VIRTUAL HUMANS FOR THE STUDY OF RAPPORT IN CROSS CULTURAL SETTINGS

Jonathan Gratch*, Anna Okhmatovskaia
University of Southern California, Institute for Creative Technologies
13274 Fiji Way, Marina del Rey, CA 90292

Susan Duncan
University of Chicago, Department of Psychology
Chicago, IL, 60637

ABSTRACT

As an increasing part of the Army's mission involves establishing rapport with diverse populations, training interpersonal skills becomes critically important. Here we describe a "Rapport Agent" that senses and responds to a speaker's nonverbal behavior and provide empirical evidence that it increases speaker fluency and engagement. We argue such agent technology has potential, both as a training system to enhance communication skills, and to assess the key factors that influence rapport in face-to-face interactions. We conclude by discussing ways the nonverbal correlates of rapport vary between Arabic and English speakers and discuss the potential of such technology to advance research and training into rapport in cross-cultural settings.

1. INTRODUCTION

Rapport is a crucial factor in establishing successful relationships. Cappella (1990) states rapport to be "one of the central, if not *the* central, constructs necessary to understanding successful helping relationships and to explaining the development of personal relationships." Rapport is argued to underlie success in negotiations (Drolet and Morris 2000; Goldberg 2005), improving worker compliance (Cogger 1982), psychotherapeutic effectiveness (Tsui and Schultz 1985), improved test performance in classrooms (Fuchs 1987), improved quality of child care (Burns, 1984) and, even, susceptibility to hypnosis (Gfeller, Lynn et al. 1987).

Rapport is correlated with characteristic nonverbal behaviors in face-to-face interactions. Participants seem tightly enmeshed in something like a dance. They rapidly detect and respond to each other's movements. Tickel-Degnen and Rosenthal (1990) equate rapport with behaviors indicating mutual attentiveness (e.g. mutual gaze), positivity (e.g. head nods or smiles) and coordination (e.g. postural mimicry or synchronized movements). Studies have also indicated that rapport can be experimentally induced or disrupted by altering the presence or character of these nonverbal signals (e.g., Bavelas, Coates et al. 2000; Drolet and Morris 2000). Such findings have encouraged the development of embodied conversational agents that can induce rapport through the appropriate generation of nonverbal behavior.

When it comes to creating synthetic agents that simulate human nonverbal behavior, research has focused on half of the equation. Systems emphasize the importance of nonverbal behavior in speech *production*. Few systems attempt the tight sense-act loops that seem to underlie rapport and, despite considerable research showing the benefit of such feedback on human to human interaction, few studies have investigated its impact in human-to-virtual human interaction (cf. Cassell and Thórisson 1999; Bailenson and Yee 2005).

The fluid, contingent nature of nonverbal behavior associated with rapport suggests that it could be induced by rapidly responding to a speaker's physical movements. This article describes a RAPPORT AGENT that attempts to create a sense of rapport simply by generating listening feedback based on superficial observable features of a speaker's bodily movements and speech prosody. We discuss the results of one study that demonstrates the RAPPORT AGENT can produce some of the beneficial social effects associated with rapport. We then describe a preliminary study that highlights the similarity and differences in the nonverbal correlates of rapport between American and Arabic speakers. Such agent technology has potential as a powerful and novel methodological tool for uncovering the key factors that influence rapport in face-to-face interactions. It also has potential as a training system to enhance communication skills, and to expose trainees to culturally varying indicators of rapport.

2. RAPPORT AGENT

The RAPPORT AGENT (Gratch, Okhmatovskaia et al. 2006) is designed to establish a sense of rapport with a human participant in "face-to-face monologs" where a human participant gives a speech (e.g., tells a story) to a silent but attentive listener. In such settings, human listeners can indicate rapport through a variety of nonverbal signals (e.g., nodding, postural mirroring, etc.) The RAPPORT AGENT attempts to replicate these behaviors through a real-time analysis of the speaker's voice, head motion, and body posture, providing rapid nonverbal feedback, *without* attending to the content of the speech. The system is inspired by findings that feelings of rapport are correlated with simple contingent behaviors between speaker and listener, including behavioral

Lowering of pitch → head nod
Raised loudness → head nod
Speech disfluency → posture/gaze shift
Speaker shifts posture → mimic
Speaker gazes away → mimic
Speaker nods or shakes → mimic

Table 1: Listening Agent Mapping Rules

mimicry (Chartrand and Bargh 1999) and backchanneling (e.g., nods) (Yngve 1970). The Rapport Agent uses a vision based tracking system and signl processing of the speech signal to detect features of the speaker and uses a set of reactive rules to drive the listening mapping displayed in Table 1. The architecture of the system is displayed in Figure 1.

