

Models of Gaze in Multi-party Discourse

David Novick

The University of Texas at El Paso

El Paso, TX 79968-0501, USA

novick@utep.edu

+1 915-747-5725

INTRODUCTION

In the last ten years or so, studies of gaze in conversational interaction have moved from cognitive studies of dyadic conversation toward implementation in avatars or embodied conversational agents and toward multi-party interaction. This position paper reviews these developments and suggests research paths related to cognitive models of gaze in multi-party interaction. These research paths would serve to extend underlying scientific models of interaction and would help develop immersive environments with agents that interacted more naturally or effectively.

DYADIC MODELS

Cognitive models of gaze (Argyle & Cook, 1976; Kendon, 1978; Beattie, 1980) and early work on agent-based computational simulation of gaze in turn-taking (Novick, Hansen & Ward, 1996) were basically dyadic. Two relatively simple patterns of gaze account for much of the observed turn-taking behavior in dyadic conversations. Over 70 percent of turn exchanges in the Letter Sequence Corpus (Novick, Hansen & Lander, 1994) used either the “mutual-break” pattern or the “mutual-hold” pattern. In 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. The value of the dyadic mutual-break model of gaze for turn-taking in dialogue between a human and an embodied conversation agent was empirically verified by van Es et al. (2002). The mutual-hold pattern is similar except that the turn recipient begins speaking without immediately looking away, although in many cases the turn recipient breaks gaze during the course of the turn.

The evidence is not clear as to why conversants use one pattern or the other, much less some other pattern entirely. Indeed, these patterns likely vary by cultural or racial group. Erickson (1979) found that whites mostly follow the mutual-gaze or mutual break pattern by looking at the speaker while listening but that blacks mostly look away when listening. For conversants who follow the mutual-gaze and mutual break patterns, Novick, Hansen & Ward (1996) found suggestive evidence that the mutual-hold pattern was associated with more difficult interactions. Vertegaal (1998) found that gaze was highly correlated with attention. Nakano et al. (2003) also explored the functions

of dyadic gaze patterns. Their findings were consistent with the mutual-break/mutual-gaze model and suggested that speakers interpret continuous gaze as evidence of non-understanding, which encourages the other conversant to elaborate until their contribution is grounded. Nakano et al. were also able to relate gaze to other non-verbal conversational behaviors.

While some of the underlying functions of gaze for conversational control in dyadic conversation have been explicated in the cognitively oriented studies previously discussed, many functions remain unclear. What is happening in the 30 percent of turns in which the conversants do not use the mutual-break or mutual-gaze patterns? Is gaze information gathered via peripheral vision used for turn-taking? So far, gaze has been modeled at a time-scale corresponding to turns, which are rarely less than a second long. Are there turn-related gaze effects at the 100-millisecond level? Similarly, are there gaze effects associated with back-channeling? Another line of inquiry would involve other functions of gaze in conversation. Some gaze behaviors, such as looking up, are generally considered turn-holding behaviors and other behaviors may communicate diverse factors such as affect.

MULTI-PARTY MODELS

The use of gaze in multi-party avatar-based video-conference systems (e.g., Vertegaal, 1999; Colburn, Cohen & Drucker, 2000) was based largely on the dyadic cognitive models of gaze discussed above. Much of this work focused on user-perceptual issues such as where users or avatars appeared to be gazing rather than on the underlying functions of gaze (e.g., Vertegaal et al., 2001). Matsusaka, Fujie and Kobayashi, (2001) used gaze models for turn-taking among robots, but cognitive models corresponding to the dyadic mutual-break and mutual-gaze patterns observed by Novick, Hansen and Ward (1996) and Nakano et al. (2003), and validated by van Es et al. (2002) apparently remain unexplored in the multi-party case. And all of the open issues from dyadic conversation have counterparts in multi-party conversation. Much of the work on gaze in multi-party interaction has involved mediated communications (e.g., Sellen, 1992; Vertegaal, 1998) rather than face-to-face interaction.

