SemDial 2011: Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue, pages 55–62.
Los Angeles, California, 21–23 September 2011.

Reducing cognitive load in in-vehicle dialogue system interaction

Kristina Lundholm Fors
GSLT and
Dept. of Philosophy, Linguistics
and Theory of Science
University of Gothenburg
kristina.lundholm@gu.se

Jessica Villing
GSLT and
Dept. of Philosophy, Linguistics
and Theory of Science
University of Gothenburg
jessica@ling.gu.se

Abstract

In-vehicle dialogue systems need to be able to adapt to the cognitive load of the user, and, when possible, reduce cognitive load. To accomplish this, we need to know how humans act while driving and talking to a passenger, and find out if there are dialogue strategies that can be used to minimize cognitive load. In this study, we have analyzed human-human in-vehicle dialogues, focusing on pauses and adjacency pairs. Our results show that when the driver is experiencing high cognitive load, the passenger's median pause times increase. We also found that, when switching to another domain and/or topic, both driver and passenger try to avoid interrupting an adjacency pair. This suggests that a dialogue system could help lower the user's cognitive load by increasing pause lengths within turns, and plan system utterances in order to avoid switching task within an adjacency pair.

1 Introduction

For safety reasons, an in-vehicle dialogue system needs to be aware of the cognitive load level of the driver and to avoid increasing it. To find out how to reduce cognitive load, we need to investigate how humans handle this task. In this paper we do this by analysing a corpus of human-human in-vehicle dialogue. We look at how a speaker uses pauses within her turn to help decrease the cognitive load level of the dialogue partner.

We also investigate how humans switch to another topic. The number of in-vehicle applications that are using a speech interface is increasing, and therefore it is necessary to be able to switch tasks without increasing the cognitive load even more. We need to find out at which point in the dialogue it is suitable to switch task or change topic if the system needs to give information to the driver. For example, the navigation system should plan its utterances and give instructions to the driver when she is mentally prepared to receive that information. If the driver's mind is occupied with something else, it is probably not a good idea to interrupt at that point. Interrupting a dialogue at a bad time might cause an even higher cognitive load level and is a safety risk.

2 Background

2.1 Pauses in dialogues

Silent intervals in dialogue can occur within a speaker's turn or between two speakers' turns. Sacks et al (1974) divide these silent intervals into *pauses*, *gaps* and *lapses*. A pause is a silence that occurs within a speaker's turn. This includes the silence at a TRP (Transition Relevance Place), when a speaker has been nominated but has not yet begun to speak. It also includes the silence at a TRP, when a speaker has stopped, but then continues to speak after the TRP. A gap is the silence that occurs at a TRP when the first speaker has not nominated another speaker, but another speaker self-nominates and there is a turn change. A lapse is the silence at a TRP, when the first speaker has stopped speaking, has not nominated a new speaker, and does not continue speaking. No other speaker takes the turn. A lapse is in part defined by its perceived length: thus, a lapse should be perceived as longer than a gap and as a discontinuity in the flow of conversation. In this paper we will focus on pauses within a speaker's turn.

2.1.1 Pause categories

Pauses that occur within a turn can have at least two functions. Firstly, they provide time for the speaker to plan what he/she is going to say. Secondly, they may also allow the speakers to negotiate who is going to take the turn. This could have an effect on pause length, where pauses that do not occur at a possible TRP should be shorter, as the speakers do not have to take speaker change into account during these pauses. Below, three different types of pauses within turns are described:

- pauses that occur within a speaker's turn, and within a syntactic unit (hence not at a possible TRP). This type of pause frequently occurs before a content word, or after a discourse marker such as *and* or *but* (van Donzel and van Beinum, 1996). These will be referred to as "pause internal within".
- pauses that occur within a speaker's turn, at a possible TRP, where speaker change does not take place. These will be referred to as "pause internal between".
- pauses that occur at the beginning of a speaker's turn, when the speaker has been nominated by the previous speaker. These will be referred to as "pause initial".

