Coordination in Conversation and Rapport

Justine Cassell, Alastair J. Gill and Paul A. Tepper
Center for Technology & Social Behavior
Northwestern University
2240 Campus Drive, Evanston, IL 60208
{justine, alastair, ptepper}@northwestern.edu

Abstract

We investigate the role of increasing friendship in dialogue, and propose a first step towards a computational model of the role of long-term relationships in language use between humans and embodied conversational agents. Data came from a study of friends and strangers, who either could or could not see one another, and who were asked to give directions to one-another, three subsequent times. Analysis focused on differences in the use of dialogue acts and non-verbal behaviors, as well as co-occurrences of dialogue acts, eye gaze and head nods, and found a pattern of verbal and nonverbal behavior that differentiates the dialogue of friends from that of strangers, and differentiates early acquaintances from those who have worked together before. Based on these results, we present a model of deepening rapport which would enable an ECA to begin to model patterns of human relationships.

1 Introduction

What characterizes the language of people who have known one another for a long time? In the US one thinks of groups of friends, leaning in towards one another, laughing, telling jokes at one another's expense, and interrupting one another in their eagerness to contribute to the conversation. The details may differ from culture to culture, but the fact of differences between groups of friends and groups of strangers are probably universal. Which characteristics, if any, reliably differentiates friends and strangers? Which can make a new friend feel welcome? An old friend feel appreciated? Advances in natural language are ensuring that embodied conversational agents (ECAs) are increasingly scintillating, emotionally and socially expressive, and personality-rich. However, for the most part, those same ECAs demonstrate amnesia, beginning every conversation with a user as if it is their first, and never getting past the stage of introductory remarks.

As the field of ECAs matures, and these systems are found on an increasing number of platforms, for an increasing number of applications, we feel that it is time to ensure that ECAs be able to engage in deepening relationships that make their collaboration with humans productive and satisfying over long periods of time. To this end, in this paper we examine the verbal and nonverbal correlates of friendship in an empirical study, and then take first steps towards a model of deepening friendship and rapport in ECAs. The current study is a part of a larger research program into linguistic and social coordination devices from the utterance level to the relationship level – how they work in humans, how they can be modeled in virtual humans, and how virtual humans can be used to teach people who wish to learn these skills.

2 Background & Theory

As people become closer, their conversational style changes. They may raise more topics in the course of a conversation, refer more to themselves as a single unit than as two people, and be more responsive to one another's talk (Cassell & Tversky, 2005; Hornstein, 1982). They also are likely to sustain eye contact longer, smile more, and lean more towards one another (Grahe & Bernieri, 1999; Richmond & McCroskey, 1995). In addition, friends appear to have fewer difficulties with lexical search, perhaps because they can rely on greater shared knowledge, and are more likely to talk at the same time, and to negotiate turn-taking in a less rigid manner, both through gaze and ges-

Proceedings of the Workshop on Embodied Language Processing, pages 41–50,
Prague, Czech Republic, June 28, 2007.

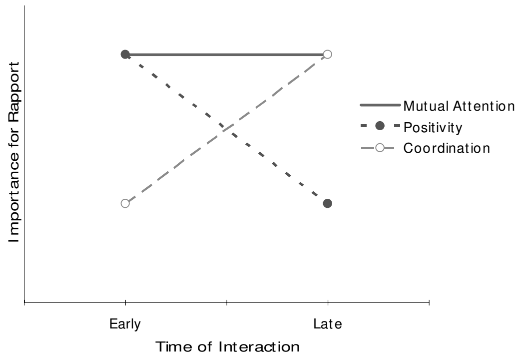

Figure 1. Three component model of rapport (from Tickle-Degen & Rosenthal, 1990).

ture (Welji & Duncan, 2005). Tickle-Degnen & Rosenthal (1990) propose a model of deepening rapport over time based on the relationship among three components: positivity, mutual attention and coordination. As shown in Figure 1, as friendship deepens, the importance of positivity decreases, while the importance of coordination increases. Attention to the conversational partner, however, is hypothesized to remain constant. That is, strangers are more likely to be polite and uniformly positive in their talk, but also more likely to be awkward and badly coordinated with their interlocutors.

As a relationship progresses and impressions have been formed and accepted, disagreement becomes acceptable and important. This may entail an increase in face-threatening issues and behaviors (cf. Brown & Levinson, 1987) accompanied by a decrease in the need to mediate these threats. At this stage in the relationship, coordination becomes highly important, so that the conversation will be less awkward and there is less likelihood of misunderstanding. Attention to one another, however, does not change. Tickle-Degnen & Rosenthal point out that these features are as likely to be expressed nonverbally (through smiles, nods, and posture shifts, for example) as verbally.

