Situationally Aware In-Car Information Presentation Using Incremental Speech Generation: Safer, and More Effective

Spyros Kousidis[1], **Casey Kennington**[1,2], **Timo Baumann**[4], **Hendrik Buschmeier**[2,3], **Stefan Kopp**[2,3], and **David Schlangen**[1]
[1]Dialogue Systems Group, [2]CITEC, [3]Sociable Agents Group – Bielefeld University
[4]Department of Informatics, Natural Language Systems Division – University of Hamburg
spyros.kousidis@uni-bielefeld.de

Abstract

Holding non-co-located conversations while driving is dangerous (Horrey and Wickens, 2006; Strayer et al., 2006), much more so than conversations with physically present, "situated" interlocutors (Drews et al., 2004). In-car dialogue systems typically resemble non-co-located conversations more, and share their negative impact (Strayer et al., 2013). We implemented and tested a simple strategy for making in-car dialogue systems aware of the driving situation, by giving them the capability to interrupt themselves when a dangerous situation is detected, and resume when over. We show that this improves both driving performance and recall of system-presented information, compared to a non-adaptive strategy.

1 Introduction

Imagine you are driving on a relatively free highway at a constant speed and you are talking with the person next to you. Suddenly, you need to overtake another car. This requires more attention from you; you check the mirrors before you change lanes, and again before you change back. Plausibly, an attentive passenger would have noticed your attention being focused more on the driving, and reacted to this by interrupting their conversational contribution, resuming when back on the original lane.

Using a driving simulation setup, we implemented a dialogue system that realises this strategy. By employing incremental output generation, the system can interrupt and flexibly resume its output. We tested the system using a variation of a standard driving task, and found that it improved both driving performance and recall, as compared to a non-adaptive baseline system.

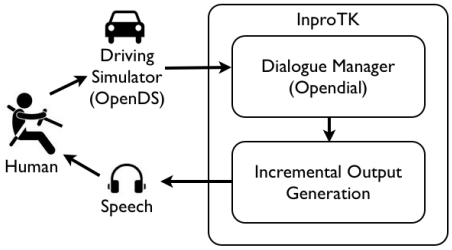

Figure 1: Overview of our system setup: human controls actions of a virtual car; events are sent to DM, which controls the speech output.

2 The Setup

2.1 The Situated In-Car System

Figure 1 shows an overview of our system setup, with its main components: a) the driving simulator that presents via computer graphics the driving task to the user; b) the dialogue system, that presents, via voice output, information to the user (here, calendar entries).

Driving Simulation For the driving simulator, we used the OpenDS Toolkit,[1] connected to a steering wheel and a board with an acceleration and brake pedal, using standard video game hardware. We developed our own simple driving scenarios (derived from the "ReactionTest" task, which is distributed together with OpenDS) that specified the driving task and timing of the concurrent speech, as described below. We modified OpenDS to pass real-time data (e.g. car position/velocity/events in the simulation, such as a gate becoming visible or a lane change) using the *mint*.tools architecture (Kousidis et al., 2013). In addition, we have bridged INPROTK (Baumann and Schlangen, 2012) with *mint*.tools via the Robotics Service Bus (RSB, Wienke and Wrede (2011)) framework.

[1]http://www.opends.eu/

Proceedings of the of the EACL 2014 Workshop on Dialogue in Motion (DM), pages 68–72,
Gothenburg, Sweden, April 26-30 2014.

Figure 2: Driver's view during experiment. The green signal on the signal-bridge indicates the target lane.

Dialogue System Using INPROTK, we implemented a simple dialogue system. The notion of "dialogue" is used with some liberty here: the user did not interact directly with the system but rather indirectly (and non-intentionally) via driving actions. Nevertheless, we used the same modularisation as in more typical dialogue systems by using a dialoge management (DM) component that controls the system actions based on the user actions. We integrated OpenDial (Lison, 2012) as the DM into INPROTK,[2] though we only used it to make simple, deterministic decisions (there was no learned dialogue policy) based on the state of the simulator (see below). We used the incremental output generation capabilities of INPROTK, as described in (Buschmeier et al., 2012).

