manoeuvre than otherwise. The setup of the study, where the participants were encouraged to talk to each other as much as possible, may have had an effect on the pausing behaviour. To get a better understanding of pausing behaviour during high and low cognitive workload, further studies are required, but we believe that looking at this behaviour in the DICO corpus may be a useful starting point.

As mentioned in Section 3.3, earlier studies have shown an increase in silent pauses during high workload. We have therefore analysed the DICO material regarding number of (long) pauses per minute during each workload type to see if the pausing behaviour differs during the varying types of workload and in what way. We have also calculated average pause length during each condition. Figure 3.1 shows the results.




(a) Number of pauses per minute




(b) Average pause length in seconds

Figure 3.1: Pausing behaviour.

The number of pauses per minute is highest, about 15.5 pauses per minute, when the workload is high – during manoeuvre-related and manoeuvre-unrelated workload. During low workload the number of pauses per second is lowest, about 13.5 pauses per minute. During state of alert the number of silent pauses is about 14 per minute. A Z test shows that pause frequency is higher during TDT episodes (i.e. manoeuvre-related and manoeuvre-unrelated workload) than during non-TDT episodes (i.e. state of alert and low workload) ($Z = 2.4, p = 0.02$). However, there is no significant difference in pause frequency between IDIS episodes (i.e. manoeuvre-related workload and state of alert) and non-IDIS episodes (i.e. manoeuvre-unrelated workload and state of alert) ($Z = 1.4, p = 0.18$). This result shows that high workload affects the frequency of pauses more than a demanding driving manoeuvre does.

The average pause length is longest when during state of alert, i.e. when IDIS signals that the driver is making a demanding driving manoeuvre, and shortest during low workload. A GLMM (Generalized Linear Mixed Model) test was conducted, to find out if there is a significant difference in pause length when TDT signals high workload and/or when IDIS signals that the driver is making a demanding driving manoeuvre. A GLMM test is analogous to the ANOVA test but allows for testing data which is not normally distributed. We found that the pauses are significantly longer when IDIS is signalling compared to when it is not signalling ($F_{(1,2268)} = 4.016, p < 0.05$). the TDT is signalling and when neither TDT nor IDIS are signalling, with a $p$-value < .05. There is no significant difference in pause length when the TDT is signalling compared to when it is not. This means that performing a demanding driving manoeuvre affects the pause length more than high workload.

These results shows that the pausing behaviour changes during various workload conditions, with more pauses when the driver's workload is high and longer pauses when the driving task is more demanding. Pausing the interaction with an in-vehicle dialogue system during manoeuvre-related workload (when the workload is high while the driver is making a demanding driving manoeuvre) may therefore be a good strategy, provided that the pausing behaviour is beneficial for the driver. It seems reasonable to assume that, by pausing the dialogue during manoeuvre-related workload, a dialogue system could avoid adding to the driver's workload. Furthermore, it may be counterproductive to give or ask for information when the driver's workload is already high, and the focus needs to be on the driving task rather than the dialogue task. It needs to be investigated in future studies whether the strategy has an effect on the workload level.

### 3.7.2 Switching topic

The next proposed strategy is to switch topic, in cases where we believe the driver would benefit from getting information that a dialogue system could give. Our review of earlier research in Section 3.3 shows that users, if they are able to choose when to interrupt an ongoing task, usually try to interrupt when the workload is low. However, in the in-vehicle environment it is not always possible to choose when to interrupt. For example, if the information that is needed is time-critical. It may even be the case that topic switches that aim to give driving-related information are more frequent during high workload, if the workload is related to the driving task. Therefore, we expect to see a different result in the DICO corpus compared to earlier studies, and expect more interruptions to the driving-related domains (i.e. the Navigation and Traffic domains) during manoeuvre-related workload and state of alert than during low workload.

We have analysed the corpus with respect to topic switches to each domain, during the different conditions measured by TDT and IDIS. We look at topic switches that occur after an ended topic, and topic switches that occur at an interruption. A comparison of these annotations tells us whether the participants, during each condition, usually end the current topic properly before switching to the next, or if they for some reason decide to leave the current topic before ending it properly and switch to another one.

### Topic switches to the Navigation domain

From a dialogue system point of view, the Navigation domain is the domain of discussion (of those included in the DICO study) that is most interesting for us regarding topic switches interrupting an ongoing discussion. If a dialogue system would take the initiative to interrupt an ongoing interaction it would most likely be to give time-critical information, such as a navigation instruction. We have analysed topic switches that are made from the Interview domain to the Navigation domain, for two reasons.

First, the Interview and Navigation domains are the two task-related domains, and therefore the discussions within these domains are more similar to a dialogue system interaction (where the user, i.e. the driver, would use the dialogue system to perform a task or ask for information), than the discussions within the task-unrelated Traffic and Other domains[1].

Second, it is only within the Interview domain that *topic-end* utterances are annotated, as described in Section 3.6. Therefore it is only when switching from the Interview domain that we can determine whether the speaker is interrupting a topic or not. The hypothesis is that topic switches that interrupt an ongoing topic discussion are more time-critical than topic switches following an ended topic, and that interrupting the dialogue therefore may be a strategy that is used to ease the driving task by giving useful information and prevent, or shorten the time of, high workload. We therefore believe that interruptions are more frequent during manoeuvre-related workload and state of alert, than during low workload.

We start by looking at the distribution of topic switches from the Interview domain to the Navigation domain, see Figure 3.2.

In total, there are 43 topic switches occurring immediately after

[1] However, one could imagine that a situation-aware dialogue system would be able to inform the driver about, for example, road blocks or other events that may affect the driving task

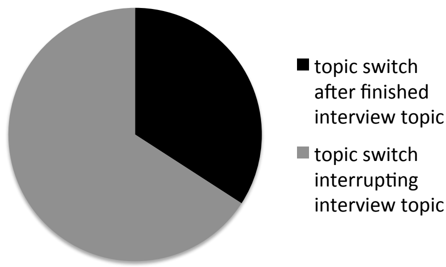



Figure 3.2: Distribution of topic switches from the Interview to the Navigation domain.

a finished interview topic and 83 topics switches interrupting an ongoing interview topic.

We see that when switching from the Interview to the Navigation domain, the speakers interrupt an ongoing interview topic more often than they finish the topic before switching domain. One reason for this is probably that the information that is given/asked for is time-critical, and hence the speaker may not have time to finish the ongoing discussion before switching domain. The participants may, for example, be close to where the next navigation instruction needs to be given (for example, a lane change or a turn at a crossing) and therefore either the driver or the passenger interrupts to clarify what to do next. Another reason may be that the participants are engaged, and possibly distracted, by the conversation and do not plan their utterances with respect to when the next navigation instruction needs to be given.

To learn more about this behaviour we have analysed the topic switches in more detail, to see if we can find a possible reason for interrupting. To do that, we have looked at topic switches during each workload type. We also want to know if the speaker's role influences the behaviour, and therefore we will compare the types of topic switches made by the drivers to those made by the passengers to see if there is a difference in behaviour. Figure 3.3 shows the distribution of topic switches from the Interview to the Navigation domain during each condition.

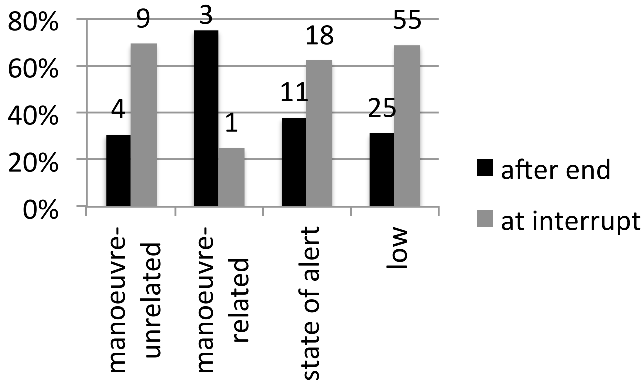


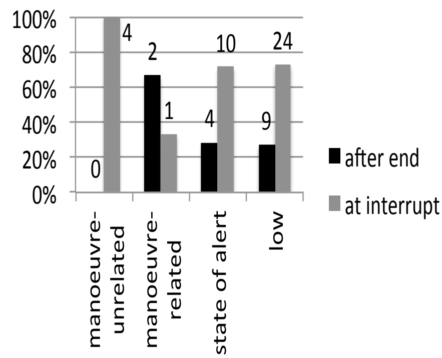

Figure 3.3: Distribution of topic switches made from the Interview domain to the Navigation domain, including number of occurancies during each condition.

The pattern regarding when to switch topic is similar during the workload types, with topic switches at an interruption being more common than topic switches after a finished topic. The proportion of *topic-interrupts* is largest during manoeuvre-unrelated and low workload (about 69% during each condition), and smallest during manoeuvre-related workload (25%). Although the data is extremely sparse, arguable so much as to be merely anecdotal in nature, if we were to make anything of these results they could be taken to indicate that the participants interrupt more during low workload than during manoeuvre-related workload, similar to the result from earlier studies in other environments (see Section 3.3). Our hypothesis was, on the contrary, that the speakers would interrupt the ongoing discussion most frequently when the information that needs to be given is time-critical, and that this would be the case more often during manoeuvre-related workload (when both TDT and IDIS is signalling) and during state of alert (when IDIS alone is signalling), than during low workload. Because of the sparse data we have not tried to compute significance. These behaviours should be investigated further and in a larger study, to confirm that workload level is a factor that should be taken into account when deciding upon system-initiated interruptions.

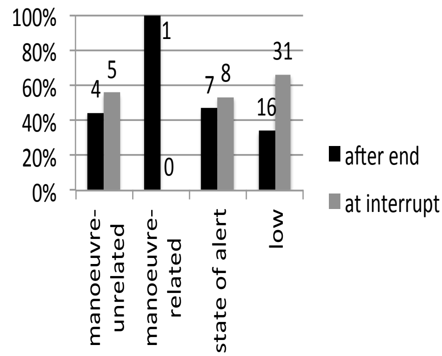
Possibly, there is a difference in behaviour depending on the role of the speaker. The participants had restrictions regarding the navigation instructions, allowing the passenger to read only one instruction at a time (see Section 1.4). Still, they had the advantage of knowing the next instruction in advance, and hence could share the instruction when they found it fitting with respect to time constraints (the driver needs to know the next manoeuvre in time to be able to perform it) and the driver's ability to receive it (with respect to, for example, workload level). We therefore analysed the driver's and the passenger's behaviour, separately. The results are shown in Figure 3.4.

When comparing topic switches made by the drivers with those made by the passengers it seems like the drivers are behaving pretty much the same regardless of workload level with more topic switches at an interruption than after an ended topic. The only exception is during manoeuvre-related workload, but only four (4) topic switches are made during this condition; three of them after an ended topic discussion (twice by the driver and once by the passenger).

However, if we are to interpret these results despite their sparsity, the passengers seem to try to adjust their behaviour depending on the driver's workload level. There are more topic switches at an interruption than after an ended topic during low workload. During manoeuvre-related and manoeuvre-unrelated workload and state of alert there is about the same distribution of topic switches at an interrupt as there is after an ended topic (with only one topic switch made during manoeuvre-related workload). This result may indicate that the passengers are taking the driver's ability to receive information into account when giving navigation instructions, by interrupting less and thereby trying to share instructions when the driver is (more) ready to receive the information although it may not always be possible due to time restrictions (for example, if the participants are close to a crossing where the driver needs to turn). The drivers do not seem to adjust their behaviour to the same degree, which may be due to the fact that the drivers are the ones who need the information and therefore ask for it when necessary regardless of workload level




(a) Topic switches made by the driver.




(b) Topic switches made by the passenger.

Figure 3.4: Topic switches made by the driver and the passenger, respectively.

and regardless of whether there is an ongoing topic discussion or not.

### Topic switches to the Interview domain

As explained in Section 3.6, we have not tried to determine whether a topic switch from the Navigation, Traffic or Other domains is interrupting an ongoing discussion or not. When studying interrupting behaviour regarding interview topics we can therefore only look at topic switches within the domain. Our hypothesis is that topic switches after an interruption are more time-critical than interruptions after an ended topic. However, topic switches to the Interview domain are not time-critical since the participants had no time constraints regarding this task and no restrictions regarding how many questions they should answer. The lack of such constraints or restrictions makes it hard to see a motivation for interrupting an interview topic to switch to another topic within the domain, unless the speaker who interrupts is still occupied with an earlier topic and wants to resume/reraise that topic. Therefore, we expect few interruptions to this domain, regardless of workload level.

As we suspected, there are few interruptions within the Interview domain. As a matter of fact, it happens only once that a speaker interrupts one interview topic to switch to another topic within the same domain. That happens when the participants are choosing between the interview questions to decide which one to discuss next, and the passenger first selects a question. The driver does not respond to that but instead interrupts and proposes another question instead. This happens during low workload, which may be one reason for the interruption. Instead of accepting the question selected by the passenger and starting to answer it, the driver proposes a topic they had mentioned, but not discussed, before. At the time, they are driving straight on with no surrounding traffic, but right before the passenger asks the question they are stopping at a pedestrian crossing. Therefore, when deciding which interview question to choose they are standing still and waiting for the pedestrians to pass. As a result, the driver did not have to pay attention to the manoeuvreing of the car at the moment, but could instead put more focus on finding an interesting question to discuss.

**Topic switches to the Traffic domain**

The participants are driving in the inner city of Gothenburg most of the time, and hence the traffic situation is often intense and a sudden event that calls for the participants attention may happen at any time. Therefore, we expected the participants, and especially the driver, to interrupt the Interview discussion rather frequently to comment about the traffic situation, and by that signal why they cannot discuss the interview question at the time.

Surprisingly, the participants interrupt an ongoing interview discussion only three times to comment on the traffic situation. Once it is the driver who interrupts, and twice the passenger. The fact that the passenger interrupts at all is rather interesting, since all other topic switches to the Traffic domain from other domains are made by the driver, and none by the passenger.

When the passenger interrupts an interview topic the reason seems to be to make the driver aware of the traffic situation, maybe because the passenger is not sure if the driver is paying enough attention to the traffic while answering the interview question. One of the interruptions made by the passenger occurs when the driver is following close behind the car in front of them, approaching a pedestrian crossing. The driver is talking about his favourite movie while moving on to the spot where the pedestrian crossing is. Two pedestrians are starting to cross the street when the car is moving towards them, the passenger then utters *"oops"* and the driver stops to let the pedestrians pass. It is unclear whether the driver had seen them before the passenger says this or even if he notices the comment or not, since he is not responding to the comment but continues to talk about the movie.

The next interruption occurs when the driver is switching to the left lane while answering an interview question, and the passenger looks to the left and says *"oops, there's a lot of traffic here"*. It is not possible here either to know if the driver is aware of the traffic situation and/or the passenger's comment, since he continues to talk about the interview question without pausing or commenting on the traffic himself.

When the driver interrupts, the passenger has just asked an interview question which the driver first starts to answer, interrupts to switch to the Traffic domain while shifting lane, and then ask for navigation directions. On this occasion it seems like the comment is made to signal that the driver is too occupied by the driving task at the moment to be able to answer.

These anecdotical observations indicate that topic switches to a domain that is related to the primary task is something that should be investigated further. In Section 3.1, we have suggested that the drivers use the Traffic domain as a marker of high workload, i.e. that they interrupt the ongoing discussion to switch to the Traffic domain when the driving task is demanding. The fact that there are so few interruptions to the Traffic domain from the Interview domain is a bit surprising and studies of this behaviour should be included in future user studies. If this behaviour is significant, we need to know the reason for this. Perhaps the urge to answer a question makes the driver less willing to interrupt even though something happens that otherwise would have attracted their attention, but if that means that they do not notice the incident or just not comment upon it we do not know.

When the passengers interrupt to switch to this domain from the Interview domain, the driver does not seem to notice the comment as they do not, for example, continue the traffic discussion by commenting or giving feedback. This may be either because they do not consider the incident important enough to comment about, or because they are so occupied by the ongoing conversation that they do not perceive the comment. This is something that should be investigated further, and be taken into account when developing situation-aware systems that warn the driver about possible hazards.

### Topic switches to the Other domain

When annotating the DICO corpus, we noted that topic switches to the Other domain seem to have two purposes; either to comment upon a sudden event (similar to comments within the Traffic domain) or to chat about subjects other than those related to the study. We expect that any interruption to this domain from the Interview domain belongs to the first category, to indicate high workload.

When analysing the interruption utterances, we see that the participants do switch to this domain to comment about sudden and unexpected events. Similar to interruptions to the Traffic domain, this happens three times, each time the driver being the one who interrupts the conversation. Once, the driver's mobile phone starts ringing, and the driver interrupts himself to comment upon that. The workload is low. Twice, the interruption is related to the TDT task. At one time the driver is thinking about what to answer to an interview question, and finds it hard to come up with an answer. The car is standing still at a traffic light, and when the traffic light switches to green the driver starts laughing and says that she accidently pushed the TDT button when she saw the traffic light switching. Apparently, the traffic light functioned as a trigger in the same way as the TDT buzzer, making the driver believe it was time to push the TDT button. The fact that the driver is not able to discriminate between the triggers may be a sign that the driver is being distracted by the interview discussion, although the TDT did not signal high workload at the time. When the driver makes the comment the car has started moving and there is a state of alert. The last topic switch to this domain occurs when the driver suddenly notices that the wrist band attaching the TDT equipment to the arm is starting to come off. The workload is low when the driver makes this comment.

These interruptions take place when the driver gets distracted by something unexpected that involves the driver rather than the passenger. The comments seem to be used to signal that something happened that makes it hard for the driver to continue the ongoing conversation.

To conclude, our investigation of topic switches interrupting discussions within the Interview domain indicated a tendency that the speakers interrupt more often than they wait until the ongoing topic is finished when switching to the Navigation domain. This in turn indicates that interrupting to switch domain may be a strategy that is used when the information that is asked for or given is more time-critical. However, is seems that the speakers try to avoid interruptions during high workload and when the driving task is demanding, and instead the participants interrupt more during low workload. Because of the sparse data we cannot tell if there is a difference in behaviour between the speakers, even though the result indicates that the passenger is more cautious when it comes to interruptions during high workload and state of alert than the driver. If so, and if we assume that this behaviour is beneficial to the driver, it may be preferable to avoid system-initiated interruptions as far as possible, and let the driver take the initiative to interrupt. Possibly, system-initiated interruptions should be allowed during low workload if it is absolutely necessary for the system to give the information immediately instead of waiting until the ongoing interaction is finished.

### 3.7.3 When to interrupt

Next we will analyse the corpus with respect to *when* to interrupt an ongoing conversation. Our hypothesis, based on earlier research (see Section 3.3) is that the participants will avoid interrupting a discussion within an adjacency pair, and instead wait until the pair is finished before interrupting. As mentioned in Section 3.6, adjacency pairs have only been annotated within the Interview domain. The type of adjacency pairs that has been annotated is *question-answer* pairs, where the first part of the pair is an interview question and the last part is a relevant answer to the question. Sometimes the topic discussion ends with the relevant answer (and hence ends with the utterance containing the last part of the adjacency pair), as in Example (10):

(10) **Passenger:** “do you have a boat” (Sw. “har du båt”)

[*domain of discussion = interview, adjacency pair-first part*]

**Driver:** “no no boat” (Sw. “nej ingen båt”)

[*domain of discussion = interview, adjacency pair-last part, topic-end*]

**Passenger:** “what about a caravan” (Sw. “husvagn då”)

[*domain of discussion = interview, adjacency pair-first part*]

However, often the discussion continues after the answer has been given, when the participants want to discuss the answer or to elaborate it a bit further, as in Example (11):

(11) **Passenger:** “we can take an easier question favourite candy” (Sw. “vi kan ta en enklare fråga favoritgodis”) [*domain of discussion = interview, adjacency pair-first part*]

**Driver:** “favourite candy yes that must be eh PAUSE salty liquorice” (Sw. “favoritgodis ja det är nog eh PAUSE salt lakrits”)

[*adjacency pair-last part*]

**Passenger:** “mhm” (Sw. “mhm”)

**Driver:** “then it’s like it doesn’t matter what it is as long as it’s salty liquorice I think” (Sw. “sen är det liksom det spelar ingen roll vad det är för bara det är salt lakrits tror jag”)

**Passenger:** “okay mm” (Sw. “okej mm”)

[*domain of discussion = interview, topic-end*]

In the latter case, the interview topic could possibly be interrupted after the adjacency pair is complete but before the topic is ended. In this investigation, we calculated the distribution of interruptions within an adjacency pair and compared to the distribution of interruptions outside a pair. Because of the sparse data we do not analyse these interruptions in relation to workload.

