COORDINATING TURN-TAKING WITH GAZE

David G. Novick, Brian Hansen and Karen Ward

Center for Spoken Language Understanding, Oregon Graduate Institute
P.O. Box 91000, Portland, OR 97291, USA

ABSTRACT

This paper explores the role of gaze in coordinating turn-taking in mixed-initiative conversation and specifically how gaze indicators might be usefully modeled in computational dialogue systems. We analyzed about 20 minutes of videotape of eight dialogues by four pairs of subjects performing a simple face-to-face cooperative laboratory task. We extend previous studies by explicating gaze patterns in face-to-face conversations, formalizing the most frequent pattern as a computational model of turn-taking, and testing the model through an agent-based simulation. Prior conversation simulations of conversational control acts relied on abstract speech-act representations of control. This study advances the computational account of dialogue through simulation of direct physical expression of gaze to coordinate conversational turns.

1. INTRODUCTION

Gaze plays a powerful and complex role in face-to-face conversation. People engaged in conversation may look at one another to monitor listener acceptance and understanding, to signal attention and interest, and to coordinate turn-taking [1, 5]. Conversely, they may look away to plan utterances or to concentrate on complex cognitive tasks. Beattie [2] found that gaze's role in turn-taking is context-specific: when the overall level of gaze is low, as in conversations between strangers [10] or when the discussion topic imposes a high cognitive load on the conversants, gaze plays a more significant role. This paper explores the role of gaze in coordinating turn-taking in mixed-initiative conversation and specifically how gaze indicators might be usefully modeled in computational dialogue systems.

Watanuki and Togawa [13] measured levels of mutual gaze in dyadic conversations. They reported that frequencies of mutual gaze and nodding were inversely related and concluded that these were complementary means of acknowledging the other participant. Their study used previously acquainted subjects discussing general topics through a video-mediated communications system. The study analyzed 70 seconds of dialogue from two pairs and generally reported results in terms of frequencies of observed behaviors, including speaking, nodding, and eye contact. The study did not address sequencing of these behaviors and did not explicate the mechanisms by which subjects used gaze to control the dialogue.

In this study, we expand upon Watanuki and Togawa's results through examination of a much larger corpus of face-to-face dialogue in a simple domain. Our analysis finds dynamic patterns of gaze associated with turn-taking; these findings enable us to begin modeling the physical actions that conversants use to coordinate turn-taking.

We formalized the most common pattern in a computational model of gaze-coordinated turn-taking that distinguishes dialogue *acts* (the symbolic and intentional contribution to the conversation) from linguistic *actions* (the realization of an act through observable communicative behaviors). We tested the model through an agent-based simulation. The system successfully generated the desired dialogues (those where no breakdowns occurred and no diagnosis was required) and ended with the expected acts and the expected beliefs in memory of each agent. Prior conversation simulations of conversational control acts had relied on the abstract representations of control acts between agents. Here we simulate the control mechanisms through the production of actions that express abstract acts in terms of physical actions.

In Section 2, we describe the laboratory task from which we developed our model. We then present the computational model and multi-agent simulation of observed dialogues in a rule-based environment. We conclude with a discussion of the results of the simulations.

2. EXPERIMENTAL PROCEDURE

To develop our computational model of gaze in conversation, we used a laboratory task designed to elicit task-oriented mixed-initiative conversations. Because we are interested in examining turn-taking, the domain was deliberately kept as simple as possible. Furthermore, we wanted the task to be relatively short so that complete conversations could be analyzed at a detailed level. We chose for our study a mental task, that of jointly reconstructing a sequence of random letters.

Two subjects were each given a card on which was printed a sequence of 17 letters and blanks (see Figure 1). The cards were prepared by producing a random sequence of letters and then introducing blanks so that (1) no more than seven letters occurred without an intervening blank, and (2) a given position was blank for at most one of the cards. Thus, the subjects had enough information to reconstruct the entire sequence but could do so only by collaborating.

```
Subject L:  O _ O S E _ A G F H C W _ E _ _ Z
Subject R:  _ S O S _ X _ G F H _ _ L E B Y _
```

Figure 1: Example letter sequences used by subjects.

