

Switching to Real-Time Tasks in Multi-Tasking Dialogue

Fan Yang and Peter A. Heeman

Center for Spoken Language Understanding
OGI School of Science & Engineering
Oregon Health & Science University
fly, heeman@cslu.ogi.edu

Andrew Kun

Electrical and Computer Engineering
University of New Hampshire
andrew.kun@unh.edu

Abstract

In this paper we describe an empirical study of human-human multi-tasking dialogues (MTD), where people perform multiple verbal tasks overlapped in time. We examined how conversants switch from the ongoing task to a real-time task. We found that 1) conversants use discourse markers and prosodic cues to signal task switching, similar to how they signal topic shifts in single-tasking speech; 2) conversants strive to switch tasks at a less disruptive place; and 3) where they cannot, they exert additional effort (even higher pitch) to signal the task switching. Our machine learning experiment also shows that task switching can be reliably recognized using discourse context and normalized pitch. These findings will provide guidelines for building future speech interfaces to support multi-tasking dialogue.

1 Introduction

Existing speech interfaces have mostly been used to perform a single task. However, we envision that next-generation speech interfaces will be able to work with the user on multiple tasks at the same time, which is especially useful for real-time tasks. For instance, a driver in a car 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 (Lemon et al., 2002; Kun et al., 2004). However, these interfaces

freely switch between different tasks without much signaling. Thus the user might be confused about which task the interface is talking about. Multi-tasking dialogues, even in the best circumstances, will be difficult for users, as users need to remember the details of each task and be aware of task switching.

In order to build a speech interface that supports multi-tasking dialogue, there needs to be a set of conventions for the user and the interface to follow in switching between tasks. To design such a set, we propose to start with conventions that are actually used in human-human 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.

In order to better understand the human conventions on task switching, we have collected the MTD corpus (Heeman et al., 2005), which consists of a set of human-human dialogues where pairs of conversants have multiple overlapping verbal tasks to perform: an ongoing task that takes a long time to finish, and a real-time task that can be done in a couple of turns but has a time constraint. This paper is focused on how conversants switch from the ongoing task to a waiting real-time task.

Previous research suggested the correlation between task switching and certain discourse context; for example, conversants try to avoid task switching in the middle of an adjacency pair (Shyrovok et al., 2007). In a preliminary study (Heeman et al., 2005), we examined the timing when conversants switched from the ongoing task to a real-time task using some pilot data, and found that

conversants did not always switch to a real-time task as soon as it arose, but instead waited for different amounts of time depending on its time constraint. In this study, we hypothesize that conversants strive to switch at an opportune place in the ongoing task, and we examine the discourse context where task switching occurs for evidence to support this hypothesis.

We are also interested in the cues that conversants use to signal task switching. Although there is a substantial body of research on how people signal topic shifts in single-tasking speech (monologue and dialogue), such as using discourse markers and prosodic cues (see Section 2.2), little research work has been done in investigating task switching in multi-tasking dialogues. In this study, we examine discourse markers and prosodic cues for their correlations with task switching. We also examine combining these cues to recognize task switching with machine learning techniques.

In Section 2, we review related literature. In Section 3, we describe the MTD corpus. In Section 4, we examine the discourse contexts in which task switching occurs. In Section 5, we examine the use of discourse markers and prosody associated with task switching. In Section 6, we examine automatic recognizing task switching with machine learning techniques. We conclude the paper in Section 7.

2 Related Research

In this section, we first describe two existing speech interfaces that allow multi-tasking dialogues. These speech interfaces, however, freely switch between tasks as soon as a new task arises, and without much signaling. We then review literature on how people signal topic shifts in single-tasking speech, which sheds light on our research of signaling task switching in multi-tasking dialogues.

2.1 Speech Interfaces for MTD

Kun et al. (2004) developed a system called Project54, which allows a user to interact with multiple devices in a police cruiser using speech. The architecture of Project54 allows for handling multiple tasks overlapped in time. For example, when pulling over a vehicle, an officer can first issue a spoken command to turn on the lights and siren, then issue spoken commands to initiate a data query, go back to interacting with the lights

and siren (perhaps to change the pattern after the vehicle has been pulled over), and finally receive the spoken results of the data query. While the current implementation of Project54 assumes that the officer initiates the task switching (e.g. the one about lights and the one about data query), the system can initiate task switching too. However, Project54 does not provide infrastructure for signaling to the officer a system-initiated switch. Any such signaling would have to be hand-coded by developers.