To produce listening behaviors, the Rapport Agent first collects and analyzes the speaker's upper-body movements and voice.

For detecting features from the participants' movements, we focus on the speaker's head movements. Watson (Morency, Sidner et al. 2005) uses stereo video to track the participants' head and incorporates learned motion classifiers that detect head nods and shakes from a vector of head velocities. Other features are derived from the position and orientation of participant's head. For example, from the head position, given the participant is seated in a fixed chair, we can infer the posture of the spine. Thus, we detect head gestures (nods, shakes, rolls), posture shifts (lean left or right) and gaze direction.

Acoustic features are derived from properties of the pitch and intensity of the speech signal (the Rapport Agent ignores the semantic content of the speaker's speech), using a signal processing package, Laun, developed by Mathieu Morales. Speaker pitch is approximated with the cepstrum of the speech signal (Oppenheim and Schafer 2004) and processed every 20ms. Audio artifacts introduced by the motion of the Speaker's head are minimized by filtering low frequency noise. Speech intensity is derived from amplitude of the signal. Laun detects speech intensity (silent, normal, loud), range (wide, narrow), and backchannel opportunity points (derived using the approach of Ward and Tsukahara 2000).

Recognized speaker features are mapped into listening animations through a set of authorable mapping rules. These animation commands are passed to the SmartBody animation system (Kallmann and Marsella 2005). This is an animation system designed to seamlessly blend animations and procedural behaviors. These animations are rendered in the Unreal Tournament™ game engine and displayed to the Speaker.

3. EVALUATION

The Rapport Agent described above could be integrated into a wide variety of embodied conversational agent applications. However there are a number of questions that need to be addressed first to ensure the suitability of such integration:

- Does the system correctly detect features of the speaker's behavior, such as head nods, shakes, pauses in speech, etc.?
- How well do behavior mapping rules approximate the behavior of human listeners?
- Is the agent's behavior judged to be natural when it is performed?
- Do listening behaviors of the agent influence human speakers' behavior and perceptions as described in social psychology literature on rapport?

Our preliminary analysis of the system's performance suggests that feature detection is reasonably accurate. We are currently collecting data on human face-to-face communication to address the second question. Finally we have conducted a formal evaluation study that focused on the last two questions. In this study we have attempted to replicate certain well-known psychological findings about social outcomes of rapport, in particular, increased motivation and engagement in communication, and improved conversational fluency.

Several studies have demonstrated increased speaker engagement when listeners provide feedback such as nods and mimicry, and when interactional synchrony between the participants can be achieved. This effect was observed in interactions between humans, as well as between humans and synthetic agents (Tatar 1997; Smith 2000).

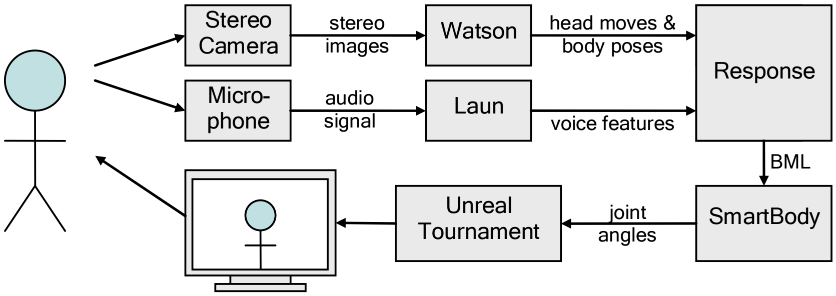


Figure 1: Rapport Agent architecture

Improved conversational fluency is another prominent characteristic of rapport interactions, and it is often explained in terms of the positive effects the listener's feedback has on the speaker. Studies show that in the absence of such feedback or when the feedback is incoherent, the speakers become disrupted, and their speech – less structured (Kraut, Lewis et al. 1982; Bavelas, Coates et al. 2000).

Our main goal was to demonstrate that nonverbal behavior displayed by the RAPPORT AGENT could not only elicit the subjective feelings of rapport in human participants, but also produce the abovementioned social outcomes that can be registered objectively.