One criticism of Tickle-Degnen & Rosenthal, and similar work, is that positive feelings for, and knowledge about, the other person are not distinguished (Cappella, 1990). That is, what might be perceived as lack of rapport could actually be a lack of familiarity with a partner's behavioral cues for indicating misunderstanding or requesting information.

This conflation may come from the fact that the word rapport is used both to refer to the phenomenon of instant responsiveness ("we just clicked") and that of deepening interdependence over time. ECA research has been divided between a focus on instant rapport (Gratch et al., 2006; Maatman, Gratch, & Marsella, 2005) and a focus on establishing and maintaining relationships over time (Bickmore & Picard, 2005; Cassell & Bickmore, 2002; Stronks, Nijholt, van der Vet, & Heylen, 2002). Perhaps due to difficulties with analyzing dyadic interdependent processes, and modeling them in computational systems, much of the work in both traditions still takes a *signaling* approach, whereby particular signals (such as nodding or small talk) demonstrate the responsiveness, extroversion, or rapport-readiness of the agent, but are decontextualized from the actions of the dyad (Duncan, 1990). Although this approach is well paired to current technological constraints, it may not adequately account for the contingency of interpersonal interaction and conversation. In addition, in none of these previous studies was there a focus on how verbal and nonverbal devices actually change over the course of a relationship, and how those devices are interdependent between speaker and listener. An instant rapport approach is useful for building systems that are initially attractive to users; but a system that signals increasing familiarity and intimacy through its linguistic and nonverbal behaviors may encourage users to stay with the system over a longer period of time.

In the current work, we concentrate how discourse and nonverbal behavior changes over time, and across the dyad, as this perspective allows us to highlight the similarities between interpersonal coordination and knowledge coordination of the kind that has been studied in both conversational analysis and psycholinguistics.

Work on conversational analysis demonstrates the importance of knowledge coordination components such as turn-taking and adjacency pairs (e.g.Goodwin, 1981; Schegloff & Sacks, 1973). Inspired by this approach, work by Clark and collaborators on grounding and conversation as joint action has made demonstrated coordination and cooperation as defining characteristics of conversation (Clark, 1996; Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986). This work has in turn, received a significant amount of attention in computational linguistics, specifically in the study of dialogue (Matheson, Poesio, & Traum, 2000; Nakano, Reinstein, Stocky, & Cassell, 2003; Traum, 1994; Traum & Dillenbourg, 1998). To develop a model of nonverbal grounding, Nakano et al. (2003) studied people giving directions with respect to a map placed in between them. In that study, we observed that when a direction-receiver looked up from the map while the direction-giver was still giving directions, the giver would initiate grounding behavior such as a repeat or a rephrase.

The literature reviewed above leads us to believe that there is an integral relationship between social and knowledge coordination. In this paper, we attempt to draw conclusions about the changes in social and linguistic coordination over the short- and long-term in a way that illuminates that potential relationship, and that is also computationally viable. In order to do this, we replicate the task we used in our earlier grounding study (Nakano et al., 2003); that is we use a direction-giving task, where half the subjects can see one another, and half are divided by a screen. Here, however, half of the subjects in each visibility condition are friends and half are strangers. And to study the potential development of rapport across the experimental period, each pair performs three subsequent direction-giving tasks.

In the next section, we discuss the experimental procedure further. In section 4, we introduce first steps towards a new computational model of rapport that incorporates conversational coordination and grounding, based on our empirical findings.

3 The Experiment

3.1 Method

Participants We collected eight task-based conversations (N = 16): in each dyad, one participant was accompanied by the experimenter and followed a specific route from one place in the rococo university building where the experiment was run to another place in the building. S/he gave the other participant directions on how to reach that location, without the use of maps or other props. The direction-receiver (Receiver) was instructed to ask the direction-giver (Giver) as many questions as needed to understand the directions. After the conversation, the Receiver had to find the location. During recruitment the Giver was always selected as someone familiar with the building, while the Receiver was unfamiliar. All subjects were undergraduate students, and were motivated by surprise gifts hidden at the target location.

Design. We manipulated long-term rapport, visibility, and subsequent route in a 2 × 2 × 3 design. We operationalized long-term rapport as a binary, between-subjects variable, with conditions Friends (self-reported as friends for at least one year) and Strangers. To study the effect of non-verbal behavior, we manipulated visibility as a second between-subject variable. To do this, half of the participants could see each other, and half were separated by a dividing panel. To study the effect of acquaintance across the experimental period, each dyad completed the task three consecutive times, going to three different locations.

Data Coding All dyads were videotaped using a six-camera array, capturing the participants' body movements from the front, side, and above, along with close-up views of their faces. From each dyad, we made time-aligned transcriptions (using Praat). Non-verbal behavior was coded using Anvil. From the transcripts, the following 9 DAMSL Dialogue Acts (Core & Allen, 1997) were coded: *Acknowledgments, Answers, Assert, Completion, Influence, Information Request, Reassert, Repeat-Rephrase,* and *Signal Non Understanding*. Non-verbal behavior in giver and receiver was coded using the following categories, based on Nakano, et al. (2003):

- *Look at Speaker* – looking at the speaker's eyes, eye region or face.
- *Look at Hearer* – looking at the hearer's eyes, eye region or face.
- *Head nod [speaker or hearer]* – Head moves up and down in a single continuous movement on a vertical axis, but eyes do not go above the horizontal axis.