Since an adjacency pair is a subset of a topic discussion, we needed to normalize the data to be able to compare the number of interruptions within and outside an adjacency pair. We therefore calculated the number of possible *TRP's* within the Interview domain. A TRP, *Transition Relevance Place*, is a point in the discussion where a speaker change may take place (Sacks et al., 1974). We calculated TRP's by counting the number of utterances - 1, meaning that we see every gap between utterances as a potential TRP [2]. We then counted the number of interruptions within and outside an adjacency pair, and calculated the distribution of interruptions within and outside a pair. The result is shown in Figure 3.15. The speakers make an almost equal amount of utterances (the passengers make 54% of all uttterances and the drivers make 46%), and therefore we do not distinguish between TRP's after each speaker's utterances. Instead, we calculate the distribution based on the total number of TRP's.

As shown in Figure 3.15, interruptions are more likely outside adjacency pairs ($\chi^2_{(1)} = 17.03, p < 0.01$). In total, the Interview domain contains 166 topic discussions containing at least one adjacency pair. These topics contain 2578 TRPs, where a speaker and/or domain change could possibly be made. 94 times a speaker interrupts the ongoing discussion to switch domain, which means that on average there is a domain switch at 3,6% of all TRPs. There are 1020 TRPs within an adjacency pair, about 40% of the total number of TRPs. At 17 of these an interruption occurs, which means that at 1,7% of the TRPs witihin an adjacency pair there is a domain switch. Outside a pair there are 1558 TRPs, at

[2] Admittedly, this is a rather crude way of defining a TRP since it rules out possible TRPs within utterances, but we will assume that it is sufficient for the purpose of this study.

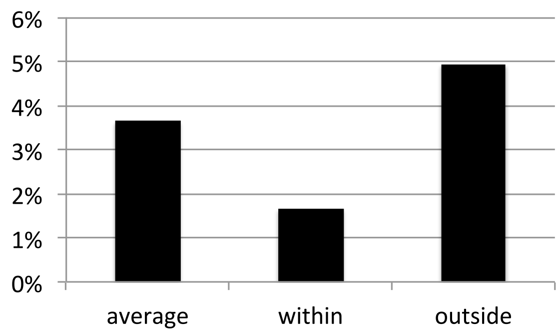



Figure 3.5: Distribution of interruptions in relation to adjacency pairs.

77 (4,9%) of these there is a domain switch.

The result shows that if the participants interrupt an ongoing conversation within the Interview domain, they more often interrupt outside an adjacency pair than within a pair. When looking at interruptions made by each speaker, we see that the behaviour is similar for both speakers although more prominent for the driver, see Figure 3.6.

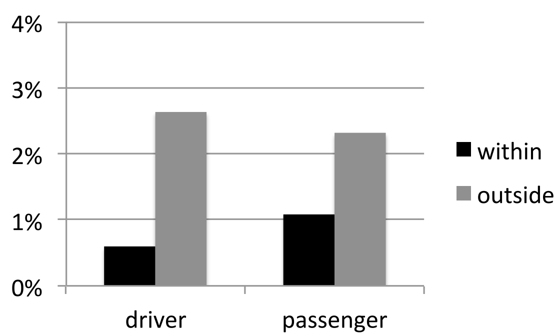



Figure 3.6: Driver and passenger interruptions within and outside an adjacency pair.

The drivers interrupt an interview topic 6 times within an adjacency pair, which means that the drivers interrupt at 0.6% of the possible TRP's that occur within a pair. Outside a pair the drivers interrupt 41 times, which is at 2.6% of the possible TRP's. The passengers interrupt 11 times within a pair, which is at about 1% of the possible TRP's. Outside a pair the passengers interrupt 36 times, which is at 2.3% of the possible TRP's.

The driver's role as the one being interviewed, and hence having to provide the last part of the adjacency pair, may be an explanation why they so rarely interrupt within a pair. Example (12) shows how the driver answers a question and then immediately switches domain:

(12) [WORKLOAD = LOW]

**Passenger:** "the last book you read" (Sw. "senast lästa bok")

[*domain of discussion = interview, adjacency pair-first part*]

**Driver:** "the last book I read" (Sw. "min senast lästa bok")

**Passenger:** "mm" (Sw. "mm")

**Driver:** "last book read or what you are reading right now" (Sw. "senast lästa bok eller vad man håller på och läser nu")

**Passenger:** "[yes] or PAUSE take what you are reading now" (Sw. "[ja] eller PAUSE ta det du läser nu då")

**Driver:** "[oops]" (Sw. "[oj]")

[*domain of discussion = traffic, topic-begin, topic-interrupt*]

**Driver:** "what I am reading now is Sophie's World" (Sw. "det jag läser nu är Sofies Värld")

[*domain of discussion = interview, topic-resume, adjacency pair-last part*]

**Driver:** "yes where do you want me to go" (Sw. "ja var vill du att jag ska köra")

[*domain of discussion = navigation, topic-begin, topic-interrupt*]

*slowing down while approaching a fork in the road*
[WORKLOAD = TDT]

**Passenger:** "straight ahead straight ahead" (Sw. "rakt fram rakt fram")

**Driver:** "straight ahead let's do that" (Sw. "rakt fram då gör vi det")

*passing the fork*
[WORKLOAD = LOW]

**Passenger:** "yes The Enemy's Enemy was the last one I read" (Sw. "ja Fiendens Fiende var den senaste jag läste")

[*domain of discussion = interview, topic-resume*]

While thinking about what to answer to the interview question, they are approaching a fork in the road. The driver is slowing down, and before he answers the question he almost stops right before it is necessary to decide which lane to choose. He still answers the question before asking where to go, although it would probably have been more convenient, from a driving point of view, to ask where to go next *before* answering the question. The reason why he still chooses to answer the question first may be that his mind is occupied by the question. Once they had decided that it was ok to mention the book he is currently reading it may be easier and less demanding to give a quick answer first, instead of interrupting the discussion before the answer had been given.

Interrupting within an adjacency pair, when the driver is focused on the discussion rather than the driving task, may not be successful even though the driver seemingly is listening to the interrupting message:

(13) *Driving straight ahead on a four-lane street, using the right lane.*
[WORKLOAD = LOW]

**Passenger**: "What do you li PAUSE fruit do you like PAUSE the most tasty fruit" (Sw. "vad tycker du f PAUSE om för frukt PAUSE godaste frukten")

[*domain of discussion = interview, topic-begin, adjacency pair-first part*]

**Driver**: "eh PAUSE yes that would be PAUSE let's see PAUSE mm PAUSE ah" (Sw. "ehm PAUSE ja de kan va PAUSE nu ska vi se PAUSE mm PAUSE äsch")

[*domain of discussion = interview*]

**Passenger:** "you should [follow] this" (Sw. "du ska [följa] den här")

[*domain of discussion = navigation, topic-begin, topic-interrupt*]

**Driver:** "[yes]" (Sw. "[ja]")

**Passenger:** "and turn right at the next" (Sw. "åsså svänga höger nästa")

[*domain of discussion = navigation*]

**Driver:** "yes" (Sw. "ja")

[*domain of discussion = navigation*]

**Driver:** "let's take pineapple PAUSE that's a tasty f" (Sw. "vi tar ananas PAUSE det är en väldigt god f")

[*domain of discussion = interview, topic-resume, adjacency pair-last part*]

> *Slowing down while passing a pedestrian crossing*
> [WORKLOAD = MANOEUVRE-RELATED]

**Passenger:** "pineapple PAUSE yes that is really tasty PAUSE it really is actually" (Sw. "ananas PAUSE ja det är väldigt gott PAUSE det är det ju faktiskt")

[*domain-of-discussion = interview, topic-end*]

> *The lanes are marked with arrows. The one in the left lane points straight ahead and the one in*

*the right lane points to the right, meaning that the right lane must turn right*
[WORKLOAD = MANOEUVRE-RELATED]

**Driver:** "I didn't listen PAUSE should I turn or s" (Sw. "nu lyssna jag inte PAUSE skulle jag svänga eller s")

[*domain of discussion = navigation, topic-resume/reraise*]

[WORKLOAD = MANOEUVRE-UNRELATED]

**Passenger:** "yes turn here" (Sw. "ja sväng här")

**Driver:** "right" (Sw. "höger")

**Passenger:** "yes right" (Sw. "ja höger")

*The driver is pointing to the right, and stays in the right lane as the lanes are divided.*
[WORKLOAD = MANOEUVRE-UNRELATED]

The driver is engaged in the interview discussion and although he is giving feedback when receiving the navigation instruction (by answering "yes" a couple of times), he is probably still occupied by the interview question and does not seem to process what he is hearing. Throughout the interaction the workload varies between manoeuvre-related and manoeuvre-unrelated workload. The passenger interrupts to give the navigation instruction after the first part of the adjacency pair (*"what fruit do you like"*) but before the second part (the answer to the question). The driver seems to be more focused on finding the answer to the question than on the navigation task. When the passenger gives the instruction they are driving in a curve and they still have some time left before they have to make a decision, which may be a reason why the driver seemingly is prioritizing the interview task over the navigation task at the moment. The driver answers the question right before it is time to decide whether to change lane or not. It is uncertain if the driver is reraising the navigation discussion at this point in the conversation (after the interview topic is ended) because the interview task thereby is finished and the driver can focus on the navigation task, or because the potential upcoming lane shift is triggering the topic switch.

In this example, the high workload may contribute to the driver's inability to perceive and process the instruction, as the focus was on the interview task rather than the navigation task at the moment. Therefore, he may not have been mentally prepared to receive any other information at the moment. Had the passenger waited until after the answer was given, the driver may have been able to shift focus from the interview task to the navigation task and grasp the instruction immediately. So, here the passenger's interruption was counterproductive.

As mentioned, the participants' different roles may have influenced their behaviour. The driver was the one being interviewed, but was also the one in charge of the driving task. The passenger was the interviewer, but was also in charge of the navigation task. That may explain why we see a different scenario when the driver interrupts an adjacency pair:

(14) *Driving in a sharp left curve, towards a crossing*
[WORKLOAD = STATE OF ALERT]

**Passenger:** "do you have a boat or a caravan or something like that" (Sw. "har du PAUSE båt eller husvagn eller nåt sånt där")

[*domain of discussion = interview, topic-begin, adjacency pair-first part*]

*The driver points at a road sign*

**Driver:** "you said downtown left here" (Sw. "du sa centrum vänster här")

[*domain of discussion = navigation, topic-resume/reraise, topic-interrupt*]

**Passenger:** "downtown yes yep" (Sw. "centrum ja jajemen")

*The driver switches lane to the left lane*

**Driver:** "no I have a house that's enough" (Sw. "nej PAUSE jag har hus det räcker")

[*domain of discussion = interview, topic-resume, adjacency pair-last part*]

In this example, the passenger had told the driver earlier to follow the signs leading towards downtown. When asking the interview question they were driving in a sharp curve, and when finishing the curve they came upon a crossing with four lanes. The driver had to decide which lane to choose, so instead of answering the question he interrupts to clarify the navigation instruction he had received earlier. Although he had to interrupt the ongoing interview discussion, he apparently heard and understood the question, as he did not have to ask to get it repeated but instead could answer it right after the interrupting topic was finished. As opposed to the driver in Example (4), the passenger was able to shift focus from the interview task to the navigation task when the driver switched topic and domain. Possibly, the fact that the passenger had done his part by providing the first part of the adjacency pair, and hence was not the one who should provide the last part, made it easier for him to shift focus than it had been for the driver.

As mentioned in Chaper 3.3, earlier research suggest that interrupting an adjacency pair is often avoided due to social conventions. It is considered impolite not to give the turn to the dialogue partner after giving the first part of a pair, or (for the dialogue partner) to ignore to give the second part and instead change topic. Shyrokov et al. (2007) also suggest that it is less cognitively demanding to resume an interrupted topic if the adjacency pair has been completed before the interruption. Our investigation shows that also in the in-vehicle environment the speakers avoid interrupting an adjacency pair. Our anecdotal examples indicate that interrupting an adjacency pair, where the driver is the one who is expected to finish the pair, may be a bad strategy from a cognitive workload point of view. Not only because it is more cognitively demanding to resume an interrupted pair, but we could also see that interruptions within a pair may make it harder to comprehend the new information that is given. If the driver is not mentally prepared to switch domain and receive new information, it may be hard to grasp the information. If the driver takes the initiative to interrupt, s/he is the one asking for the information and in those cases there is a better chance that the domain switch will be successful. This result indicates that avoiding system-initiated interruptions within an adjacency pair, and instead letting the system make a domain shift either before a question is asked or after the answer is given, may be a good strategy to avoid dialogue-induced cognitive workload.

The results presented in this section indicate that the participants, if they need to interrupt an Interview topic, try to avoid interruptions within an adjacency pair. It seems like domain shifts to the Navigation and Interview domains are usually planned ahead to a higher degree than domain shifts to the Traffic and Other domains. 85% of the interruptions to the Navigation domain occur outside an adjacency pair, which may be compared to the average 60% of all interruptions occuring outside an adjacency pair. Due to the sparse data we do not know if the difference is significant, but it is something that should be investigated further. If there is a significant difference in behaviour, it may be a strategy that is used to minimize cognitive workload and maximize the chance that the dialogue partner can receive the information. It remains to be investigated if this is a good strategy, and if so it may also be a useful strategy for an in-vehicle dialogue system. As far as possible, system utterances should be planned to make sure that the driver is mentally prepared to receive the information, and we believe that a good strategy would be to let the system consider both the driver's cognitive workload level and the urgency of the information before interrupting. If the driver's mind is occupied with something else, it can be hard to absorb and interpret information in a correct way and interrupting, whether to pause or to switch topic, at a bad time may therefore be a safety risk if the interruption itself causes an even increased workload level.

### 3.7.4 How to switch topic

Now that we know a little bit more about both the reasons for interruption and when to interrupt, we need learn more about *how* to interrupt an ongoing conversation to switch to another topic [3]. As mentioned in Section 3.5, we believe that topic switches that are made during high workload are more often spontaneous while they are more often planned ahead during low workload. For example, in Section 2.7.1 we learned that topic switches to the Traffic domain are usually made during high workload and may serve to signal that the speaker's workload is high, and hence are not planned ahead but instead forced by the situation.

We have analysed the corpus with respect to type of sequencing move within each domain and during the various types of workload, to see if and how the behaviour differs. First of all, we need to know more about the two types of sequencing moves we are interested in; *standard phrases* and *domain specific phrases*(see Section 3.6). All in all, the drivers make 187 sequencing moves and the passengers make 283. The most common phrase is a domain specific phrase, 79% of the sequencing moves are domain specific and 21% are standard phrases. Table 3.1 and 3.2 shows the most common standard phrases used by the drivers and the passengers, respectively.

| | Interview | Navigation | Traffic | Other |
|---|---|---|---|---|
| oops (*oj*) | 0 | 1 | 8 | 3 |
| alright (*jaha*) | 6 | 4 | 0 | 0 |
| let's see (*ska vi se*) | 2 | 5 | 0 | 0 |

Table 3.1: Standard phrases for driver sequencing utterances

*"Oops"* is the most common sequencing word for the drivers, used when switching to the Traffic or Other domains. It is used when there is a sudden traffic incident, or when something unexpected happens that in some way concerns the speaker. For example, it was uttered when the TDT buzzer almost fell of a

[3] Parts of the results presented in this section have been published earlier (Villing et al., 2008)

| | Interview | Navigation | Traffic | Other |
|---|---|---|---|---|
| alright (*jaha*) | 6 | 1 | 0 | 2 |
| let's see (*ska vi se*) | 7 | 9 | 0 | 0 |
| okay | 4 | 1 | 0 | 0 |

Table 3.2: Standard phrases for passenger sequencing utterances

driver's hand (annotated as a sequencing move to the Other domain), and when switching to the Traffic domain when the driver accidently pushed the brake pedal instead of the gas pedal. The phrase was only used once by a passenger, it happened when the passenger dropped the mobile phone that was used for keeping track on time to know when to switch driver/passenger roles. Hence, it is used by the passenger when the incident concerned a task managed by the passenger, in the same way as it is used by the driver when the incident concerns a driver-related task.

The Swedish phrases corresponding to *"Let's see"* and *"alright"* are commonly used by both speakers. The phrases are usually used to switch to the task-related Interview and Navigation domains.

We now move on to the domain-specific phrases. The domains included in the DICO study may be categorized as task-related (including the Interview and Navigation domains) versus task-unrelated (including the Traffic and Other domains) or driving-related (including the Navigation and Traffic domains) versus driving-unrelated (including the Interview and Other domains). The different natures of the domains (driving-related vs. driving-unrelated, task-related vs. task-unrelated) may affect how the speakers express themselves when switching to the domains. To see if the behaviours differ depending on domain and possibly get a hint on why it differs, we have looked at each domain separately. We calculated the distribution of utterances over the grammatical categories for each domain, and separated the driver and the passenger utterances to see if the speaker's role affects how they express themselves when switching domain. The distribution of domain-specific phrases within the Interview domain can be seen in Figure 3.7.

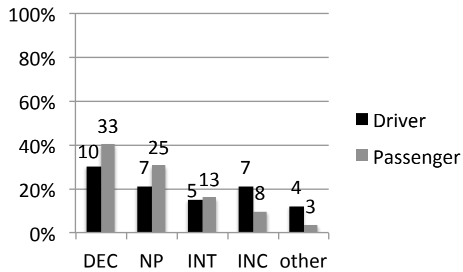



Figure 3.7: Domain-specific phrases when switching to the Interview domain.

Both speakers use mostly *declarative*, *noun*, *interrogative* and *incomplete* phrases, and also a few *verb*, *answer* and *imperative* sequencing phrases. Within the Interview domain, declarative phrases are typically used to either resume an earlier raised question, for example, "*I was supposed to name a favourite actor*" (Sw. "*favoritskådis skulle jag ju säga*"), or raise a question without actually using an interrogative phrase, for example, "*...and at the same time I want you to tell me the historical person you admire the most*" (Sw. "*...och samtidigt ska du säga den historiska person som du beundrar mest*"). Incomplete sequencing phrases are usually uttered when the driver resumes to an earlier question (for example, "*chocolate is incredibly tasty and*" (Sw. "*choklad är ju otroligt gott och*")), to remind the passenger of the topic even though they do not yet have the full answer to the question. Noun phrases are usually used by the passengers to suggest possible topics to discuss (for example, "*hobbies, books, movie*" (Sw. "*fritidsintressen, böcker, film*")) or for both of them to refer to a previously discussed topic by naming a key phrase within the topic (for example, *Wolfmother*, referring to a heavy metal band).

Figure 3.8 shows the distribution of domain-specific phrases when switching to the Navigation domain.

When switching to the Navigation domain, *declarative* and *interrogative* phrases are most common, followed by *incomplete* and *imperative* phrases and a few *noun*, *verb*, *answer* and *adverbial* phrases.

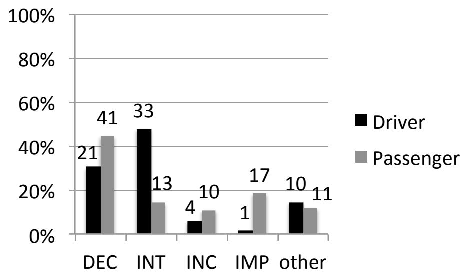



Figure 3.8: Domain-specific phrases when switching to the Navigation domain.