3.1. Experimental Setup

In evaluating the system we adapted the "McNeill lab" paradigm (McNeill 1992) for studying gesture research. In this research, one participant, the Speaker, has previously observed some incident, and describes it to another participant, the Listener. Here, we replaced the Listener with the RAPPORT AGENT system, but used a cover story to make the subjects believe that they interacted with a real human. Our participants were told that the study evaluates an advanced telecommunication device, specifically a computer program that accurately captures all movements of one person and displays them on the screen (using an Avatar) to another person. According to the cover story, we were interested in comparing this new device to a more traditional telecommunication medium such as video camera, which is why one of the participants was sited in front of the monitor displaying a video image, while the other saw a life-size head of an avatar (see Figure 2).

The subjects were randomly assigned to one of two conditions labeled respectively "responsive" and "unresponsive". In a *responsive condition* the Avatar was controlled by the RAPPORT AGENT, as described earlier. The Avatar therefore displayed a range of nonverbal behaviors intended to provide positive feedback to the speaker and to create an impression of active listening.

In an *unresponsive condition* the Avatar's behavior was controlled by a pre-recorded random script and was independent of the Speaker's or Listener's behavior. The script was built from the same set of animations as those used in responsive condition, excluding head nods and shakes. Thus, the Avatar's behavioral repertoire was limited to head turns and posture shifts.

3.2. Subjects

The participants were 30 volunteers from among employees of USC's Institute for Creative Technologies. Two subjects were excluded from analysis due to an unforeseen interruption of experimental procedure. The final sample size was 28: 16 in a responsive and 12 in an unresponsive condition.

3.3. Procedure

Each subject participated in an experiment twice: once in a role of a Speaker and once as a Listener. The order was selected randomly.

While the Listener waited outside of the room, the Speaker watched a short segment of Sylvester and Tweety cartoon, after which s/he was instructed to describe the segment to the Listener. The participants were told that they would be judged based on the Listener's story comprehension. The Speaker was encouraged to describe the story in as much detail as possible. In order to prevent the Listener from speaking back we have emphasized the distinct roles assigned to participants, but did not explicitly prohibit the Listener from talking. No time constraints were introduced.

After describing the cartoon (during which time the Speaker was sitting in front of the Avatar), the Speaker was asked to fill out a short questionnaire collecting the subject's feedback about his experience with the system. Then the participants switched their roles and the procedure was repeated. A different cartoon from the same series and of similar length was used for the second round.

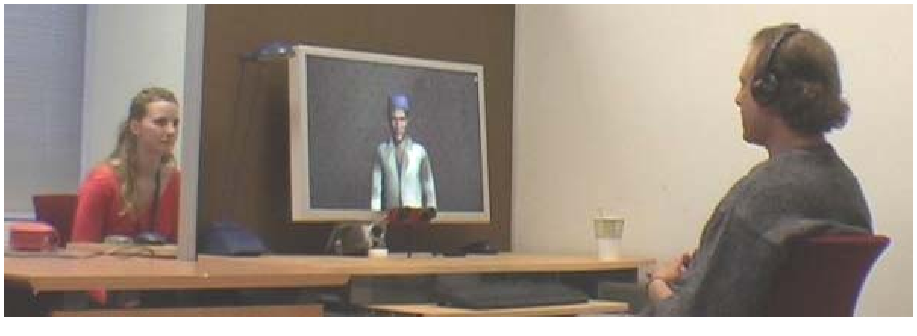


Figure 2: Experimental setup: The Speaker and the Listener separated by a screen. The Listener (left) can hear the Speaker (right) and see a video image of him/her. The Speaker instead sees an Avatar allegedly controlled by the Listener's behavior. In fact, the Avatar is controlled by the RAPPORT AGENT.

At the end of the experiment, both participants were debriefed. The experimenter collected some informal qualitative feedback on their experience with the system, probed for suspicion and finally revealed the goals of the study and experimental manipulations.

3.4. Dependent Variables

To measure subjects' engagement and motivation we have looked at the duration of their interaction with the system, assuming that under no time constraints imposed the subjects would spend more time talking if they are more engaged and willing to communicate. In particular, we measured *total time* it took the subject to tell the story, total *number of words* in the subject's story (independent of individual differences in speech rate), and the *number of "meaningful" words* (lexical and functional) in the subject's story.

