Context Restoration in Multi-Tasking Dialogue

Fan Yang
Center for Spoken Language Understanding
Oregon Health & Science University
20000 NW Walker Rd., Beaverton OR, USA
fly@cslu.ogi.edu

Peter A. Heeman
Center for Spoken Language Understanding
Oregon Health & Science University
20000 NW Walker Rd., Beaverton OR, USA
heeman@cslu.ogi.edu

ABSTRACT

In this paper we conduct an exploratory experiment on context restoration in multi-tasking dialogue and report our preliminary findings. We examine a corpus of human-human dialogues, in which pairs of conversants, using speech, work on an ongoing task while occasionally completing real-time tasks. We investigate whether the conversants, when returning to the ongoing task, make any effort to restore the context. First, we identify two types of actions, utterance restatement and information review, as possible restorations. Second, from a statistical analysis, we find that these actions are used more often when returning to the ongoing task, and hence seem to play a role in context restoration. Our findings will help to build a foundation for future speech interfaces that support multi-tasking dialogue.

Author Keywords

Context Restoration, Multi-Tasking Dialogue

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces—*Voice I/O*

INTRODUCTION

Speech interfaces have been used to perform a single task at a time, where the user finishes with one task before moving on to the next. However, we envision that future speech interfaces will be able to work with the user on multiple tasks at the same time, which is especially needed for real-time tasks. For instance, a driver might use a speech interface to catch up on emails, while occasionally checking upcoming traffic conditions, and receiving navigation instructions.

Several speech interfaces that allow multi-tasking dialogues have been built [5, 6]. However, these interfaces freely switch between different tasks, with little effort to help the user with task switching. Thus the user might be confused about which task the interface is talking about and the progress of the task. Multi-tasking dialogues, even in the best circumstances, will be difficult for users, as users need to be aware of task switching and remember the details of each task.

In order to build a speech interface that supports multi-tasking dialogue, we need to determine a set of conventions that the user and interface can follow in task switching. To determine such a set, we propose to start with conventions that are actually used in human-human speech conversations, which are natural for users to follow and probably efficient in problem-solving. Multi-tasking dialogues, where multiple independent topics overlap with each other in time, regularly arise in human-human conversation: for example, a driver and a navigator in a car might be talking about their summer plans, while occasionally interjecting road directions or conversation about what music to listen to.

We have collected the MTD corpus [4], which consists of a set of human-human dialogues where pairs of conversants have multiple overlapping verbal tasks to perform: an ongoing task type that takes a long time to finish, and a real-time task type that can be done in a couple of turns but has a time constraint. In previous research, we examined how conversants switch from the ongoing task to a real-time task [8], which we refer to as *task interruption*. We now, in this paper, examine the opposite problem: how conversants switch from the real-time task back to the ongoing task. More specifically, we focus on how conversants re-establish the conversation on the ongoing task after they complete a real-time task, which we refer to as *context restoration*.

Multi-tasking dialogues have received only minor attention in the discourse analysis literature. Grosz and Sidner briefly discussed task interruption in their discourse structure theory [3]. When a task interruption happens, a focus space is created and pushed on top of the discourse stack (which contains the focus space of the ongoing task). There is an impenetrable separation between the interruption space and the interrupted ongoing task so that the interruption cannot access the ongoing task. When the interrupting task is completed, its focus space is popped off and the ongoing task becomes salient. However, Grosz and Sidner did not address whether or how conversants restore the context of the ongoing task after completing the interrupting task.

In the rest of the paper, we first briefly describe the MTD corpus. We then examine the behavior of context restoration, and finally discuss our findings.

IUI'09, February 8-11, 2009, Sanibel Island, Florida, USA.

CORPUS

In order to fully understand multi-tasking human-human dialogue, we collected the MTD corpus, in which pairs of subjects perform overlapping verbal tasks [4].

Design of Tasks

A pair of conversants, or players, work on two types of tasks via conversation: an on-going task, which takes a long time to finish, and a real-time task, which has a time constraint and just takes a couple turns to complete.

In the ongoing task, the pair of players work together to form as many poker hands as possible, where a poker hand consists of a full house, flush, straight, or four of a kind. Each player has three cards in hand, which the other cannot see. Players are separated so that they cannot see each other. Players take turns drawing an extra card and then discarding one, until they find a poker hand, for which they earn 50 points; they then start over to form another poker hand. To discourage players from rifling through the cards to look for a specific card without talking, one point is deducted for each picked-up card, and 10 points for a missed or incorrect poker hand.

