Experimental modeling of human-human multi-threaded dialogues in the presence of a manual-visual task

Alexander Shyrokov, Andrew Kun
Electrical and Computer Engineering
University of New Hampshire
{shirokov, andrew.kun}@unh.edu

Peter Heeman
Center for Spoken Language Understanding
Oregon Health and Science University
heeman@cslu.ogi.edu

Abstract

We discuss the design and preliminary results of an experiment for modeling human-human multi-threaded dialogues. We found that participants tend to complete adjacency pairs in dialogues before switching to a new dialogue thread. We also have indications that, in the presence of a manual-visual task, the difficulty of the task influences switching between dialogue threads.

1 Introduction

Humans can carry on multi-threaded dialogues in which several dialogue threads overlap in time. Humans can do this while they are involved in manual-visual tasks, such as driving. For example a driver can discuss the weather with one passenger in the car, while periodically talking to another passenger about directions. However, it is an unsolved problem how to enable human-computer spoken multi-threaded interaction, especially while the human participant is involved in a manual-visual task. Our major hypothesis is that this problem can be solved by applying models of human-human interactions to human-computer interactions.

In this paper, we describe an experimental approach to model human-human spoken interactions in the presence of a manual-visual task, specifically driving a simulated car. We performed experiments with pairs of participants who were involved in an ongoing task but periodically needed to switch to an interrupting task. In the ongoing task one of the participants drove a simulated car and received verbal navigation instructions from the other participant who had a map of the simulated

world but was not in the driving simulator. The interrupting task was initiated by a visual stimulus presented to the driver in the simulator and it had to be completed verbally. The driver had to initiate the switch to the new dialogue thread verbally.

We were interested in three elements of the model of this human-human interaction. First, we investigated how the urgency of the interrupting task affects the timing of the interrupting task. We hypothesized that more urgent interruptions will be dealt with more quickly.

Next we looked in which dialogue state participants choose to initiate a switch to the interruption dialogue thread. We define the state of the dialogue in terms of whether the speakers are in the midst of an adjacency pair.

Finally, we explored the relationship between driving task difficulty and how quickly participants initiated an interruption. From our previous experiments we know that driving task difficulty has a significant influence on the performance of spoken tasks in the simulator. Therefore, we expect that driving task difficulty (and in general manual-visual task difficulty) has to be incorporated in our model. We hypothesized that participants will respond to interruptions more quickly when the driving task is less difficult.

2 Related Research

We investigate the use of multi-threaded dialogues similarly to cognitive load studies in which participants switch between two separate manual-visual tasks (McFarlane, 1999). In our prior work we explored the timing of switches between dialogue threads in human-human conversations, depending on the urgency of the interrupting task (Heeman, 2005). We found that some participants varied the place within a dialogue where they switch to the

interrupting task, depending on the urgency of the interrupting task. However, the tasks were artificial, that of playing a card game and determining whether a player has a certain colored shape on their computer screen. Furthermore, only gross discourse structure was examined, rather than the local discourse phenomena of adjacency pairs.

3 Experiment

Two participants took part in each session. One was assigned the role of a police officer, and the other was the dispatcher. The police officer operated the driving simulator, while the dispatcher sat in another room. Participants used headsets and microphones to communicate with each other. This task was related to the ongoing work at the University of New Hampshire on the Project54 system. The Project54 system integrates devices in police cruisers and provides a speech user interface to these devices (Kun, 2004). Our use of navigation as the ongoing task was inspired by the Map Task experiments (Anderson, 1991).

3.1 Ongoing task

We conducted our experiments in a high-fidelity driving simulator with a 180º field of view and a motion base, as shown in Figure 1. The simulator presented a city scenario with two-lane (one lane for each direction) roads (7 meters wide). The city consisted of sixteen intersections organized in a four-by-four grid, as shown in Figure 2. The limits of the area were marked with construction barrels. The officer was instructed not to drive past the barrels. Participants were not allowed to travel faster than 30 mph and they were required to stop at every stop sign, in order to lower the possibility of motion sickness (Mourant, 2000).

