Now, where was I?
Resumption strategies for an in-vehicle dialogue system

Jessica Villing

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

Abstract

In-vehicle dialogue systems often contain more than one application, e.g. a navigation and a telephone application. This means that the user might, for example, interrupt the interaction with the telephone application to ask for directions from the navigation application, and then resume the dialogue with the telephone application. In this paper we present an analysis of interruption and resumption behaviour in human-human in-vehicle dialogues and also propose some implications for resumption strategies in an in-vehicle dialogue system.

1 Introduction

Making it useful and enjoyable to use a dialogue system is always important. The dialogue should be easy and intuitive, otherwise the user will not find it worth the effort and instead prefer to use manual controls or to speak to a human.

However, when designing an in-vehicle dialogue system there is one more thing that needs to be taken into consideration, namely the fact that the user is performing an additional, safety critical, task - driving. The so-called 100-car study (Neale et al., 2005) revealed that secondary task distraction is the largest cause of driver inattention, and that the handling of wireless devices is the most common secondary task. Even if spoken dialogue systems enables manouvering of devices without using hands or eyes, it is crucial to adjust the interaction to the in-vehicle environment in order to minimize distraction from the interaction itself. Therefore the dialogue system should consider the cognitive load of the driver and adjust the dialogue accordingly. One way of doing this is to continously measure the cognitive workload level of the driver and, if the workload is high,

determine type of workload and act accordingly. If the workload is *dialogue-induced* (i.e. caused by the dialogue itself), it might be necessary to rephrase or offer the user help with the task. If the workload is *driving-induced* (i.e. caused by the driving task), the user might need information that is crucial for the driving task (e.g. get navigation instructions), or to pause the dialogue in order to enable the user to concentrate on the driving task (Villing, 2009). Both the driver and the system should be able to initiate interruptions.

When the interaction with a dialogue system has been interrupted, e.g. because the user has not answered a question, it is common that the system returns to the top menu. This means that if the user wants to finish the interrupted task she has to restart from the beginning, which is both time-consuming and annoying. Instead, the dialogue system should be able to either pause until the workload is low or change topic and/or domain, and then resume where the interruption took place. However, resumption of an interrupted topic needs to be done in a way that minimizes the risk that the cognitive workload increases again. Although a lot of research has been done regarding dialogue system output, very little work has been done regarding resumption of an interrupted topic. In this paper we will analyse human-human in-vehicle dialogue to find out how resumptions are done in human-human dialogue and propose some implications for resumption strategies in a dialogue system.

2 Related work

To study resumption behaviour, Yang (2009), carried out a data collection where the participants were switching between an ongoing task (a card game) and a real-time task (a picture game). The participants randomly had to interrupt the ongoing task to solve a problem in the real-time task. When studying the resumption behaviour after an

interruption to the real-time task they found that the resuming utterance contained various amounts and types of redundant information depending on whether the interruption occured in the middle of a card discussion, at the end of a card or at the end of a card game. If the interruption occured in the middle of a card discussion it was possible to make a distinction between *utterance restatement* (repeat one's own utterance, repeat the dialogue partners utterance or clarification of the dialogue partners utterance) and *card review* (reviewing all the cards on hand although this information had already been given). They found that the behaviour is similar to grounding behaviour, where the speaker use repetition and requests for repetition to ensure that the utterance is understood.

3 Data collection

A data collection has been carried out within the DICO project (see, for example, (Larsson and Villing, 2007)) to study how an additional distraction or increase in the cognitive load would affect a driver's dialogue behaviour. The goal was to elicit a natural dialogue (as opposed to giving the driver a constructed task such as for example a math task) and make the participants engage in the conversation.

The participants (two female and six male) between the ages of 25 and 36 drove a car in pairs while interviewing each other. The interview questions and the driving instructions were given to the passenger, hence the driver knew neither what questions to discuss nor the route in advance. Therefore, the driver had to signal, implicitly or explicitly, when she wanted driving instructions and when she wanted a new question to discuss. The passenger too had to have a strategy for when to change topic. The reasons for this setup was to elicit a natural and fairly intense dialogue and to force the participants to frequently change topic and/or domain (e.g. to get driving instructions). The participants changed roles after 30 minutes, which meant that each participant acted both as driver and as passenger. The cognitive load of the driver was measured in two ways. The driver performed a Tactile Detection Task (TDT) (van Winsum et al., 1999). When using a TDT, a buzzer is attached to the driver's wrist. The driver is told to push a button each time the summer is activated. Cognitive load is determined by measuring hit-rate and reaction time. Although the TDT task in itself

might cause an increased workload level, the task is performed during the whole session and thereby it is possible to distinguish high workload caused by something else but the TDT task.