2.2 Adjacency pairs and turn-taking

Levinson (1983) states 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 (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. However, 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 draw inferences and assume that the second speaker is sulking, not interested, is being deliberately rude or did not understand (Bridge, 2002). Consequently, dialogue partners strive to follow this rule of turn-taking as far as they can and will not break it unless necessary.

In-vehicle dialogue is rather special, since the driver is busy with a safety critical task and therefore must consider the dialogue task as secondary. The passenger, on the other hand, is not directly involved with the driving task but is aware of the traffic situation and is thereby able to adapt the dialogue in order to make the driving task easier. This makes it interesting to look at the turn-taking behaviour to find out in which cases the rule is followed and in which cases it is violated.

3 The DICO project

The corpus used for this paper is developed in the Vinnova funded DICO project (Larsson and Villing, 2007)). The DICO project aimed at developing a proof-of-concept demo system, with fully integrated multimodality. This allows the user to choose among the modalities to interact with the system and thereby choose the modality that is most suitable for the task or situation at hand.

4 Related work

Research on speaker's cognitive load show that pause duration tends to increase during high cognitive load (Cappella, 1979; Villing, 2009). When a speaker is showing signs of high cognitive load, it is reasonable to expect the other participant in the conversation to adjust their speech to reduce cognitive load. We are therefore interested in the pause patterns of the other speaker (the passenger in our study). Edlund et al (2009) have shown that speakers tend to align their pause lengths, that is, two speakers in a dialogue will make pauses of approximately equal length,

Topic switch and interruption in relation to adjacency pairs have been examined in a user study described in Shyrokov (2007). The study showed that humans strive to avoid interruption in the middle of an adjacency pair. The authors suggest that one reason for this is that a finished adjacency pair makes a simpler discourse context to resume to, compared to an interrupted adjacency pair that might force the resuming dialogue partner to repeat (parts of) the discourse.

5 Method

To study human-human in-vehicle dialogue a data collection was carried out within the DICO project. The aim was to elicit a fairly natural dialogue with frequent topic and/or domain shifts, in order to study dialogue strategies during varying levels of cognitive workload. There were 8 subjects between the ages 25-36 participating in the study. The subjects drove a car in pairs while performing two tasks, one navigation task and one memory task. To carry out the navigation task, instructions were given to the passenger who was only allowed to give (and look at) one instruction at a time (for example, "turn right in the next crossing", "keep straight on for 500 meters"). The memory task consisted of a list of 52 interview questions about personal information such as "where were you born", "what fruit do you like the most", "who is your favourite actor". The questions were given to the passenger who could choose freely among the questions. The participants were told to interview each other and ask as many questions as they managed, and to try to remember as much data as possible since they should be tested after the ride. The reason for this was to elicit a fairly intense and absorbing conversation. They drove for one hour and were told to switch roles after half the time, so that all participants acted as both driver and passenger.

The driver had an additional task. In order to measure cognitive workload, a TDT (Tactile Detection Task) equipment was used. It consist of a buzzer that is attached to the wrist, and a button attached to the index finger. Each time the buzzer is activated (which is done randomly every 2-5 seconds) the driver should press the button. Workload is measured based on reaction time and hit-rate. The TDT therefore enables measurement of cognitive workload that is not related to the driving task, for example the workload that is caused by the dialogue itself.

Workload was also measured using an IDIS system (Broström et al., 2006). IDIS determines workload based on the driving behaviour, for example steering wheel movements or sudden changes in speed. The output from IDIS was shown as a red light (high workload) or a green light (low workload), which was captured by the camera heading towards the road.

5.1 Transcription and coding

For transcription and coding of the material the ELAN transcription tool [1] was used.

Due to technical problems, one driver/passenger pair had to be removed and therefore the corpus contains 3 hours of dialogue. All in all 3590 driver utterances and 4382 passenger utterances were transcribed and coded.