Research into cognitive models of multi-party discourse has produced some results with respect to gaze and turn-taking.

Parker (1988) showed that small group discussions are primarily made up of sequences of two-way conversations. This is because the floor is passed in discussions using highly visual means such as eye contact; if people naturally look at the last person to speak, then that person is at an advantage for speaking first when the next opportunity to change speakers comes up. Another early cognitive model of multi-party discourse (Novick, Walton & Ward, 1996) extended Clark and Schaefer's (1987, 1989) model of dialogue structure based on contribution trees. Analysis of transcripts of multi-party conversations indicated that multiple conversants created more complex contribution structures, and that these structures could be explicated by distinguishing primary and secondary evidence of understanding and by extending the definitions of presentation and acceptance to account for collaborative acceptance.

More recent work in multi-party turn-taking has included gaze based on cognitive models. For example, ICT's Mission Rehearsal Exercise (ICT, undated), serves as a multi-modal test-bed for a dialogue model that is founded on a layered model of cognitive functions in conversational interaction (Traum, 2002). There is also related work in multi-party turn-taking that does not include gaze as a factor. For example, some researchers (e.g., Dignum & Vreeswijk, 2003), have looked at turn-taking in multi-party dialogues from what one might call an engineering standpoint. This work involved the design of multi-agent or other multi-party communications environments in which multi-party turn-taking mechanisms were based on analytical approaches to efficiency rather than on human-inspired, cognitive models of turn-taking.

Many interesting issues remain open with respect to turn-taking and gaze in multi-party interaction. And many of these issues remain from those identified by Novick, Walton and Ward (1996): How is the content of a presentation affected by the presence of multiple hearers, each of whom a speaker may wish to leave with a different interpretation of the act? How does the level of evidence required by speakers change when there are several hearers present? Do speakers require stronger evidence of understanding because they cannot watch everyone at once? (In the mediated three-person case studied by Vertegeal (1998), speakers typically distributed their gaze over both of the other conversants.) Or do speakers require less evidence from any one hearer as long as they receive enough total evidence to convince themselves that they were understood? Do speakers aggregate acceptance of their presentations or do they require independent levels of acceptance from each addressee? How do the non-verbal presentations (e.g., raised eyebrows) and acceptances (e.g., continued attention) of face-to-face discourse function in the multiparty setting? How do participants in face-to-face discourse adapt their conversational skills in the presence of multiple targets for mutual gaze?

That these issues of gaze and turn-taking in multiparty interaction remain open suggests the continued salience of cognitive modeling as a research approach. It is not clear that the field is presently capable of producing results in the style of van Es et al. (2002) for multi-party interaction because the underlying cognitive multi-party gaze models have yet to be sufficiently articulated. Research approaches for developing such models might extend the methodology of Nakano et al. (2003), which related gaze to other non-verbal communicative elements, from the dyadic to the multi-party case. This would include new rounds of corpus collection, measuring gaze either through analysis of multi-camera video recordings or directly through eye-tracking, classifying behaviors into communicative acts, and finding patterns in the sequences of acts.

Intuitively, there should be a significant difference in gaze behaviors between two-person and three-person conversations, a nearly equally significant difference when moving to a four-party conversation, and similar but increasingly smaller differences as parties are added to the interaction. At some points there may be discontinuities, as the interaction moves from conversation to meeting to presentation. One might hypothesize that these discontinuities would correspond to changes in the physical arrangement of the parties, and that these arrangements would have corresponding design implications for immersive multi-agent environments.

ACKNOWLEDGMENTS

This paper benefited significantly from helpful comments and suggestions from Karen Ward, Nigel Ward and the students of UTEP's Interactive Systems Group (<http://www.cs.utep.edu/isg/>).