Workload was also measured by using an IDIS system (Broström et al., 2006). IDIS determines workload based on the driver's behaviour (for example, steering wheel movements or applying the brake). What differs between the two measurements is that the TDT measures the actual workload of each driver, while IDIS makes its assumptions based on knowledge of what manouvres are usually cognitively demanding.

The participants were audio- and videotaped, the recordings are transcribed with the transcription tool ELAN[1], using an orthographic transcription. All in all 3590 driver utterances and 4382 passenger utterances are transcribed. An annotation scheme was designed to enable analysis of utterances with respect to topic change for each domain.

Domain and topic was defined as:

- *interview* domain: discussions about the interview questions where each interview question was defined as a topic

- *navigation* domain: navigation-related discussions where each navigation instruction was defined as a topic

- *traffic* domain: discussions about the traffic situation and fellow road-users where each comment not belonging to a previous event was defined as a topic

- *other* domain: anything that does not fit within the above domains where each comment not belonging to a previous event was defined as a topic

Topic changes has been coded as follows:

- *begin-topic*: whatever → new topic

   - I.e., the participants start discussing an interview question, a navigation instruction, make a remark about the traffic or anything else that has not been discussed before.

- *end-topic*: finished topic → whatever

---

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

- A topic is considered finished if a question is answered or if an instruction or a remark is confirmed.

- *interrupt-topic*: unfinished topic → whatever

  - An utterance is considered to interrupt if it belongs to another topic than the previous utterance and the previous topic has not been ended with an *end-topic*.

- *resume-topic*: whatever → unfinished topic

  - A topic is considered to be resumed if it has been discussed earlier but was not been finished by an *end-topic* but instead interrupted with an *interrupt-topic*.

- *reraise-topic*: whatever → finished topic

  - A topic is considered to be reraised if it has been discussed before and then been finished with an *end-topic*.

The utterances have been categorised according to the following schema:

- DEC: declarative

  - (*"You are a Leo and I am a Gemini", "This is Ekelund Street"*)

- INT: interrogative

  - (*"What do you eat for breakfast?", "Should we go back after this?"*)

- IMP: imperative

  - (*"Go on!"*)

- ANS: *"yes"* or *"no"* answer (and variations such as *"sure, absolutely, nope, no way"*)

- NP: bare noun phrase

  - (*"Wolfmother", "Otterhall Street"*)

- ADVP: bare adverbial phrase

  - (*"Further into Karlavagn Street"*)

- INC: incomplete phrase

  - (*"Well, did I answer the", "Should we"*)

Cognitive load has been annotated as:

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

- **high**: high workload according to IDIS

- **low**: low workload according to IDIS

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

4 Results

The codings from the DICO data collection has been analysed with respect to interruption and resumption of topics (*interrupt-topic* and *resume-topic*, respectively). Interruption can be done in two ways, either to pause the dialogue or to change topic and/or domain. In the DICO corpus there are very few interruptions followed by a pause. The reason is probably that both the driver and the passenger were strongly engaged in the interview and navigation tasks. The fact that the driver did not know the route elicited frequent switches to the navigation domain done by both the driver and the passenger, as can be seen in Figure 1. Therefore, we have only analysed interruption and resumption from and to the interview and navigation domains.
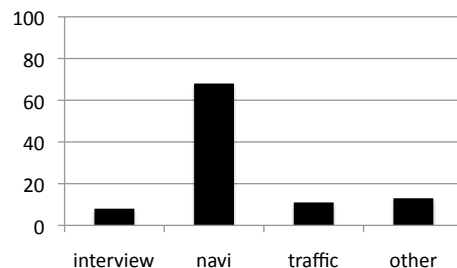

Figure 1: Distribution of utterances coded as *interrupt-topic* for each domain, when interrupting from an interview topic.