5.1.1 Cognitive workload

Cognitive workload, as mentioned, has been measured in two ways.

Workload according to the TDT is annotated as:

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

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

Workload according to IDIS as annotated as:

- *High*: the IDIS sensor is showing red, indicating high workload
- *Low*: the IDIS sensor is showing green, indicating low workload

5.1.2 Pauses

We are interested in finding out how pausing patterns are affected when the conversation partner is experiencing high cognitive load. The hypothesis is that, in line with the research described in section 4, the driver will make longer pauses during high cognitive load, and that this will cause the passenger to also exhibit longer pause durations. If the passenger adjusts his/her pauses to the cognitive load level of the driver, this behaviour may be applicable in dialogue systems, to reduce the cognitive load of the user.

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

Pauses were identified manually and acoustically with the help of ELAN. Since pauses were identified manually, no silence threshold was set. They were then categorized into the different categories described in 2.1.1. Not all of the material at hand has been investigated; we chose to analyze part of the material first, to then decide whether we would move on to analyzing all pauses in the material. The main reason for this is that manually identifying and categorizing pauses is very time-consuming.

5.1.3 Adjacency pairs

We want to investigate if there is a place in a natural human-human dialogue that is more suited for making a topic or domain shift. When performing a task that is cognitively demanding, such as driving a car, it is probably even more important not to interrupt at a bad time. Therefore, the DICO corpus has been coded with *topic-shifts* and *adjacency pairs* with the purpose to investigate where interruptions take place in a human-human dialogue.

Codings:

- *Topic*: each interview question
  - *begin-topic*: the beginning of a topic that has not been discussed earlier. For example, the actual interview question or a general comment about a question.
  - *end-topic*: the utterance that ends a topic. For example, the answer to a question.
  - *interrupt-topic*: an utterance that interrupts the dialogue partner or change topic (if the speaker interrupts herself) without ending current topic with an *end-topic*.
- *Adjacency pair (question-answer)*: beginning with the utterance where an interview question is asked, ending with the first relevant answer to that question.

Since the annotations of adjacency pairs are sequencies, they might contain only two turns, the pairs:

(1) Passenger: "`What is your occupation?`"
Driver: "`HMI expert is what my card says`"

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

(2) P: "`What star sign are you?`"
D: "`This is not where I should turn, is it?`"
P: "`No it's not`"
D: "`Ok`"
P: "`Star sign?`"
D: "`Scorpio`"

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 it is answered. Therefore, there might be occasions such as in example (2) where the interview question, due to an interruption, is asked twice. In this case, 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 the answer, so that the annotations include how the pair is interrupted and how it is reraised.

The hypothesis is that, although the in-vehicle environment might force the speakers to sometimes change their normal dialogue strategies, interruptions are typically not done within an adjacency pair. We believe that for courtesy and cognitive load reasons the speaker as far as possible strive to complete an adjacency pair before interrupting. However, if necessary, speakers might interrupt during the small talk that sometimes follows an answered question.

Adjacency pairs are coded only within the interview domain. The interview domain contains explicit questions, and the interaction is comparable to a human-computer interaction where the driver (who is being interviewed) has the role of the user and the passenger (who is interviewing) has the role of the system. Furthermore, only the adjacency pairs that contain the actual intervew question has been coded. Often, the participants continue to talk about an interview topic after the answer has been given but this conversation is considered to be small talk and not necessary for the task.

Regarding the navigation task, neither the passenger nor the driver knew the entire route. The passenger was only allowed to give one instruction at a time and was not allowed to look further in the navigation instructions. The instructions were furthermore not easy to understand, since the test leaders wanted both participants to be engaged in the discussion and the interpretation of the instructions to elicit frequent domain shifts. This is not comparable to a user interacting with a navigation system, and therefore there is no sense in coding the navigation domain.