Lemon et al. (2002) also explored multi-tasking in a dialogue system. They built a multi-tasking dialogue system for a human operator to direct a robotic helicopter on executing multiple tasks, such as searching for a car and flying to a tower. The system keeps an ordered set of active dialogue tasks, and interprets the user utterance in terms of the most active task for which the utterance makes sense. Conversely, during the system's turn of speaking, it can produce an utterance for any of the dialogue tasks. Thus the system does not take into account the user's cost of task switching. The system switches to a new task as soon as it arises, instead of at an opportune place to minimize the user's effort. Moreover, the system does not signal when it switches between tasks. As with the approach of Kun et al. (2004) to multiple devices, it is unclear whether an actual user will be able to understand such conversations. The user might become confused about which task the system is on.

2.2 Signaling Topic Shifts in STP

Although speech interfaces have not used cues to signal task switching, researchers have found various cues that people naturally use in single-tasking speech to signal topic shifts. These cues are a good starting point from which to study how people signal task switching in multi-tasking dialogue.

Signaling topic shifts in single-tasking speech is about signaling the boundary of related discourse segments that contribute to the achievement of a task. Two types of cues have been identified for signaling topic shifts. The first type is discourse markers (Moser and Moore, 1995; Schiffrin, 1987; Grosz and Sidner, 1986; Passonneau and Litman, 1997; Bangerter and Clark, 2003). Discourse markers can be used to signal the start of a new discourse segment and its relation to other discourse segments. For example, "now" might signal moving on to the next topic, while "well" might signal

a negative or unexpected response.

The second type of cue is prosody. In read speech, Grosz and Hirschberg (1992) studied broadcast news and found that pause length is the most important factor that indicates a new discourse segment. Ayers (1992) found that pitch range appears to correlate more closely with hierarchical topic structure in read speech than in spontaneous speech. In spontaneous monologue, Butterworth (1972) found that the beginning of a discourse segment exhibited slower speaking rate; Swerts (1995), and Passonneau and Litman (1997) found that pause length correlates with discourse segment boundaries; Hirschberg and Nakatani (1996) found that the beginning of a discourse segment correlates with higher pitch. In human-human dialogue, similar behavior was observed: the pitch value tends to be higher for starting a new discourse segment (Nakajima and Allen, 1993). In human-computer dialogue, Swerts and Ostendorf (1995) found that the first utterance of a discourse segment correlates with slower speaking rate and longer preceding pause. Clearly, prosody is used to signal topic shifts in single-tasking speech.

3 The MTD 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. Details of the corpus collection can be found in (Heeman et al., 2005).

3.1 Design of Tasks

Conversants work on two types of tasks via conversation: an on-going task that takes a long time to finish and a real-time task that just takes a couple turns to complete but has a time constraint.

In the ongoing task, a 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 simply 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 poker hand or incorrect poker hand. To complete



Figure 1: The game display for players

this game, players converse to share card information, explore and establish strategies based on the combined cards in their hands (Toh et al., 2006).

The poker game is played on computers. The game display, which each player sees, is shown in Figure 1. 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.

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 for the picture game 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 know that there is a waiting picture game without taking the 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 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.

3.2 Corpus Annotations

We transcribed and annotated ten MTD dialogues totaling about 150 minutes of conversation. The dialogues were by five 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 total 180 picture game prompted, 8 were never started by players¹. Thus the corpus contains 172 picture games.

The ongoing task can naturally be 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 105 game segments and 690 card segments in the corpus. 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 got a poker hand of 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 an segment for a picture game. As can be seen, players switched from the ongoing poker-playing to a picture game. After the picture game was completed, the conversation on the poker-playing resumed.

4 Where to Switch

In a preliminary study (Heeman et al., 2005), we found that players did not always switch to a real-time task as soon as it arose, but instead waited for different amounts of time depending on the time constraint of the real-time task. We thus hypothesize that players strive to switch at an opportune place in the ongoing task (poker-playing). There are three types of places where a player could suspend the poker playing and switch to a waiting picture game: (G) immediately after completing a poker game (at the end of a game), (C) immedi-