The poker game is played on computers. Figure 1 shows the computer display for a player. The player with four cards can click on a card to discard it. The card disappears from the screen, and an extra card is automatically dealt to the other player. The player with four cards clicks the "Done Poker Hand" button to start a new game once they find a poker hand.

The real-time task is a picture game. From time to time, the computer generates a prompt for one player to find out whether the other has a certain picture on the bottom of the display. The picture game has a time constraint of 10, 25 or 40 seconds, which is (pseudo) randomly determined. The players get 5 points if the correct answer is given in time. The overall goal of the players is to earn as many points as possible from the two games.

To alert the player to the picture game, two solid bars flash above and below the player's cards. Thus the player will be aware of a waiting picture game without taking attention away from the poker game. The color of the flashing bars depends on how much time is remaining: green for 26-40 seconds, yellow for 11-25 seconds and red for 0-10 seconds. The player can see the exact amount of time left in the heading for the picture game. In Figure 1, the player needs to find out whether the other has a blue circle, with 6 seconds left.

Corpus Statistics

We transcribed and annotated twelve MTD dialogues totaling about 180 minutes of conversation. The dialogues were by six pairs of players, all native American-English speakers. Each pair participated in two sessions and each session lasted about 15 minutes. During each session, 9 picture games (3 for each time constraint) were prompted for each player. Of the 216 prompted picture games, 8 were never started by players (although in the post-experiment survey

Figure 1. The game display for players

all players reported that they never ignored a picture game on purpose.) Thus the corpus contains 208 picture games.

The ongoing task can be naturally divided into individual poker games, in which the players successfully complete a poker hand. Each poker game can be further divided into a sequence of card segments, in which players discuss which card to discard, or a poker hand is found. In total, there are 120 game segments and 781 card segments. As well, we grouped the utterances involved in each picture game into segments. Figure 2 shows an excerpt from an MTD dialogue with these annotations. Here b7 is a game segment in which players formed a poker hand of a flush; and b8, b10, b11, b12 and b14, inside of b7, are card segments. Also embedded in b7 are b9 and b13, each of which is a segment for a picture game. As can be seen, players switch from the ongoing poker-playing to a picture game. After the picture game is completed, the players resume the conversation on the poker-playing.

CONTEXT RESTORATION

In this section, we investigate the resumption of an interrupted task, i.e. how players in the MTD corpus re-establish the conversation on poker playing after an interruption from the picture game. In our previous research, we distinguished three types of interruptions based on the discourse context where the interruption occurs: (1) interruption at the end of

Figure 2. An excerpt of an MTD dialogue

a poker game: immediately after players achieving a poker hand, (2) interruption at the end of a card discussion: immediately after discarding a card, and (3) embedded interruption inside a card discussion: where players are deciding which card to discard. We use this same distinction in investigating context restoration: are players returning to the ongoing task that was interrupted (1) at the end of a poker game, (2) at the end of a card discussion, or (3) in the middle of a card discussion. Where the ongoing task is interrupted will undoubtedly affect how much context needs to be restored.

In the Middle of a Card Discussion

We start by investigating context restoration in the middle of a card discussion, because these have the deepest context in the poker playing. We explore the corpus to look for signs of context restoration behavior after an embedded interruption. We do so by examining the informational redundancy [7] of the first non-trivial utterance after completing a picture game. We define *trivial utterances* as those that are just a stall (such as "uh" and "um") or a simple acknowledgement (such as "okay", "uh-huh", and "alright"). In the rest of this paper, we use utterance to refer to non-trivial utterance when there is no ambiguity.

Probably due to the simplicity of the task setup, we find that in a lot of cases, after completing an embedded picture game, players continue poker playing without a clear indication of context restoration. As shown in Example 1, A suspends B's question and inserts an interruption of picture game. However, right after the completion of the picture game, A gives the answer; the dialogue on poker playing continues as if the interruption never happened.

Example 1: Continuation

B: what do you have to make a high straight with?
> B: you got a red circle?
> A: no

A: I have a ten of diamonds and an ace of clubs

We do find two types of utterances at the beginning of a resumption that are informational redundant, as listed below.