4.1 Redundancy

The easiest way of resuming an interrupted topic in a dialogue system is to repeat the last phrase that was uttered before the interruption. One disadvantage of this method is that the dialogue system might be seen as tedious, especially if there are several interruptions during the interaction. We wanted to see if the resuming utterances in human-human dialogue are redundant and if redundancy has anything to do with the length of the interruption. We therefore sorted all utterances coded

as *resume-topic* in two categories, those which contained redundant information when comparing with the last utterance before the interruption, and those which did not contain and redundant information. As a redundant utterance we counted all utterances that repeated one or more words from the last utterance before the interruption. We then counted the number of turns between the interruption and resumption. The number of turns varied between 1 and 42. The result can be seen in Figure 2.
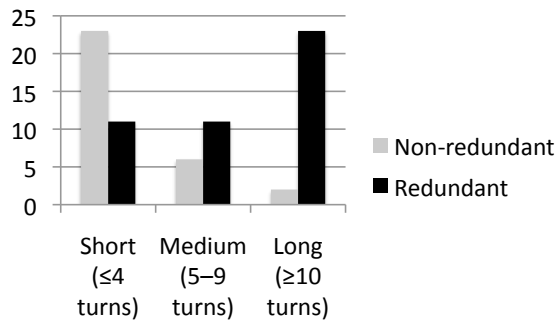

Figure 2: Number of redundant utterances depending on length of interruption.

As can be seen, there are twice as many non-redundant as redundant utterances after a short interruption (≤ 4 turns), while there are almost solely redundant utterances after a long interruption (≥ 10 turns). The average number of turns is 3,5 when no redundancy occur, and 11,5 when there are redundancy. When the number of turns exceeds 12, there are only redundant utterances.

4.2 Category

Figure 3 shows the distribution, sorted per category, of driver utterances when resuming to an interview and a navigation topic. Figure 4 shows the corresponding figures for passenger utterances.
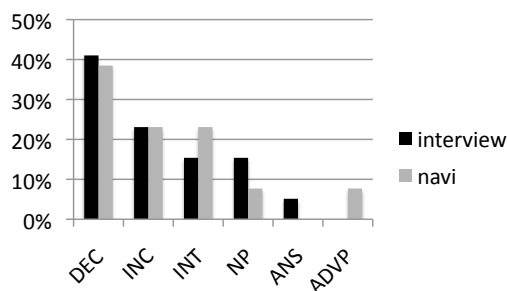

Figure 3: Driver resuming to the interview and navigation domains.

The driver's behaviour is similar both when resuming to an interview and a navigation topic. Declarative phrases are most common, followed by incomplete, interrogative (for interview topics) and noun phrases.
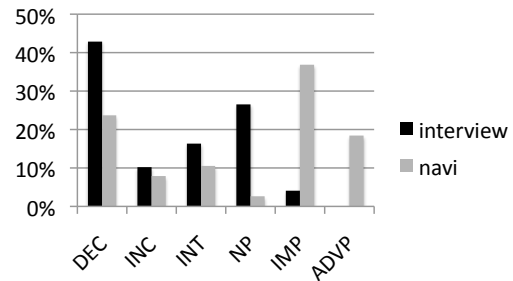

Figure 4: Passenger resuming to the interview and navigation domains.

When looking at the passenger utterances we see a lot of variation between the domains. When resuming to an interview topic the passenger uses mostly declarative phrases, followed by noun phrases and interrogative phrases. When resuming to a navigation topic imperative phrases are most common, followed by declarative phrases. Only the passenger use imperative phrases, probably since the passenger is managing both the interview questions and the navigation instructions and therefore is the one that is forcing both the interview and the navigation task through.

4.3 Workload level

The in-vehicle environment is forcing the driver to carry out tasks during high cognitive workload. To minimize the risk of increasing the workload further, an in-vehicle dialogue system should be able to decide when to interrupt and when to resume a topic depending on the driver's workload level.

The figures in this section shows workload level and type of workload during interruption and resumption to and from topics in the interview domain. When designing the interview and navigation tasks that were to be carried out during the data collection, we focused on designing them so that the participants were encouraged to discuss as much as possible with each other. Therefore, the navigation instructions sometimes were hard to understand, which forced the participants to discuss the instructions and together try to interpret them. Therefore we have not analysed the workload level while interrupting and resuming topics in the navigation domain since the result might be

misleading.