Declarative phrases are often used by the passengers to give an instruction (for example, *"we continue straight ahead here"* (Sw. *"vi fortsätter rakt fram"*), *"this is where we should turn"* (Sw. *"här ska vi svänga"*)), and by both participants when they have discussed a navigation instruction and later resume/reraise that discussion when looking for the spot mentioned in the instruction. For example, *"we are driving on Fridhem Street now"* (Sw. *"Fridhemsgatan är det nu"*), or *"we are supposed to turn into something called Mårten Krakow Street"* (Sw. *"vi ska in på någon Mårten Krakowgatan"*). Imperative phrases are almost only used when switching to the Navigation domain (except for 1 occasion when switching to the Other domain), which is not surprising since this is the domain where one speaker (the passenger) gives instructions to the other (the driver). Rather, we would have expected more imperative phrases from the passenger when giving the instructions, but the result shows that they use declarative phrases more often than they use imperative phrases.

Noun phrases are not as commonly used when switching to this domain as they are when switching to the Interview domain. We may speculate that the speaker avoids bare noun phrases when switching to this potentially safety-critical domain, because they want to be more clear about what they mean, to avoid confusion and possibly high workload from trying to interpret the meaning of the utterance (Breitholtz, 2014). For example, by saying *"we are driving on Fridhem Street now"* the speaker is informing the dialogue partner about what s/he knows. If the speaker instead would have said *"Fridhem Street"* the dialogue partner might have been confused about whether the speaker is stating what s/he knows, or if s/he is looking for the street. Interrogative phrases are usually used by the drivers to clarify an instruction. For example, *"should I turn here" (Sw. "ska jag svänga här")*, or by the passengers when they are unsure of how to interpret an instruction or do not know exactly where they are. For example, *"where are we now, does the street change name here" (Sw. "var är vi nu, byter gatan namn här")*.

Figure 3.9 shows the distribution of domain specific phrases for the Traffic domain.

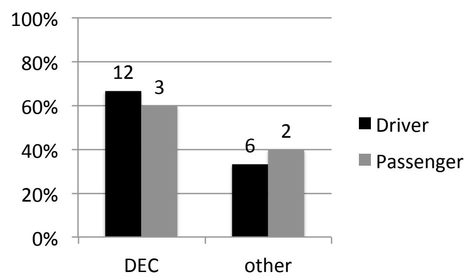



Figure 3.9: Domain-specific phrases when switching to the Traffic domain.

It is mostly the drivers who comment upon the traffic, but both speakers mostly use *declarative* phrases when switching to this domain. They also use a few *noun*, *interrogative*, *incomplete*, *verb* and *adverbial* phrases. The driver seems to use declarative phrases to inform the passenger of traffic-related events that occupies his or her mind at the moment, like in the example mentioned in Section 3.1 where the driver interrupts an interview topic to say *"squeeze myself in here"* when turning in to a narrow street.

Figure 3.10 shows the distribution of domain specific phrases for the Other domain.

Both drivers and passengers mostly use *declarative* phrases to switch to the Other domain, and a few *noun*, *interrogative*, *incomplete* and *imperative* phrases. Often the declarative phrases seem to be used

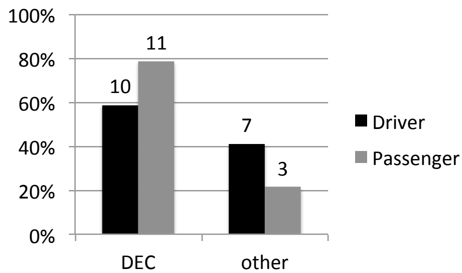
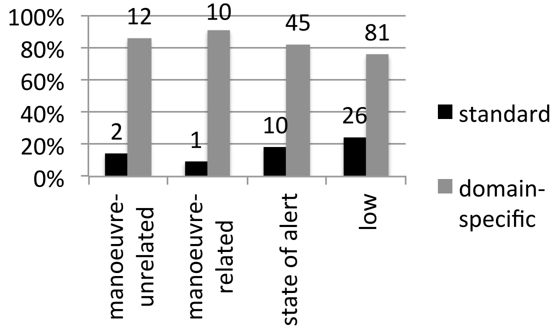



Figure 3.10: Domain-specific phrases when switching to the Other domain.

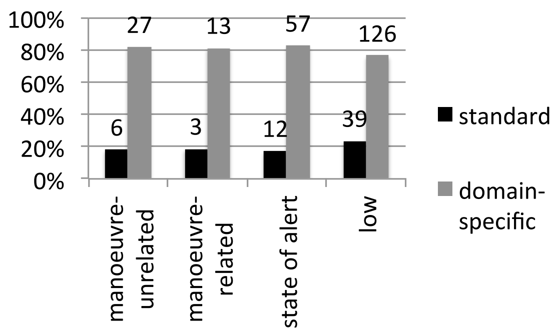
in the same way as they are when switching to the Traffic domain; to inform the dialogue partner about what it is that caught their attention or why they need to pause the ongoing conversation for a moment. For example, *"it was a harder task than we could have imagined" (Sw. "det var en svårare uppgift än vi hade kunnat tänka oss")* (about performing the navigation task) or *"now I have to turn pages" (Sw. "nu måste jag vända blad")* (when looking for the next navigation instruction).

Earlier studies (see Section 3.3) indicate that standard phrases are used more often when the workload is low (and the speaker is more able to plan how to switch topic), and that domain-specific phrases are used more often during high workload (when the speaker may not be as able, or have time, to plan their utterances as they are during low workload). We wanted to see if we could find similar differences in behaviour in the in-vehicle environment, and therefore calculated the number of standard phrases and domain-specific phrases during various workload types. The hypothesis is that the speakers will use more standard phrases and less domain-specific phrases during low workload than they do during high workload (see Section 3.5). Even though the number of occurences is too small to draw any strong conclusions, it may still be interesting to compare the annotations to get an idea about the behaviour. Figure 3.11 shows the result for each speaker and workload type.

Although the results are tentative and not tested for significance




(a) Sequencing moves made by the drivers.




(b) Sequencing move types made by the passengers.

Figure 3.11: Distribution of types of sequencing moves for each workload type.

because of the small number of occurrences, it seems pretty clear that both speakers use mostly domain-specific sequencing phrases. This means that usually they do not use a general opening phrase to announce that a domain switch is coming, but instead are more head-on and get to the point more directly by using a phrase that is specific for the domain they are switching to.

However, when looking at the driver's utterances we may see a trend pointing at more standard phrases when the workload is low compared to when the workload is high or there is a state of alert, and more domain-specific phrases when the workload is high or there is a state of alert compared to when the workload is low. There are less variations when looking at the passengers' sequencing moves. Since we only measure the driver's workload (see Section 1.4), we do not know if the passengers' use of sequencing moves varies depending on their own workload level and type in a way that is similar to the drivers' behaviour. The result shown here may indicate that the passengers do not adjust their behaviour according to the driver's workload, although it is hard to tell from this small material. However, a cognitive workload-aware dialogue system would get information about the driver's workload and hence it would be possible to adjust the way the system expresses itself depending on the driver's workload level and type. It remains to be investigated if it would be more beneficial for the driver if the system varies type of sequencing move based on the driver's workload level, or if one type of sequencing move would always be preferable over the other when the system takes the initiative to switch domain.

Apart from looking at sequencing moves in relation to workload level, it may also be interesting to know if the behaviour differs depending on which domain the speaker is switching to. We want to know if the participants prefer one type of sequencing move when switching to a certain type of domain, and therefore we looked at the use of sequencing moves when switching to each domain to see if a categorization of domains would be useful in this context. A categorization of driving-related versus driving-unrelated domains may be useful here, since the hypothesis is that the participants will use more domain-specific phrases when switching to the (time-critical and/or event-driven) driving-related domains than they do when switching to the driving-unrelated domains (see Section 3.5). Consequently, the participants should use relatively more standard phrases and less domain-specific phrases when switching to the non time-critical Interview domain compared to when switching to the driving-related domains. The purpose of switching to the Other domain varies, and so does the level of how time-critical the domain switch is. Therefore it is hard to make any assumptions about the behaviour when switching to this domain. Figure 3.12 shows the result.

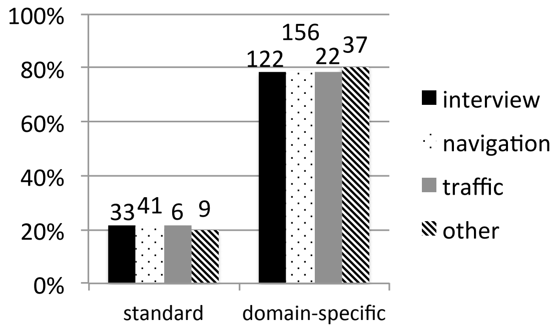



Figure 3.12: Type of sequencing moves when switching to each domain.

We found no difference in behaviour when looking at sequencing moves in relation to type of domain. The distribution of types of sequencing moves are equal within all domains, with domain-specific phrases being more common than standard phrases. Therefore, these results do not indicate that an adjustment in the use of sequencing moves in relation to domain of discussion is useful. However, this is a small study and the result may be different if studying a larger population.

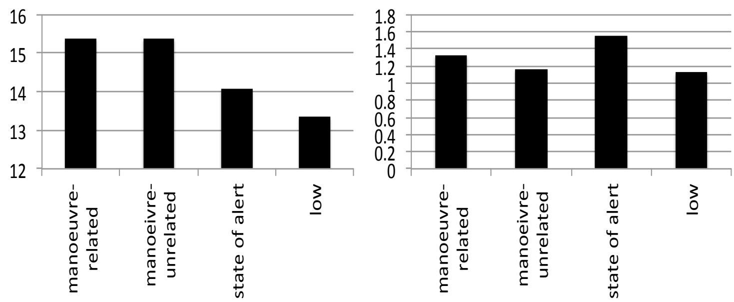
To summarize this study, the tentative results shown in Figures 3.11 and 3.12 indicate that workload may influence the use of type of sequencing move more than type of domain does.

## 3.8 Implications for a dialogue system

We have looked at interruption strategies in human-human dialogue. Specifically, we have looked at silent pauses and domains switches as a way of managing the driver's cognitive workload. In this chapter, we will suggest how these strategies could be applicable in an in-vehicle dialogue system. Following the reasoning in Section 1.4.2, we will propose strategies for such systems by learning from the participants' behaviour when discussing topics within the Navigation and Interview domains. As the passenger is the one providing the navigation instructions and asking the interview question, we map the passenger to the system and the driver to the user.

### 3.8.1 Pausing the dialogue

Our analysis of pausing behaviour showed that the participants pause more often when TDT is signalling (during manoeuvre-related and manoeuvre-unrelated workload), and that the pause length is longer when IDIS is signalling (during manoeuvre-related workload and state of alert), see Figure 3.13, presented in Section 3.7.1 but repeated here for convenience.




(a) Number of pauses per minute (b) Average pause length in seconds

Figure 3.13: Pausing behaviour.

These results indicate that when the driver is interacting with a dialogue system while driving, the system should be aware of the driver's workload level and driving manoeuvres. The result concerning pause frequency can be taken as motivation that a system should pause the dialogue during high workload until the workload is low.

**Strategy 3.1:**

- if workload = manoeuvre-related OR manoeuvre-unrelated
    - pause until workload = low

Using this strategy, a (made up) interaction during manoeuvre-related workload would look like this:

(15) WORKLOAD = LOW
*The driver starts an interaction with the dialogue system.*

```
system:  What do you want to do?
driver:  Change radio station
```

WORKLOAD = MANOEUVRE-RELATED

*the system pauses*

WORKLOAD = LOW

```
system:  What frequency or preset station
do you want to change to?
```

The results shown in Figure 3.13 shows that the pause frequency is not significantly higher during state of alert, but when pausing, the pause length is significantly longer. Therefore, we suggest that the system should not take the initiative to pause during this condition, since it may distract the driver more than an uninterrupted dialogue would. However, if the driver pauses during state of alert, the system should probably pause too. Basically, this means that if the driver does not respond to a question or a request from the system, the system should not repeat its utterance immediately but wait until the workload is low.

**Strategy 3.2:**

- if workload = state of alert:
  - do not repeat utterance until workload = low

(16) WORKLOAD = LOW
*The driver starts an interaction with the dialogue system.*

```
system:  What do you want to do?
```

WORKLOAD = STATE OF ALERT

```
driver:  Change radio station.

system:  What frequency or preset station
do you want to change to?
```

*The driver does not answer, and therefore the system does not repeat the question.*

*Pause n seconds.*

WORKLOAD = LOW

```
system:  What frequency or preset station
do you want to change to?
```

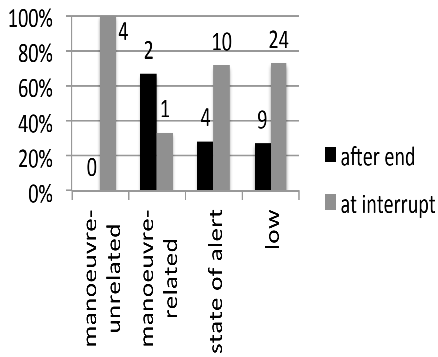
In Example (16), the driver answers the system's first question (*"what do you want to do"*) although there is a state of alert. Therefore, the system continues the interaction and asks the next question (*"what frequency or preset station do you want to change to"*). Now, the user does not answer, and typically the system would, after a certain amount of time, repeat a question that has not been answered[4]. But due to the state of alert condition, the system does not repeat the unanswered question right away but pauses until the workload is low before repeating it.

### 3.8.2 Switching domain

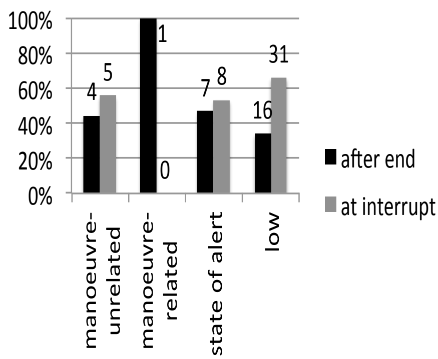
Although the data is sparse, our analysis of topic switches hinted that the participants often interrupt to give or ask for navigation instructions. However, during high workload it seems like the passenger, to a higher degree than the driver, tries to wait until

[4]This is for example how Apple's Siri operates (when connected to a power source).

the ongoing topic discussion is finished before switching topic, if possible. See Figure 3.14 (presented in Section 3.7.2 but repeated here for convenience).




(a) Topic switches made by the driver.




(b) Topic switches made by the passenger.

Figure 3.14: Topic switches made by the driver and the passenger, respectively.

Therefore, a possible dialogue strategy would be that during high workload, a dialogue system should avoid interrupting an ongoing interaction to switch domain. However, our results also indicate that the driver asks for instructions whenever they want the information, regardless of whether they have to interrupt an ongoing discussion to do so or not, and therefore we believe that it may sometimes be necessary to interrupt a topic if it is urgent to give the information. Therefore, the system should be aware of both the workload level and the urgency of the information that needs to be given:

**Strategy 3.3:**

- when discussing topic T1, if information on topic T2 is received
  - ∗ if workload = manoeuvre-related OR manoeuvre-unrelated;
    - · if information on topic T2 is urgent, interrupt topic T1 and switch to T2
    - · else finish T1 before moving on to T2

When the manoeuvre-related workload is high and the information that needs to be given is not urgent, the dialogue would look like this:

(17) WORKLOAD = LOW
*The driver starts an interaction with the dialogue system.*

```
system:  What do you want to do?
```

WORKLOAD = MANOEUVRE-RELATED
*The next navigation instruction is labelled "not urgent".*

```
Driver:  Change radio station.
system:  What frequency or preset station
do you want to change to?
Driver:  88.5
System:  Ok, tuning to 88.5.  At the next
crossing, make a left turn.
```

If the information that needs to be given is urgent, the dialogue would instead look like this:

(18) WORKLOAD = LOW
*The driver starts an interaction with the dialogue system.*

```
system:  What do you want to do?
```

WORKLOAD = MANOEUVRE-RELATED
*The next navigation instruction is labelled "urgent".*

```
driver:  Change radio station.
system:  At the next crossing, make a left
turn.
driver:  Ok.
system:  What frequency or preset radio
station do you want to change to?
driver:  88.5
system:  Ok, tuning to 88.5.
```

### 3.8.3 Adjacency pairs

In Section 3.7.4 we investigated interruption strategies in relation to adjacency pairs. We found that the participants avoid interrupting a topic within an adjacency pair. See Figure 3.15 (presented in Section 3.7.3 but repeated here for convenience).

Based on this, we propose that the system avoids interrupting within an adjacency pair. A possible dialogue strategy would be to not interrupt a topic until a raised question has been answered:

**Strategy 3.4:**

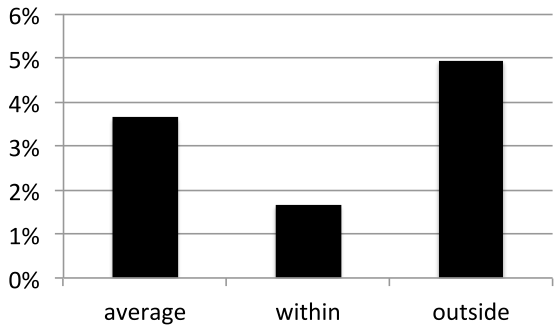



Figure 3.15: Distribution of interruptions in relation to adjacency pairs.

- when discussing topic T1, if information on topic T2 is received and urgent
  - ∗ if a system-initiated question has not been answered
    - · wait until the user has answered, then switch to T2
  - ∗ if a user-initiated question has not been answered
    - · answer the question, then switch to T2

Using this strategy, the system will wait for the user to answer a question before switching domain:

(19) *The driver starts an interaction with the dialogue system*

```
System:  What do you want to do?
Driver:  Make a call.
System:  Who do you want to call?
```

*The next navigation instruction is labelled "urgent".*

```
Driver:  John.
System:  At the next crossing, make a left
turn.
```

```
Driver:  Ok.

System:  Do you want to call the mobile
phone or the home number?

Driver:  The home number.

System:  Ok.  Calling John's home number.
```

After asking the question (*"who do you want to call"*), the system receives urgent information that needs to be given to the driver. However, instead of switching domain immediately, the system waits until the driver has answered the question.

If the system receives urgent information after the driver has asked a question, the system answers the question first before switching domain:

(20) *The driver starts an interaction with the dialogue system*

```
Driver:  Do I have any new messages?
```

*The next navigation instruction is labelled "urgent".*

```
System:  No, you haven't received any new
messages.

System:  At the next crossing, make a right
turn.

Driver:  Ok.
```

The driver asks a question (*"do I have any new messages"*), and after that the system receives urgent information. The system answers the question first (*"no, you haven't received any new messages"*), before switching domain.

### 3.8.4 Sequencing moves

In Section 3.7.4, we presented preliminary results concerning how an interruption should be made when interacting with a dialogue system. When analysing the DICO corpus we found that the participants most often use domain-specific phrases, without explicitly announcing the domain switch, similar to Example (18) (*"At the next crossing, make a left turn"*). However, during low workload we could see indications that standard phrases were used more often than during high workload and state of alert, see Figure 3.16 (presented in Section 3.7.4 but repeated here for convenience).

Using a standard phrase before switching domain can be seen as an implicit way of announcing the domain switch. A possible dialogue strategy for notifying the driver of an upcoming system-initiated interruption would be to use a standard phrase before switching domain:

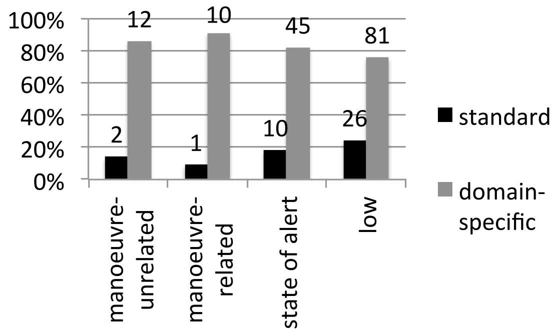
**Strategy 3.5:**

- when discussing topic T1, if information on topic T2 is received and urgent
  * interrupt topic T1 and switch to T2 using standard-phrase

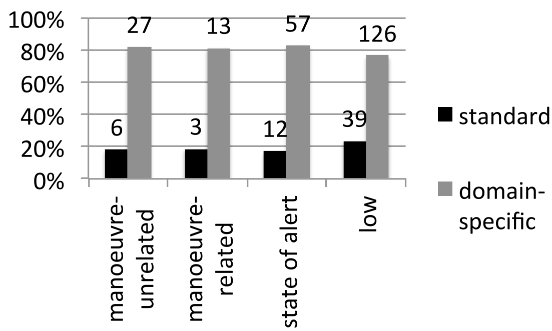
(21) WORKLOAD = LOW
*The driver starts an interaction with the dialogue system*

```
System:  What do you want to do?
Driver:  Change radio station.
```

*The driver is approaching a crossing and the next navigation instruction is labelled "urgent".*




(a) Sequencing moves made by the drivers.