To assess conversational fluency we have used two groups of measures: speech rate and the amount of speech disfluencies (Alibali, Heath et al. 2001). For speech rate we distinguished between *overall speech rate* (all words per second) and *fluent speech rate* (lexical and functional words per second). To measure the amount of disfluencies, we useed *disfluency rate* (disfluencis per second) and *disfluency frequency* (a ratio of the number of disfluencies to total word count).

Subjective sense of rapport was measured through self-report using forced-choice items of the questionnaire, for instance: "Did you feel you had a connection with the other person?". Additionally the questionnaire included several open-ended questions, which were used as a source of qualitative data.

3.5. Hypotheses

The following hypotheses were formulated in terms of measured variables:

H1a: Total time to tell the story will be higher in responsive condition.

H1b: The recorded stories will be longer in responsive condition in terms of both total word count and the number of lexical and functional words.

H2a: Overall and fluent speech rate will be higher in responsive condition.

H2b: Disfluency rate and disfluency frequency will be higher in an unresponsive condition.

H3a: The subjects in responsive condition will be more likely to report rapport on the questionnaire.

3.6. Results

Non-parametric statistical criteria were used to evaluate the differences between two conditions. Table 3 summarizes the data on duration of interaction and speech fluency.

Variable	Responsive	Unresponsive	Sig[a].
Total time	188.68	98.50	0.001*
N words	432	300	0.015*
N meaningful words	411	288	0.007*
Speech rate	2.55	2.77	0.074
Fluent sp. Rate	2.42	2.60	0.174
Disfluency rate	0.13	0.21	0.001*
Disfluency frequency	0.05	0.08	0.026*

[a] Using Mann-Whitney U statistics to compare medians

* p < .05

Table 3: Duration of interaction and fluency of speech

Consistent with H1a and H1b, the subjects in responsive condition talked significantly longer both in terms of overall time and word count. Moreover, the increase in word count was associated with the higher number of lexical and functional words.

Consistent with H2b, the disfluency rate was higher in unresponsive condition. The same was true for the disfluency frequency. Contrary to H2a, the subjects in unresponsive condition tended to speak faster, not slower. This finding, however, is non-significant for both the overall speech rate and fluent speech rate.

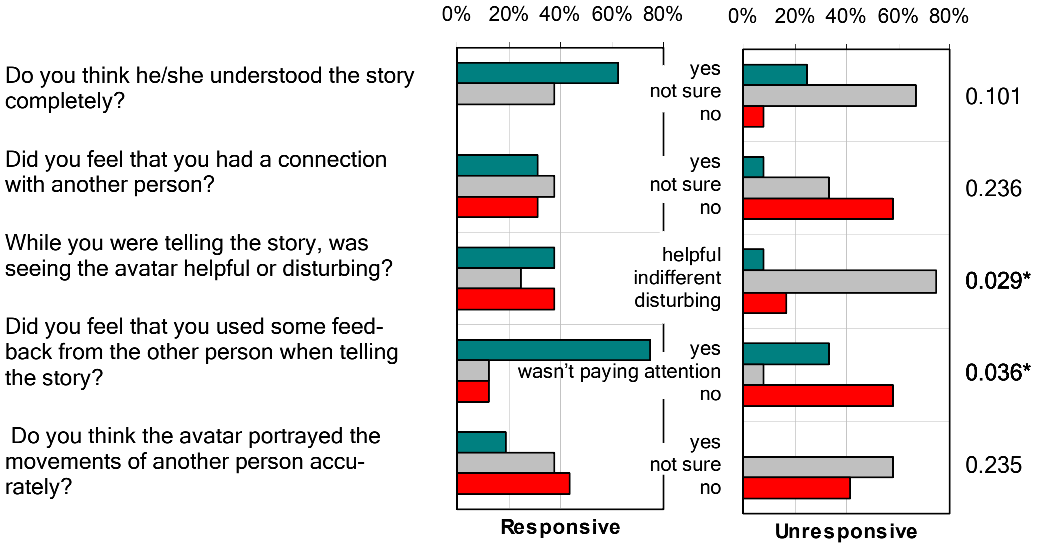
Questionnaire data is presented in Figure 3. Several trends are worth mentioning:

- Subjects in the responsive condition were more likely to feel that they had a connection with their conversational partner, and to form an impression that the listener understood them. They also reported that they used the listener's feedback when they were telling the story.
- Most subjects did not consider the avatar to be an accurate representation of a real listener; those few who did – all belonged to the responsive condition.
- Opinions on the helpfulness of the avatar were markedly different across the conditions. The subjects in responsive condition found the avatar to be either helpful or disturbing. In unresponsive condition 75% of the speakers had indifferent attitude.