3.2 Results

We first provide basic statistics on the experimental manipulations and then examine the role of friendship and visibility on verbal and non-verbal behavior.

Basic Statistics: Overall, we find that Friend dyads use a significantly greater number of turns per minute than Strangers ($t(6)= 2.45$, $p<.05$, two tail), however, there is no difference in the mean number of seconds it took for dyads to complete the task. This lack of significance may have been due to variance among the dyads, since the mean length was 847 seconds for friends and 1049 for

DV	Source	DF	DF Total	F Ratio
ACK	V*F	1	20	10.64**
COMP	V*F	1	4	9.78*
SNU	V*F	1	20	3.31†
	Rte	2	20	3.38*

Note: †p<0.08; *p < 0.05; **p < 0.01;

Table 1: Verbal behavior

Abbreviation: ACK=Acknowledgment; COMP=Completion; SNU=Signal Non Understanding; Sources abbreviated as: F = Friendship; V = Visibility; Rte = Route

strangers. Given the instructional nature of the task, this means that Friends were more likely to intervene in the direction-giving than were Strangers, even though – for most of the dyads – friends appear to take less time to finish. No difference was found in turns per minute for Visible and Non-visible dyads; nor is there a difference in length in seconds. For routes, there is no difference in turns per minute, however for the length of the route in seconds there is a difference ($F(2,21)=10.66$; $p<.006$) such that the mean length of Route 1 is 165 seconds; Route 2 is 395 seconds; Route 3 is 387. For this reason, all statistics below are normalized as a function of the length of that dyad's data in seconds, and graphs are plotted to show least squares mean.

Verbal and Nonverbal behavior: We examine the relationship between friendship and visibility of both Giver and Receiver across the three route tasks. Each of the DAMSL dialogue act variables and Non-verbal behavior variables was entered as the dependent variable in building mixed method models using the JMP statistical package (Version 6, SAS Institute Inc., Cary, NC, 1989-2005); Speaker (direction-giver or receiver), Visibility, Friendship and Route were entered as predictor variables; experimental dialogue number was also entered as a source of random variance. We report the results in Tables 1 and 2 (for DAMSL and Non-verbal behavior variables respectively).

Verbal Behavior: In terms of overall variance explained, we find that Acknowledgments is best accounted for by the model (Adjusted R Square of 0.91), whilst Completion is least well accounted for (Adjusted R^2 of 0.06).

Turning first to main effects, for Visibility, Visible-Givers use Acknowledgements, Assert, Influence, and Reassert dialogue acts more frequently than Non-visible Givers (post-hoc t tests at $p <.05$)

Visible-Receivers use Acknowledgement, Repeat-rephrase, Signal Non Understanding these features more frequently than Non-visible Receivers. (post-hoc t tests at $p <.05$)

For Friendship, no differences were found for production of DAMSL acts by givers. For receivers, receiver-strangers use more acknowledgements than receiver friends.

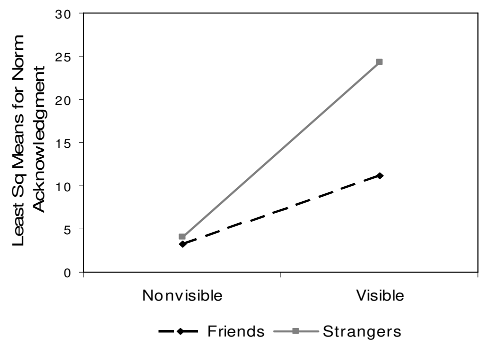

Figure 2: Giver Acknowledgment by condition

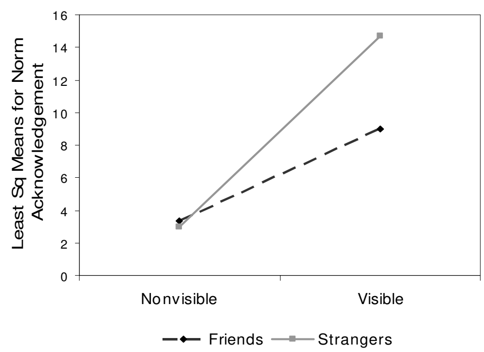

Figure 3: Receiver Acknowledgment by condition

These main effects are mediated by an interaction between Visibility×Friendship for Acknowledgements. Here, as shown in Figure 2 and 3 we see that in the nonvisibility condition, there is no difference in the use of acknowledgements per second between friends and strangers; on the other hand, strangers use more acknowledgements in the visible condition ($p<.05$). A very similar interaction was found for Signal Non Understanding (at the trend level of $p<.08$).