(b) Sequencing move types made by the passengers.

Figure 3.16: Distribution of types of sequencing moves for each workload type.

```
System:  Let's see.  At the next crossing,
make a left turn.

Driver:  Ok.

Aystem:  Alright.  What frequency or preset
radio station do you want to change to?

Driver:  88.5

System:  Ok, tuning to 88.5.
```

## 3.9 Conclusions and future work

When comparing pausing behaviour during all workload types, we found that the participants made more pauses during TDT, and longer pauses during IDIS. The next step is to test if pausing the dialogue during high workload would be a good strategy for preventing, or shortening the duration time of, high workload. We believe that especially during manoeuvre-related workload, pausing the dialogue until the workload is low would be a useful strategy. Pausing the secondary task during these circumstances would probably make it easier for the driver to concentrate on the driving task. To further fine-tune the pausing behaviour of a cognitive workload-aware dialogue system, we would also suggest investigating what types of driving manoeuvres trigger silent pauses. Since the participants in our study made fewer pauses during state of alert compared to the number of pauses per minute during high workload, it seems like all driving manoeuvres that are assumed to be demanding do not affect the driver in such a way that the driver needs to pause the dialogue. If we know more about which manoeuvres are demanding enough to make the driver want to pause, we may be able to prevent manoeuvre-related workload if the dialogue system adjusts dialogue behaviour already during state of alert, when necessary.

As indicated in Section 3.7, it appears that when interrupting the dialogue, interrupting within an adjacency pair may be counterproductive if the dialogue partner is not mentally prepared for the interruption, and especially not to receive information from another domain than the one currently discussed. However, we need to know more about this behaviour to know which strategy is best from both a cognitive workload and safety point of view. From a cognitive workload point of view, it seems preferable to avoid interruption within an adjacency pair and instead finish the pair before switching domain. From a safety point of view, we risk that the delay that results from waiting until the pair is finished means that the information that needs to be given (for example, a navigation instruction) becomes more time-critical which may lead to a possibly risky driving manoeuvre.

Our analysis of sequencing moves in Section 3.7.4 showed that the participants most often did not use a standard phrase to introduce an upcoming interruption and domain switch, but instead used a domain-specific phrase. However, during low workload it may be more common using standard phrases compared to during high workload and state of alert. It should be investigated further whether the driver would benefit from being notified of an upcoming interruption, and, if so, a notification should come in the form of an earcon[5] or a sequencing phrase.

## 3.10 Summary

In this chapter, we have looked at interruption behaviour in dialogue. We started with an investigating of two possible strategies for managing the driver's workload; interruption by pausing the dialogue, and interruption by switching domain. We analysed these behaviours during various workload types. We also investigated when to interrupt to switch domain, by analysing interruptions in relation to adjacency pairs. Finally, we wanted to know how to interrupt an ongoing discussion, and therefore analysed types of sequencing moves and grammatical category of the interrupting utterance during various workload types. Based on these findings, we proposed dialogue strategies for an in-vehicle dialogue system. We concluded with a discussion of future work issues.

[5] An earcon is a short and distinctive sound (for example, a beep), that is used to inform the user of an error, an alarm or some other event. See, for example, Sumikawa (1985).

# Chapter 4

# Resumption in dialogue

## 4.1 Introduction

In the previous chapter, we suggested that if the driver is interacting with the dialogue system during manoeuvre-related high workload, the interaction should be interrupted. We proposed two possible strategies; either to pause the dialogue to let the driver concentrate on manoeuvreing the car, or to switch topic if the driver could benefit from getting information that a dialogue system could give, such as navigation instructions.

Both these strategies mean that the ongoing interaction is interrupted and potentially a task is unfinished. If the task needs to be finished, the interrupted topic needs to be resumed in some way. If so, we first of all need to know *when* to resume an interrupted topic. If the interruption was in form of a domain switch, the interrupting topic probably needs to be finished before a resumption of the first topic. However, it may not always be suitable to resume an interrupted topic immediately when the interrupting topic is finished. Therefore, we need to know when to resume both after a pause and after a domain switch, and if the system should take the initiative to resume or if it should wait for the user to do it.

The next step is to know *how* to resume the interaction. A possible strategy when a topic has been interrupted is to abort the task that was interrupted, and return to the top menu. This means that if the user wants to finish the interrupted task, the task needs to be restarted from the beginning. However, this strategy is both time-consuming and annoying for the user. The strategy is especially unfortunate if the user is performing a parallel (and possibly cognitively demanding) task such as driving. Instead, the dialogue system should be able to resume the interaction where the interruption took place. However, resumption of an interrupted topic needs to be done in a way that minimizes the risk of the cognitive workload increasing yet again. Resuming an interrupted topic in a bad way may confuse the driver and cause the workload to increase because of the dialogue itself.

## 4.2 Background

Research regarding resumption of an interrupted task often deals with interruptions when using a graphical user interface, and focus on the time and effort it takes to resume the interrupted task. The research therefore aims at finding out how to interrupt the interaction in a way that makes it faster and less cognitively demanding to resume the interrupted task. For example, Paul et al. (2015) found that if the user gets a notification about an upcoming interruption, it is easier for them to later resume the interrupted task. The notification helps them maintain an awareness of the status of the multiple tasks they are performing, and therefore facilitates the resumption of an interrupted task.

When it comes to spoken interaction, and especially spoken interaction in vehicles, it is often difficult to give a notification before an interruption. If the system needs to interrupt to give information it may be a good idea to notify the driver of an upcoming interruption, but if the interruption is due to changes in the traffic environment that call for the driver's immediate attention it may be difficult to give a notification signal before the interruption. Possibly, an interruption signal could itself be distracting, if it steals the driver's attention away from what is happening in the traffic.

Edlund et al. (2014) conducted a user study where the participants used a car simulator while listening to and answering questions about the content of a text read aloud by an experimenter. Randomly, they were interrupted by a signal notifying that they needed to perform a second task, in which the experimenter instead read a sequence of eight numbers which the participants had to repeat. The authors found that the place of interruption influences the response time. In the cases where the interruption signal appeared in the middle of a phrase the response time was short and the speaker stopped as fast as possible, while if the interruption signal occurs at the end of a phrase the speaker preferred to finish the phrase before switching task. They also found that the length of the interruption played a role when resuming the interrupted task. When resuming after a short interruption, when the interrupted topic is still fresh in the memory, the participants usually did not signal resumption, or used a lexical cue phrase such as *"ok"* or *"right"*. Resumption after a longer interruption was almost always signalled by a filled pause, presumably to prepare the dialogue partner that they were about to resume an interrupted task.

Grice's Maxim of Quantity, "do not make your contribution more informative than is required" (Grice, 1975), tells us that in a conversation we do not repeat what has already been said. Walker (1996), however, argues that people sometimes violate this redundancy constraint, for example due to limited resources. Even though something has been said earlier in the conversation, it does not mean that this information is still instantly available in the working memory of both dialogue partners. Breitholtz and Villing (2008) argue that giving redundant information may help lowering the cognitive workload of a user, if the redundant information provides the user with arguments for why the utterance is made.

In addition to studying interruption strategies, Yang et al. (2011) also studied resumption behaviour when they carried out the user study mentioned in Section 3.3. The participants played a card game and were randomly interrupted by a time-critical secondary task which they had to carry out before resuming the ongoing card game. When studying the resumption utterances the authors found that they contained various amounts and types of redundant information depending on whether the interruption occurred in the middle of a card discussion, at the end of a card or at the end of a card game. Hence, the place of an interruption, and not only the duration, play a role when resuming. If the interruption occurred in the middle of a card discussion, there was a distinction between *utterance restatement* (repeat one's own utterance, repeat the dialogue partners utterance or clarification of the dialogue partner's utterance) and *card review* (reviewing all the cards on hand although this information had already been given). This behaviour is similar to grounding behaviour, where the speaker uses repetition and requests for repetition to ensure that the utterance is understood.

## 4.3 Research questions

In Section 4.1 we outlined three research questions that are relevant when dealing with resumption of an interrupted dialogue system interaction: *when* to resume, *how* to resume and *who* should resume.

Regarding when to resume an interrupted interaction, the most obvious answer may be to resume the interrupted topic when the workload is low and the driver therefore seems to be able to interact with the system without being distracted by, for example, a demanding driving manoeuvre or a complicated traffic situation. However, depending on the topic that was interrupted, it may be more beneficial for the driver to resume the topic while the workload is still high, if we believe the interrupted topic could help the driver with the driving task. The interrupted topic may, as in Example (22), be a navigation instruction that needs to be given:

(22) [WORKLOAD = MANOEUVRE-RELATED]

**Passenger:** "we should go to Hjalmar Branting Street later [this err]" (Sw. "vi ska till Hjalmar Brantingsgatan [sen eh]") [*domain of discussion = navigation, topic-begin*]

[WORKLOAD = MANOEUVRE-UNRELATED]

**Driver:** "[it feels like] I am forgetting to click" (Sw. "[det känns som jag] glömmer att klicka") [*domain of discussion = other, topic-interrupt, topic-begin*]

**Passenger:** "to the right so it must be over here at the traffic lights" (Sw. "till höger så det är väl framme här vid trafikljusen") [*domain of discussion = navigation, topic-interrupt, topic-resume/reraise*]

The driver is interrupting the passenger's utterance, to comment that she thinks she sometimes forgets to click the TDT button when the buzzer activates. This comment may be seen as a sign of high workload; the driver comments about something that is occupying her mind at the moment and takes her attention away from the task at hand since she forgot to click, and we can see that workload is in fact high during this interaction. However, they are approaching a traffic light where they have to decide which lane to follow, and the passenger resumes the navigation instruction so that the driver will know where to go next. In this scenario, the driver probably benefits more from resuming the navigation discussion to get the full instruction, instead of pausing until the workload is low. Therefore, we may have to consider more factors than the workload level when deciding when to resume an interrupted topic.

Concerning how to resume an interrupted task, resumption may be done by simply repeating the last phrase that was uttered before the interruption, regardless of place and length of the interruption. Earlier research (see Section 4.2) suggests that giving redundant information by repeating (parts of) earlier utterances may help preventing high workload. However, one disadvantage of always doing this after an interruption is that the dialogue system may be seen as tedious and even annoying, especially if there are several interruptions during the interaction.

## 4.4 Hypotheses

We have analysed the DICO corpus to investigate resumption behaviour during in-vehicle dialogue. When it comes to in-vehicle dialogue, the primary task of driving should be the priority for the participants. The hypothesis regarding when to resume an interrupted topic is therefore that resumption behaviour may differ depending on domain of discussion and its relationship to the driving task. If the interrupted topic is a topic within a driving-unrelated domain (i.e. the Interview and Other domains), we believe the participants will try to resume the topic during low workload. These topics are not necessary to discuss in order to be able to perform the driving task, and therefore it is probably unsuitable to discuss these matters while the driver is making a demanding driving manoeuvre or his/her mind is occupied with something else. Furthermore, the conversation risks adding to the cognitive workload, and therefore we expect to see fewer resumptions of topics within driving-unrelated domains during high workload and state of alert. By contrast, if the interrupted topic is a topic within a driving-related domain, it may be important to finish the topic in order for the driver to perform the driving task. We therefore believe that the participants will resume driving-related topics during manoeuvre-related workload and state of alert, and that we will see more resumptions of driving-related than driving-unrelated topics during these conditions.

In the matter of how to resume an interrupted topic, we will look at redundant information within the resuming utterance, in relation to the length of interruption. The in-vehicle environment puts heavy demands on the driver, requiring them to split their attention between the driving task and the dialogue task. Hence, the driver may need help remembering the details about an interupted task. Still, our hypothesis is that redundant information is not necessary after a short interruption, but after a longer interruption the resuming speaker probably needs to repeat some of the information. We will therefore investigate redundancy in resumption utterances with respect to the number of turns of the interrupting topic, to see if redundant information is more common within utterances resuming a topic after a long interruption compared to utterances resuming a topic after a short interruption.

We also want to know about the structure of the resuming utterances, to learn how humans express themselves when resuming an interrupted topic. We will therefore look at the grammatical structures of the utterances, in a similar way as we investigated the grammatical structures of domain specific sequencing moves (see Section 3.7).

## 4.5 Method

Apart from annotations of domain of discussion and workload type (see Section 1.4), the DICO corpus has been annotated with topic shifts (see Section 3.6). The utterances we are interested in here are those annotated as *topic-resume*. We have analyzed utterances that resume topics within the Interview and Navigation domains. As mentioned in Section 3.6, we cannot make a distinction between *topic-resume* and *topic-reraise* within the Navigation domain, and we therefore include all utterances that refer to an earlier discussed topic within this analysis, and annotate them as *topic-resume/reraise*. We discard the Traffic and Other domains from this analysis, since the topics within these domains are rarely resumed; none of the topics within the Traffic domain is resumed, and only two (2) topics are resumed within the Other domain.

An utterance annotated as *topic-resume* is the first utterance after an interruption that contains a reference of some sort to the interrupted topic. Resumption utterances may be preceded by a sequencing move (see Section 3.6), and in those cases the sequencing move and the resumption phrase are annotated separately as *seq-std-phrase/seq-domain specific-phrase* and *topic-resume*, as in Example (23) where the driver interrupts a navigation discussion to comment about his driving:

(23) **Driver:** "it's probably here towards downtown" (Sw. "det är nog här mot centrum") [*domain of discussion = navigation*]

**Passenger:** "probably" (Sw. "kan det nog vara")

**Driver:** "I feel I'm driving badly now LAUGHS I'm driving a bit jerkily" (Sw. "jag känner att jag kör ganska dåligt nu LAUGHS jag kör lite ryckigt") [*domain of discussion = traffic, topic-interrupt*]

**Passenger:** "LAUGHS yes" (Sw. "LAUGHS ja")

**Pasenger:** "okay" (Sw. "okej") [*sequencing-std-phrase*] "follow the sign towards downtown" (Sw. "följ skylt mot centrum") [*domain of discussion = navigation, topic-resume/reraise*]

### 4.5.1 Redundancy

When resuming an interrupted topic, the speaker may want to repeat (parts of) the information that has been given earlier, to remind the dialogue partner about the interrupted topic and what has already been agreed upon. This can be done in different ways, either by repeating one or more words that have been uttered before, or by paraphrasing an earlier utterance. For example, Walker (1992) exemplified this with an excerpt of a dialogue between a radio talk show host and a caller:

(24) **Caller:** uh 2 tax questions. One: since April 81 we have had an 85 year old mother living with us. Her only income has been social security plus approximately $3000 from a certificate of deposit and I wonder what's the situation as far as claiming her as a dependent or does that income from the certificate of deposit rule her out as a dependent?
**Talk show host:** Yes it does.
**Caller:** It does.
**Talk show host:** Yup that knocks her out.

Both the caller's last utterance ("it does") and the talk show host's last utterance ("yup that knocks her out") are redundant, although in different ways. The caller repeats parts of the talk show host's earlier utterance, while the talk show host's last utterance paraphrases what he has already said. The reasons for giving redundant information may vary, but in this thesis we are interested in redundancy that seems to have the purpose of reminding the dialogue partner of an earlier, interrupted, discussion with the aim of resuming that discussion.

For this investigation, we have only looked for utterances where one or more words are repeated. The reason for this is that when annotating the corpus we found very few examples of redundancy in the form of paraphrasing, and therefore it was not meaningful to analyse this phenomenon. As a redundant utterance we counted all utterances that repeated one or more content words[1] when compared to earlier utterance. Since the DICO corpus contains rather few silent pauses and the pauses are relatively short, we have not looked at redundancy after a pause, but only redundant information within utterances that resume a topic after a topic switch. We sorted all utterances coded as *topic-resume* (within the Interview domain) and *topic-resume/reraise* (within the Navigation domain) into two categories: those which contained redundant information when comparing with the two last utterances made before the interruption (to include both speakers' last utterances), and those which did not contain such redundant information. This annotation has not been tested for reliability. Still, it is fairly straightforward compared to many other annotations, since it relies only on finding at least one content word which appears in at least one of the two utterances before the interruption and in the resuming utterance.

We counted the number of turns between the interruption and the resumption, i.e. the total number of turns within the interrupting dialogue. A turn is here defined as everything being said by the same speaker before a speaker change. The number of turns varied between 1 and 46 (average 7.9, median 4.5) in the Navigation

[1]Content words are words that refer to an object, action or characteristics in some way (for example, nouns, verbs and adjectives), in contrast to function words that serve primarily to express grammatical relationships between words (for example, pronouns, prepositions and articles) (Winkler, 2007).

domain and 1 and 50 (average 9.1, median 5) in the Interview domain. We categorized the interruptions into two categories; short interruption (containing $\leq$4 turns), and long interruption (containing $\geq$5 turns). This threshold was chosen since it seemed to separate the data into two coherent clusters.

### 4.5.2 Grammatical category

Resumption utterances have been classified in the same way as domain-specific sequencing phrases (see Section 3.7), according to the following schema:

- DEC: declarative sentence
  - for example, *"I keep on towards Linné place here" (Sw. "jag kör mot Linnéplatsen här")*
- INT: interrogative sentence
  - for example, *"are we going to Järntorget again or what" (Sw. "skulle vi mot Järntorget igen eller")*
- IMP: imperative sentence
  - for example, *"keep on straight ahead" (Sw. "fortsätt rakt fram")*
- ANS *yes* or *no* answer
- NP: bare noun phrase
  - for example, *"sniper" (Sw. "prickskytt")*
- ADVP: bare adverbial phrase
  - for example, *"via the entrance here" (Sw. "via påfart här")*
- INC: inomplete phrase
  - for example, *"keep left after" (Sw. "håll till vänster efter")*

## 4.6 Results and discussion

We have analysed resumption behaviour within the DICO corpus. More specifically, we have analysed resumption behaviour in relation to workload and also looked at how the participants resume an interrupted topic with respect to redundant information in, and grammatical category of, the resuming utterance. Similarly to when analysing the interruption behaviour, we will indicate statistical significance, but in the cases where the data is too sparse we will show and discuss trends without drawing any definitive conclusions.

As mentioned in Section 3.6, it is only possible to determine if a topic is interrupted within the Interview domain. We found that when switching to the Navigation domain from the Interview domain the participants interrupt more frequently than they wait until the ongoing discussion is finished, see Figure 4.1 (presented in Section 3.7 but repeated here for convenience).

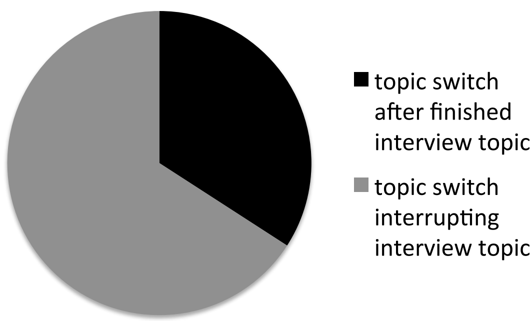



Figure 4.1: Distribution of topic switches from the Interview to the Navigation domain.