Not all of the differences in self-report measures have reached statistical significance, and thus additional data may be needed to support these findings.

3.7. Discussion

The results obtained for the duration of interaction (word count and time) fully support our predictions, and are also consistent with some findings mentioned earlier (Smith 2000). The subjects spent more time talking to a responsive agent, and produced longer stories. What is important to note here is that there were significantly more "meaningful" words in these stories, suggesting that the increase in quantity of speech was not associated with a decreased quality.



* Used Chi-square statistics was to compare frequency distributions in 2 conditions (significant at p < .05)

Figure 3: Summary of subjects' responses to selected questions

We believe that longer interaction times and increased speech production reflects the subjects' willingness to communicate with the listener (represented by an avatar). The nonverbal behavior generated by the Rapport Agent was intended to create an impression of an engaged and attentive listener and encourage the speaker. During the debriefing procedure after the experiment two subjects in the unresponsive condition (tested in different sessions) pointed out that they intentionally kept their stories short because the listener seemed to be uninterested. This observation brings to light an important consideration in the design of embodied conversational agents: human observers tend to interpret not only the nonverbal clues displayed by the agent, but the absence of clues as well. The unresponsive agent in our experiment differed from a responsive one only by the absence of head nods, and randomized timing of posture and gaze shifts, so there weren't any specific behaviors that conveyed lack of interest or boredom. And yet, at least some subjects saw these signs in the agent's behavior. This suggests that one must carefully model the nonverbal behavior in embodied agents, since not only inappropriate behaviors, but sometimes just the lack of behaviors can produce undesirable effects in human observers depending on the context.

There is additional evidence that speakers were more engaged in conversation with a responsive agent, which is based on observations we made during the experiment. Several subjects in a responsive condition responded verbally to the feedback provided by the agent. In particular they could say "yes" and nod after the agent nodded. Or they could ask "Did you get it so far?", and then continue only after the agent nodded. This was not observed in an unresponsive condition. Since the experiment was built as a one-way communication, and such spontaneous interactions were actually discouraged by an instruction, they indicate a potential power of the system in producing social effects. These observations require further elaboration and formal experimental verification. Additional data on speaker's engagement may be obtained from analyzing gaze and gesturing behavior, which we will do in future studies.

Our hypothesis regarding speech fluency was only partially supported: there was support for the amount of disfluencies (H2b), but not for speech rate (H2a). This suggests that speech rate may have a more complex relationship with conversational fluency than we believed. Indeed, speaking quickly does not necessarily imply speaking fluently. Particularly, in our study an increase in speech rate in the unresponsive condition was mainly due to more frequent inclusion of pause fillers, indicating that the subjects in this condition talked fast but with many disfluencies.

It is important to keep in mind that speech rate can be affected by a number of factors, in particular emotional. It is possible that the subjects in the unresponsive condition spoke faster because they felt uncomfortable and were trying to complete the task as quickly as possible. It was previously shown that synthetic agents can elicit anxiety in human users (Rickenberg and Reeves 2000) and, particularly, that unresponsive virtual audience produces greater anxiety in the speaker (Pertaub, Slater et al. 2001). Our results for the unresponsive condition are consistent with these findings.

The results on self-reported feelings of rapport did not reach statistical significance, however the observed trends are consistent with our predictions. Increasing the sample size and using more fine-grained scales (compared to just "yes/no/unsure") may help obtain more conclusive results.

We have demonstrated that the RAPPORT AGENT exerts certain effects on the human speaker. However in order to further improve the system we need to know what it is about generated listening behavior that is responsible for these effects. Could the same results be achieved by manipulating the overall amount of movement displayed by the agent, or type of movement is important? Is it the mere occurrence of certain behaviors, or their timing that matters? How the results would change if the subjects believed they were talking to a computer and not to another human? We have already performed some additional analysis and are planning to gather more data to address these questions.