Route is only a main effect predictor of Signal Non Understanding as used by receivers, who produce it significantly more frequently during the third route task than the first. Since signaling one's lack of understanding is potentially face-

DV	Source	DF	DF Total	F Ratio
Look At Speaker	Rte	2	10	18.03***
Hearer Nod	SPKR*F*Rte	2	20	5.14*
Speaker Nod	SPKR*F*Rte	2	20	4.21*

*p<0.05;**p<0.01; ***p<0.001

Table 2. Non Verbal Behaviors.

Sources abbreviated as: SPKR = Speaker; F = Friendship; V = Visibility; Rte = Route,

threatening, this result may indicate that both friends and strangers become more comfortable with one another by the third route.

Nonverbal Behavior: Variance explained by the non-verbal models is the greatest for Look At Speaker (Adjusted R^2 0.83) and least for Speaker Nod (0.38). With respect to the main effects resulting from the analysis of the non-verbal behaviors, we find the following.

Visibility: Givers nod more in the visible condition when the receiver is speaking than they do in the Non-visible condition.

Route: For both givers and receivers, there is an increase in use of Look At Speaker and Look At Hearer, over time; in both cases significantly greater instances of these variables occurred during Route task 2 and 3, compared to Route 1. Once again, these results may indicate increasing coordination in conversational behavior for both Friends and Strangers.

In fact, in the case of head nods, we note an interesting pattern of coordination between speaker and hearer head-nods across the routes that differs for friends and strangers. For friends, both Receiver and Giver head nods in response to Receiver talk reduce in frequency between the first and second routes (Giver $t(8)=-2.36$; $p<0.05$; Receiver $t(8)=-2.28$; $p<0.05$). For strangers, no such accommodation over time occurs. Conversely, for friends when the Giver is speaking, both giver and receiver head nods increase over the three routes (significant only for Receiver $t(8)=2.38$; $p<0.05$). For strangers, however, head nods decrease (Giver $t(8)=-2.80$; $p<0.05$, Receiver $t(8)=3.92$; $p<0.01$). This means that speaker and hearer are increasingly coordinated across the routes, particularly when they are friends.

Interaction of verbal and nonverbal behavior

So far we have concentrated on how individual verbal and nonverbal behaviors differ across conditions. However, this does not take account of the interactive nature of the task and the focus of this paper. We therefore examine how specific responsive nonverbal behaviors (looking at speaker/hearer and head nods) co-occur before, during, or after the DAMSL variables. Examination of the residuals of chi square analysis was used to identify co-occurrence of DAMSL dialogue acts with nonverbal behavior for each Speaker (Giver or Receiver) and condition (Friend/Stranger, Visible /Nonvisible). Significant over-use or underuse of these verbal/nonverbal co-occurrences was then compared using the log-likelihood statistic (Rayson, 2003) to dialogues in the other conditions (e.g., Giver-Friend-Visible with Giver-Friend-Nonvisible, and Giver-Stranger-Visible for Head-nods, and just Friends with Strangers for the Gaze data). This technique, which we used in our earlier grounding experiment (Nakano et al., 2003) allows us insight into the probable causality of the behaviors of speaker and hearer, across verbal and nonverbal behavior.

When direction-givers are speaking

Head-nods. Givers did not nod significantly more or less frequently across Friends/Strangers conditions when they were speaking.

Gaze. More than in friendship dialogues, when strangers are speaking, and the direction-giver is acknowledging, the direction-receiver is likely to look at the Giver ($G^2=17.14$; $p<0.0001$).

More than in friendship dialogues, in Stranger dialogues, both before and after the direction-giver asserts something, the Receiver is likely to look at the Giver ($G^2=5.09$; $p<0.05$, and $G^2=4.16$, $p<0.05$, respectively).

More than in friendship dialogues, both before and during the Giver's use of Repeat-Rephrase utterances, the Receiver is likely to look at the Giver ($G^2=35.02$; $p<0.0001$, and $G^2=60.74$; $p<0.0001$, respectively).

More than in friendship dialogues, both before and during the Giver's use of Info-Request dialogue acts, the Receiver is likely to look at the Giver ($G^2=39.01$; $p<0.0001$, and $G^2=9.60$; $p<0.01$, respectively).

This means that right after a direction receiver looks at the direction-giver, the giver produces an Assertion, a Repeat-Rephrase, or an Information

Request. As with Nakano et al., the stranger's gaze towards the direction-giver can be seen as a signal of non-understanding and, in these contexts, it evokes one of these three grounding responses from the direction-giver.