We have analysed resumption utterances to the Interview domain, and resumption/reraising utterances to the Navigation domain. As mentioned, it is hard to determine if a topic within the Traffic and Other domains is being interrupted (see Section 3.7), which also makes it hard to determine if a topic is being resumed. However, when annotating the corpus we noticed that the participants do not seem to either resume or reraise a topic within the Traffic domain, since they do not return to discuss a certain traffic-related event after the domain shift following the initial discussion. The comments within the Traffic domain are often, at least in part, made to signal high workload (see Section 3.1) and the short comments within this domain are usually not elaborated or discussed further. The same goes for interrupted topics within the Other domain - only two topics in total within this domain are returned to after a domain shift. However, we found that topics within the Interview and Navigation domains are resumed/reraised to a higher degree. These topics are discussed in order to solve a task, and an interrupted discussion may need to be resumed for the participants to be able to finish the task if the interruption takes place before all necessary information has been shared. This is probably the main reason why it is more common that the participants resume/reraise these topics than the topics within the task-unrelated domains.

### 4.6.1 When to resume an interrupted topic

#### Workload level when resuming an interrupted topic

We have analysed the DICO corpus with respect to resumption utterances in relation to the driver's workload level, to see if there is a difference in the distribution of resumption utterances when comparing the different types of workload. To normalize for the different duration times of the workload types, we have calculated the frequency of resumption/reraising utterances during each workload type within the Navigation and Interview domains.

**Navigation domain** Looking at anecdotal data, we noticed that often the resumption/reraising of a navigation topic is made when the participants are approaching the point where they (think they) should make the next manoeuvre. For example, the driver may ask *"is this where I should turn" (Sw. "ska jag svänga här")* or the passenger may repeat an instruction, for example, *"the round-*

*about straight ahead then" (Sw. "rondellen rakt fram här då").* It appears that resumption/reraising of a navigation instruction may be associated with demanding traffic situations, and that resumption/reraising is done either by the passenger to refresh the driver's memory about a previous instruction or by the driver to clarify a previous instruction. To investigate this, we counted the number of resumption/reraising utterances during each workload type, and to compensate for the various duration times we normalized by calculating the number of resumption/reraising utterances per minute for each workload type. By doing this, we wanted to see if the frequency is higher during high workload or state of alert.

Figure 4.2 shows the number of resumption/reraising utterances per minute when resuming/reraising a navigation topic. Each bar is labelled with the number of such utterances during each workload type.

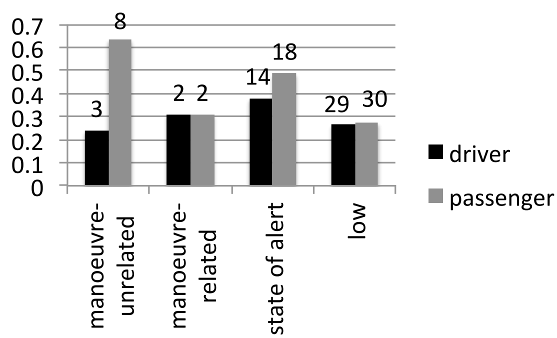



Figure 4.2: Resume/reraise utterances per minute when resuming/reraising a Navigation topic during various workload types.

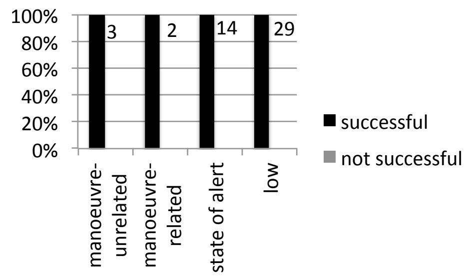
All in all, there are 106 resuming/reraising utterances made to the Navigation domain. The result indicates that the frequency of resumption/reraising utterances for the passengers is highest during state of alert and manoeuvre-unrelated workload, and lowest during low workload. For the drivers, the differences are smaller but the frequency is highest during state of alert and lowest during manoeuvre-unrelated workload.

Although it is hard to know from the sparse data, the result indicates that both speakers may tend to resume/reraise a navigation topic more frequently when there is a demanding situation, i.e. either when the workload is high or when the driver performs a demanding driving manoeuvre. As mentioned, the highest frequency of resumption/reraising utterances made by the passenger is during manoeuvre-unrelated workload. A possible explanation may be that the passenger notices that the driver is distracted by something and assumes that they are thinking about where to go next, possibly because they are approaching a point where the next navigation manoeuvre is to take place. This is, however, only speculations, and we need more data from video recordings to be able to know if this is the case.

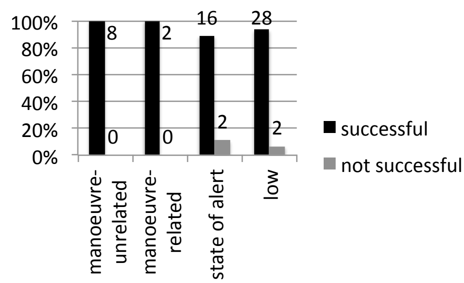
When the driver resumes/reraises a navigation topic during this condition we may speculate that the reason for the increased workload is distraction caused by confusion about the navigation task. When the driver makes a manoeuvre that causes IDIS to signal (i.e. during manoeuvre-related workload and state of alert), it is probably evident also for the passenger (sitting next to the driver and observing both the driving manoeuvre and the traffic situation) that the driving task is more demanding. Therefore we might have expected to see the highest frequency of resumption utterances made by the passenger during these conditions rather than during manoeuvre-unrelated workload, assuming that the passenger would think that the driver needs help remembering where to go next.

Now that we know a little more about when the participants resume/reraise a navigation topic, we may also investigate the success rate of these resumption/reraising attempts. The result shown above indicates that the need to resume/reraise a navigation topic is highest when the workload is high and/or during state of alert (assuming that the preceding navigation instruction has not been performed yet), but is this a good strategy? To learn more about that, we classified all resumption/reraising utterances immediately followed by at least one more utterance (made by the dialogue partner) within the resumed topic as successful. The resumption utterances that are not immediately followed by at least one more utterance within the resumed topic are counted as unsuccessful.

Figure 4.3 shows the distribution of successful and unsuccessful utterances when resuming a navigation topic during each workload type, for both speakers.




(a) Resumption utterancess made by the driver.




(b) Resumption utterances made by the passenger.

Figure 4.3: Success rate when resuming a navigation topic.

The drivers are successful in all their attempts to resume/reraise a navigation topic, during all workload types, which indicates that the passenger is attentive to the driver's questions and comments regarding the navigation task and responds to them immediately. This is mostly the case also when the passenger attempts to resume/reraise a navigation topic and the driver is the one who is supposed to respond to it, but not all of these attempts are successful. During the two high workload types, the passenger's resumption attempts are successful. Two out of 18 attempts fail during state of alert, and two out of 30 attempts fail during low workload.

The unsuccessful resumption/reraising attempts made by the passenger during low workload were made when the driver was answering an interview question and had not yet finished their utterance. When trying to resume, the passenger interrupted the interview discussion within an adjacency pair, which may be a reason why the attempts were not successful (see Section ). One of the passengers who did this seemed to realize that is was better to let the driver answer the interview question before trying to resume again, and made a new attempt after the interview question had been answered:

(25) [WORKLOAD = LOW]

**Driver:** "melon is good but er well" (Sw. "nä melon är gott men eh ja") [*domain of discussion = interview*]

**Passenger:** "turn here then" (Sw. "av här då") [*domain of discussion = navigation, topic-interrupt, topic-resume*]

**Driver:** "grapes do they count as fruit" (Sw. "vindruvor är det en frukt eller") [*domain of discussion = interview, topic-resume*]

**Passenger:** "grapes are fruit yes I think so" (Sw. "vindruvor är väl en frukt ja det tycker jag")

**Driver:** "yes i think it's good" (Sw. "ja nä tycker jag är gott")

**Passenger:** "then towards Djurgårds Street towards Majorna so let's go to Majorna then" (Sw. "så ska man mot Djurgårdsgatan mot Majorna så vi kör Majorna då")

**Driver:** "then let's turn here" (Sw. "då svänger vi här")

It is unclear whether the driver noticed the passenger's navigation instruction, since he does not respond to it. However, when responding to the second resumption attempt the driver used a similar phrase as the passenger had used when trying to resume the first time, which may indicate that the driver had heard the comment even though he did not comment upon it the first time.

The other passenger who made an unsuccessful attempt to resume a navigation topic during low workload did not want to let the driver continue the interview discussion, even though the driver did not respond to the first attempt to switch topic and domain. Instead, the passenger insisted on switching to the navigation domain:

(26) [WORKLOAD = LOW]

**Passenger:** "favourite candy" (Sw. "favoritgodis") [*domain of discussion = interview*]

**Driver:** "yes that's a tricky one I am actually a chocolate kind of person so" (Sw. "a den är svår alltså jag är ju chokladmänniska i och för sig så att")

**Passenger:** "Ekelund Street do you have do you see anything like that" (Sw. "Ekelundsgatan har du hittar du nåt sånt") [*domain of discussion = navigation, topic-interrupt, topic-resume*]

**Driver:** "chocolate is chocolate is incredibly good and er" (Sw. "choklad är ju choklad är ju otroligt gott å eh") [*domain of discussion = interview, topic-resume*]

[WORKLOAD = MANOEUVRE-UNRELATED]

**Passenger:** "yes here is Ekelund Street PAUSE at the end of the street take right into Otterhälle Street" (Sw. "aa här är Ekelundsgatan PAUSE i slutet på gatan kör höger in på Otterhällegatan") [*domain of discussion = navigation, topic-interrupt, topic-resume*]

**Driver:** "what did you say" (Sw. "vad sa du")

**Passenger:** "at the end of the street that's here take right into Otterhälle Street" (Sw. "i slutet på gatan här då kör höger in på Otterhällegatan")

**Driver:** "absolutely" (Sw. "jajemen")

The driver is occupied by the interview question and does not seem to be prepared to receive other information at the time. When the passenger insists of changing topic and domain the driver asks *"what did you say"*, which may indicate that his mind is still occupied with the interview question and therefore is not paying enough attention to the passenger's utterance to be able to grasp what he is saying. The workload increases after the passenger's first attempt to resume the navigation topic, perhaps because the driver gets distracted by the change of topic, but since he does not respond to that resumption attempt we cannot now if the increased workload is due to the interview question, the resumption attempt or something else.

The unsuccessful resumption attempt that was made during state of alert was made when the participants were looking for the street they should turn into. The driver did not respond to the resumption utterance, but made a comment about the surroundings. However, the driver followed the instruction given by the passenger, so it seems like he perceived the instruction although he did not respond to it:

(27) [WORKLOAD = LOW]

**Passenger:** "take left into Alfhems Street" (Sw. "kör vänster in på Alfhemsgatan") [*domain of discussion = navigation*]

*The driver is looking at a car standing still in front of them and gesticulates with his hands.*

**Driver:** "where is he going" (Sw. "var ska han in") [*domain of discussion = traffic, topic-begin*]

**Passenger:** "he probably wants to park here PAUSE or maybe wait for somebody" (Sw. "han ska nog parkera här antagligen PAUSE eller vänta på någon kanske")

[WORKLOAD = STATE OF ALERT]

**Passenger:** “left here into Alfhems Street” (Sw. “här vänster in på Alfhemsgatan”) [*domain of discussion = navigation, topic-resume/reraise*]

**Driver:** “Slobbo’s gym” (Sw. “Slobbos gym”) [*domain of discussion = other, topic-begin*]

*The driver turns left into Alfhem Street.*

As a comparison, let us take a look at a successful resumption attempt made by the passenger, during manoeuvre-unrelated workload:

(28) [WORKLOAD = LOW]

**Passenger:** “then we are PAUSE going to Ekelund Street PAUSE it’s probably just straight ahead [too]” (Sw. “sen så ska vi PAUSE in på Ekelundsgatan PAUSE det är nog bara rakt fram [också]”) [*domain of discussion = navigation*]

**Driver:** “[it’s] straight ahead here right” (Sw. “[det] är rakt fram här va”)

WORKLOAD = STATE OF ALERT
*Pause for 1 second*

**Driver:** “there are many cars now” (Sw. “nu blir det många bilar”) [*domain of discussion = traffic*]

**Passenger:** “yep” (Sw. “japp”)

[WORKLOAD = LOW]
*Pause for 4 seconds*

**Passenger:** “er” (Sw. “ehm”)

[WORKLOAD = MANOEUVRE-UNRELATED]
*Pause for 4 seconds.*

WORKLOAD = MANOEUVRE-RELATED

**Passenger:** "then we will in the end of the s PAUSE treet we will turn right into Otterhälle Street PAUSE" (Sw. "sen så ska vi i slutet på g PAUSE atan ska vi svänga in höger på Otterhällegatan PAUSE")

WORKLOAD = LOW

**Passenger:** "can we look at for that should be pretty soon here it looks like but it's probably yes we can take" (Sw. "kan vi hålla in utkik på det skall vara rätt så snart här ser det ut som men det är väl ja vi kan ta") [*domain of discussion = navigation, topic-resume/reraise*]

**Driver:** "but I must turn right first" (Sw. "men jag ska ju svänga höger först")

There is a lot going on in this situation. The participants are discussing the navigation instruction and are not sure where to go next. The driver makes a comment about the traffic (*"there are many cars now"*), as if to indicate that the driving task is demanding at the moment. The passenger responds to the comment and after that both speakers pause, assumingly to let the driver concentrate on the driving task. The passenger then utters *"er"*, a filled pause, as he is reading the navigation instructions, which also may function as a signal to the driver that the passenger is about to make an utterance. Then there is another pause before the passenger resumes the navigation discussion. As mentioned, the passenger's workload is not measured, but the disfluent talk (i.e hesitation (*"s PAUSE treet"*), self-correction ("look at* for that") and incomplete sentences ("but it's probably*", "yes we can take*")) indicates that the passenger's workload level is high too. We know that the driver's workload is high (manoeuvre-unrelated workload followed by manoeuvre-related workload) and we can therefore assume that the driving task is still demanding. However, it seems like the passenger resumes the navigation discussion at the right time, since the driver is able to answer immediately and continue the navigation discussion. It seems like the passenger's strategy is successful; he pauses the dialogue during the most intense part of the driving and then resumes the navigation discussion in order to help the driver figure out where to go next. We can see that the driver's workload increases, and the resumption attempt is successful.

**Interview domain** Since topics within the Interview domain are neither time- nor safety-critical, our hypothesis is that the participants will discuss these topics during low workload when the driving is less demanding. We therefore expect resumption utterances to be less frequent during the two high workload types and state of alert, and more frequent during low workload.

Figure 4.4 shows the number of resumption utterances per minute when resuming an Interview topic.

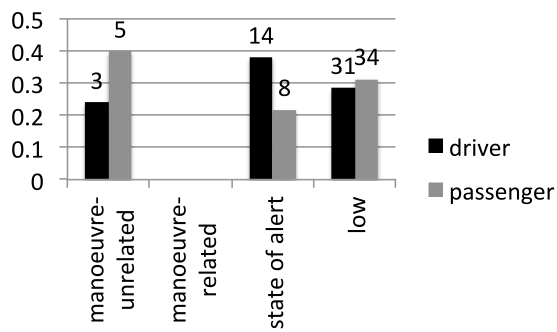



Figure 4.4: Resumption utterances per minute when resuming an Interview topic during various workload types.

All in all, there are 95 resumption utterances made to the Interview domain. The drivers resume interview topics most frequently during state of alert and low workload, and least frequently during manoeuvre-related workload (no attempts) and manoeuvre-unrelated workload. The passengers resume these topics most frequently during manoeuvre-unrelated and low workload, and least frequently during manoeuvre-related workload (no attempts) and state of alert.

The results indicate first of all that no resumptions are made during manoeuvre-related workload, which is consistent with our hypothesis. The driver does not seem to avoid resumption during state of alert. This is perhaps somewhat surprising, but indicates that as long as the manoeuvre is not demanding enough to affect the workload, the driver is capable of discussing matters that are not related to the driving task while making the manoeuvre. The reason why the frequency of resumption utterances for the driver is even higher during state of alert than during low workload we do not know. However, it may have to do with the fact that the frequency of interruptions is higher during state of alert than during low workload, and hence the number of topics to be resumed are higher too.

The passenger, on the other hand, seems to avoid resumption of interview questions when the driving task is evidently more demanding (i.e. during manoeuvre-related workload and state of alert), but not during manoeuvre-unrelated workload. This makes sense since the passenger is able to watch both the traffic situation and the driver's manoeuvres, but has less direct access to the driver's workload level. Therefore, the passenger may not always be aware that the driver's workload is high, if the traffic situation is less intense and it is not yet time to give the next navigation instruction, and therefore resumes the interview discussion. Also, when looking at the utterances made during this condition, we see that some of the resumption attempts made by the passengers are made when the driver seems to indicate in some way that they are ready to talk about something else other than the navigation task. See, for example, Example (29). The participants are unsure of where to go next due to a complicated navigation instruction, but decides to continue along the street:

(29) [WORKLOAD = LOW]

**Passenger:** "you should go straight ahead in any case" (Sw. "då ska du köra rakt fram i vilket fall") [*domain of discussion = navigation*]

**Driver:** "okay then let's continue further" (Sw. "okej då kör vi vidare")

> [WORKLOAD = MANOEUVRE-UNRELATED]
> *Pause for 1 second.*

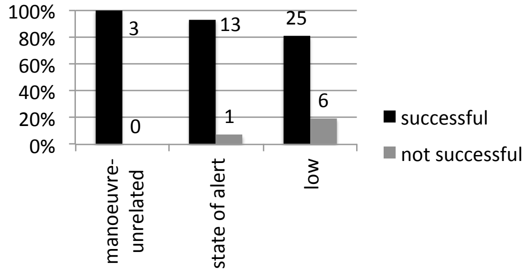
**Passenger:** "what fruit did you say grapes you said" (Sw. "vad sa du för frukt vindruvor sa du") [*domain of discussion = interview, topic-resume*]

It is hard to say why the workload increases in this situation, as the participants have decided on a plan regarding how to deal with the complicated navigation instruction. Perhaps the driver is still thinking about the instruction and tries to figure out what it means. After the short pause, that together with the last comment from the driver may be interpreted as an ending of the navigation discussion, the passenger resumes the earlier interrupted interview discussion, and the driver seems to agree on the domain shift.

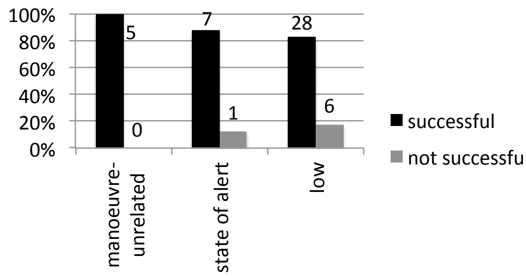
Figure 4.5 shows the number of successful and unsuccessful utterances when resuming an interview topic during each workload type, for both speakers.

Both speakers are successful in all their attempts to resume an interview topic during manoeuvre-unrelated workload, and in all but one attempt each when resume a topic during state of alert. Least successful are the attempts to resume an interview topic during low workload - about 20% of the these attempts fail. None of the speakers tries to resume an interview topic during manoeuvre-related workload.

In the cases where the speaker fails to resume an interview topic, the participants are discussing the next navigation instruction and the dialogue partner is not prepared to switch to the Interview domain until it is clear where they should go next. See Example (30), where the passenger has proposed to discuss signs of the Zodiac before they interrupted the interview discussion to switch to the navigation domain:




(a) Resumption utterances made be the driver.




(b) Resumption utterances made be the passenger.

Figure 4.5: Success rate when resuming an interview topic.