Utterance Restatement: The first utterance after the interruption is a restatement of the last utterance before the interruption. This can be further divided into three subcategories: (A) self-repetition: the player repeats (part of) their own utterance, as shown in Example 2; (B) other-repetition: the player repeats (part of) the other's utterance, as shown in Example 3; and (C) clarification: the player asks for a repetition with a clarification question, as shown in Example 4.

Example 2: Self-Repetition

B: I have three clubs right now
> B: do you have a yellow square?
> A: yes

B: I have three clubs
B: do you have any clubs?

Example 3: Other-Repetition

B: I have jack and two queens
> B: um do you have a yellow plus sign?
> A: yes

A: a jack and two queens
A: I have a ten

Example 4: Clarification

A: I have a six of clubs a nine of spades and a four of diamonds
B: okay
> B: okay how about a uh red cross
> A: no
> B: okay

B: four diamonds six something?
A: clubs

Card Review: The player re-communicates the information of all cards in hand, which has been communicated before, as shown in Example 5. Two conditions must be satisfied for an utterance to be a card review: the utterance informs about all of the cards in the player's hand, and this information has already been communicated.

Example 5: Card Review

A: so I got a ten of spades
B: alright
> B: and do you have a red circle?
> A: um yes
> B: I mean no no a blue circle
> A: oh yes

A: and okay I have a queen of spades a ten of I mean a queen of diamonds a ten of spades a king of clubs and a two of clubs

After all 115 cases of embedded interruptions, we find in total 34 cases of utterance restatement (20 self-repetitions, 4

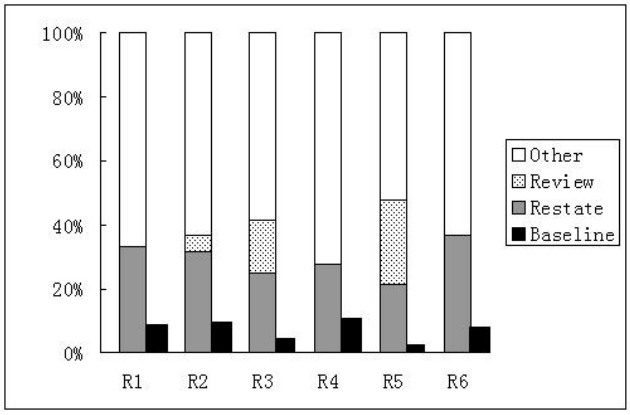

Figure 3. Restoration in the middle of a card discussion

other-repetitions, and 10 clarifications) and 9 cases of card review. Figure 3 shows the rate of each category, aggregated on each pair of players (R1-R6).

The rate of performing utterance restatement, calculated as the number of embedded interruptions that are followed by an utterance restatement divided by the total number of embedded interruptions, ranges from 22% to 37% among the player pairs. To make sense of these numbers, we annotate each utterance in the poker playing as whether it is a restatement of the previous one within a card segment, and calculate the baseline as the rate of performing utterance restatement without being interrupted by a picture game. From Figure 3, we can see that for all six pairs of players, the rates of performing utterance restatement after an embedded interruption are higher than the baseline, and this is statistically significant ($t(5) = 13.52$, $p < 0.001$). This suggests that utterance restatement after an embedded interruption is not a random behavior, but it is part of the resumption to the ongoing task.

We next examine card review, which seems not to be a common behavior in all player pairs. The pairs R1, R4, and R6 never perform it in resuming poker playing, and R2 only performs it once. The pairs with the highest rates are R5 and R3, with 26% (6/23) and 17% (2/12) respectively. Interestingly, these two pairs have the lowest rates of performing utterance restatement, with 22% and 27% respectively. This suggests that card review might be complementary to utterance restatement for context restoration, although more data is needed to validate this hypothesis. In future work, we will also compare these results to a baseline of where card review is used in the middle of card discussion but not immediately following an interruption.

After Interruption at the End of Card

We now examine context restoration after interruptions at the end of a card discussion. Again, we examine the first non-trivial utterance that follows an interruption, which is the first non-trivial utterance in a card segment.

We find that players mostly just continue the poker playing without a clear indication of being affected by the interruption. Some players perform an act similar to card review after an interruption at the end of a card discussion, in that they communicate all of the cards in hand. However, the version here differs as it includes the new card just picked up, which has not been communicated before. Thus we refer to this act as *card review + new card*.