Type of workload is determined by analysing the TDT and IDIS signals described in 3. Workload is considered to be dialogue-induced when only the TDT is indicating high workload (since the TDT indicates that the driver is carrying out a task that is cognitively demanding but IDIS is not indicating that the driving task is demanding at the moment), driving-induced when both the TDT and IDIS is indicating high workload (since the TDT is indicating that the workload level is high and IDIS is indicating that the driving task is demanding) and possibly driving-induced when only IDIS is indicating high workload (since IDIS admittedly is indicating that the driving task is demanding but the TDT indicates that the driver's workload is low, it could then be that this particular driver does not experience the driving task demanding even though the average driver does) (Villing, 2009).

The data has been normalized for variation in workload time. The diagrams shows the distribution of interruption and resumption utterances made by the driver and the passenger, respectively.
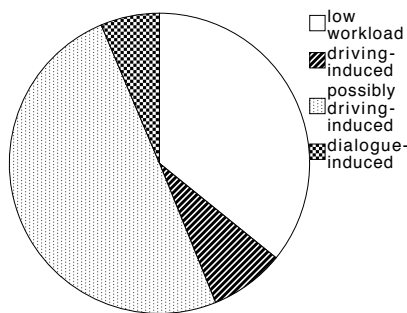

Figure 5: Workload while the driver is interrupting an interview topic.
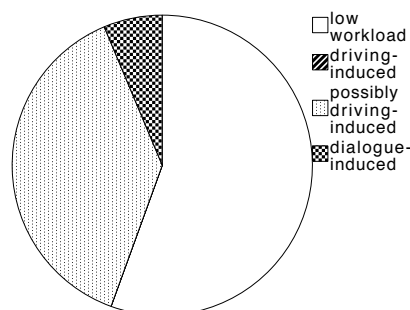

Figure 6: Workload while the passenger is interrupting an interview topic.

Figures 5 and 6 show driver workload level while the driver and the passenger (respectively)

are interrupting from the interview domain. The driver most often interrupts during a possible driving-induced or low workload, the same goes for the passenger but in opposite order. It is least common for the driver to interrupt during dialogue- or driving-induced workload, while the passenger rarely interrupts during dialogue-induced and never during driving-induced workload.
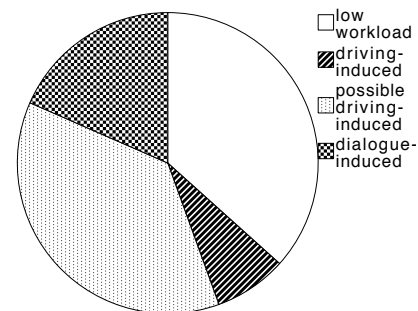

Figure 7: Workload while driver is resuming to the interview domain.
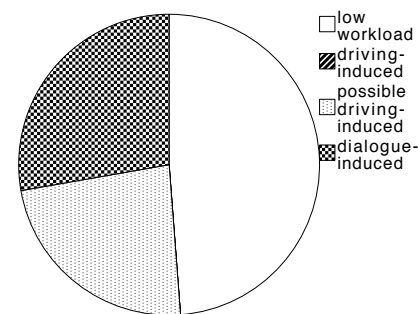

Figure 8: Workload while passenger is resuming to the interview domain.

Figures 7 and 8 show workload level while the driver and the passenger (respectively) are resuming to the interview domain. The driver most often resumes while the workload is low or possibly driving-induced, while the passenger is mostly resuming during low workload and never during driving-induced workload.

5 Discussion

For both driver and passenger, the most common way to resume an interview topic is to use a declarative utterance, which is illustrated in Figure 3. When studying the utterances in detail we can see that there is a difference when comparing information redundancy similar to what Yang (2009) describe in their paper. They compared grade of

redundancy based on where in the dialogue the interruption occur, what we have looked at in the DICO corpus is how many turns the interrupting discussion contains.

As Figure 2 shows, if the number of turns is about three (on average, 3,5), the participants tend to continue the interrupted topic exactly where it was interrupted, without considering that there had been any interruption. The speaker however often makes some sort of sequencing move to announce that he or she is about to switch domain and/or topic, either by using a standard phrase or by making an extra-lingustic sound like, for example, lipsmack or breathing (Villing et al., 2008). Example (1) shows how the driver interrupts a discussion about what book he is currently reading to get navigation instructions:

(1) Driver: `What I read now is Sofie's world.`
    Driver (*interrupting*): `Yes, where do you want me to drive?`
    Passenger: `Straight ahead, straight ahead.`
    Driver: `Straight ahead. Alright, I'll do that.`
  Passenger (*resuming*): `Alright [sequencing move]. Enemy of the enemy was the last one I read. [DEC]`