(30) [WORKLOAD = STATE OF ALERT]

**Passenger:** "and then we should cross Ascheberg Street PAUSE which is not this one but the one a little bit further ahead PAUSE so we should go straight ahead a bit" (Sw. "så ska vi korsa Aschebergsgatan PAUSE vilket är inte den som är här utan lite längre fram PAUSE så vi skall rakt fram en bit") [*domain of discussion = navigation*]

WORKLOAD = LOW
*Pause for 4 seconds.*

**Passenger:** "star" (Sw. "stjärn") [*domain of discussion = interview, topic-resume*]

**Driver:** "it's not this one" (Sw. "det är inte den här det") [*domain of discussion = navigation, topic-interrupt*]

**Passenger:** "nope" (Sw. "nä")

**Driver:** "nope" (Sw. "nä")

**Passenger:** "star sign" (Sw. "stjärntecken") [*domain of discussion = interview, topic-resume*]

**Driver:** "Scorpio" (Sw. "skorpion")

The passenger gives the navigation instruction, to which the driver does not respond but instead there is a pause for 4 seconds. It is possible that the passenger interprets the silence as an acceptance from the driver that the navigation discussion is ended, at least the passenger seems to think so as he switches domain to the Interview domain and tries to resume the earlier discussion about star signs. However, the attempt is unsuccessful as the driver interrupts to further clarify the instruction, and after that the next resumption attempt is successful. This example shows that it is important to confirm that an utterance is understood before switching to another topic and/or domain, especially if the instruction that has been given has not being performed yet.

### 4.6.2 How to resume an interrupted topic

Regarding how to express oneself when resuming an interrupted topic, we are curious about two things. First, we want to learn more about what information to include in the resuming utterance, and therefore we have analysed resumption utterances with respect to redundant information. Second, we want to know more about the grammatical structure of the resuming utterances, and have therefore categorized the utterances according to grammatical category.

### Redundancy in resuming utterances

As mentioned above, when resuming an interrupted dialogue with a dialogue system, the probably easiest strategy for system-initiated resumption is to simply repeat the last utterance made by the system. But when humans resume an interrupted dialogue, do they always repeat parts of what was said before the interruption? Or could the length of the interruption influence the occurrence of redundant information, so that repeating some of the information that has been shared may be of less importance after a short interruption when the interrupted dialogue is still easily accessible in the dialogue partners memory? To find out more about this we have analysed the corpus with respect to redundancy in resumption utterances (see Section 4.5 for more information about the method used).

**Navigation domain** Figure 4.6 shows the distribution of redundant and non-redundant utterances when resuming a navigation topic. The bars are labelled with number of utterances.

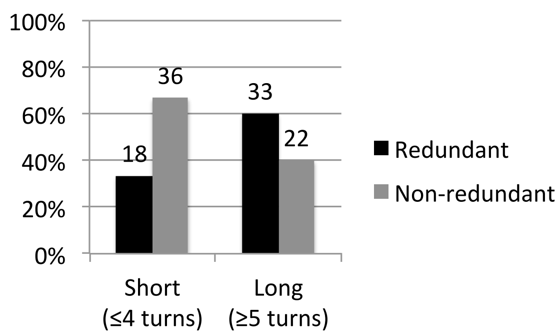



Figure 4.6: Redundancy when resuming a navigation topic, depending on length of interruption

As shown in Figure 4.6, in the Navigation domain long interruptions are more likely to result in redundant resumptions ($\chi^2_{(1)} = 7.78, p < 0.01$). 33% of the resuming utterances after a short interruption contained redundant information when comparing with the utterances made before the interruption. 67% contained no redundant information. After a long interruption, 60% of the resuming utterances contained redundant information, and 40% did not. The average number of turns when no redundancy occurs is 5.1, and when there is redundancy the average number of turns is 7.9.

**Interview domain** Figure 4.7 shows the distribution of redundant and non-redundant utterances when resuming an interview topic. The bars are labelled with number of utterances.

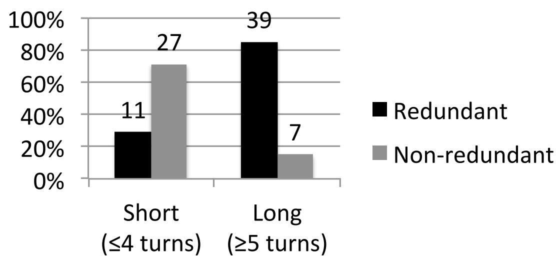



Figure 4.7: Redundancy when resuming an Interview topic, depending on length of interruption

As shown in Figure 4.7, also in the Interview domain long interruptions are more likely to result in redundant resumptions ($\chi^2_{(1)} = 26.93, p < 0.01$). On average, there are 9.3 turns during an interruption. 29% of the resuming utterances after a short interruption (i.e. $\leq$4 turns) contained redundant information when comparing with the utterances made before the interruption, 71% contained no redundant information. After a long interruption (i.e $\geq$5 turns), 85% of the resuming utterances contained redundant information, and 15% did not. The average number of turns when no redundancy occur is 5.0, if we remove the two extremes (of 30 and 33 turns, respectively), the average number of turns is about 3.5. When there is redundancy the average number of turns is about 12.0.

So, after a short interruption the speaker tends to continue the interrupted topic where it was interrupted, without considering that there has been any interruption. Example (31) shows, for example, how the passenger interrupts a discussion about holiday trips to give a navigation instruction:

(31) [WORKLOAD = LOW]

**Passenger:** "yes there are some nice roads to drive there" (Sw. "ja det är ju en del fina bergsvägar där å köra") [*domain of discussion = interview*]

**Driver:** "yes it was a bit exciting" (Sw. "ja det var lite spännande") [*last driver utterance before interruption*]

**Passenger:** "yes I can imagine that" (Sw. "ja det kan jag tänka mig") [*last passenger utterance before interruption*] "turn to Järntorget here then" (Sw. "svänger vi upp mot Järntorget här då") [*domain of discussion = navigation, topic-interrupt, topic-resume/reraise*]

**Driver:** "yes" (Sw. "yes")

*Pause for 10 seconds.*

**Driver:** "no they have different you get something else to think about there than" (Sw. "nä dom har lite andra där får man lite annat att tänka på än") [*domain of discussion = interview, topic-resume*]

After a long interruption (on average, 12 turns) the resuming speaker usually makes a redundant utterance, repeating one or more content words from the last utterance before the interruption. See Example (32), where boldface indicates shared lexical content across utterances:

(32) [WORKLOAD = MANOEUVRE-UNRELATED]

**Passenger:** "points for motivation it says here well should you motivate why you read that then or" (Sw.

"poäng för motivering står det här ja ska du motivera varför du läste det kanske då eller") [*domain of discussion = interview, last passenger utterance before interruption*]

**Driver:** "well I think it's really because PAUSE I have always had eh an **interest** in **computing** and **technology** so" (Sw. "tja det är väl egentligen PAUSE jjag har alltid haft eh **intresse** av **data** och **teknik** egentligen") [*last driver utterance before interruption*]

**Passenger:** "turn right into Vasaplatsen is that here PAUSE no it's Grönsakstorget there" (Sw. "höger in på Vasaplatsen är det här det PAUSE nej det är Grönsakstorget där") [*domain of discussion = navigation, topic-interrupt , topic-begin*]

*navigation discussions, all in all 21 turns*

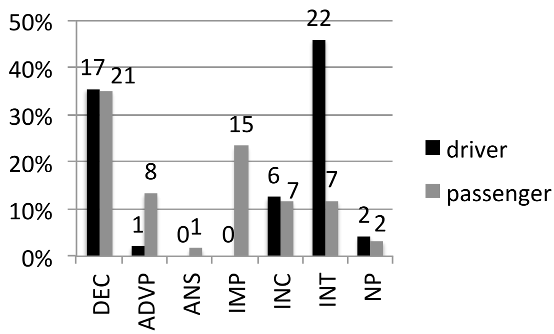
**Driver:** "yes no as I said [*sequencing-move*]I have always been **interested** in **computing** and **computers** and **technology** and stuff like that" (Sw. "ja nä som sagt jag har alltid varit **intresserad** av **dator** och **datorer** och **teknik** och så där") [*domain of discussion = interview, topic-resume*]

Here, several content words ("interested", "computing", "technology") appear in both the resuming and the preceding utterances, providing evidence of redundancy.

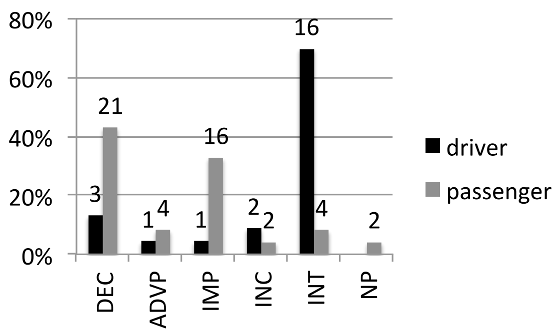
#### Grammatical category of resuming utterance

Finally, we have investigated how the participants express themselves with respect to grammatical category of the utterance (see Section 4.5) when resuming an interrupted topic within the Interview and Navigation domains. We want to find out what type of grammatical category the speakers use, and compare to the grammatical category that is used when beginning a new topic that has not been discussed before, to see if the behaviours differ. We have included topics beginning and being resumed/reraised both after an ended topic and after an interrupted topic.

**Navigation domain** Figures 4.8a and 4.8b show the distribution of grammatical categories when resuming and beginning a navigation topic.




(a) Topic-resume




(b) Topic-begin

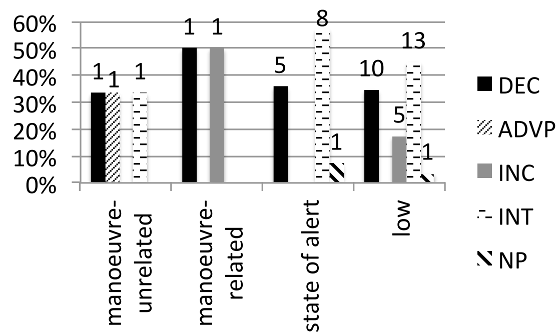
Figure 4.8: Distribution of grammatical categories when resuming and beginning a navigation topic, for each speaker.

When resuming a topic within the Navigation domain, the driver mostly uses an interrogative or a declarative phrase, typically to clarify an instruction or repeat key phrase in the latest instruction. For example, *"you said left towards downtown here" (Sw. "du sa centrum vänster här")*, *"this is Götaberg Street" (Sw. "det här är Götabergsgatan")*. Incomplete phrases are typically self-interrupted interrogative phrases such as *"should I" (Sw. "ska jag")*, which the passenger answers by clarifying which way to go. The passenger also uses declarative phrases, and also imperative phrases which are used as a reminder of the last instruction, such as *"keep straight on" (Sw. "fortsätt rakt fram")* or *"turn right into Vasa Street" (Sw. "sväng höger in på Vasagatan")*. Adverbial phrases, such as *"via the slip road here" (Sw. "via påfarten här")* or *"down there" (Sw. "ner där")*, are also used by the passengers in a similar way as the imperative phrases. Imperative and adverbial phrases seem to be used to clarify that they have reached the point where the driver needs to make a manoeuvre.

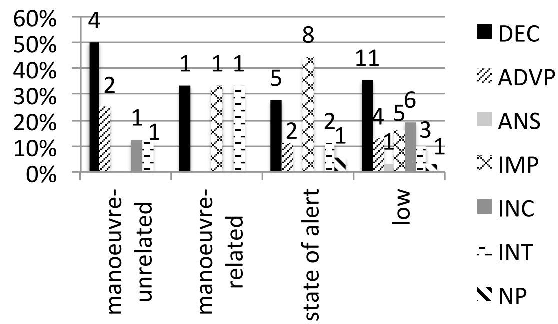
When comparing the way the participants express themselves when beginning a new topic to how they express themselves when resuming a topic, we see that especially the driver's behaviour differs. Both when beginning a new topic and when resuming an earlier topic the driver most often uses interrogative phrases, but the use of declarative phrases is not as common when starting a new topic as it is when resuming a topic. The purpose of using declarative phrases when resuming a navigation topic seems to be to inform the passenger that they have reach a certain point, and with that indirectly ask for the next instruction. For example, the driver often states where they are by using a declarative phrase (as in the example above when the driver declares *"this is Götaberg Street" (Sw. "det här är Götabergsgatan")*. When doing so, the passenger often responds by giving the next instruction. The passenger's behaviour does not differ when comparing new topics to resumed topics. Similar to when resuming a topic, when beginning a new topic they most often use declarative and imperative phrases.

The purpose of switching to the navigation domain seem to be the same for both speakers both when beginning a new topic and when resuming an earlier discussed topic, with the driver eliciting information and the passenger giving information. It may be more common for the driver to elicit information in a more indirect way when resuming a topic compared to when starting a new discussion, but this needs to be further investigated. The passenger, however, seems to express themselves in a similar way both when resuming and when starting a new topic.

Figure 4.9 shows the distribution of grammatical categories during various workload types when resuming a navigation topic.




(a) Resumption utterances made be the driver.




(b) Resumption utterances made be the passenger.

Figure 4.9: Distribution of grammatical categories during various workload types when resuming a topic within the Navigation domain.

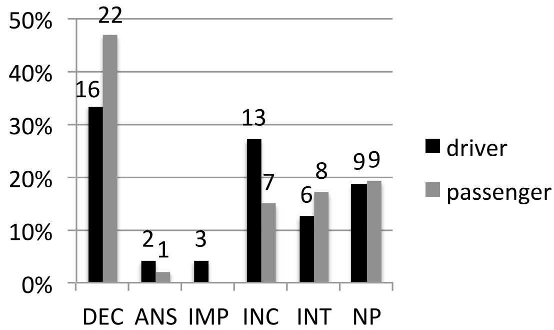
With so few instances distributed over several grammatical categories it is not possible to draw any strong or even moderately strong conclusions, so our analysis will be rather sketchy. The purpose of making this comparison is to look for indications of different behaviours during various workload types, and see if it seems meaningful to investigate this further in future studies.

Figure 4.8 told us that declarative phrases are most common to use for both speakers when resuming a navigation topic (together with interrogative phrases for the drivers). Figure 4.9 indicates that the use of declarative phrases seem to be rather equally distributed irrespective of workload type, for both speakers. The most common phrase used by the drivers during state of alert and low workload is an interrogative phrase. During the two high workload types an interrogative phrase is only used once, during manoeuvre-unrelated workload. Overall, there are few resumption attempts made during the two high workload types.

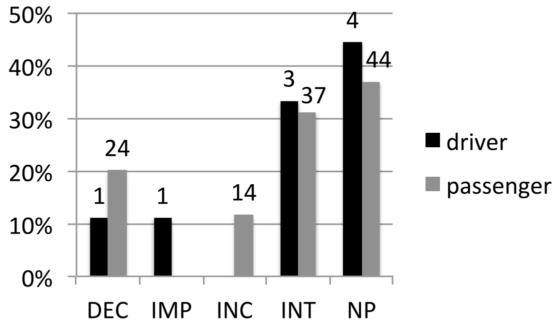
The most common phrase type to use during manoeuvre-unrelated and low workload, when the driving manoeuvre being performed is less demanding, is a declarative phrase. During state of alert, when the driving task is more demanding, the most common phrase to use is an imperative phrase which is also one of three phrases used during manoeuvre-related workload.

The result indicates that when the driver is making a more demanding driving manoeuvre the passenger tends to use imperative phrases more than declarative phrases, and when the driver is not making a demanding driving manoeuvre we see the opposite pattern. It seems as if the driving task is less demanding the passenger is reasoning about the navigation instructions using declarative phrases (for example, *"this is Ekelund Street"*, *"we will end up at Älvsborg Bridge if we go this way"*, *"now Olivedal Street is coming up I think"*), while they give more clear commands that may be easier to follow when the driving task is more demanding (for example, *"take right into Viktoria Street"*, "continue twohundred and fifty meters").

**Interview domain** Figures 4.10a and 4.10b show the distribution of grammatical categories when resuming and beginning an interview topic.




(a) Topic-resume




(b) Topic-begin.

Figure 4.10: Distribution of grammatical categories when resuming and beginning an interview topic.

Within the Interview domain, both speakers most often use a declarative phrase when resuming an interrupted topic, for example, *"okay you are Leo and I am Gemini" (Sw. "jaha lejon är du och jag är tvilling")*. Noun phrases, incomplete phrases and interrogative phrases are also used. Often, the participants use incomplete phrases and noun phrases to repeat parts of the interview question. For example, the driver once resumed a discussion about winter hobbies by saying *"my hobbies during the winter were w" (Sw. "min vintersysselsättningen va v")*, and the passenger resumed a discussion about professions by saying *"dream career" (Sw. "drömyrke")*. The noun phrase is also often a key phrase in the dialogue partner's earlier answer, as in the example we have mentioned briefly before where the participants are discussing their favourite band:

(33) [WORKLOAD = LOW]

**Driver:** "Wolfmother if you know it" (Sw. "Wolfmother om du känner till det") [*domain of discussion = interview*]

**Passenger:** "never heard of it LAUGHS let's see if I can remember that" (Sw. "aldrig hört talas om LAUGHS ska vi se om jag kommer ihåg det")

**Driver:** "LAUGHS are gonna be hard to remember there eh later" (Sw. "LAUGHS blir svåra att komma ihåg där eh senare")

**Passenger:** "eh turn left into Götaberg Street is what we should do next" (Sw. "eh kör vänster in på Götabergsgatan ska vi göra sen") [*domain of discussion = navigation, topic-interrupt*]

*navigation and traffic discussions, all in all about 10 turns*

**Passenger:** "Wolfmother" (Sw. "Wolfmother") [*domain of discussion = interview, topic-resume, grammatical category = NP*]

**Driver:** "Wolfmother is their name" (Sw. "Wolfmother heter de")

The name of the band was unfamiliar to the passenger, but probably since it is a somewhat peculiar name it stands out and all the passenger had to do to resume the interview discussion after the interruption was to repeat the name of the band. The discussion was interrupted once more, and once again the passenger resumed by uttering the single word *"Wolfmother"*.

The second most common phrase that the drivers are using is an incomplete phrase, which may have to do with the fact that the driver primarily needs to concentrate on the driving task and therefore is less able to express themselves in a correct way.

In the cases where the driver resumes with a yes/no answer, the answer is always followed by an explanation in some form, usually either a noun phrase or, as in Example (34), a declarative phrase:

(34) [WORKLOAD = STATE OF ALERT]

**Passenger:** "do you have a boat or a caravan or something like that" (Sw. "har du båt eller husvagn eller nåt sånt där") [*domain of discussion = interview*]

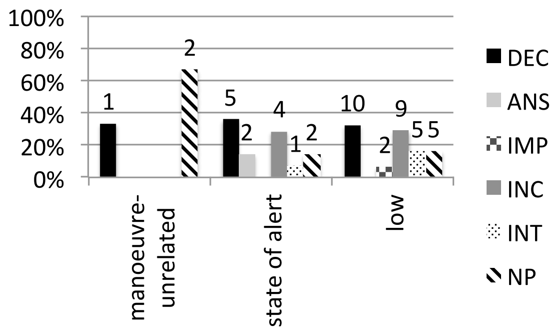
**Driver:** "you did say downtown left here" (Sw. "du sa centrum vänster här") [*domain of discussion = navigation, topic-interrupt*]

**Passenger:** "downtown yes that's right" (Sw. "centrum ja jajemen")

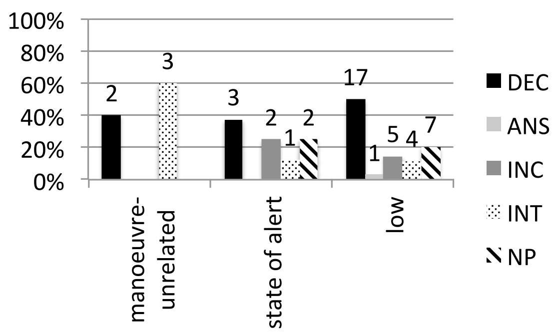
**Driver:** "no PAUSE" (Sw. "nej PAUSE") [*domain of discussion = interview, topic-resume, grammatical category = ANS*] "I have a house" (Sw. "jag har hus") [*domain of discussion = interview, topic-resume, grammatical category = DEC*]

As it is the passenger who is in charge of the interview task and is holding the form with the interview questions (see Section 1.4), the driver seldom starts a new interview topic. It is therefore hard to make any comparisons of the behaviour when starting a new topic compared to when resuming an earlier topic. The passenger is more active both when starting and resuming an interview topic. When starting a new topic they most often use either a noun phrase or an interrogative phrase, compared to a declarative phrase when resuming a topic. Noun phrases are used in the same way as interrogative phrases; to propose a new question to discuss, while resuming delarative phrases are often used to summarize information that has been given before the interruption, as in the example above (*"you are Leo and I am Gemini" (Sw. "jaha lejon är du och jag är tvilling")*).