3 Experiment

We evaluated the adaptation strategy in a driving simulation setup, where subjects performed a 30 minute, simulated drive along a straight, five-lane road, during which they were occasionally faced with two types of additional tasks: a lane-change task and a memory task, which aim to measure the driving performance and the driver's ability to pay attention to speech while driving, respectively. The two tasks occured in isolation or simultaneoulsy.

The Lane-Change Task The driving task we used is a variant of the well-known lane-change task (LCT), which is standardised in (ISO, 2010): It requires the driver to react to a green light positioned on a signal gate above the road (see Figure 2). The driver (otherwise instructed to remain in the middle lane) must move to the lane indicated by the green light, remain there until a tone is sounded, and then return again to the middle lane. OpenDS gives a *success* or *fail* result to this task depending on whether the target lane was reached within 10 seconds (if at all) and the car was in the middle lane when the signal became visible. We also added a speed constraint: the car maintained 40 km/h when the pedal was not pressed, with a top speed of 70 km/h when fully pressed. During a Lane-change, the driver was to maintain a speed of 60 km/h, thus adding to the cognitive load.

[2]OpenDial can be found at `http://opendial.googlecode.com/`.

Table 1: Experiment conditions.

Lane Change	Presentation mode	Abbreviation
Yes	CONTROL	CONTROL_LANE
Yes	ADAPTIVE	ADAPTIVE_LANE
Yes	NO_TALK	NO_TALK_LANE
No	CONTROL	CONTROL_EMPTY

The Memory Task We tested the attention of the drivers to the generated speech using a simple true-false memory task. The DM generated utterances such as *"am Samstag den siebzehnten Mai 12 Uhr 15 bis 14 Uhr 15 hast du 'gemeinsam Essen im Westend mit Martin' "* (on Saturday the 17th of May from 12:15–14:15 you are meeting Martin for Lunch). Each utterance had 5 information tokens: day, time, activity, location and partner, spoken by a female voice. After utterance completion, and while no driving distraction occurred, a confirmation question was asked by a male voice, e.g. *"Richtig oder Falsch? – Freitag"* (Right or wrong? – Friday). The subject was then required to answer true or false by pressing one of two respective buttons on the steering wheel. The token of the confirmation question was chosen randomly, although tokens near the beginning of the utterance (day and time) were given a higher probability of occurrence. The starting time of the utterance relative to the gate was varied randomly between 3 and 6 seconds before visibility. Figure 3 gives a schematic overview of the task and describes the strategy we implemented for interrupting and resuming speech, triggered by the driving situation.

3.1 Conditions

Table 1 shows the 4 experiment conditions, denoting if a lane change was signalled, and what presentation strategy was used. Each condition appeared exactly 11 times in the scenario, for a total of 44 *episodes*. The order of episodes was randomly

CONTROL *am Samstag den siebzehnten Mai um 12 Uhr hast du 'Besprechung mit Peter'*

ADAPTIVE *am Samstag den siebzehn-* *den siebzehnten Mai …*

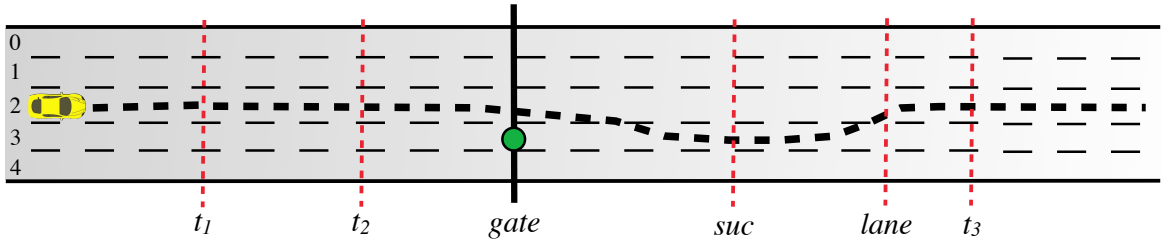

Figure 3: Top view of driving task: as the car moves to the right over time, speech begins at t_1, the gate with the lane-change indicator becomes visible at t_2, where in the adaptive version speech pauses. Successful lane change is detected at `suc`; successful change back to the middle lane is detected at `lane`, and resumes. (If no change back is detected, the interruption times out at t_3). All red-dotted lines denote events sent from OpenDS to the Dialogue Manager.

generated for each subject. With this design, subjects perceive conditions to be entirely random.