4. CROSS-CULTURAL COMPARISON

In order for the RAPPORT AGENT system to fulfill its potential as a assist to training in cross-cultural settings, it is necessary to address the issue of cross-cultural variability in how rapport is manifested and maintained. The beneficial effect of rapport on the outcomes of interaction is assumed to be a cultural universal. However, cross-cultural similarities and differences in rapport-related behaviors (verbal and nonverbal), have yet to be systematically investigated and described. A study currently underway has the immediate goal of establishing a baseline description, for members of different cultures, of natural conversational interaction in groups of individuals who feel rapport with one another. The study compares members of an Arabic-speaking culture (Egyptian) engaged in free conversation with a comparable group of American English speakers. The two groups are found to interact comparably in many ways, however, differences were observed in tendency to overlap with one another in speaking turn, in production of meaningful gestures as "backchannel" contributions to the ongoing conversational exchange, and in tendency to intrude on others' speaking turn.

The study draws on previous findings by Welji & Duncan (in preparation) of differences in the interactional styles of friend vs. stranger dyads (American) engaged in a quasi-controlled narrative discourse elicitation. The differences included greater content detail and tendency to digress in the friend dyads, longer interactions by friends, more of a tendency on the part of speakers to involve their listeners in their discourses, both verbally and nonverbally (with interactive gestures), and for listeners to interrupt. It was further found that, on for every dimension of behavior quantified for comparison across the two groups of dyads, the range of variability on that dimension for friends was found to encompass the range for strangers. In other words, stranger dyads appeared to constrain each of their behaviors more toward the mean than friends did.

It was assumed that such differences between friends and strangers are a function of different levels of experienced rapport. Grahe & Shermam (2006) have shown that friends tend to manifest more rapport than strangers.) The behaviors typical of the friend dyads, therefore, are taken to characterize rapport-ful interactions generally, among members of American culture. For the current exploration of cross-cultural differences in rapport-related behaviors, therefore, sample groups were solicited in which all members of the group were well-acquainted and friends.

4.1. Participants and Elicitation.

Five natives of Egyptian Arabic culture, all native speakers of Arabic, volunteered to be videotaped engaged in free conversation. They are academics, long-term colleagues in a Middle Eastern Studies department at an American university; four professors and one graduate student, age range of 30 years to mid-fifties. Four native English-speaking Americans, all graduate students in the social sciences at an American university in their late 20s, also volunteered. Each group was videotaped separately, with boom microphones present, for 35 minutes, conversing on a variety of topics of the group members' own choosing. In the intervals excerpted for the pilot analysis summarized below, the Arabs conversed about jewelry and traditional Arab clothing; the Americans conversed about the phenomenon of "love at first sight." These particular excerpts were selected as representative intervals during which all members of each group participated significantly in the interaction.

4.2. Analysis.

The speech of all participants in both conversational groups was transcribed in detail by a native speaker of the language. The speech transcriptions include disfluencies such as false starts, filled pauses and hesitations and the transcripts display the interleaved structure of speaking turns, intrusions and overlaps. The transcribed speech was furhter annotated to reflect co-occurrence of nonverbal behaviors with speech such as nodding and gesticulation, including "representational" gestures, those depictive of imagistic content relevant to the ongoing speech.

4.3. Results.

Since the samples of Arabic- and English-language free conversation are as yet small, the observations we report here are to be taken as pointers to dimensions of interactional style that merit further systematic analysis in more extensive samples of natural discourse from the two cultural groups.

Similarities between Egyptians and Americans. Both groups appeared equally comfortable in the videotaping environment. There was no difference between the groups in terms of animatedness or tendency to manifest behaviors generally thought to be related to nervousness, such as self-touching and soft-voiced or disfluent, stammering speech. The conversations of both groups were free-flowing and lively, accompanied by friendly joking and laughter. There was the same amount of joint group laughter and individual laughter in both groups. There was also no difference between the groups in tendency to give verbal backchannel feedback (e.g., "mm-hm" or "ooh") or nonverbal feedback in the form of nodding. There was, further, no difference between the two groups in terms of each speaker's tendency to gesticulate along with his or her individual speaking turns. Rates of coverbal gesturing did not differ, nor did the types of representational gesturing produced by each group differ. Gestures of the Egyptian Arabic speakers were "co-expressive" of meanings in their co-occurring speech to the same extent as those of the American English speakers. These observations about spontaneous, coverbal gesturing are in keeping with the findings of McNeill (1992, 2005) and many others (e.g., McNeill 2000) that the tendency to gesture along with speech is a linguistic universal. The current study confirms that this is true for conversational discourse as well as for the more well-studied narrative discourse (story telling).