The hypothesis is that interruptions are typically not done within an adjacency pair. That is, the interrupting speaker does not switch to another topic or domain before a relevant answer has been given. However, if necessary she interrupts during the small talk that sometimes follows an answered question.

6 Results

6.1 Pauses

A total number of 143 pauses in the passenger's speech were investigated. The least common pause type was the kind of pause that occurs at the beginning of a speaker's turn, at a TRP (*pause initial*). They made up 14% of the pauses. The other two pause types had approximately the same frequency: *pauses internal within* 41% and *pauses internal between* 45%.

Since pause distribution is normally positively skewed, median values are more appropriate to describe central tendencies than mean values (Heldner and Edlund, 2010). In Figure 1 the passenger's median pause lengths are shown, divided into different pause categories and with/without high cognitive load for the driver.

What we can see in Figure 1 is that when the driver is experiencing high cognitive load, the passenger's median pause length is longer in all three pause categories. However, this difference is not statistically significant in this rather small sample.

Also noticeable is that the pause type with the longest median length, the pause that occurs at a possible TRP but where turn change does not occur, is the one type of pause where speakers need time both to plan their utterance and to negotiate possible speaker change.

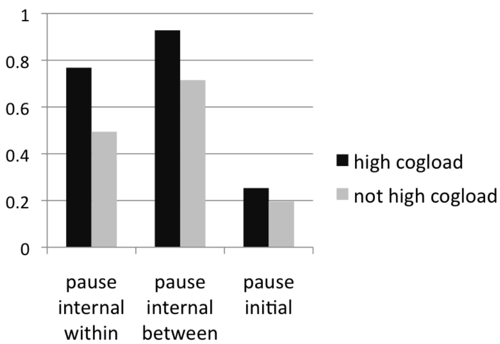


Figure 1: Passenger's median pause length in seconds

6.2 Adjacency pairs

When looking at how a new topic is introduced (i.e the speaker makes a *begin* or an *interrupt* utterance), we found that the passengers makes the most *begin* utterances, 136 compared to 27 for the drivers. This is not surprising since the questionaire and the navigation instructions were given to the passenger, who consequently were in charge of these tasks. They make, however, an almost equal amount of *interrupt* utterances, 61 for the drivers and 71 for the passengers. Figure 2 shows how the interruptions in the interview domain is divided between those which occur within an adjacency pair and those which occur within a topic but before or after an adjacency pair, respectively.

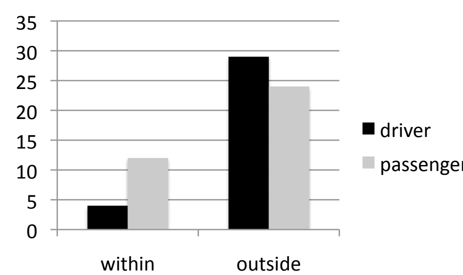


Figure 2: Number of interruptions in the interview domain, within and outside an adjacency pair.

We can see that the number of interruptions within an adjacency pair is about a fifth compared to interruptions outside a pair, and even less when it comes to passenger utterances. There is no significant difference in behaviour between driver and passenger. An one-sample binomial test revealed that the difference between interruptions within and outside an adjacency pair is significant at $\alpha < .01$.

Figure 3 shows which domain is the target domain when the speaker is interrupting an interview topic within an adjacency pair and within topic, respectively.

We can see that there is a similar behaviour for both conditions. The navigation domain is the most common domain to interrupt to for both the driver and the passenger.