Figure 4.11 shows the distribution of grammatical categories during various workload when resuming an interview topic.




(a) Resumption utterances made be the driver.




(b) Resumption utterances made be the passenger.

Figure 4.11: Distribution of grammatical categories during various workload when resuming a topic within the Interview domain.

Similar to the Navigation domain, declarative phrases are most common to use (see Figure 4.10), and for both speakers the distribution of declarative phrases are rather equal between the workload types. This may indicate that the use of declarative phrases within the Interview domain is not affected by high workload or demanding manoeuvres (in contrast to what we could see when looking at the passenger's resumption utterances within the Navigation domain), but more data is needed to draw such conclusions. None of the speakers use incomplete phrases during high workload, but since there are only a few resumption utterances made during this condition we do not know if this behaviour is significant.

## 4.7 Implications for a dialogue system

We have looked at resumption strategies in human-human dialogue, when resuming topics within the Navigation and Interview domains during various workload types. In this section, we will suggest how these strategies could be applicable in an in-vehicle dialogue system. Similar to what we did in Section 3.8, we will propose strategies for in-vehicle dialogue systems by learning from the participants' behaviour when discussing topics within the Navigation and Interview domains. As the passenger is the one providing the navigation instructions and asking the interview questions, we map the passenger to the system and the driver to the user.

### 4.7.1 Resumption during high and low workload

Our investigation of resumption strategies during various workload types indicates that the participants use different strategies depending on whether the topic is within the Navigation or the Interview domain. Navigation topics are often resumed when the driver's workload is high or there is a state of alert, while Interview topics are often resumed during low workload or state of alert. This indicates that an in-vehicle dialogue system should consider not only the driver's workload level and type, but also the type of application that is to be resumed when deciding when to resume an unfinished topic.

**Resumption of a navigation topic** Our results when analysing resumption behaviour indicate that resumption of a navigation topic by the passenger is most frequent during high workload or state of alert, see Figure 4.12 (presented in Section 4.6.1 but repeated here for convenience).

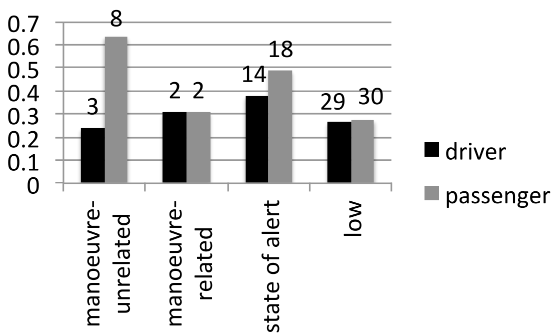



Figure 4.12: Utterances per minute when resuming/reraising a Navigation topic during various workload types.

In Section 3.8, we proposed that the dialogue should be paused during high workload and state of alert, until the workload is low. Inspired by the result presented above, we now propose a modification of this strategy. System-initiated resumption of the latest navigation instruction can be done during these conditions, provided that the latest navigation instruction has not yet been executed by the driver.

**Strategy 4.1:**

- if workload $\neq$ low AND the latest navigation instruction has not been executed
  - ∗ repeat last navigation instruction

Using this strategy, a (made up) interaction during manoeuvre-related workload or state of alert would look like this:

(35) [WORKLOAD = LOW]

```
System:  In 1 kilometer, turn left.
```

*The driver continues along the road, and is now close to a crossing.*

[WORKLOAD = MANOEUVRE-RELATED, MANOEUVRE-UNRELATED OR STATE OF ALERT]

```
System:  Continue 400 meters, then turn left.
```

In this example, the navigation system has given a navigation instruction on where to make a left turn. When the driver is approaching a crossing, the workload increases or there is a state of alert, possibly due to confusion of whether this is the crossing where s/he should turn left. The system takes the initiative to resume to the Navigation domain to repeat the last navigation instruction, in order to clarify that this is not the crossing where the driver should turn.

**Resumption of a driving-unrelated topic** We propose that system-initiated resumption of a driving-unrelated topic should be done during low workload. Our results indicate that the passenger more frequently resumes an Interview topic during manoeuvre-unrelated or low workload, see Figure 4.13 (presented in Section 4.6.1 but repeated here for convenience).

We propose that system-initiated resumption of a driving-unrelated task should be done only during low workload:

**Strategy 4.2:**

- if workload $\neq$ low AND there is an unfinished driving-unrelated task
  - wait until workload = low
  - then resume any unfinished driving-unrelated task

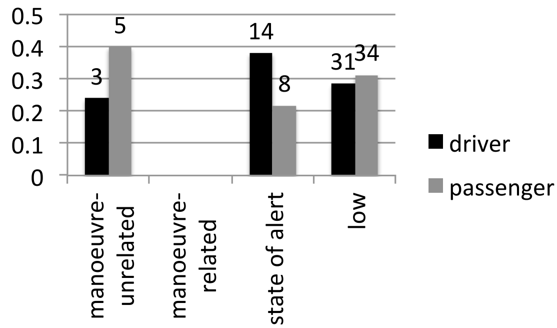



Figure 4.13: Frequency of utterances resuming/reraising an Interview topic during various workload types.

Using this strategy, we elaborate Example (22) from Section 3.8 where the interaction with the radio application is interrupted, and then resumed:

(36) [WORKLOAD = LOW]
*The driver starts an interaction with the dialogue system.*

```
System:  What do you want to do?
```

[WORKLOAD = MANOEUVRE-RELATED]
*The next navigation instruction is labelled "urgent".*

```
Driver:  Change radio station.
System:  At the next crossing, make a left
turn.
Driver:  Ok.
```

*The driver is making the left turn.*
*Pause n seconds.*
[WORKLOAD = LOW]

```
System:  What frequency or preset radio station
do you want to change to?
```

```
Driver:  88.5
System:  Ok, tuning to 88.5.
```

The interaction with the radio application is interrupted by the system, which gives an urgent navigation instruction during manoeuvre-related workload. When the instruction has been given and the workload has decreased, the system takes the initiative to resume the interrupted interaction with the radio application.

### 4.7.2 Redundancy in resumption utterances

When analysing resumption utterances with respect to the length of the interruption, we found that after a short interruption the participants often did not repeat information that had already been given. After a long interruption, they often repeated (parts of) the information to remind the dialogue partner of the topic that had been interrupted. See Figures 4.14 and 4.15 (presented in Section 4.6.2 but repeated here for convenience).

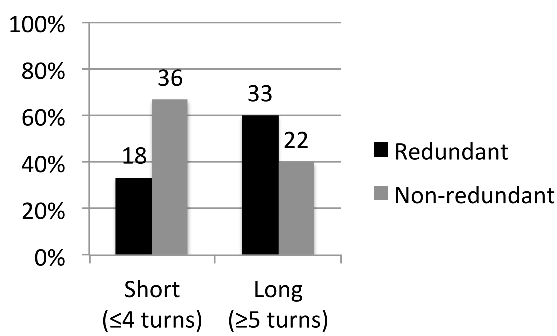



Figure 4.14: Redundancy when resuming a navigation topic, depending on length of interruption.

The behaviour is similar within both domains, and therefore we find it plausible to assume that the resumption strategy when it comes to redundancy within the resuming utterance is the same

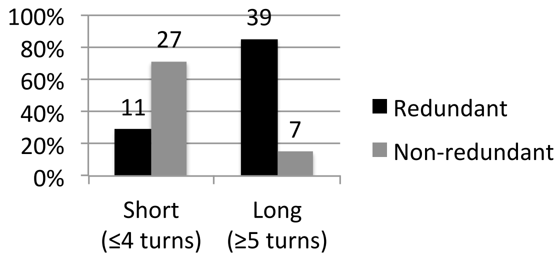



Figure 4.15: Redundancy when resuming an interview topic, depending on length of interruption.

for both driving-related and driving-unrelated domains. We suggest that after a short interruption, the dialogue system may resume the interrupted interaction without repeating earlier utterances, while after a long interruption key words within the latest utterances should be included in the resuming utterance.

**Strategy 4.3:**

- if resuming a topic n turns after it was interrupted
  - then if $n \leq 4$
    - resume using next utterance
  - else if $n \geq 5$
    - resume using key words from latest utterances

Using this strategy, a (made up) interaction when resuming an interrupted interaction after a short interruption would look like this:

(37) WORKLOAD = LOW
*The driver starts an interaction with the dialogue system.*

```
System:  What do you want to do?
```

WORKLOAD = MANOEUVRE-RELATED
*The driver is close to a crossing and the next navigation instruction is labelled "urgent".*

```
Driver:  Call John.
System:  At the next crossing, make a left
turn.
Driver:  Ok.
System:  The work number or the mobile phone?
Driver:  Work number.
System:  Ok, calling John's work number.
```

The system interrupts to give a navigation instruction and the driver confirms it, all in all two turns. When resuming interaction with the phone application, the system continues with the next question instead of repeating information that has already been given. After a long interruption, it may be necessary to repeat key words from the interrupted dialogue:

(38) WORKLOAD = LOW
*The driver starts an interaction with the dialogue system.*

```
System:  What do you want to do?
```

WORKLOAD = MANOEUVRE-RELATED
*The driver is close to a crossing and the next navigation instruction is labelled "urgent".*

```
Driver:  Call John.
system:  Make a left turn now.
driver:  Ok.
```

*The driver makes the left turn.*

```
System:  Continue 3 kilometers, then turn
right.
Driver:  Ok.
System:  You will need a refill in less than
10 kilometers.
```

```
Driver:  Ok.

System:  Do you want to call John using the
work number or the mobile phone number?

Driver:  Work number

system:  Ok, calling John's work number.
```

After a long interruption, the keywords *"call"'* and *"John"* are included in the resuming utterance to remind the driver about what has been agreed upon so far.

### 4.7.3 Grammatical category

The results reported in Section 4.6 indicate that the passengers most often use declarative phrases when resuming both navigation and interview topics. However, during high workload and state of alert the behaviour seems to change, and the passenger use more imperative phrases when resuming a navigation topic and more interrogative phrases when resuming an interview topic.

**Grammatical category when resuming a navigation topic** Overall, when resuming a navigation topic, the passenger most often uses delarative phrases, but during state of alert (and possibly manoeuvre-related workload) they may use imperative phrases more often than declarative phrases, although the evidence for this is quite weak. See Figures 4.16 and 4.17 (presented in Section 4.6 but repeated here for convenience).

If these results hold up, then an in-vehicle dialogue system may adjust its utterance according to the driver's workload type, and use imperative phrases during manoeuvre-related workload and state of alert and declarative phrases during low workload:

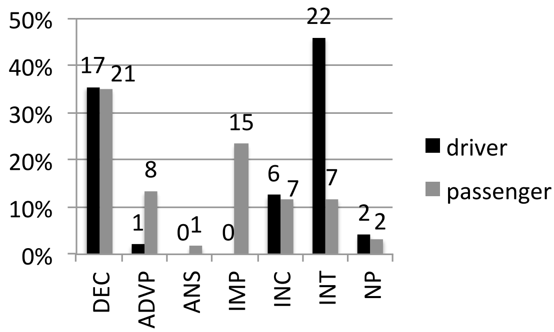



Figure 4.16: Distribution of grammatical categories when resuming a navigation topic.

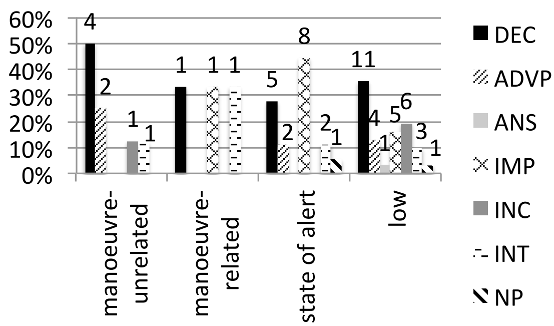



Figure 4.17: Distribution of grammatical categories during various workload when the passenger is resuming a navigation topic.

**Strategy 4.4:**

- if resuming a driving-related topic
  - then if workload = manoeuvre-related OR state of alert
    - use imperative phrase
  - else if workload = low
    - use declarative phrase

Using this strategy, the interaction during manoeuvre-related workload and state of alert would look like this:

(39) WORKLOAD = MANOEUVRE-RELATED OR STATE OF ALERT

```
System:  Turn left now.
```

During low workload, the interaction would look like this:

(40) WORKLOAD = LOW

```
System:  This is where you should turn left.
```

**Grammatical category when resuming driving-unrelated topics**
When resuming interview topics, the passengers most often use declarative phrases, similar to when resuming navigation topics. However, during manoeuvre-unrelated workload they more often use interrogative phrases. See Figures 4.18 and 4.19 (presented in Section 4.6 but repeated here for convenience).

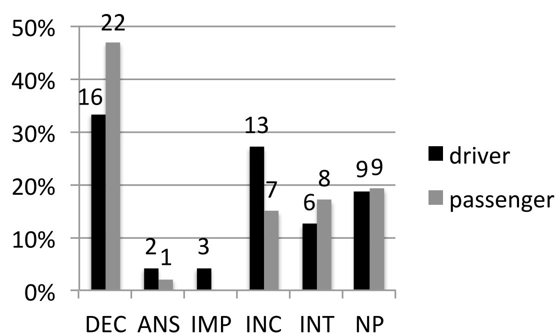



Figure 4.18: Distribution of grammatical categories when resuming an interview topic.

**Strategy 4.5:**

– if resuming a driving-unrelated topic

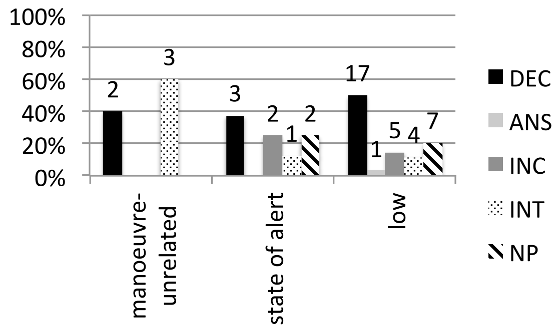



Figure 4.19: Distribution of grammatical categories during various workload when the passenger is resuming an interview topic.

- ∗ then if workload = manoeuvre-unrelated
  - · use interrogative phrase
- ∗ else if workload = low
  - · use declarative phrase

(41) Using this strategy, the system would resume a driving-unrelated application using by asking a question about what to do next:
WORKLOAD = MANOEUVRE-UNRELATED

```
System:  Which frequency or preset radio station
do you want to change to?
```

During low workload, the system would instead use a declarative phrase to remind the user of the interrupted task.

(42) WORKLOAD = LOW

```
System:  You wanted to change radio station to another
frequency or preset radio station.
```

## 4.8 Conclusions and future work

We started our investigation of resumption strategies by looking at the behaviour when resuming a topic within the Navigation and Interview domains during all types of workload. The results indicate that navigation topics are resumed more often during high workload and state of alert, when it is reasonable to believe that the driver could benefit from resuming the latest navigation instruction – either by repeating the instruction or by further clarifying when to make the next manoeuvre.

A common strategy in today's navigation systems seems to be to repeat a navigation instruction within certain intervals. For example, to give the instruction the first time within a certain distance from a crossing where the driver should turn, then repeat the instruction when the distance is divided into half and finally repeat the instruction one last time just before it is time to make the turn (this is a made up example with fictitious intervals). Future research should include taking the driver's workload as well as the traffic situation into account when deciding when to repeat a navigation instruction. It may be a better strategy to repeat a navigation instruction if the driver's workload is high (including both manoeuvre-related and manoeuvre-unrelated workload) in combination with a potentially confusing traffic siuation (for example, if there are more than one crossing within a short distance from where the driver should turn), than to wait until the driver is within a certain distance from where the next manoeuvre is to be performed.

Also when resuming driving-unrelated tasks it should be investigated if taking the driver's workload into account when deciding when to resume is a good strategy to avoid increased workload. Our results when analysing the resumption behaviour when resuming a topic within the Interview domain indicate that these topics are often resumed during low workload and state of alert. Future studies should be made to investigate if it is a good strategy to resume driving-unrelated tasks (for example, entertainment systems like the phone or the radio application) during these conditions. Especially, it should be investigated if it is a good strategy to let the system take the initiative to resume tasks during state of alert. Our study indicates that the driver most often resumes interview topics during this condition, and it is probably preferable to let the driver take the initiative for any resumption of an interrupted interaction while the driving task is demanding. However, the result indicates that the driver may be able to interact with the system during at least some of the manoeuvres that are assumed to be demanding, and future studies of this behaviour should include an investigation of whether it is possible to distinguish between suitable and unsuitable driving manoeuvres with respect to simultaneous dialogue system interaction.

Although quite uncommon, the unsuccessful resumption attempts, within both domains, seem to be made when the driver is busy with an ongoing task - either answering an interview question or trying to find a certain street or interpreting a navigation instruction. In these situations, the passenger often resumes the topic without using any standard phrase or in other ways signal that s/he is about to switch domain. The best strategy is probably to avoid domain switches during these conditions, but if it is necessary (for example, to give time-critical information), a possible strategy to help increase the success rate may be to use a notification signal of some sort. This strategy may be useful in any situation where the system needs to interrupt an ongoing task, in order to facilitate the later resumption of the task. As mentioned in Section 4.2, Paul et al. (2015) found that a notification signal announcing an upcoming interruption helps also when resuming the interrupted task; the notification helps them remember the status of the multiple tasks they are performing, and therefore makes it easier to remember the interrupted task and continue where it was interrupted. Future testing of this behaviour should also include what type of notification signal is preferable for an in-vehicle dialogue system, see Section 3.9.

Regarding redundancy when resuming an interrupted topic, this behaviour remains to be tested when interacting with a dialogue system. These tests should include trying to find the threshold regarding the length of a short and a long interruption, respectively. Our results indicate that in human-human dialogue a short interruption is about 4 turns long, after which the participants usually did not repeat information that had been given before the interruption. When interacting with a dialogue system the threshold may be different.

Also, redundancy should be investigated after a pause, to see if and how the pause length influences the use of redundant information when resuming the dialogue. We also need to know more about exactly what information is repeated, and if the place of interruption play a role in how much of the dialogue is repeated. Yang et al. (2011) found that different techniques were used depending on whether a discussion was interrupted in the beginning, middle or end of a card game that the participants were playing. There might be similar differencies in in-vehicle dialogue too, together with differencies depending on whether it is the system or the driver who is interrupting the interaction. In our study, we only looked at content words in the resuming utterance. To avoid tedious repetitions if the dialogue is interrupted several times, studies should also be made regarding paraphrasing as an alternative to repeating the exact words from the interrupted discussion. Perhaps paraphrasing could also be used to further emphasize important information.

Finally, we have analysed the grammatical category of resumption utterances during various workload. The results indicate that declarative phrases are most common for the passengers to use, both when resuming navigation and interview topics. However, during state of alert (and possibly manoeuvre-related workload although more data is needed to know this), they seem to use imperative phrases more often than declarative phrases when resuming a navigation topic, and interrogative phrases may be more common than declarative phrases when resuming an interview topic during manoeuvre-unrelated workload. This is interesting results and should be further investigated, to see if the use of different types of grammatical categories during various workload types can make it easier to understand and interpret information correctly.

## 4.9 Summary

In this chapter, we have analysed resumption behaviour. We analysed resumption utterances with respect to when a topic is resumed within the Navigation and Interview domains in relation to the driver's workload type. We have also investigated how the participants express themselves when resuming topics within these domains, both regarding redundancy in the resumption utterance when compared to the last utterances before the interruption, and regarding the grammatical categories of the resumption utterances. Based on the analyses, we proposed dialogue strategies to use for an in-vehicle dialogue system. Finally, we discussed future work issues.

# Chapter 5

# Conclusions and future work

In this thesis, we have addressed the problem of in-vehicle dialogue system interaction interfering with the driving task, as the interaction with the system may distract the driver. We have discussed the concept of a cognitive workload-aware system, i.e. a system that is able to adjust its dialogue behaviour to the driver's workload level in order to prevent or lower high workload.