Differences between the groups. The transcripts of the conversations for both groups were divided up by speaker turn and within speaker turn, very roughly, by "utterance." An utterance was the full or partial expression of a unitary idea, most typically an idea expressed in a single "breath group." The proportions of such utterances that *were* versus *were not* overlapped either all or in part by utterances of one or more other group members were roughly inverse for Egyptians and the Americans. That is, roughly 80% of all the Americans' utterances were overlapped by speech from another group member, whereas only roughly 20% of the Egyptians' utterances were. This suggests that, in American rapport-ful interactions it is acceptable to talk over another speaker while in an Egyptian interaction it may be less so. Correlated with this tendency toward turn-overlap is a higher rate of interruptive intrusions into the speaking turns of others, on the part of Americans. There were almost five times as many instances of such intrusions by Americans, compared to the Egyptians. Finally, a very noticeable gesticulatory behavior of the Egyptians that occurred a half dozen times in the interval sampled but was unattested in the American sample was the production of depictive gestures not accompanied by any speech on the part of the gesturer and directed toward one of the other conversation participants. An example of this is when one participant, listening to an exchange about a piece of jewelry shaped like a crown, produces a gesture depicting the shape of a crown, directed twoard the head of the participant wearing the jewelry in question. McNeill (1985) has noted that gestures on the part of listeners in an interaction are extremely rare. However in this sample of Egyptian conversation we see several instances of just this behavior. It seems reasonable to link this finding to the finding that Egyptians are reluctant to speak over one another. Perhaps the production of a theme-related depictive gesture is felt to be less intrusive.

Discussion. In this small-sample analysis of free conversation among friends who have rapport with one another, a handful of differences between Americans and Egyptian Arabs in interaction style emerge. Though many dimensions of verbal and nonverbal behavior that are relevant for the maintenance of interactions appear to be quite similar across the two groups, the differences are of the type that may be significant for the success of cross-cultural interactions. The observed similarities and differences described must be confirmed on the basis of samples in which factors such as participant age and status differences, motivational states, interlocutor relationships, topic matter, and so on are controlled. In the case of the two samples examined here, it is reasonable to ask if a group of five professional colleagues (the Egyptians) are sufficiently similar in inter-individual status to a cohort of graduate students to support generalizing the findings of differences in interaction style. Nevertheless, the differences observed here begin to suggest dimensions of interaction style that one might explore further in more comprehensive samples and might also manipulate experimentally in contexts of human-agent interactions.

5. CONCLUSIONS

Simulation technology has moved beyond the training of mechanistic skills (e.g., driving a tank) towards the more ambitious problem of interpersonal skills training. This article discusses a Rapport Agent that can mimic the nonverbal behavior that people exhibit when they have established rapport. Just as some studies suggest people can induce rapport in others through the judicious use of nonverbal signals, we have empirical evidence that the Rapport Agent can lead human subjects to exhibit some of the social benefits of rapport, at least within the context of face-to-face dialogues – human participants spoke much longer and more fluently.

The nonverbal correlates of rapport vary based on the social context that surrounds the participants. One important context is culture. Although many aspects of social interaction are universal, people clearly exhibit important differences in their face-to-face nonverbal behavior. We reported the results of a preliminary study where clear similarly and differences were seen between Arab and English speakers. For example, although there were strong similarities in backchannel behaviors, Arab speakers were far less willing to speak over each other.

Such agent technology has potential as a powerful and novel methodological tool for uncovering the key factors that influence rapport in face-to-face interactions. It also has potential as a training system to enhance communication skills, and to expose trainees to culturally varying indicators of rapport.

ACKNOWLEDGEMENTS

This work was sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM), and the content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

REFERENCES

Alibali, M. W., D. C. Heath, et al. (2001). "Effects of visibility between speaker and listener on gesture production: some gestures are meant to be seen." Journal of Memory and Language **44**: 169-188.

Bailenson, J. N. and N. Yee (2005). "Digital Chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments." Psychological Science **16**: 814-819.

Bavelas, J. B., L. Coates, et al. (2000). "Listeners as Co-narrators." Jurnal of Personality and Social Psychology **79**(6): 941-952.