The dispatcher had a map with four marked locations that the officer had to visit. In order to ensure that the officer and the dispatcher engaged in a dialogue with each other, some city streets were also blocked with construction barrels, as shown in Figure 2. The barrel locations changed dynamically depending on the officer's location. The officer had to explain to the dispatcher if a street was open, so the dispatcher could make corrections to his/her instructions.
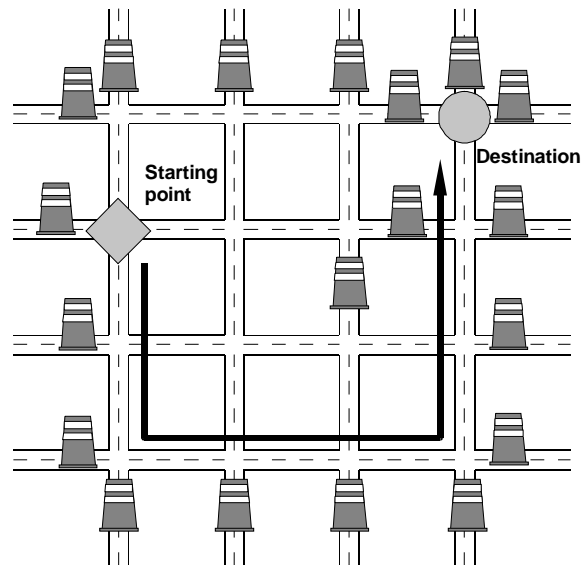

Figure 1. Driving simulator.

Figure 2. Blocked streets and possible path.

3.2 Interrupting task

Periodically the officer was presented with a visual stimulus. The officer then had to tell the dispatcher about the visual stimulus. Visual stimuli consisted of a text message and a progress bar. We used two different text messages for the interrupting task to make sure that the participants shift their attention from the ongoing task.

A progress bar was used to inform the officer about the urgency of the stimulus. Visual stimuli had one of two urgency levels. Officers had to respond to "urgent" visual stimuli (47% of all visual stimuli) within 10 seconds. For "non-urgent" visual stimuli officers had 20 seconds to respond. If the officer failed to inform the dispatcher about a visual stimulus within these time limits, the car would stop moving for 10 seconds (these car break-downs were controlled by the experimenter). Participants were told to complete the ongoing task as fast as

possible, and car break-downs provided an additional incentive to inform the dispatcher about visual stimuli quickly.

3.3   Procedure and participants

Participants were given an overview of the simulator, and were trained to perform the ongoing task and then both tasks. Training took about 10 minutes. Participants then performed the actual experiment which lasted about 30 minutes. At the end, participants completed questionnaires and received a debriefing. The experiment was completed by ten participants (five pairs) between 20 and 43 years of age. The average age of the participants was about 30 years and 30% were female.

4   Analysis and Results

We recorded the speech of all participants, as well as the car position. Vehicle data was collected at 10 Hz, resulting in about 90,000 vehicle data points for 2.5 hours of driving. We also recorded the time the visual stimuli appeared and synchronized these times with the audio recording of the participants. The five pairs of participants were presented with a total of 286 visual stimuli.

We analyzed three aspects of the data. First we looked at the average response time of the officer to urgent and non-urgent visual stimuli. We found no significant difference in response time depending on the urgency of the interruption (one tail t-test p=0.434), possibly because participants did not realize that some interruptions were more urgent than others.
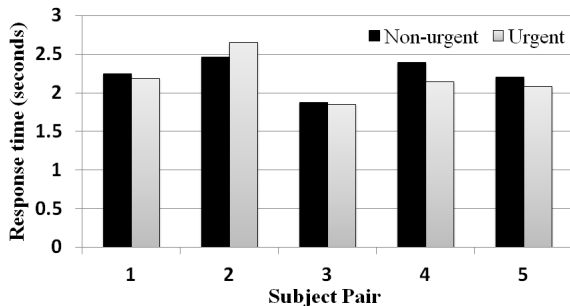

Figure 3. Average response times.

Figure 3 shows the plot of average response times for different participants. The response times are slower (average around 2.8 seconds for all cases) than reported by Tsimhoni et al. (2001) (average 1.3 seconds), who investigated reading messages on a heads-up display while driving. A rea-

sonable explanation for this is that in our experiment the officer was engaged in verbal communication with the dispatcher and did not pay as close attention to the messages as the participants in the study of Tsimhoni et al. Even more likely, the officer was complying with established conventions in human-human dialogue, and so waited for a suitable point in the interaction. This waiting for an opportunity to speak slowed down his/her response.

We next analyzed what dialogue states allow people to initiate a dialogue thread switch. Figure 4 shows a model of the local dialogue state of the ongoing task, based on sequences of adjacency pairs (Schegloff, 1973). In the first part of an adjacency pair, either the dispatcher or the police officer speaks (e.g. poses a question). We denote the first part with "a" when the dispatcher speaks and with "e" when the officer speaks. After a pause (denoted with "b" after the dispatcher speaks and "f" after the officer speaks), the dialogue continues with the second part of the adjacency pair. The second part is denoted with "c" when the officer speaks and with "g" when the dispatcher speaks. Finally, when the second part ends, and before the next first part begins, we have a pause in the dialogue, denoted with "d."
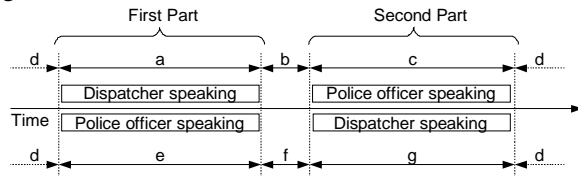

Figure 4. Adjacency pairs.