Each subject was instructed to memorize the sequence on his or her own card. The cards were then put out of sight and the subjects were instructed to work together to reconstruct the complete sequence. Subjects were given as much time as they wanted to complete the task. Preliminary testing showed this task is difficult enough that subjects sometimes make mistakes, but not so difficult that they are likely to become frustrated. Nevertheless, the domain of discourse is simple enough to permit modeling of conversants' goals and belief states at a detailed level. Although the task imposes a high cognitive load on the participants, the utterances themselves tend to be very simple linguistically, often consisting of only the name of the next letter. While the domain of this conversational task is quite simple, it nonetheless captures some of the flavor of many real-world conversations. Each conversant attempts to reach mutual understanding cooperatively in the face of missing information.

A total of four pairs of subjects (two male, two female) each completed two letter sequences for a total of eight dialogues. Subjects were unacquainted adults and native speakers of American English and were matched by gender and approximate age. During the experiment subjects were seated approximately three feet apart, at a 90-degree angle to each other with a low coffee table between them. Three cameras were used to record the sessions. One camera was positioned in front of the subjects to capture the interaction as a whole. The other two cameras were positioned to the sides, one focused on the face of each subject to capture gaze direction and facial expressions.

Two transcripts were created for each dialogue. One transcriber prepared a detailed record of the hand gestures, gaze direction, and utterances of each speaker. A second transcriber, working from the video tapes and the detailed transcripts, prepared an explication in a narrative style similar to that suggested by Cook [3]. Each set of transcripts was checked by the other transcriber with discrepancies resolved informally. Complete transcripts and narratives for the dialogues are reported in [8]. Figure 2 shows the verbal and gaze behaviors for a brief excerpt of one dialogue**.**

Event Number	Left Subject		Gaze			Right Subject	
	Eyes	Verbal	L		R	Verbal	Eyes
38		O	<	<	<		
39	to		>		<		
40			>	**>**	>		away
41			>	>	>	S	
42		O	>	>	>	O	
43	away		<	**<**	<		to
44		C	<	<	<		
45		No	<	<	<		
46		S E	<	<	<		
47	to		>		<		
48			>	**>**	>		away
49			>	>	>	X	
50	away		<	**<**	<		to
51		A	<	<	<		

Figure 2: This brief excerpt of a transcript of a letter-sequence conversation shows verbal and gaze behaviors. The actual transcript used in the study included columns showing actions of the conversants' face and body. The markers in the L and R gaze columns indicate the direction of the left and right subjects' gaze. The markers in the middle gaze column indicate cases where one subject is looking at the other while the other is looking away; bold arrows in this column indicate the first event for which this pattern holds.

3. ANALYSIS

For the purposes of this study, a turn is defined as a period of speech from one speaker without verbal contribution from the other. This definition has several limitations [9], two of particular concern being overlapping speech and backchannel responses. When overlapping speech occurred, we coded a turn change if the original speaker stopped talking and the other continued speaking. Also, our definition of turn treats verbal and non-verbal back-channel responses differently. For example, a verbal "OK" counts as a turn but a head nod does not.

While transcribing the sessions, we observed that several patterns of gaze direction were associated with turn-taking. Two sequences in particular occurred fairly frequently. The first we term the "mutual-break" pattern: as one conversant completes an utterance, he or she looks toward the other. Gaze is momentarily mutual, after which the other conversant breaks mutual gaze and begins to speak. Overall, nearly 42 percent of turn exchanges followed this pattern, which is, in fact, consistent with the brief section of labeled multimodal data excerpted by Watanuki and Togawa in [13]. The second pattern, the "mutual-hold" pattern, is similar except that the turn recipient begins speaking without immediately looking away, although in many cases the turn recipient broke gaze during the course of the turn. This pattern was seen in 29 percent of turn exchanges.

In most cases, the conversants required more than one pass through the letter sequence before they were satisfied that they had reconstructed the sequence correctly. We found that each pass could be divided into two phases: in the main phase, the conversants attempted to construct the letter sequence; in a diagnostic phase, they determined whether they had succeeded and how to proceed. Our initial observations suggested that the "mutual-break" pattern seemed more prevalent in the main phase of the pass. An analysis of these sections of the interactions revealed that the gaze patterns were indeed different for the two different phases (t=2.11, df=25, p=0.045). Turn exchange coincided with the "mutual-break" pattern 57 percent of the time during the main phase of the interaction, but only 31 percent of the time during diagnosis.