3.2 Dependent Variables

The dependent variables for the Memory task are (a) whether the subject's answer was *correct* (true/false), and (b) the *response delay*, which is the time from the end of the clarification question to the time the true or false button was pressed. For the driving task, the dependent variables are the OpenDS performance measurements *success/failure* (as defined above) and *reaction time* (time to reach the target lane).

3.3 Procedure

After signing a consent form, subjects were led into the experiment room, where seat position and audio level were adjusted, and were given written instructions. Next, the OpenDS scenario was initiated. The scenario started with 10 successive lane-change signal gates without speech, for driving training. An experimenter provided feedback during training while the subjects familiarized themselves with the driving task. Following the training gates came a clearly-marked "START" gate, signifying the beginning of the experiment to the subjects (at this point, the experimenter left). There was a "FINISH" gate at the end of the scenario. The whole stretch of road was 23 km and took approximately 30 minutes to complete. After the driving task, the subjects were given a questionnaire, which asked them to identify the information presentation strategies and assign a preference.

Table 2: Subjects' judgement of task difficulty.

Diff.	Freq.
4 (easy)	8
3	7
2	1
1 (hard)	1

Table 3: Subjects' system preference.

Preference	Freq.
ADAPTIVE	3
CONTROL	9
Neither	5

4 Results

In total, 17 subjects (8 male, 9 female, aged 19-36) participated in the study. All of the subjects were native German speakers affiliated with ANONYMIZED University. As reported in the post-test questionnaire, all held a driving license, two had previous experience with driving simulators and only one had previous experience with spoken dialogue systems. Table 2 shows the subjects' assessment of difficulty, while Table 3 shows their preference between the different strategies. Most subjects found the task relatively easy and either prefer the speech *not* to adapt or have no preference.

Memory task The overall percentages of correct answers to the system's recall questions (across all subjects) are shown in Table 4. We see that the subjects' performance in this task is considerably better when the system adapts to the driving situation (ADAPTIVE_LANE condition) rather than speaking through the lane change (CONTROL_LANE condition). In fact, the performance in the ADAPTIVE_LANE condition is closer to the control upper

Table 4: Performance in memory task per condition.

Condition	Percentage
CONTROL_EMPTY	169/180 (93.9%)
ADAPTIVE_LANE	156/172 (90.7%)
CONTROL_LANE	150/178 (84.3%)

Table 5: Success in driving task per condition (as reported by OpenDS).

Condition	Success
NOTALK_LANE	175/185 *(94.6%)*
ADAPTIVE_LANE	165/174 *(94.8%)*
CONTROL_LANE	165/180 *(91.7%)*

bound (CONTROL_EMPTY condition). We tested significance of the results using a generalized linear mixed model with CONDITION and SUBJECT as factors, which yields a p-value of 0.027 when compared against a null model in which only SUBJECT is a factor. No significant effects of between-subjects factors *gender*, *difficulty* or *preference* were found. In addition, the within-subject variable *time* did not have any significant effect (subjects do not improve in the memory task with time).

The average response delay (from the end of the recall question to the button press) per condition across all subjects is shown in Figure 4. Subjects reply slower to the recall questions in the CONTROL_LANE condition, while their performance in the ADAPTIVE_LANE condition is indistinguishable from the CONTROL_EMPTY condition (in which there is no distraction). Additionally, there is a general decreasing trend of response delay with time, which means that users get acquainted with the task (type of information, format of question) over time. Both factors (condition and time) are significant (repeated measures ANOVA, 2x2 factorial design, $F_{condition} = 3.858, p = 0.0359, F_{time} = 4.672, p = 0.00662$). No significant effects were found for any of the between-subject factors (gender, difficulty, preference).

Driving task The success rate in the lane-change task per condition is shown in Table 5. Here too we find that the performance is lower in the CONTROL_LANE condition, while ADAPTIVE_LANE does not seem to affect driving performance, when compared to the NOTALK_LANE condition. The effect is significant ($p = 0.01231$) using the same GLMM approach and factors as above.