If the number of turns is higher than ten (on average, 11,5) the resuming speaker makes a redundant utterance, repeating one or more words from the last utterance before the interruption. See example (2):

(2) Driver: `Actually, I have always been interested in computers and technology.`
    Passenger (*interrupting*): `Turn right to Vasaplatsen. Is it here? No, this is Grönsakstorget.`
    Driver: `This is Grönsakstorget. We have passed Vasaplatsen.`
    .
    .
    . `(Discussion about how to turn around and get back to Vasaplatsen, all in all 21 turns.)`
  Driver (*resuming*): `Well, as I said [sequencing move]. I have always been interested in computer and computers and technology and stuff like that. [DEC]`

The passenger often uses a bare noun phrase to resume, the noun phrase can repeat a part of the interview question. For example, after a discussion about wonders of the world, which was interrupted by a discussion about which way to go next, the passenger resumed by uttering the single word "wonders" which was immediatly understood by the driver as a resumption to the interview topic. The noun phrase can also be a key phrase in the dialogue partner's answer as in example (3) where the participants discuss their favourite band:

(3) Driver: `I like Wolfmother, do you know about them?`
    Passenger: `I've never heard about them. [...] You have to bring a cd so I can listen to them.`
    Driver (*interrupting*): `Where was I supposed to turn?`
    .
    .
    . `(Navigation discussion, all in all 13 turns.)`
  Passenger (*resuming*): `[LAUGHS]Wolfmother. [NP]`

When resuming to the navigation domain, the driver mostly uses a declarative phrase, typically to clarify an instruction. It is also common to use an interrogative phrase or an incomplete phrase such as *"should I..."* which the passenger answers by clarifying which way to go. The passenger instead uses mostly imperative phrases as a reminder of the last instruction, such as *"keep straight on"*.

When the speakers interrupts an interview topic they mostly switch to the navigation domain, see Figure 1. That means that the most common reason for the speaker to interrupt is to ask for or give information that is crucial for the driving task (as opposed for the other and traffic domains, which are mostly used to signal that the speaker's cognitive load level is high (Villing et al., 2008)). As can be seen in Figures 5 and 6, the driver mostly interrupts the interview domain during a possible driving-induced workload while the passenger mostly interrupts during low workload. As noted above (see also Figure 3), the utterances are mostly declarative (*"this is Ekelund Street"*), interrogative (*"and now I turn left?"*) or incomplete (*"and then..."*), while the passenger gives additional information that the driver has not asked for explicitly but the passenger judges that the driver might need (*"just go straight ahead in the next crossing"*, *"here is where we should turn towards Järntorget"*). Hence, it seems like the driver interrupts to make clarification utterances that must be answered immediately, for example, right before a

crossing when the driver has pressed the brakes or turned on the turn signal (and therefore the IDIS system signals high workload which is interpreted as driving-induced workload) while the passenger take the chance to give additional information in advance, before it is needed, and the workload therefore is low.

Figure 7 shows that the driver mostly resumes to the interview domain during low or possible driving-induced workload. Since the IDIS system makes its assumption on driving behaviour, based on what the average driver finds cognitively demanding, it might sometimes be so that the system overgenerates and indicates high workload even though the driver at hand does not find the driving task cognitively demanding. This might be an explanation to these results, since the driver often resumes to an interview topic although he or she is, for example, driving through a roundabout or pushing the brakes. It is also rather common that the driver is resuming to an interview question during dialogue-induced workload, perhaps because she has started thinking about an answer to a question and therefore the TDT indicates high workload and the IDIS does not. The passenger mostly resumes to the interview domain during low workload, which indicates that the passenger analyses both the traffic situation and the state of mind of the driver before he or she wants to draw the drivers attention from the driving task.

6 Implications for in-vehicle dialogue systems

In this paper we point at some of the dialogue strategies that are used in human-human dialogue during high cognitive load when resuming to an interrupted topic. These strategies should be taken under consideration when implementing an in-vehicle dialogue system.