The letter sequence task imposes significant requirements on the memory capacity of the conversants, often leading to misremembered letters and the need for self- and other-repair. Since it was our intent to model the behavior of agents in the absence of these impediments to smooth interaction, we examined the question of whether either pattern was associated with more successful conversations. We noted a trend (p=0.11) toward a negative correlation between the use of the "mutual-break" pattern and the number of turns in the conversations. Conversely, the "mutual-hold” pattern correlated positively with the number of turns taken (t=2.55, df=18, p=0.02). At the same time, there was a slight but not statistically significant negative correlation between use of the “mutual-break” gaze model and turn length, indicating that it was not associated with longer turns. In the absence of significant differences in turn length, the use of more turns suggests either a greater need for certainty on the part of the conversants, greater use of repair, or both. These observations suggest that the “mutual-break” gaze pattern was used more frequently when the conversation was proceeding smoothly (fewer turns required to complete the task) and mutual gaze was preferred where conversants were having difficulty (more turns required).

Based on our observations, we focused on the “mutual-break” gaze model for our initial computational implementation. It was observed more frequently during the conversations overall and especially when the conversations were proceeding smoothly. Consequently, this first model of turn-taking is not complete, as we include only one gaze pattern. Moreover, gaze is clearly not the only turn-taking mechanism available to speakers. Turn-taking occurs easily in conversations in which neither speaker looks at the other, for example while they are speaking over a telephone or are engaged in a task (such as driving) which requires their visual attention. Sacks, Schegloff, and Jefferson [11] related turn-taking to syntactically-defined “transition relevance points” and Duncan [4] identified six distinct prosodic, linguistic, or gestural turn-taking cues. This wide variety of turn-taking signals makes a precise account of gaze's role in conversational control elusive.

4. COMPUTATIONAL SIMULATION

To validate the “mutual-break” gaze model, we built and tested a working implementation of a rule set encompassing the letter-sequence domain, basic speech acts, and turn-taking. The domain rules capture the discourse planning and interactions that are peculiar to a particular domain; in this case, the domain rules implement this task's goals of confirming and exchanging information about individual letters and substrings of letters. The turn-taking rules and speech act rules attempt to capture domain-independent notions of performing speech acts (assert, request, and acknowledge) and of controlling the floor.

In our model, conversation is seen as an attempt to establish and build upon mutual knowledge using speech acts. This synthesis was first proposed by Novick [6] to explain conversational control or “meta-locutionary” acts. More recently, Traum and Hinkelman [12] developed a similar model of mutuality maintenance (or grounding) as part of their theory of “conversation acts.”

Until this study, our computational agents have communicated only in abstract speech-act representations. Because our first concern has been to develop and validate our basic models, we relied on human coders to translate the physical actions to abstract speech and dialogue control acts. Now, beginning with the turn-taking rules, we augment the abstract acts with rules reasoning more directly about physical actions. In this instance, instead of simulating turn exchange via acts like give_turn and accept_turn, we can use the actions look_at and look_away_from. This change in level of abstraction allows us to provide computational agents with more realistic reasoning processes for understanding human conversational phenomena.

The computational model was tested in *saso* [7], a rule-based Prolog tool for modeling multi-agent problems involving simultaneous actions and subjective belief. The conversants are represented by computational agents that communicate using the acts defined in the model. Saso uses forward-chaining control to simulate the parallel execution of multiple rule-based agents with separate belief spaces. A user-defined set of extended STRIPS-style operators is used to specify the behavior of agents in terms of preconditions and effects. Agents communicate through acts; when an agent performs an act, clauses representing the action are posted to the memories of all agents in the system. An initial state corresponding to the initial beliefs and intentions of the conversants was defined. The simulation was allowed to run to completion and the results were compared with the transcripts of the original conversation.

The system successfully generated the desired conversations (those where no breakdowns occurred and no diagnosis was required) and ended with the expected acts and beliefs in the memory of each agent. An excerpt from the summary trace is shown in Table 1 with the active gaze operators in bold type in distinct columns. This excerpt shows the steps that agents A and B go through in order to arrive at a mutual understanding of an entire sequence of three letters; the agent’s sequences are indicated in the table’s headers. Most operators represent individual intentions and are not shared with the other. Gaze operators and operators prefaced with “do_” (such as do_assert and do_request) are public and available to the other.

5. CONCLUSIONS

The simulation demonstrated that computational agents could successfully coordinate their conversation using gaze actions to signal their intention to give or keep the turn. However, our implementation is preliminary because gaze is only one indicator of turn change [4] and thus gaze alone is not sufficient to model all turn-taking behavior. Nevertheless, we found that the gaze pattern we modeled was a reasonably reliable indicator of turn handoff primarily when the task was going smoothly. When problems arose, other turn-taking mechanism were preferred. Because our computational model does not yet address misunderstanding or conversational repair, we did not implement mechanisms based upon those patterns.