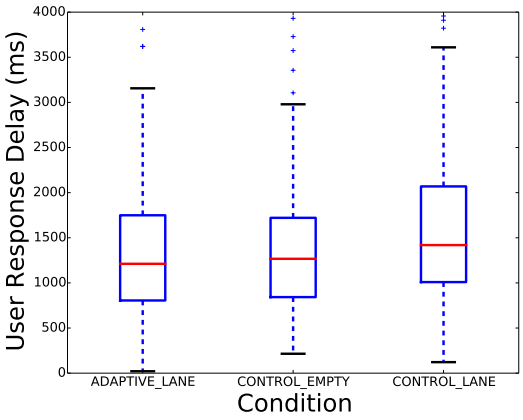

Figure 4: User answer response delay under three conditions.

5 Discussion, Conclusions, Future Work

We have developed and tested a driving simulation scenario where information is presented by a spoken dialogue system. Our system has the unique ability (compared to today's commercial systems) to adapt its speech to the driving situation: it interrupts itself when a dangerous situation occurs and later resumes with an appropriate continuation. Using this strategy, information presentation had no impact on driving, and dangerous situations no impact on information recall. In contrast, a system that blindly spoke while the driver was distracted by the lane-change task resulted in worse performance in both tasks: subjects made more errors in the memory task and also failed more of the lane-change tasks, which could prove dangerous in a real situation.

Interestingly, very few of the subjects preferred the adaptive version of the system in the post-task questionnaire. Among the reasons that they gave for this was their inability to control the interruptions/resumptions of the system. We plan to address the issue of control by allowing future versions of our system to accept user signals, such as speech or head gestures; it will be interesting to see whether this will impact driving performance or not. Further, more sophisticated presentation strategies (e.g., controlling the complexity of the generated language in accordance to the driving situation) can be tested in this framework.

Acknowledgments This research was partly supported by the Deutsche Forschungsgemeinschaft (DFG) in the CRC 673 "Alignment in Communic-

ation" and the Center of Excellence in "Cognitive Interaction Technology" (CITEC). The authors would like to thank Oliver Eckmeier and Michael Bartholdt for helping implement the system setup, as well as Gerdis Anderson and Fabian Wohlgemuth for assisting as experimenters.

References

Timo Baumann and David Schlangen. 2012. The InproTK 2012 release. In *NAACL-HLT Workshop on Future directions and needs in the Spoken Dialog Community: Tools and Data (SDCTD 2012)*, pages 29–32, Montréal, Canada.

Hendrik Buschmeier, Timo Baumann, Benjamin Dosch, Stefan Kopp, and David Schlangen. 2012. Combining incremental language generation and incremental speech synthesis for adaptive information presentation. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 295–303, Seoul, South Korea.

Frank A. Drews, Monisha Pasupathi, and David L. Strayer. 2004. Passenger and cell-phone conversations in simulated driving. In *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*, pages 2210–2212, New Orleans, USA.

William J. Horrey and Christopher D. Wickens. 2006. Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors*, 48:196–205.

ISO. 2010. Road vehicles – Ergonomic aspects of transport information and control systems – Simulated lane change test to assess in-vehicle secondary task demand. ISO 26022:2010, Geneva, Switzerland.

Spyros Kousidis, Thies Pfeiffer, and David Schlangen. 2013. MINT.tools: Tools and adaptors supporting acquisition, annotation and analysis of multimodal corpora. In *Interspeech 2013, Lyon, France*. ISCA.

Pierre Lison. 2012. Probabilistic dialogue models with prior domain knowledge. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 179–188, Seoul, South Korea.

David L Strayer, Frank A Drews, and Dennis J Crouch. 2006. A comparison of the cell phone driver and the drunk driver. *Human Factors*, 48:381–91.

David L Strayer, Joel M Cooper, Jonna Turrill, James Coleman, and Nate Medeiros. 2013. Measuring cognitive distraction in the automobile. Technical report, AAA Foundation for Traffice Safety.

J Wienke and S Wrede. 2011. A middleware for collaborative research in experimental robotics. In *System Integration (SII), 2011 IEEE/SICE International Symposium on*, pages 1183–1190.