¹Although in the post-experiment survey all players reported that they never ignored a picture game on purpose

b7	A (u18):	I got two diamonds and one spade
	B (u19):	I have a pair of nines
	A (u20):	alright go
	B (u21):	no spades
	A (u39):	let's go for diamonds
	B (u22):	alright
b9	A (u23):	do you have a blue circle
	B (u24):	yes
b10	A (u25):	ok I got two diamonds
b11	B (u26):	I have jack
	B (u27):	I have two diamonds but I have or sorry one diamond and jack queen or jack ten nine
	A (u28):	I got an eight but
	B (u29):	eight
	A (u30):	um still doesn't give us a straight
	B (u31):	are we going for diamonds you said
	A (u32):	diamonds two diamonds
	A (u33):	I got two diamonds that is
	B (u34):	alright I'll get rid of one of my nines
b12	A (u35):	I got three diamonds you got two right
	B (u36):	I have one diamond
	A (u37):	oh only one ok
	A (u38):	I got three diamonds
b13	A (u39):	do you have a green plus
	B (u40):	green plus no
b14	A (u41):	I got three diamonds
	B (u42):	I have two diamonds
	A (u43):	yeah
b15	A (u44):	ok I have two clubs
	B (u45):	I have two clubs

Figure 2: An excerpt of an MTD dialogue

ately after discarding a card (at the end of a card), and (E) embedded inside a card segment, where players are deciding which card to discard. In this section, we examine where task switching occurs.

4.1 Time Constraint and Place of Switching

We first examine the place of switching under different time constraints. As shown in Table 1, for the time constraint of 10s, 75% of the task switching was embedded inside a card segment, 23% at the end of a card, and 2% at the end of a game; for the time constraints of 25s and 40s², 46% embedded inside a card segment, 33% at the end of a card, and 21% at the end of a game. The difference in the places of switching between the time constraint of 10s and 25s/40s is statistically significant ($\chi^2(2) = 15.92, p < 0.001$). The time constraint of 10s requires players to start a picture game very quickly in order to complete it in time. On the other hand, when given 25s or 40s, players are in a less hurry to switch. Compared with 10s, when players had 25s or 40s, the percentage of switching embedded inside a card segment decreases by 29%, while at the end of a card increases by 10%, and at the end of a game increases by 19%. These results suggest that when given more time, players try to switch at the end of a game or a card.

²We combined the time constraints of 25s and 40s because 25s seemed to be sufficient for most players.

Table 1: Time constraint and place of switching

	E	C	G	Total
10s	42 (75%)	13 (23%)	1 (2%)	56 (100%)
25/40s	54 (46%)	38 (33%)	24 (21%)	116 (100%)

Table 2: Waiting time and place of switching

	E	C	G	Total
$\leq 3s$	47 (69%)	18 (27%)	3 (4%)	68 (100%)
$> 3s$	49 (47%)	33 (32%)	22 (21%)	104 (100%)

4.2 Waiting Time and Place of Switching

We next examine the place of task switching from the perspective of waiting time. Waiting time refers to the time interval between when a picture game is prompted to a player and when the player actually starts the picture game. Our question is: if players wait at least a certain amount of time, where would they switch tasks? We arbitrarily choose a time amount of 3 seconds. We assume that when the waiting time is shorter than 3s, the player starts the picture game as soon as he or she notices it without significant waiting; in other words, based on human reaction time, if players are going to respond to it right away, they should be able to do so within 3s. The results are shown in Table 2. When the waiting time is shorter than 3s, 69% of the task switching is embedded inside a card segment, 27% at the end of a card, and only 4% at the end of a game; when longer than 3s, 47% is embedded inside a card segment, 32% at the end of a card, and 21% at the end of a game. The difference in the places of switching is statistically different ($\chi^2(2) = 11.88, p = 0.003$). When the waiting time is longer than 3s, the percentage of switching inside a card decreases by 22%, while switching at the end of a card increases by 5%, and at the end of a game increases by 17%. These results suggest that players wait for the end of a game or a card to switch to a picture game.

4.3 Discussion

We examined the discourse context of task switching, and found that 1) when given more time, players intend to switch to a picture game at the end of a (poker) game or a card; and (2) if players wait, they are waiting for the end of a (poker) game or a card to switch to a picture game. These results suggest that players strive to switch to a picture game at the end of a (poker) game or a card.

In fact, we also observed that after a picture game that is at the end of a game, players smoothly

start a new poker game as if nothing had happened; after a picture game that is at the end of a card, players might sometimes remind each other what cards they have in hands; while after a picture game that is in the middle of a card segment, players might even repeat or clarify the previous utterances that were said before the interruption. It is thus reasonable to assume that switching embedded inside a card segment is the most disruptive, followed by at the end of a card, and at the end of a game is the least. Our experiment results hence suggest that players strive to switch to a real-time task at a less disruptive place in the ongoing task. This is consistent with Clark and Wilkes-Gibbs (1986), that conversants try to minimize collaborative effort.