For friends, on the other hand, gaze towards the speaker evokes the next segment of the directions, and is therefore functioning as a signal of understanding. That is, more than in stranger dialogues, both before and during the Giver's use of Influence dialogue acts (utterances such as "turn right"), Receivers are more to look at the Giver (G^2=4.77; p<0.05, and G^2=31.92; p<0.0001, respectively).

When direction-receivers are speaking

Head-nods. As shown in Figure 4, Strangers used more head nods than Friends during their use of Acknowledgment dialogue acts in the visible condition (G^2 = 10.48, p<.01), however they do not differ from friends in the nonvisible condition (G^2 = 0.01, *ns*).

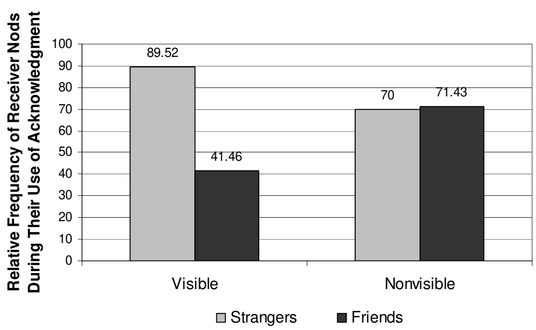

Figure 4: Receiver nods during Acknowledgment

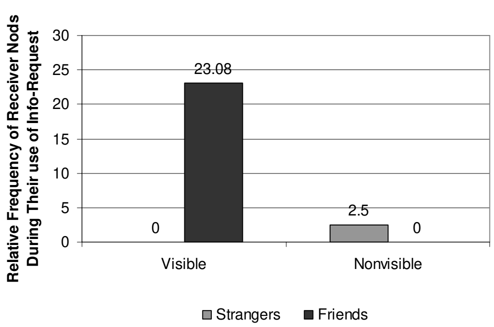

Figure 5: Receiver nods during Info Request

Conversely, as shown in Figure 5, when receivers are making an Info-Request in the visible condition (G^2=14.13, p<.001), Friends nod much more often than Strangers; but do not differ from Strangers in the nonvisible condition (G^2 = 1.44, *ns*). Once again, here the friends are marking their understanding, by nodding, even while they request further information.

Gaze. Before the Receiver's use of Acknowledgment dialogue acts in Stranger dialogues, the Giver is more likely to look at the Receiver (G^2=10.79; p<0.01). After the Receiver has used an Acknowledgment in a Stranger dialogue, s/he is more likely to look at the Giver a (G^2=14.79; p<0.001). This means that among strangers the giver and receiver are likely to engage in mutual gaze around the acknowledgement dialogue act.

During and after a Repeat-Rephrase dialogue act in Friends dialogues, the Receiver is more likely to look at the Giver (G^2=10.37; p<0.01 and G^2=6.72; p<0.01, respectively). Before the Receiver uses a Repeat-Rephrase dialogue act in a Friends dialogues, the Giver is more likely to look at the Receiver (G^2=9.08; p<0.01). This means that among friends, giver and receiver engage in mutual gaze around the repeat and rephrase dialogue act.

3.3 Discussion

Our analysis of verbal and non-verbal behaviors reveals consistent differences across the short term, comparing subsequent direction-giving tasks, and across the long term, comparing strangers to friends.

Strangers – Knowledge coordination

With respect to the co-ordination of verbal and nonverbal behavior, it is apparent that, among strangers, the Receiver's use of Acknowledgments is strongly associated with characteristic gaze patterns of signaling non-understanding. In these Stranger dyads, the Giver looks at the Receiver to signal the need for feedback. The Receiver then nods to emphasize comprehension while uttering the Acknowledgment (e.g., "okay"), and then looks back at the Giver. This pattern is very specific to Strangers, and in the case of the Receiver's use of head nod, specific to the visible condition. Similarly, among strangers, when the Giver acknowledges the receiver's correct understanding, the Receiver looks at the Giver. This pattern of gaze request, and grounding response, happens repeatedly and often (Acknowledgements being used more frequently by strangers), ensuring coordination among strangers, but at the cost of frequent explicit requests. Of course, since Acknowledgements are generally backchannel utterances, used to indicate mutual understanding, one explanation for the higher frequency of acknowledgements, head nods and eye gaze by strangers, especially in the earlier tasks, is over-generation aimed at showing attention. Although over-generation could achieve these goals, it can also result in creating a false impression of mutual understanding and it is notable that these behaviors decrease over time.

We also find that in the Strangers condition, Receivers are more likely to look at the Giver before and during the Giver's use of Repeat-Rephrase (i.e., repeating back to the Receiver some earlier information), and also before and during the Giver's use of Info-Request acts (that is the Giver asking the Receiver a question such as "do you get that?").

From the frequent and repeated use of Acknowledgments and gaze (implying something like "okay… are you sure you're okay… really?"), to the Receiver's gaze-anticipation of the Giver's Repeat-Rephrase and Info-Request, we infer a much more effortful interaction for Strangers, and one that, in fact, for most dyads, takes longer.