To be able to do that, the dialogue system needs to know not only the driver's workload level, but also what type of workload the driver is experiencing. Different types of workload may require different dialogue strategies, and therefore the system needs to be able to distinguish between workload types. We presented a taxonomy of the workload types that we believe are most important to consider when studying simultaneous driving and interaction with a dialogue system. We distinguish between manoeuvre-related workload (i.e related to the driving task), manoeuvre-unrelated workload (i.e related to a secondary task), and state of alert (i.e driving behaviour correlating with high workload).

We have analysed the DICO corpus, containing in-vehicle human

/ human dialogue, and based on that presented a method for distinguishing between workload types by analysing the driver's workload level and driving behaviour.

As an inspiration for how a cognitive workload-aware system could use dialogue strategies to lower the driver's workload level, we have analysed the DICO corpus to learn more about dialogue strategies used by humans.

The corpus is annotated with four domains of discussion; Interview, Navigation, Traffic and Other, and it is therefore possible to study the participants behaviour when switching between the domains of discussion. We found that the passengers and the drivers often interrupt the dialogue, in different ways. We therefore investigated interruption behaviour in relation to the driver's workload. We found that when the driver's workload is related to the driving task, the participants either pause the dialogue to let the driver concentrate on the driving task, or switch domain to give or ask for navigation instructions. Our analysis showed that the participants pause significantly more often during high workload, and they make significantly longer pauses when the driving task is more demanding. The participants frequently interrupt to switch domain in order to give time-critical information. Our results indicate that interruptions to switch domain are made more often during low workload than during high workload and state of alert.

To learn more about when it is suitable to interrupt an ongoing discussion, we analysed interruption behaviour in relation to adjacency pairs. Earlier research have shown that interruptions are avoided within an adjacency pair, and we therefore wanted to know if the behaviour is the same in the in-vehicle environment where certain information (for example, navigation instructions) may be both time- and safety-critical. We found that the participants behaviour in our study of in-vehicle dialogue is similar to other studies; interruptions are significantly more likely to occur outside an adjacency pair.

We also wanted to know how the participants express themselves when interrupting to switch domain, with respect to what type of sequencing move they are using. By doing this, we wanted to see if the speaker prepare their dialogue partner of the upcoming domain switch by notifying them using a standard phrase like *"let's see"* before they switch domain, or if they switch without giving any notification. We were also interested in the grammatical structure of the interrupting utterance, to see if it differs from a non-interrupting domain switch. We found that the speakers seem to use domain-specific phrases (i.e. domain switch without notification) more often than standard phrases during all workload types. However, there may be a tendency that they use slightly more standard phrases during low workload compared to during high workload and state of alert. The behaviour seems to be similar regardless of which domain of discussion the speaker switch to. When using domain-specific phrases, declarative phrases seem to be most common within all domains of discussion.

When an in-vehicle task has been interrupted before it is finished, the user may want to resume the interrupted task in order to finish it. We therefore continued our investigation of in-vehicle dialogue by analysing human / human resumption behaviour. We started by looking at the behaviour with respect to the driver's workload type. The result, although not tested for significance, indicates that the frequency of passengers' resumption attempts is highest during manoeuvre-unrelated workload and state of alert, and lowest during low workload. The drivers resume navigation topics most frequently during state of alert, and least frequent during manoeuvre-unrelated workload.

When looking at the success rate when resuming a navigation topic, the result indicates that the passengers are successful in most of their attempts but fail a few times during state of alert and low workload. The anecdotal data indicates that passengers are good at perceiving and adjusting to the driver's workload level, but that they sometimes try to resume a topic when the driver is not yet ready to end the ongoing discussion. These tentative results emphasize the importance of further studies of this behaviour, to make it possible for an in-vehicle dialogue system to give information to the driver when they are mentally prepared to receive it.

Similar to our investigation of interruption behaviour, we wanted to know how the participants express themselves when resuming a topic. Specifically, we wanted to know if and when information that has already been shared is repeated in the resuming utterance, to learn how to make in-vehicle dialogue systems less repetitive. Therefore, we analysed the corpus with respect to redundancy. We wanted to know if the resuming utterance usually contains redundant information, and therefore compared the resuming utterance after a short (less than 5 turns) interruption and after a long (5 or more turns) interruption with the last utterances before the interruption. We found that it is significantly more likely that an utterance contain redundant information after a long interruption, in both the Navigation and Interview domains.

We also analysed resumption utterances with respect to grammatical catagory, and compared to the behaviour when beginning a new topic. In this way, we wanted to see if humans express themselves differently when referring to a topic that has recently been discussed, compared to when they start a new topic. Our results indicate that the drivers most often use interrogative phrases when beginning a new navigation topic, and interrogative and declarative phrases when resuming it. The passengers use mostly declarative and imperative phrases both when beginning a new topic and when resuming it. When beginning an Interview topic, both speakers most often use noun phrases or interrogative phrases, and declarative phrases when resuming it.

We completed our investigations of human-human interruption and resumption behaviour by proposing dialogue strategies for system-initiated interruption and resumption of dialogue. When the driver is experiencing high workload, we proposed to let the system take the initiative to pause the dialogue. During state of alert, we suggested that the system continues the interaction unless the driver pauses, then the system should pause too until the workload is low. However, if the driver's workload is high or there is a state of alert and the latest navigation instruction has not been executed yet, we propose to let the system take the initiative to switch topic and/or domain to repeat the latest navigation instruction. If the information is urgent, we suggest that the system interrupts any ongoing topic to give the information. If the information is not urgent the system should instead wait until the topic is finished. However, since our analysis of human-human interaction showed that interruptions are avoided within an adjacency pair, we suggested that the system should not interrupt an ongoing topic until any unanswered questions have been answered, either by the driver or by the system itself.

When resuming a topic, we suggest that the system do not repeat information that has already been given if the interruption was short, but instead continue the interrupted topic with the next utterance. After a long interruption, some information probably needs to be repeated to remind the driver of what has been agreed upon, and therefore we suggest that key words from the latest utterances are included in the resuming utterance.

Finally, we proposed that the system uses a standard phrase when switching domain, to notify the driver of the upcoming domain switch. Our results, although tentative, indicates that passengers adjust their utterances according to the driver's workload when resuming an interrupted topic. We therefore suggested that the system imitates this behaviour, by using an imperative phrase when giving a navigation instruction during high workload or state of alert and a declarative phrase during low workload. When resuming a driving-unrelated topic, we suggested that the system uses an interrogative phrase when seeking information from the driver during high workload, and a declarative phrase during low workload.

Finally, we concluded with a discussion of how to further investigate interruption and resumption behaviour in future work. We proposed to carry out user tests to evaluate the dialogue strategies of a cognitive workload-aware system that we have described, and to test the strategies with a larger group of participants to possibly get significant results.

Given the small dataset used in this thesis, one could argue that the exact results found are perhaps not the most important contribution. Instead, one may focus instead on the overall method of collecting and exploring a corpus of in-vehicle human-human dialogue, the research questions raised and the hypotheses formulated in the various studies presented here, the way they have been investigated by correlating facts about the dialogue interaction with measures of cognitive load and driving manoeuvres, and the use of the results from these studies as a basis for formulating dialogue strategies for managing cognitive load in in-vehicle dialogue systems.

# Bibliography

Allwood, J., Cerrato, L., Dybkjaer, L., Jokkinen, K., Navarretta, C., and Paggio, P. (2004). The mumin multimodal coding scheme. Technical report, Center for Sprogteknologi, Copenhagen University.

Bailey, B. P. and Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Trans. Comput.-Hum. Interact.*, 14(4):1–28.

Bailey, B. P. and Konstan, J. A. (2006). On the need for attention aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Journal of Comput. Human Behaviour*, 22(4):709–732.

Bernsen, N. O. and Dybkjaer, L. (2001). Exploring natural interaction in the car. In *CLASS Workshop on Natural Interactivity and Intelligent Interactive Information Representation*, volume 2.

Berthold, A. (1998). Repräsentation und verarbeitung sprachlicher indikatoren für kognitive ressourcenbeschränkungen [representation and processing of linguistic indicators of cognitive resource limitations]. Master's thesis, Department of Computer Science, University of Saarbrücken, Germany.

Breitholtz, E. (2014). *Enthymemes in Dialogue: A micro-rhetorical approach*. PhD thesis, University of Gothenburg.

Breitholtz, E. and Villing, J. (2008). Can aristothelian enthymems decrease the cognitive load of a dialogue system user? In *LonDial 2008, the 12th SEMDIAL workshop*.

Bridge, D. (2002). Towards conversational recommender systems: A dialogue grammar approach. In Aha, D., editor, *Proceedings of the ECCBR'02 Workshop (Technical Report)*, Mixed-Initiative Case-Based Reasoning, Aberdeen, Scotland.

Brookhuis, K. A. and de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis & Prevention*, 42(3):898–903.

Broström, R., Engström, J., Agnvall, A., and Markkula, G. (2006). Towards the next generation intelligent driver information system (idis): The volvo cars interaction manager concept. In *Proceedings of the 2006 ITS World Congress, London*.

Bühler, D., Vignier, S., Heisterkamp, P., and Minker, W. (2003). Safety and operating issues for mobile human-machine interfaces. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 227–229. ACM.

Byron, D. K. and Heeman, P. A. (1997). Discourse marker use in task-oriented spoken dialog. In *Proceedings of Eurospeech*, volume 97, pages 2223–2226. Citeseer.

De Waard, D. (1996). *The measurement of drivers' mental workload*. Groningen University, Traffic Research Center.

DeMain, B. (2012). When the car radio was introduced, people freaked out. *mental_floss*.

Dozza, M. (2013). What factors influence drivers' response time for evasive maneuvers in real traffic? *Accident Analysis & Prevention*, 58(0):299 – 308.

Edlund, J., Edelstam, F., and Gustafson, J. (2014). Human pause and resume behaviours for unobtrusive humanlike in-car spoken dialogue systems. *EACL 2014*, page 73.

Engström, J., Åberg, N., Johansson, E., and Hammarbäck, J. (2005). Comparison between visual and tactile signal detection tasks applied to the safety assessment of in-vehicle information systems. In *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pages 232–239. Citeseer.

Esbjörnsson, M., Juhlin, O., and Weilenmann, A. (2006). Drivers using mobile phones in traffic: An ethnographic study of interactional adaptation. *International Journal of Human Computer Interaction, Special issue on: In-Use, In-Situ: Extending Field Research Methods.*

Ferguson, G. M., Allen, J. F., Miller, B. W., and Ringger, E. K. (1996). The design and implementation of the trains-96 system: A prototype mixed-initiative planning assistant. trains technical note 96-5. Technical report, Computer Science Department, University of Rochester.

Fors, K. L. and Villing, J. (2011). Reducing cognitive load in in-vehicle dialogue system interaction. In *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue*, pages 55–62.

Gärtner, U., König, W., and Wittig, T. (2002). Evaluation of manual vs. speech input using a driver information system in real traffic. In *Proceedings of the First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*.

Grice, H. P. (1975). *Logic and Conversation*, chapter Syntax and Semantics: Speech Acts, pages 41–58. Acad Press:NY.

Harms, L. (1991). Variation in drivers' cognitive load. effects of driving through village areas and rural junctions. *Ergonomics*, 34(2):151–160.

Iqbal, S. and Bailey, B. P. (2006). Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 1489–1492.

Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., and Bailey, B. P. (2005). Towards an index of opportunity: Understanding changes in mental workload during task execution. In *Proceedings of CHI 2005*, Take a Number, Stand in Line (Interruptions & Attention 1), pages 311–320.

Khawaja, M. A., Ruiz, N., and Chen, F. (2008). Think before you talk: An empirical study of relationship between speech pauses and cognitive load. In *Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat*, OZCHI '08, pages 335–338. ACM.

Kousidis, S., Kennington, C., Baumann, T., Buschmeier, H., Kopp, S., and Schlangen, D. (2014). Situationally aware in-car information presentation using incremental speech generation: safer and more effective. In *Proceedings of the EACL 2014 Workshop on Dialogue in Motion (DM)*, pages 68–72.

Krause, M., Knott, V., and Benger, K. (2015). Implementing the tactile detection task in a real road experiment to assess a traffic light assistant. In *The Eight International Conference on Advances in Computer-Human Interactions*.

Kun, A., Miller, W.T., I., and Lenharth, W. (2002). Project54: introducing advanced technologies in the police cruiser. In *Vehicular Technology Conference, 2002. VTC Spring 2002. IEEE 55th*, volume 2, pages 675–678.

Kun, A. L., Palinko, O., Medenica, Z., and Heeman, P. A. (2013). On the feasibility of using pupil diameter to estimate cognitive load changes for in-vehicle dialogue systems. In *Proceedings of Interspeech 2013*.

Larsson, S. (2002). *Issue-Based Dialogue Management*. PhD thesis, Department of Linguistics, Goteborg University.

Levinson, S. C. (1983). *Pragmatics*. Cambridge Textbooks in Linguistics. Cambridge University Press.

Lindström, A., Villing, J., Larsson, S., Seward, A., Åberg, N., and Holtelius, C. (2008). The effect of cognitive load on disfluencies during in-vehicle spoken dialogue. In *Proceedings of Interspeech 2008*.

Maciej, J. and Vollrath, M. (2009). Comparison of manual vs. speech-based interaction with in-vehicle information systems. *Accident Analysis and Prevention*, 41(5):924–930.

Mahr, A., Feld, M., Moniri, M. M., and Math, R. (2012). The contre (continuous tracking and reaction) task: A flexible approach for assessing driver cognitive workload with high sensitivity. In *Adjunct Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'12)*, pages 88–91.

Mattes, S. (2003). The lane-change-task as a tool for driver distraction evaluation. *Quality of Work and Products in Enterprises of the Future*, pages 57–60.

McGlashan, S., Burnett, D. C., Carter, J., Danielsson, P., Ferrans, J., Hunt, A., Lucas, B., Porter, B., Rehor, K., and Tryphonas, S. (2004). Voice extensible markup language (voicexml) version 2.0.

McTear, M. F. (2004). *Spoken Dialogue Technology - Toward the Conversational User Interface*. Springer Verlag, London.

Medenica, Z. and Kun, A. L. (2012). Data synchronization for cognitive load estimation in driving simulator-based experiments. In *Cognitive Load and In-Vehicle Human-Machine Interaction*.

Merat, N. and Jamson, A. H. (2005). The effect of stimulus modality on signal detection: Implications for assessing the safety of in-vehicle technology. *Human Factors: The Journal of the Human Factors and Ergonomics Society*.

O'Donnell, R. D. and Eggemeier, F. T. (1986). *Workload assessment methodology.*, chapter Handbook of perception and human performance. Volume II, cognitive processes and performance. John Wiley & Sons.

Palinko, O., Kun, A. L., Shyrokov, A., and Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, pages 141–144, New York, NY, USA. ACM.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, 116:220–244.

Patten, C. J. D., Kircher, A., Östlund, J., and Nilsson, L. (2003). Using mobile telephones: cognitive workload and attention resource allocation. *Accident Analysis & Prevention*, 36:341–350.

Paul, C. L., Komlodi, A., and Lutters, W. (2015). Interruptive notifications in support of task management. *International Journal of Human-Computer Studies*.

Pieraccini, R., Dayanidhi, K., Bloom, J., Dahan, J.-G., Phillips, M., Goodman, B., and Prasad, K. V. (2004). Multimodal conversational systems for automobiles. *Communications of the ACM*, 47.

Pretschner, A., Broy, M., Kruger, I. H., and Stauner, T. (2007). Software engineering for automotive systems: A roadmap. In *2007 Future of Software Engineering*, pages 55–71. IEEE Computer Society.

Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *language*, pages 696–735.

Salvucci, D. D. and Bogunovich, P. (2010). Multitasking and monotasking: the effects of mental workload on deferred task interruptions. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 85–88. ACM.

Sangani, K. (2013). Voice recognition comes of age. *Engineering and Technology Magazine*.

Schegloff, E. A. and Sacks, H. (1973). Opening up closings. *Semiotica*, 7:289–327.

Schriffrin, D. (1987). *Discourse Markers*. Cambridge University Press.

Shriberg, E. (2001). To "errrr" is human: ecology and acoustics of speech and disfluencies. *Journal of the International Phonetic Association*, 31:153–169.

Shriberg, E. (2005). Spontaneous speech: How people really talk and why engineers should care. In *Ninth European Conference on Speech Communication and Technology*.

Shyrokov, A., Kun, A., and Heeman, P. (2007). Experiments modeling of human-human multi-threaded dialogues in the presence of a manual-visual task. In *Proceedings of 8th SIGDial*, pages 190–193.

Sumikawa, D. A. (1985). Guidelines for the integration of audio cues into computer user interfaces. Technical report, Lawrence Livermore National Lab., CA (USA).

Sweller, J., Merrienboer, J. V., and Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10:251–296.

Tashev, I., Seltzer, M., Ju, Y.-C., Wang, Y.-Y., and Acero, A. (2009). Commute ux: Voice enabled in-car infotainment system. In *Mobile HCI*, volume 9.

Traum, D. R. and Larsson, S. (2003). The information state approach to dialogue management. In *Current and new directions in discourse and dialogue*, pages 325–353. Springer.

van Winsum, W., Martens, M., and Herland, L. (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.

Villing, J. (2009a). Dialogue behaviour under high cognitive load. In Healey, P., Pieraccini, R., Byron, D., Young, S., and Purver, M., editors, *Proceedings of the SigDial 2009 Conference*, pages 322–325.

Villing, J. (2009b). In-vehicle dialogue management - towards distinguishing between different types of workload. In *Proceedings of SiMPE, Fourth Workshop on Speech in Mobile and Pervasive Environments*, pages 14–21.

Villing, J., Holtelius, C., Larsson, S., Lindström, A., Seward, A., and Åberg, N. (2008). Interruption, resumption and domain switching in in-vehicle dialogue. In Nordstrom, B. and Ranta, A., editors, *Proceedings of GoTAL, 6th International Conference of Advances in Natural Language Processing*, volume 5221, pages 488–499.

Villing, J. and Larsson, S. (2006). Dico: a multimodal menu-based in-vehicle dialogue system. In *brandial06, Proceedings of the 10th Workshop on the Semantics and the Pragmatics of Dialogue*.

Villing, J. and Larsson, S. (2011). Speech, buttons or both? a comparative study of an in-car dialogue system. In *Proceedings of the workshop Cognitive Load and In-vehicle Human-Machine Interaction*.

Vollrath, M. and Maciej, J. (2008). In-car distraction study, final report. Technical report, the HMI laboratory of the Technical University of Brunswick.

Walker, M. A. (1992). Redundancy in collaborative dialogue. In *Proceedings of the 14th Conference on Computational Linguistics - Volume 1*, COLING '92, pages 345–351, Stroudsburg, PA, USA. Association for Computational Linguistics.

Walker, M. A. (1996). The effect of resource limits and task complexity on collaborative planning in dialogue. *Artificial Intelligence Journal*, 85.

Weng, F., Varges, S., Raghunathan, B., Ratiu, F., Pon-Barry, H., Lathrop, B., Zhang, Q., Scheideck, T., Bratt, H., Xu, K., Purver, M., Mishra, R., Raya, M., Peters, S., Meng, Y., Cavedon, L., and Shriberg, L. (2006). Chat: A conversational helper for automotive tasks. In *Proceedings of International Conference on Spoken Language Processing (Interspeech/ICSLP)*, pages 1061–1064.

Winkler, E. G. (2007). *Understanding Language*. Continuum Publishing.

Yang, F., Heeman, P. A., and Kun, A. (2008). Switching to realtime tasks in multi-tasking dialogue. In *Proceedings of International Conference on Computational Linguistics*, pages 1025–1032.

Yang, F., Heeman, P. A., and Kun, A. L. (2011). An investigation of interruptions and resumptions in multi-tasking dialogues. *Computational Linguistics*, 37(1):75–104.

Young, K. and Regan, M. (2007). *Distracted driving*, chapter Driver distraction: A review of the literatur, pages 379–405. Sydney, NSW: Australasian College of Road Safety.