When looking at the video recordings we found that both within and outside an adjacency pair all interruptions take place immediatly before a crossing or a road sign and therefore were time critical. Thus there is no difference in behaviour depending on where the interruption takes place, the interruptions that take place within an adjacency pair is not more time critical than those which take place outside. We therefore wanted to know if the behaviour differed depending on who is interrupted. We distinguish between *self interrupt* and *dialogue partner interrupt*. Within an adjacency pair, a self interrupt occur when the speaker that gives the first part of the pair is the one that is switching topic before the pair is finished. Since the passenger is the one that is interviewing it is only he or she that can do a self interrupt within an adjacency pair. A dialogue partner interrupt occur when the speaker that has not started the pair is the one that switches topic before the pair is finished, i.e. only the driver can do a dialogue partner interrupt within an adjacency pair. Outside an adjacency pair, a self interrupt occur when the speaker that is switching topic is the speaker that uttered the last utterance before the switch. A dialogue partner interrupt occur if the dialogue partner is the one that uttered the last utterance before the switch. Table 1 shows the result.

	Self interrupt		**Dp interrupt**	
	driver	**pass**	**driver**	**pass**
within	0	12	4	0
outside	10	10	19	14

Table 1: Interrupted speaker within and outside an adjacency pair.

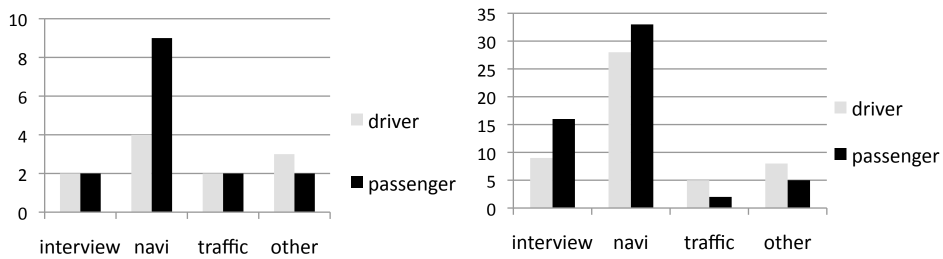
We can see that within an adjacency pair, the passenger interrupts more often that the driver does. Outside, both speakers interrupt themselves equally often, while the driver interrupts her dialogue partner more often than the passenger do. The results can only be seen as tentative, as this is a small sample and the results are not significant.

7 Discussion

In this rather small sample, it is still possible to discern a tendency for pauses to become longer when cognitive load is detected in the dialogue partner (the driver). This is in line with our hypothesis and also with previous research, which shows that pauses become longer when a speaker is experiencing high cognitive load, and that speakers tend to adjust their pauses to become more equal in length to the dialogue partner's pauses. A possible application of these results would be to construct in-vehicle dialogue systems so that they are able to adapt in a similar way; that is, when they detect cognitive load in the driver they should increase pause lengths to reduce cognitive load in the driver. It is however important that pause lengths are not too long. Pauses that exceed the expected length could lead the driver to think that there is some problem in the conversation (Roberts et al., 2006), or that there are technical problems with the dialogue system. If pauses are too long, we risk an increased rather than a decreased cognitive load level.

Since our measure of cognitive load did not allow a quantification of cognitive load, but merely a detection of high cognitive load versus *not* high cognitive load, we have not been able to investigate how pause lengths vary with variation in high cognitive load.

Moving forward, we now intend to investigate all of the material with regards to pauses, to see if the tendencies shown in part of the material are visible in all of it. It is very possible that a more in-depth analysis of pause lengths would reveal more interesting pause patterns. For example, in future research we plan to investigate what happens at the transition to high cognitive load, and compare pause lengths just before and at the beginning of high cognitive load. It is reasonable to believe that pause length is perceived as relative to the pauses that have occurred just previously, as opposed to relative to a perceived



(a) Within an adjacency pair. (b) Outside an adjacency pair.

Figure 3: Target domain when interrupting an interview topic.

mean pause length for the whole conversation.