In line with Welji and Duncan (2005), we found evidence that the task may demand additional cognitive resources for Strangers, with the Receiver in the Strangers dialogues breaking gaze at the Giver to apparently consult some internal representation of the space just described by the Givers Assert (e.g., "you'll find some blue couches"), before returning attention, and gaze, once again to the Giver.

We also note a greater use overall of Acknowledgment and Completions by Strangers and in visible situations; Receivers in the visible situations also use more Signal Non Understanding. Taken together, these findings indicate that coordination and achieving mutual understanding is more effortful for Strangers: Friends use fewer dialogue acts such as Acknowledgment, Completion, and Signal Non Understanding, indicating that there is less need to negotiate understanding, and that they are more likely to have some kind of shared representation. Because of this, the Friends dialogues and task performance would appear to be more efficient, with less grounding required and less mutual gaze around their use of Acknowledgments, Info-Requests and Repeat-Rephrase.

The fact that Friends are better able to calibrate the task than Strangers is also demonstrated by the results found for Route. Both Friend and Stranger dyads increase their gaze towards one another from Route 1 to Route 2. But Friends shift the way they use head nods over the course of the three routes. They begin in Route 1 by producing them in conjunction with Receiver talk (acknowledgment, request for further information, repeating directions back). However, by Route 2, the friends are nodding when the direction-giver speaks, marking that they don't need further information but have understood on the first try. On the contrary, Strangers continue to nod just as much with receiver talk, and decrease their nods with giver talk; perhaps since by Route 2, it is clear that Strangers don't understand on the first try.

Tickle-Degnen and Rosenthal (1990) predict greater coordination as a relationship progresses. We found better coordination, but that was revealed, paradoxically, through fewer coordination devices and fewer dialogue acts in each turn, both comparing from Route 1 to Route 3, and comparing Strangers to Friends.

Friends – Positivity

In the Friends dialogues, we find a notable collocation of non-verbal behavior and the Receiver's use of Repeat-Rephrase utterance (i.e., repeating the Giver's utterance back to ensure correct interpretation). This is in contrast to the findings for Stranger dyads which found nonverbal behaviors found in conjunction with the Giver's reactive use of Repeat-Rephrase – i.e., the Giver's questioning of the Receiver's understanding – perhaps after a breakdown in mutual understanding. In the Friend dyads, it is the Receiver who proactively checks correct understanding of the Giver's utterance before the interaction continues.

Tickle-Degnen & Rosenthal predict a reduction in the importance of positivity as rapport increases over time in a relationship. We found some evidence to support this, since such questioning of the Giver in itself may be viewed as face-threatening behavior. However, in the Friends dialogues, this Repeat-Rephrase appears anticipated – or sanctioned – by the Giver who looks at the Receiver prior to the utterance. Further, during and after the Receivers' use of the Repeat-Rephrase utterance, they also look at the Giver, which again would be expected to be viewed as a threat to face.

Similarly, the Receiver gazes at the Giver before and during the Giver's use of Influence dialogue acts (explicit commands, such as "turn left"). Such direct gaze, along with a reduced number of mediating dialogue acts such as Acknowledgments, appears to indicate that Friends dialogues are less concerned with avoiding face-threatening behavior, and as such would appear less concerned with maintaining positivity during the interaction.

Note that, almost paradoxically, Friends demonstrate their increased ability to coordinate their interaction through a diminished use of explicit coordination devices. This speeds up the interaction, and reduces the number of overall dialogue acts.

And, finally, differences between Friends and Strangers are vastly diminished when the interlocutors cannot see one another. This leads us to believe that nonverbal behaviors in addition to gaze and head nods may be playing a role in how Friends coordinate with one another; an advantage which is taken away when they can only hear one another's voices.

4 Towards a Computational Model

In the short-term context of conversation, maintenance of mutual attention and incremental coordination of beliefs are requisites for grounding and turn-taking. In prior computational systems, grounding has been achieved by marking the status of conversational contributions as provisional (ungrounded) or shared (grounded). Conversational actions by either the user or the system can trigger updates that change provisional information to shared. Acknowledgements, for example, are explicit ways of achieving grounding, but moving on to the next stage of the task is equally effective, as it presupposes that prior utterances have been taken up (Traum, 1994). In a model such as this, grounding occurs at the turn level. In order to handle the multimodal phenomena that participate in grounding in face-to-face conversation, as Nakano et al. (Nakano et al., 2003) have shown, a model of knowledge coordination needs to have more frequently updated access to potential grounding events. In that implementation, we continuously polled for inputs, so as to capture the updates in grounding that occur between typical linguistic segments. We believe that the focus on time and process that allowed us to look at events of a smaller granularity in our earlier work on nonverbal grounding behavior will also allow us to extend up to events of a larger granularity, such as stages in a relationship. That is, we believe that the results described in earlier sections of this paper can be taken into account in a computational system by maintaining a model of the state of shared and private information across several interactions (several years, if possible). In this way, the shared history of two interlocutors (the user and the system) can be translated into patterns of linguistic behavior, such as reduced use of acknowledgements, and reduced positivity, with increased interruption and information requests. This is similar to Cheng, Cavedon & Dale (2004)'s approach to direction-giving. In this approach, the system maintains a history of places it has given directions to before. Using this *task history*, it is able to generate shorter directions at later stages in the dialogue. In our implementation, however, the very style of the interaction is modified by the shared history of the user and the system. In the sense that we are modifying the linguistic style of the dialogue based on psychological attributes, our approach is similar to work by Mairesse & Walker (2007) and Isard et al. (2006). In both cases, a broad set of natural language generation parameters is employed to generate language that differs along a personality dimension, based on a number of previous empirical studies. In the current approach, however, the features that are modified derive from the interde-