REFERENCES

1. Argyle, M., and Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press.
2. Beattie, G. (1980). 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.
3. Clark, H. & Schaefer, E. (1987). Collaborating on Contributions to Conversations, *Language and Cognitive Processes*, 2: 19- 41.
4. Clark, H. & Schaefer, E. (1989). Contributing to Discourse, *Cognitive Science*, 13: 259-294.
5. Colburn, A., Cohen, M. F., and Drucker, S. (2000). *The Role of Eye Gaze in Avatar Mediated Conversational Interfaces*, MSR-TR-2000-81. Microsoft Research.
6. Dijkstra, T., and Van Heuven, W. J. B. (1998). The BIA-model and bilingual word recognition. In J. Grainger and A. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189-225). Mahwah, NJ: Erlbaum.

7. Dignum, F. & Vreeswijk, G.A.W. (2003). Towards a test bed for multi-party dialogues. In F. Dignum, editor, *Advances in Agent Communication*, LNAI, Springer Verlag.
8. Erickson, F. (1979). Talking down: Some cultural sources of miscommunication in interracial Interviews, in Wolfgang, A. (ed.), *Nonverbal behavior: Applications and cultural implications*, Academic Press, 99-126.
9. ICT (undated). Mission Rehearsal Exercise. Available at http://www.ict.usc.edu/disp.php?bd=proj_mre.
10. Kendon, A. (1978). 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.
11. Matsusaka, Y., Fujie, S., and Kobayashi, T. (2001). Modeling of conversational strategy for the robot participating in the group conversation, *7th European Conference on Speech Communication and Technology (Eurospeech 2001)*, Aalborg, Denmark, September 3-7, 2001, 2173-2176.
12. Nakano, Y., Reinstein, G., Stocky, T., and Cassell, J. (2003). Towards a Model of Face-to-Face Grounding, *Proceedings of Association for Computational Linguistics*, Sapporo, Japan, July 7-12, 2003, 553-561.
13. Novick, D., Hansen, B., and Lander (1994). T. *Letter-sequence dialogues*. Technical Report CSE 94-007, Department of Computer Science and Engineering, Oregon Graduate Institute of Science and Technology, 1994.
14. Novick, D., Hansen, B., & Ward, K. (1996). Coordinating turn-taking with gaze, *Proceedings of ICSLP-96*, Philadelphia, PA, October, 1996, 3, 1888-91.
15. Novick, D., Walton, L., & Ward, K. (1996). Contribution graphs in multiparty conversations, *Proceedings of the International Symposium on Spoken Dialogue (ISSD-96)*, Philadelphia, PA, October, 1996, 53-56.
16. Parker, K. (1988). Speaking turns in small group interaction: A context-sensitive event sequence model. *Journal of Personality and Social Psychology*, 54(6), 965-971.
17. Sellen, A. (1992). Speech patterns in video-mediated conversations, *Proceedings of CHI'92*, Monterey, CA, 49-59
18. Traum, D. (2002). Ideas on Multi-layer Dialogue Management for Multi-party, Multi-conversation, Multi-modal Communication: Extended Abstract of Invited Talk, in *Computational Linguistics in the Netherlands 2001: Selected Papers from the Twelfth CLIN Meeting*, 1-7.
19. van Es, I., Heylen, D., van Dijk, B., and Nijholt, A. (2002). Making agents gaze naturally - Does it work? *Proceedings AVI 2002: Advanced Visual Interfaces*, Trento, Italy, May 2002, 357-358.
20. Vertegaal, R. (1998). *Look Who's Talking to Whom*. PhD Thesis, Cognitive Ergonomics Department, Twente University, The Netherlands.
21. Vertegaal, R., Slagter, R., Van der Veer, G.C., and Nijholt, A. (2001). Eye Gaze Patterns in Conversations: There is More to Conversational Agents Than Meets the Eyes, *CHI 2001 Conference on Human Factors in Computing Systems*, Seattle, WA, March/April 2001, 301-307.