To make the dialogue natural and easy to understand the dialogue manager should consider which domain it will resume to and the number of turns between the interruption and resumption before deciding what phrase to use as output. For example, the results indicate that it might be more suitable to use a declarative phrase when resuming to a domain where the system is asking the user for information, for example when adding songs to a play list at the mp3-player (cf. the interview domain). If the number of turns are 4 or less, it probably does not have to make a redundant utterance at all, but may continue the discussion where it was interrupted. If the number of turns exceeds 4 it is probably smoother to let the system just repeat one or more keywords from the interrupted utterance to make the user understand what topic should be discussed, instead of repeating the whole utterance or even start the task from the beginning. This will make the system feel less tedious which should have a positive effect on the cognitive workload level. However, user tests are probably needed to decide how much redundant information is necessary when talking to a dialogue system, since it may well differ from talking to a human being who is able to help the listener understand by, for example, emphasizing certain words in a way that is currently impossible for a computer. When resuming to a domain where the system has information to give to the user it is suitable to make a short, informative utterance (e.g. *"turn left here", "traffic jam ahead, turn left instead"*).

Finally, it is also important to consider the cognitive workload level of the user to determine when - and if - to resume, and also whether the topic that is to be resumed belongs to a domain where the system has information to give to the user, or a domain where the user gives information to the system. For example, if the user is using a navigation system and he or she is experiencing driving-induced workload when approaching e.g. a crossing, it might be a good idea to give additional navigation information even though the user has not explicitly asked for it. If the user however is using a telephone application it is probably better to let the user initiate the resumption. The DICO corpus shows that it is the passenger that is most careful not to interrupt or resume when the driver's workload is high, indicating that the system should let the user decide whether it is suitable to resume during high workload, while it is more accepted to let the system interrupt and resume when the workload is low.

When resuming to the interview domain the driver (i.e. the user) mostly uses declarative phrases, either as an answer to a question or as a redundant utterance to clarify what was last said before the interruption. Therefore the dialogue system should be able to store not only what has been agreed upon regarding the interrupted task, but also the last few utterances to make it possible to interpret the user utterance as a resumption.

It is common that the driver utterances are incomplete, perhaps due to the fact that the driver's primary task is the driving and therefore his or her mind is not always set on the dialogue task. Lindström (2008) showed that deletions are the most common disfluency during high cognitive load, which is supported by the results in this paper. The dialogue system should therefore be robust regarding ungrammatical utterances.

7 Future work

Next we intend to implement strategies for interruption and resumption in the DICO dialogue system. The strategies will then be evaluated through user tests where the participants will compare an application with these strategies with an application without them. Cognitive workload will be measured as well as driving ability (for example, by using a Lane Change Task (Mattes, 2003)). The participants will also be interviewed in order to find out which version of the system is more pleasant to use.

References

Robert Broström, Johan Engström, Anders Agnvall, and Gustav Markkula. 2006. Towards the next generation intelligent driver information system (idis): The volvo cars interaction manager concept. In *Proceedings of the 2006 ITS World Congress*.

Staffan Larsson and Jessica Villing. 2007. The dico project: A multimodal menu-based in-vehicle dialogue system. In H C Bunt and E C G Thijsse, editors, *Proceedings of the 7th International Workshop on Computational Semantics (IWCS-7)*, page 4.

Anders Lindström, Jessica Villing, Staffan Larsson, Alexander Seward, Nina Åberg, and Cecilia Holtelius. 2008. The effect of cognitive load on disfluencies during in-vehicle spoken dialogue. In *Proceedings of Interspeech 2008*, page 4.

Stefan Mattes. 2003. The lane-change-task as a tool for driver distraction evaluation. In *Proceedings of IGfA*.

V L Neale, T A Dingus, S G Klauer, J Sudweeks, and M Goodman. 2005. An overview of the 100-car naturalistic study and findings. In *Proceedings of the 19th International Technical Conference on Enhanced Safety of Vehicles (ESV)*.

W van Winsum, M Martens, and L Herland. 1999. The effect of speech versus tactile driver support messages on workload, driver behaviour and user acceptance. tno-report tm-99-c043. Technical report, Soesterberg, Netherlands.

Jessica Villing, Cecilia Holtelius, Staffan Larsson, Anders Lindström, Alexander Seward, and Nina Åberg. 2008. Interruption, resumption and domain switching in in-vehicle dialogue. In *Proceedings of GoTAL, 6th International Conference on Natural Language Processing*, page 12.

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

Fan Yang and Peter A Heeman. 2009. Context restoration in multi-tasking dialogue. In *IUI '09: Proceedings of the 13th international conference on Intelligent user interfaces*, pages 373–378, New York, NY, USA. ACM.