5 How to Switch

In Section 2.2, we discussed how people use certain cues, such as discourse markers and prosody, to signal topic shifts in single-tasking speech. This suggests that people might also signal task switching in multi-tasking dialogues. In this section, we examine how players signal that they are switching from the ongoing task to a real-time task with discourse markers and prosody.

5.1 Task Switching and Discourse Markers

Close examination of the MTD corpus found that “oh” was the most frequently used discourse marker when switching to a picture game. Another discourse marker, “wait” (including “wait a minute”), was often used together with “oh” in the way of “oh wait”. Thus we examined the use of “oh” and “wait” in switching to a picture game.

Players used the discourse markers “oh” or “wait” 14.5% (25/172) of the time in switching to a picture game. In poker playing, 5.7% (238/4192) of utterances contain the words “oh” or “wait”, and only 4.6% (32/690) of card segments are initiated with the two discourse markers (i.e. the first utterance of a card segment has “oh” or “wait” at the very beginning). Players have a statistically higher percentage of using “oh” or “wait” at task switching than in poker playing ($\chi^2(1) = 22.89, p < 0.001$) or to initiate a card segment ($\chi^2(1) = 21.84, p < 0.001$).

5.2 Task Switching and Prosody

To understand the prosodic cues in initiating a topic, traditionally researchers compared the

prosody of the first utterance in each segment with other utterances (e.g. (Nakajima and Allen, 1993; Hirschberg and Nakatani, 1996)). This approach encounters two problems here. First, the words in an utterance might affect the prosody. For example, the duration and energy of “bat” are usually larger than “bit”. Thus a large amount of data are required to balance out these differences. Second, in the MTD corpus, players typically switch to a picture game by using a yes-no question, such as “do you have a blue circle”, while most forward utterances (c.f. Core and Allen 1997) in the ongoing task are statements or proposals. As questions have very different prosody than statements or proposals, a direct comparison is further biased.

Examination of the MTD corpus found that 86% (148/172) of the picture games were initiated by “do you have ...” with optional discourse markers at the beginning. While in the poker game, players used “do you have ...” 108 times to ask whether the other had certain cards, such as “do you have a queen?” This observation inspired us to compare the prosody of the phrase “do you have” in switching to a picture game and during poker-playing.³ This avoids comparing prosody of different words or of different types of utterances.

We measure pitch, energy (local root mean squared measurement), and duration of each case of “do you have”. We aggregate on each player and calculate the average values. The results are shown in Table 3. The second and third columns show the average pitch of the phrase “do you have” for task-switching (SWT) and poker-playing (PKR) respectively. When switching to a picture game, players’ pitch is statistically higher than poker-playing ($t(9) = 4.15, p = 0.001$). In fact, for all ten players, the average pitch of “do you have” in switching to a picture game is higher than in poker-playing. These results show a strong correlation between task switching and higher pitch.

We next examine the correlation between energy and task switching. The fourth and fifth columns in Table 3 show the average energy of the phrase “do you have” for task switching and poker-playing respectively. We do not find a statistically significant difference ($t(9) = 0.80, p = 0.44$). We also examine the duration of “do you have”. The sixth and

³Note that most cases of “do you have” in poker-playing are not at the beginning of a card segment. It is also interesting to compare the prosody of “do you have” of initiating a picture game and of initiating a card segment. However, we do not have enough data for the latter.

Table 3: Average prosodic values for each player

Player	pitch (Hz)		energy		duration (s)	
	SWT	PKR	SWT	PKR	SWT	PKR
4A	136	123	383	266	0.28	0.38
4B	178	156	466	506	0.32	0.30
5A	164	152	357	367	0.37	0.25
5B	214	182	231	153	0.36	0.28
6A	144	126	414	370	0.32	0.21
6B	122	117	564	496	0.25	0.23
8A	238	199	973	1061	0.36	0.21
8B	150	143	246	180	0.33	0.35
9A	109	102	538	465	0.44	0.59
9B	125	122	702	814	0.33	0.24

Table 4: Pitch (Hz) and place of switching

Player	E	C & G	PKR
4A	137	131	123
4B	180	173	156
5A	167	161	152
5B	219	206	182
6A	146	143	126
6B	124	121	117
8A	245	233	199
8B	152	140	143
9A	110	108	102
9B	130	117	122

seventh columns in Table 3 show the results. We do not find a statistically significant difference either ($t(9) = 1.03, p = 0.33$). These results do not support that energy or duration (i.e. speaking rate) is correlated to task switching.