Table 1 shows the rate of performing *card review + new card* after an interruption at the end of a card for each pair respectively. The baseline is the rate of performing this action at the beginning of a card segment (excluding the first card segment of a poker game) not following an interruption. Player pairs R1 and R5 never perform this action at all. R3 and R4 perform this action only once after interruptions at the end of a card discussion and have a very low overall rate of performing this action during poker playing. However, for player pairs that have a higher overall rate of using this action in the conversation, especially R6, the rate of performing this action after an interruption at the end of a card is significantly higher than the baseline ($\chi^2(1) = 6.81, p = 0.01$). This suggests that if players adopt *card review + new card* in conversation, they tend to use it after an interruption at the end of a card probably for context restoration.

Table 1. After interruptions at the end of a card

	R1	R2	R3	R4	R5	R6
Card Review +	0%	19% (3/16)	9% (1/11)	8% (1/13)	0%	42% (5/12)
Baseline	0%	5%	1%	3%	0%	5%

After Interruption at the End of Game

For interruptions at the end of a poker game, we do not identify any behavior that we associate with context restoration. This is not surprising because there is really no context that needs to be carried over to the next game.

CONCLUSION

In this paper, we conduct an exploratory experiment on context restoration in multi-tasking dialogue. In the MTD corpus, we find that when returning to an interrupted task, players have a higher rate of performing utterance restatement and card review. Utterance restatement resumes the conversation from the point where it was interrupted. Depending on whether it is the same speaker who resumes the ongoing task and who said the last utterance before the interruption, it takes the form of a self-repetition, other-repetition, or request-repetition if clarification is needed. This type of context restoration seems to use the *grounding* mechanism [1], where speakers use repetition (and requests for repetition) to ensure that the utterance is understood well enough for current purposes. The other type of context restoration is card review, or more generally, information review, which refreshes the information state. It re-synchronizes the conversation by summarizing the most critical information that the other speaker might have forgotten.

Probably due to the simplicity of the tasks in the MTD corpus, we only find two forms of context restoration: utterance

restatement and card review. In a more complex domain, we speculate that conversants will perform context restoration more frequently when returning to an interrupted task, and will use even higher-level summarization, such as the agreements or decisions that have been made. We hypothesize that context restoration is related to factors such as task complexity and conversants' memory capacity, and plan to use reinforcement learning to further study the underlying causes [2].

The research described in this paper will help to build a foundation for future speech interfaces that support multi-tasking dialogue. By understanding people's conventions in task interruptions [8] and context restoration, we can implement these conventions into a speech interface to allow natural and smooth task switching in multi-tasking human-computer dialogue. For example, the system can do the restoration for the user when resuming the ongoing task, because the user might need to be reminded of what happened while the system has perfect memory. The system can either restate the last utterance, or review the critical information, to help the user re-construct the attentional states. Future research is needed to determine when such restoration is needed, and what techniques (restatement, review, or even others) to use.

ACKNOWLEDGEMENT

This research was funded by the National Science Foundation under IIS-0326496.

REFERENCES

1. H. H. Clark and E. F. Schaefer. Contributing to discourse. *Cognitive Science*, 13:259–294, 1989.
2. M. S. English and P. A. Heeman. Learning mixed initiative dialog strategies by using reinforcement learning on both conversants. In *Proceedings of the HLT/EMNLP*, pages 1011–1018, 2005.
3. B. J. Grosz and C. L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.
4. P. A. Heeman, F. Yang, A. L. Kun, and A. Shyrokov. Conventions in human-human multithreaded dialogues: A preliminary study. In *Proceedings of IUI*, pages 293–295, San Diego CA, 2005.
5. A. L. Kun, W. T. Miller, and W. H. Lenharth. Computers in police cruisers. *IEEE Pervasive Computing*, 3(4):34–41, October-December 2004.
6. O. Lemon and A. Gruenstein. Multithreaded context for robust conversational interfaces: Context-sensitive speech recognition and interpretation of corrective fragments. *ACM Transactions on Computer-Human Interaction*, 11(3):241–267, 2004.
7. M. A. Walker. The effect of resource limits and task complexity on collaborative planning in dialogue. *Artificial Intelligence Journal*, 85:181 – 243, 1996.
8. F. Yang, P. A. Heeman, and A. Kun. Switching to real-time tasks in multi-tasking dialogue. In *Proceedings of CoLing*, pages 1025–1032, Manchester, UK, August 2008.