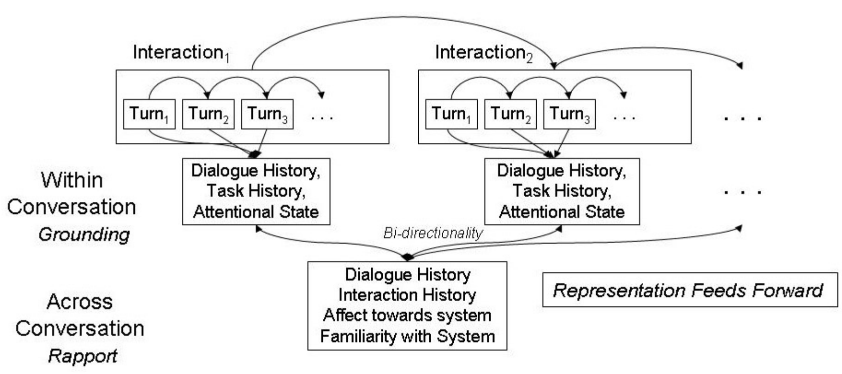

Figure 6. Proposed architecture for modeling coordination within and across conversation

pendence of the system with a particular user.

Some of the features that are present in the conversations of friends, such as interjections and completion of one another's utterances, are still beyond current computational abilities, as they would require online, real-time processing and understanding of utterances with incremental planning and generation of responses. We are interested in pursuing this feature of the system as dialogue technologies improve.

5 Conclusion

In this paper, we have compared direction-giving between friends and strangers, and within these two groups we have compared three subsequent direction-giving episodes. In order to determine the effect played specifically by nonverbal behavior in short- and long-term rapport, half of our participants could see one another, while the other half were divided by a screen. Our experimental and analytic methodology drew from both the social psychological, conversational analysis, and conversation as joint action traditions. Consequently, our results were able to demonstrate the ways in which the verbal and nonverbal devices that index rapport relate to the role those same devices play in knowledge coordination. Based on this commonality, we proposed a computationally viable model of deepening friendship within and across subsequent tasks that extends our previous work on grounding in face-to-face interaction. The work we have presented here therefore differs substantially from previous work on rapport and relationship building in embodied conversational agents. We did not start out with a definition of rapport but instead investigated those behaviors that characterize dyads who have self-identified as friends or strangers. And rather than looking at rapid assessment of rapport (the feeling of "clicking") we looked at the long-term version: acquiring a sense of mutual interdependence. Finally, rather than looking at how to get ECAs to engage users into establishing a relationship, or into letting down their guard, we examined those behaviors that characterize the dyadic interaction at each stage.

All of these topics, however, are clearly interrelated, and future research will benefit from taking a greater number of them into account in both data analysis, and the implementation of ECAs. Future research in our own lab will also have to be more explicit about how to implement the computational model that we have started to lay out here. Additional subjects in a similar experiment will no doubt facilitate that task.

As we increasingly understand better how conversation changes when people come to know one another, we expect to apply these results to our ongoing research on virtual peers that can teach children with autism how to sustain interpersonal relationships (Tartaro & Cassell, 2006) and to our work on building the survey interviewers of the future, who can both engage their survey-takers and keep them honest (Cassell & Miller, in press). More generally, however, we hope to increasingly implement ECAs who will stick around for the long haul.

Acknowledgments

Thanks to Kristina Striegnitz , Will Thompson, Tara Latta and Nate Cantelmo for their help, and Darren Gergle for his superior statistical knowledge. We are grateful to Motorola for funding that supported some of the research reported here.

References

Bickmore, T., & Picard, R. (2005). Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Transactions on Computer Human Interaction (ToCHI)*, *12*(2), 293-327.

Brown, P., & Levinson, S. (1987). *Politeness: Some Universals in Language Usage*. Studies in International Sociolinguistics. New York: Cambridge University Press.