We coded each presentation of a visual stimulus with "a" through "g" based on where it happened with respect to the model in Figure 4. Each presentation resulted in the eventual initiation of an interruption (switch to the interrupting task). We also coded the interruption initiated by the officer based on where it happened with respect to the model in Figure 4. Note that the officer could have ignored the visual stimulus, but this happened only 5 out of 286 times, hence we did not further consider these cases. This left us with 7 x 7 = 49 possible types of interruption. In this paper, we decided to focus on interruptions in which the stimulus occurred during the first part of an adjacency pair ("a" or "e") as this is the point in the local discourse structure that is the most embedded.

192

When a stimulus is presented during the officer's first part ("e") 10% of the time the officer interrupts his/her own first part ("ee"). In 25% of the cases he/she completes the first part and then introduces the interruption ("ef"). In about 1% of the cases the officer introduces the interruption during the dispatcher's second part ("eg"). Most often, in 51% of the cases, the officer waits until after the adjacency pair is over ("ed"). In about 11% of the cases the officer introduces the interruption during the first part of the next adjacency pair when the dispatcher is speaking ("ea"). Finally, in 3% of the cases he/she interrupts after the dispatcher's first part in the next adjacency pair ("eb").

When the stimulus is presented while the dispatcher is speaking the first part ("a"), the officer interrupts immediately in about 23% of the cases ("aa") and after the first part in about 26% of the cases ("ab"). Again, most often, 40% of the time, the interruption came after the adjacency pair was over ("ad"). In about 2% of the cases each, the interruption came in the next adjacency pair during or after the officer's first part.

The above data shows that the officer often waited to initiate the interrupting task until after the adjacency pair was done. This might account for the difference between the average response times in this study and the one reported by Tsimhoni et al (2001).

Finally, we also looked at the average response time of officers during difficult and easy driving conditions. We defined difficult driving as driving within a radius of 10 meters of the center of an intersection. We found that officers on average responded slower under difficult driving conditions, however, our findings were not statistically significant. Note that the officers spent only about 8% of their time driving through the intersections and thus, on average this resulted in 5 visual stimuli out of 57 being presented in difficult driving conditions.

5   Conclusion and Future Directions

In this paper, we tried to determine the conventions that humans follow in initiating a switch to a new dialogue thread. We found that when the stimulus to signal the interruption was in the first part of an adjacency pair, participants either immediately interrupted the first part, or waited until the conclusion of the adjacency pair. This might indicate that participants were trying to avoid having the first part of an adjacency pair pending during a thread switch, so that there is a simpler discourse context to resume to. However, more analysis is needed to fully explore this issue, including examining other stimulus points, and distinguishing between different types of adjacency pairs.

Our analysis also shows that we need to further revise our task setup. We need to revise the experimental setup so that the urgency of the interrupting task is more realistic. We also need to better balance the easy with the difficult driving segments in order to better understand the impact of driving difficulty.

Acknowledgements

References

A. Anderson, M. Bader, E. Bard, E. Boyle, G. M. Doherty, S. Garrod, S. Garrod, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson and R. Weinert. 1991. *The HCRC Map Task Corpus,* Language and Speech, 34:351-366.

Peter Heeman, Andrew L. Kun, Fan Yang and Alexander Shyrokov. 2005 *Conventions in Human-Human MultiThreaded Dialogues: A Preliminary Study,* IUI'05.

Andrew L. Kun, W. Thomas Miller, III and William H. Lenharth. 2004. *Computers in police cruisers,* IEEE Pervasive Computing, 3(4):34-41.

D. McFarlane. 1999. *Coordinating the Interruption of People in Human-Computer Interaction,* Angela Sasse and Chris Johnson, Eds. Human–Computer Interaction.

Ronald R. Mourant and Thara R. Thattacherry. 2000. *Simulator Sickness in a Virtual Environments Driving Simulator*, Proceeding of the 44th Annual Meeting of the Human Factors and Ergonomics Society, 534-537.

O. Tsimhoni, P. Green and H. Watanabe. 2001. *Detecting and Reading Text on HUDS: Effects of Driving Workload and Message Location,* Paper presented at the 11th Annual ITS America Meeting, Miami, FL.

E. A. Schegloff and H. Sacks. 1973. *Opening up closings,* Semiotica VIII: 4: 290-327.