Capella, J. N. (1990). "On defining conversational coordination and rapport." Psychological Inquiry **1**(4): 303-305.

Cassell, J. and K. R. Thórisson (1999). "The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents." International Journal of Applied Artificial Intelligence **13**(4-5): 519-538.

Chartrand, T. L. and J. A. Bargh (1999). "The Chameleon Effect: The Perception-Behavior Link and Social Interaction." Journal of Personality and Social Psychology **76**(6): 893-910.

Chiu, C., Y. Hong, et al. (1995). Gaze direction and fluency in conversational speech**:** Unpublished manuscript.

Cogger, J. W. (1982). "Are you a skilled interviewer?" Personnel Journal **61**: 840-843.

Drolet, A. L. and M. W. Morris (2000). "Rapport in conflict resolution: accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts." Experimental Social Psychology **36**: 26-50.

Fuchs, D. (1987). "Examiner familiarity effects on test performance: implications for training and practice." Topics in Early Childhood Special Education **7**: 90-104.

Gfeller, J. D., S. J. Lynn, et al. (1987). "Enhancing hypnotic susceptibility: interpersonal and rapport factors." Journal of Personality and Social Psychology **52**: 586-595.

Grahe, J. and R. Sherman (2006). "Examining the good judge in a group setting using a Social Relations Analysis." Annual Meeting of the Midwestern Psychological Association, Chicago, IL.

Goldberg, S. B. (2005). "The secrets of successful mediators." Negotiation Journal **21**(3): 365-376.

Gratch, J., A. Okhmatovskaia, et al. (2006). Virtual Rapport. 6th International Conference on Intelligent Virtual Agents, Marina del Rey, CA, Springer.

Kallmann, M. and S. Marsella (2005). Hierarchical Motion Controllers for Real-Time Autonomous Virtual Humans. 5th International Working Conference on Intelligent Virtual Agents, Kos, Greece, Springer.

Kraut, R. K., S. H. Lewis, et al. (1982). "Listener Responsiveness and the Coordination of Conversation." Journal of Personality and Social Psychology: 718-731.

McNeill, D. (1985). "So you think gestures are nonverbal." Psychological Review 92: 359-371

McNeill, D. (1992). Hand and mind: What gestures reveal about thought. Chicago, IL, The University of Chicago Press.

McNeill, D. (2000). Language and Gesture, Cambridge University Press.

McNeill, D. (2005). Gesture and Thought, University of Chicago press.

Morency, L.-P., C. Sidner, et al. (2005). Contextual Recognition of Head Gestures. 7th International Conference on Multimodal Interactions, Torento, Italy.

Nass, C. and B. Reeves (1996). The Media Equation, Cambridge University Press.

Oppenheim, A. V. and R. W. Schafer (2004). From Frequency to Quefrency: A History of the Cepstrum. IEEE Signal Processing Magazine. **September:** 95-106.

Pertaub, D.-P., M. Slater, et al. (2001). "An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience." Presence: Teleoperators and Virtual Environments **11**(1): 68-78.

Rickenberg, R. and B. Reeves (2000). The effects of animated characters on anxiety, task performance, and evaluations of user interfaces,. SIGCHI conference on Human factors in computing systems, The Hague, The Netherlands.

Smith, J. (2000). GrandChair: Conversational Collection of Family Stories. Cambridge, MA, Media Lab, MIT.

Tatar, D. (1997). Social and personal consequences of a preoccupied listener. Department of Psychology. Stanford, CA, Stanford University**:** Unpublished doctoral dissertation.

Tickle-Degnen, L. and R. Rosenthal (1990). "The Nature of Rapport and its Nonverbal Correlates." Psychological Inquiry **1**(4): 285-293.

Tsui, P. and G. L. Schultz (1985). "Failure of Rapport: Why psychotheraputic engagement fails in the treatment of Asian clients." American Journal of Orthopsychiatry **55**: 561-569.

Ward, N. and W. Tsukahara (2000). "Prosodic features which cue back-channel responses in English and Japanese." Journal of Pragmatics **23**: 1177-1207.

Welji, H. and S. Duncan (in preparation). Social resonance in interaction: A comparison of Friends and Strangers.

Yngve, V. H. (1970). On getting a word in edgewise. Sixth regional Meeting of the Chicago Linguistic Society.