Cappella, J. N. (1990). On Defining Conversational Coordination and Rapport. *Psychological Inquiry*, *1*(4), 303-305.

Cassell, J., & Bickmore, T. (2002). Negotiated Collusion: Modeling Social Language and its Relationship Effects in Intelligent Agents. *User Modeling and Adaptive Interfaces*, *12*, 1-44.

Cassell, J., & Miller, P. (in press). Is it Self-Administration if the Computer Gives you Encouraging Looks? In F. G. Conrad & M. F. Schober (Eds.), *Envisioning the Survey Interview of the Future*. New York: John Wiley & Sons.

Cassell, J., & Tversky, D. (2005). The Language of Online Intercultural Community Formation. *Journal of Computer-Mediated Communication*, *10*(2), article 2.

Cheng, H., Cavedon, L., & Dale, R. (2004, 28th-29th August). Generating Navigation Information Based on the Driver's Route Knowledge. *Paper presented at the COLING 2004 Workshop on Robust and Adaptive Information Processing for Mobile Speech Interfaces*. Geneva,Switzerland.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington DC: American Psychological Association.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1-39.

Core, M., & Allen, J. (1997). Coding Dialogue with the DAMSL Annotation Scheme. *Proceedings of the AAAI Fall Symposium on Communicative Action in Humans and Machines*. Boston, MA.

Duncan, S., Jr. (1990). Measuring Rapport. *Psychological Inquiry*, *1*(4), 310-312.

Goodwin, C. (1981). *Conversational Organization: Interaction between speakers and hearers*. New York: Academic Press.

Grahe, J. E., & Bernieri, F. J. (1999, Win). The importance of nonverbal cues in judging rapport. *Journal of Nonverbal Behavior*, *23*(4), 253-269. http://www.springeronline.com

Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., van der Werf, R. J., et al. (2006). Virtual Rapport. *Proceedings of the 5th International Conference on Interactive Virtual Agents (IVA)*. Marina del Rey, CA.

Hornstein, G. A. (1982). *Variations in conversational style as a function of the degree of intimacy between members of a dyad.* Unpublished Doctoral, Clark University.

Isard, A., Brockmann, C., & Oberlander, J. (2006). Individuality and alignment in generated dialogues. *Proceedings of the 4th International Natural Language Generation Conference* (pp. 22-29). Sydney, Australia.

Maatman, M., Gratch, J., & Marsella, S. (2005). Natural Behavior of a Listening Agent. *Paper presented at the 5th International Conference on Interactive Virtual Agents (IVA)*. Kos, Greece.

Mairesse, F., & Walker, M. (2007). PERSONAGE: Personality generation for dialogue. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*. Prague.

Matheson, C., Poesio, M., & Traum, D. (2000). Modelling Grounding and Discourse Obligations Using Update Rules. *Proceedings of the 1st Annual Meeting of the North American Association for Computational Linguistics (NAACL2000)*. Seattle, WA.

Nakano, Y. I., Reinstein, G., Stocky, T., & Cassell, J. (2003, July 7-12). Towards a Model of Face-to-Face Grounding. *Proceedings of the Annual Meeting of the Association for Computational Linguistics* (p. 553–561). Sapporo, Japan: Association for Computational Linguistics

Rayson, P. (2003). *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison.* Unpublished doctoral thesis, Lancaster University, Lancaster.

Richmond, V. P., & McCroskey, J. C. (1995). Immediacy. In *Nonverbal Behavior in Interpersonal Relations* (pp. 195-217). Boston: Allyn & Bacon.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica*, *8*, 289-327.

Stronks, B., Nijholt, A., van der Vet, P., & Heylen, D. (2002). Designing for friendship: Becoming friends with your ECA. *Proceedings of the Embodied conversational agents - let's specify and evaluate them!* (pp. 91-97). Bologna, Italy: ACM Press.

Tartaro, A., & Cassell, J. (2006, August 28 - September 1). Authorable virtual peers for autism spectrum disorders. *Proceedings of the Combined workshop on Language-Enabled Educational Technology and Development and Evaluation for Robust Spoken Dialogue Systems at the 17th European Conference on Artificial Intellegence.* Riva Del Garda, Italy.

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry*, *1*(4), 285-293.

Traum, D. R. (1994). *A Computational Theory of Grounding in Natural Language Conversation.* University of Rochester, Rochester, NY.

Traum, D. R., & Dillenbourg, P. (1998). Towards a Normative Model of Grounding in Collaboration. *Proceedings of the ESSLLI-98 workshop on Mutual Knowledge, Common Ground and Public Information*. Saarbrucken, Germany.

Welji, H., & Duncan, S. (2005). Collaboration and Narration: The role of shared knowledge in the speech and gesture production of friends and strangers. *Paper presented at the International Society of Gesture Studies Conference*. Lyon, France.