In the cases where humans need to interrupt the topic that is currently discussed, our results show that both drivers and passengers tries to avoid interrupting within an adjacency pair. This is also supported by the study described by Shyrokov et al (2007). In the cases were they do interrupt it is to ask for or give time critical information about the navigation task. As described in Section 2.2, the reason why we do not interrupt within an adjacency pair unless absolutely necessary could be due to convention. 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 subject. Shyrokov et al (2007), as reported in Section 4, also suggested that it is less cognitively demanding to resume to an interrupted topic if the adjacency pair has been completed before the interruption. In addition to this, we suggest that when performing two cognitively demanding task such as driving a car and being interviewed simultaneously, it might increase the cognitive load even more to interrupt an adjacency pair to shift topic. Interrupting before an answer has been given would be cognitively demanding for the responding dialogue partner, since her mind is set on finding the answer to the question and therefore is not mentally prepared to receive other information. The following example from the corpus illustrates how difficult it can be to receive new and unexpected information while trying to come up with an answer to a question:

(3) P: `What fruit do you like... the most tasty fruit.`
D: `Eeh... Well it is... Let's see... mm... oh... well...`
P [interrupting]: `You should follow this road and then turn right.`
D [answering]: `Yes.`
D [resuming]: `Let's take pineapple. That is a very tasty fruit.`
P: `Pineapple? Yes, that is very tasty.`
D: `Sorry, I didn't listen. Should I turn here?`

Table 1 showed who is interrupted within and outside an adjacency pair. This is a small study with only a few participants, but the results may still indicate which strategy humans use during high cognitive load. The driver avoid to interrupt within an adjacency pair, probably because she is occupied with finding an answer to the question and therefore is not paying attention to the navigation task. The passenger, instead, interrupts three times as often, indicating that her mind is not as occupied with the interview task once the question is raised and therefore can pay attention also to the navigation task. Outside an adjacency pair they interrupt both themselves and their dialogue partner almost equally often. The reason might be that when an answer has been given both the driver and the passenger are more focused on the navigation task and less on the small talk.

8 Conclusion

Further research on a larger sample needs to be done, but the results reported in this paper indicate that a dialogue system could help lowering the user's cognitive load by changing dialogue strategy.

This can be done by

- increasing pause length within a turn
- avoid switching task within an adjacency pair

The turn-taking rule of not aborting an adjacency pair holds even if it is urgent to switch topic. In human-human dialogue, the dialogue partners strive to complete an adjacency pair, and therefore a dialogue system should do the same. If, for example, the navigation system needs to give an instruction this should not be done at a fixed distance (for example, always give an instruction at X meters before a crossing), instead the dialogue system should plan its utterances so that such an instruction is given before an adjacency pair is started or after it is finished.

In further research, the location of pauses within syntactic units should be investigated, to see if cognitive load has an impact on the placement of pauses. It would also be necessary to map pause times in more detail, to be able to apply these results within dialogue systems.

References

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

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*.

Joseph N. Cappella. 1979. Talk-silence sequences in informal conversations I. *Human Communication Research*, 6(1):3–17.

Jens Edlund, Mattias Heldner, and Julia Hirschberg. 2009. Pause and gap length in face-to-face interaction. In *Proceedings of Interspeech*, pages 2779–2782.

Mattias Heldner and Jens Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555 – 568.

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 17th International Workshop on Computational Semantics (IWCS-7)*.

Steven C Levinson. 1983. *Pragmatics*. Cambridge Textbooks in Linguistics. Cambridge University Press.

Felicia Roberts, Alexander L. Francis, and Melanie Morgan. 2006. The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in conversation. *Speech Communication*, 48(9):1079–1093.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

Emanuel A Schegloff and Harvey Sacks. 1973. Opening up closings. *Semiotica*, 7:289–327.

Alexander Shyrokov, Andrew Kun, and Peter Heeman. 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.

Monique E. van Donzel and Florian J. Koopmans van Beinum. 1996. Pausing strategies in discourse in dutch. In *Proceedings of the Fourth International Conference on Spoken Language, (ICSLP 96)*, pages 1029–1032.

Jessica Villing. 2009. Dialogue behaviour under high cognitive load. In *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 322–325. Association for Computational Linguistics.