5.3 Signaling Place of Switching

To better understand how pitch is used in signaling task switching, we next examine its correlation with place of switching, i.e., switching at the end of a game, at the end of a card, or embedded inside a card segment. Because there are relatively less data for switching at the end of a game (see Table 1 and 2), we combine switching at the end of a game and at the end of a card (C & G) as a category.

Table 4 shows the average pitch of “do you have” when switching to a picture game embedded inside a card segment, at the end of a card or game segment, and during poker-playing. The difference between these three conditions is statistically significant ($F(2, 9) = 15.61, p < 0.001$). Switching embedded inside a card segment has a statistically higher pitch than switching at the end of a card or game segment ($t(9) = 5.54, p < 0.001$), which in turn has a statistically higher pitch than during poker-playing ($t(9) = 2.91, p = 0.01$).

5.4 Discussion

Consistent with previous research on topic shifts in single-tasking speech, our experiments show that switching to a real-time task correlates with the use of certain discourse markers and prosodic variations. It is not surprising that “oh” and “wait” correlate with task switching. Task switching involves a sudden change of the conversation topic, and previous research found that conversants use “oh” to mark a change of state in orientation or awareness (Heritage, 1984). “Wait” is used to mark a discontinuity in the ongoing topic, which is also required by task switching. Thus people may use these discourse markers to signal switching to a real-time task. In terms of prosodic variations, we find that task switching correlates with higher pitch. This suggests that pitch is used to signal switching to a real-time task.

Our experiments have also shown that pitch correlates to place of switching. As discussed in Section 4.3, task switching embedded inside a card segment is the most disruptive, switching at the end of a card is less, and at the end of a game is the least. Our results show that switching embedded in a card segment has a higher pitch than switching at the end of a card or a game, which in turn has a higher pitch than non-switching (poker-playing). This suggests that the degree of disruptiveness corresponds to the value of pitch: the more disruptive place to switch, the higher is the pitch.

From our results we speculate that pitch is used to divert the hearer from the ongoing task, signaling an unexpected event (c.f. (Sussman et al., 2003)). When task switching is more disruptive, the speaker uses higher pitch; probably because the hearer has a stronger expectation of the next utterance to be in the context of poker-playing. The use of higher pitch serves as a cue that the hearer should suspend the ongoing context and interpret the utterance in a new context. According to the theory of *least collaborative effort*, the effort of raising the pitch by the speaker is probably to reduce the effort of recognizing and processing the task switching by the hearer (Clark and Wilkes-Gibbs, 1986).

6 Machine Learning Experiment

In the previous sections, we showed the correlation of various cues with task switching. In this section, we conduct a machine learning experiment to determine whether we can reliably recognize task

switching using these cues. For the reasons given in Section 5.2, we limit our experiment to the 256 cases of “do you have”, 148 for task switching and 108 for poker playing. We train a decision tree classifier (C4.5) to discriminate task switching from poker playing. We use 5-fold cross validation to evaluate the performance. We use decision tree learning because its output is interpretable and we have found its performance comparable to other discriminative classifiers for this task.

The feature set includes 1) discourse context: whether the utterance before “do you have” is the end of a poker game, the end of a card segment, or in the middle of a card segment⁴; 2) cue word: whether the “do you have” follows the cue word “oh” or “wait”; and 3) normalized pitch: the pitch of “do you have” divided by the average pitch of the speaker during the dialogue.

The decision tree learning obtains an accuracy of 83% in identifying whether a “do you have” initiates a task switching or belongs to poker playing; and the recall, precision, and F measure for task switching are 90%, 82%, and 86% respectively. As a baseline, if we blindly assume that all cases of “do you have” are for task switching, we have an accuracy of 58%. Thus decision tree learning with the three features has 43% relative error reduction over the baseline.

To examine the structure of the decision tree, we build a single tree from all 256 cases of “do you have”. We find that the decision tree first examines the normalized pitch; if it is greater than 1.085, it is a task-switch. Otherwise, if the discourse context is at the end of a game, then it is for task switching; if the discourse context is embedded in a card segment, it is for poker playing; if the discourse context is at the end of a card: if normalized pitch is higher than 0.975 then it is for task switching, otherwise for poker playing. Interestingly, the feature of cue word is not used in the tree.