Future work will include modeling misunderstanding and repair in general; as part of that work, we will examine more closely turn-taking mechanisms used during repair episodes. We will also model other routine and non-routine conversational control strategies found in our protocols as we continue to lexicalize and instantiate abstract acts in terms of the physical actions seen in human-human conversation.

6. ACKNOWLEDGMENTS

This research was supported by NSF awards IRI-9314959, IRI-9110797 and GER-9354959, by the Defense Advanced Projects Research Agency, and by the Center for Spoken Language

Event #	Agent A: [blank,i,s]		Agent B: [o,blank,s]	
	Non-gaze operators	Gaze operators	Gaze operators	Non-gaze operators
1.	goal_obtain_next_subsequence: [blank]			goal_confirm_next_subsequence: [o]
2.	goal_obtain_next_letter: blank			goal_confirm_next_letter: o
3.	goal_request_next_letter: blank			goal_assert_next _letter: o
4.	do_request: blank	**look_at: B**		
5.				request_received: blank
6.			**look_away_from: A**	goal_respond_next_letter: o
7.	other_look_away_from: A			do_assert: o
8.	assertion_received: o			
9.	do_acknowledgement: o			
10.	last_letter_obtained: o			acknowledgment_ received: o
11.	next_subsequence_obtained.: [blank]			last_letter_confirmed: o
5.	goal_confirm_next_subsequence: [i,s]			next_subsequence_confirmed: [o]
6.	goal_confirm_next_letter: i			goal_obtain_next _subsequence: [blank]
7.	goal_assert_next_letter: i			goal_obtain_next _letter: blank
8.				goal_request_next_letter: blank
9.			**look_at: A**	do_request: blank
10.	request_received: blank	**look_away_from: B**		
11.	do_assert: i			other_look_away_from: B

Table 1: Summary trace expert illustrating gaze operators

Understanding. We thank Kenneth D. Rubesh for his implementation of the gaze operators in the simulation.

7. REFERENCES

1. Argyle, M., and Cook, M. *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press, 1976.

2. Beattie, G. The role of language production processes in the organization of behavior in face-to-face interaction. In B. Butterworth (Ed.) *Language Production*, Vol. 1, 69-107, 1980.

3. Cook, G. Transcribing infinity: Problems of context presentation. *Journal of Pragmatics*, 14, 1-24, 1990.

4. Duncan, S. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23(2), 283-292, 1972.

5. Kendon, A. Looking in conversations and the regulation of turns at talk: A comment on the papers of G. Beattie and D. R. Rutter et al. *British Journal of Social and Clinical Psychology*, 17, 23-24, 1978.

6. Novick, D. *Control of mixed-initiative discourse through meta-locutionary acts: A computational model*. Technical Report No. CIS-TR-88-18, Department of Computer and Information Science, University of Oregon, 1988.

7. Novick, D. Modeling belief and action in a multi-agent system. In B. Ziegler & J. Rozenblit (Eds.), *AI, Simulation and Planning in High Autonomy Systems*. Los Alamitos, CA: IEEE Computer Society Press, 1990.

8. Novick, D., Hansen, B., and Lander, T. *Letter-sequence dialogues*. Technical Report CSE 94-007, Department of Computer Science and Engineering, Oregon Graduate Institute of Science and Technology, 1994.

9. O'Connell, D., Kowal, S., and Kaltenbacher, E. Turn-taking: a critical analysis of the research tradition. *Journal of Psycholinguistic Research*, 19(6), 345-373, 1990.

10. Rutter, D., Stephenson, G., Ayling, K., and White, P. The timing of looks in dyadic conversation. *British Journal of Social and Clinical Psychology*, 17, 17-21, 1978.

11. Sacks, H., Schegloff, E., and Jefferson, G. A simplest systematics for the organization of turn-taking for conversation. *Language*, 30, 696-735, 1974.

12. Traum, D., and Hinkelman, E. *Conversation acts in task-oriented spoken dialogue*. Technical Report 425, Department of Computer Science, University of Rochester., 1992.

13. Watanuki, K., and Togawa, F. Some signals of emotional arousal: Analysis of conversations using a multimodal database, *Proceedings of EuroSpeech ‘95*, Madrid, Sept., 1995, 1165-1168, 1995.