The performance and structure of the learned tree suggest that discourse context and normalized pitch are useful features for discriminating task switching.

7 Conclusion

In this paper we have described an empirical study of human-human multi-tasking dialogues, where people perform multiple verbal tasks overlapped

⁴Card and game segments can be determined fairly accurately from the mouse clicks even without the speech.

in time. We first examined the place of task switching, i.e. where players suspend the ongoing task and switch to a real-time task. Our analysis showed that people strive to switch at a less disruptive place. We then examined the cues to signal task switching. We found that task switching correlates with certain discourse markers and prosodic variations. More interestingly, the more disruptive the switching is, the higher is the pitch. We thus speculate that pitch is used by the speaker to help the listener be aware of task switching and understand the utterance. Finally, our machine learning experiment showed that discourse context and pitch are useful features to reliably identify task switching.

Acknowledgement

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

References

- Ayers, Gayle M. 1992. Discourse functions of pitch range in spontaneous and read speech. Presented at the Linguistic Society of America Annual Meeting.
- Bangerter, Adrian and Herbert H. Clark. 2003. Navigating joint projects with dialogue. *Cognitive Science*, 27:195–229.
- Butterworth, Brian. 1972. Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*, 4:75–87.
- Clark, Herbert H. and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognitive Science*, 22:1–39.
- Core, Mark G. and James F. Allen. 1997. Coding dialogues with the DAMSL annotation scheme. In *Working Notes: AAAI Fall Symposium on Communicative Action in Humans and Machines*, pages 28–35, Cambridge.
- Grosz, Barbara J. and Julia Hirschberg. 1992. Some intonational characteristics of discourse structure. In *Proceedings of 2nd ICSLP*, pages 429–432.
- Grosz, Barbara J. and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- Heeman, Peter A., Fan Yang, Andrew L. Kun, and Alexander Shyrokov. 2005. Conventions in human-human multithreaded dialogues: A preliminary study. In *Proceedings of IUI (short paper session)*, pages 293–295, San Diego CA.
- Heritage, John. 1984. A change-of-state token and aspects of its sequential placement. In Atkinson, J. M. and J. Heritage, editors, *Structures of social action: Studies in conversation analysis*, chapter 13, pages 299–345. Cambridge University Press.
- Hirschberg, Julia and Christine H. Nakatani. 1996. A prosodic analysis of discourse segments in direction-giving monologues. In *Proceedings of 34th ACL*, pages 286–293.
- Kun, Andrew L., W. Thomas Miller, and William H. Lenharth. 2004. Computers in police cruisers. *IEEE Pervasive Computing*, 3(4):34–41, October–December.
- Lemon, Oliver, Alexander Gruenstein, Alexis Battle, and Stanley Peters. 2002. Multi-tasking and collaborative activities in dialogue systems. In *Proceedings of 3rd SIGdial*, Philadelphia PA.
- Moser, Megan and Johanna D. Moore. 1995. Investigating cue selection and placement in tutorial discourse. In *Proceedings of 33rd ACL*, pages 130–135.
- Nakajima, Shin'ya and James F. Allen. 1993. A study on prosody and discourse structure in cooperative dialogues. Technical report, Rochester, NY, USA. technical report.
- Passonneau, Rebecca J. and Diane J. Litman. 1997. Discourse segmentation by human and automated means. *Computational Linguistics*, 23(1):103–139.
- Schiffirin, Deborah. 1987. *Discourse Markers*. Cambridge University Press.
- Shyrokov, Alexander, 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.
- Sussman, E., I. Winkler, and E. Schrgöer. 2003. Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bulletin & Review*, 10(3):630–637.
- Swerts, Marc and Mari Ostendorf. 1995. Discourse prosody in human-machine interactions. In *Proceedings of ESCA workshop on spoken dialogue systems: theories and applications*, pages 205–208, Visgo Denmark.
- Swerts, Marc. 1995. Combining statistical and phonetic analyses of spontaneous discourse segmentation. In *Proceedings of the 12th ICPHS*, volume 4, pages 208–211.
- Toh, Siew Leng, Fan Yang, and Peter A. Heeman. 2006. An annotation scheme for agreement analysis. In *Proceedings of 9th ICSLP*, pages 201–204, Pittsburgh PA.