Appears in the International Conference on Intelligent Virtual Agents, Paris, France 2007

Creating Rapport with Virtual Agents

Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy
University of Southern California
Institute for Creative Technologies
13274 Fiji Way, Marina del Rey, CA, 90405, USA
gratch, nwang, gerten, fast at ict.usc.edu

Abstract. Recent research has established the potential for virtual characters to establish rapport with humans through simple contingent nonverbal behaviors. We hypothesized that the contingency, not just the frequency of positive feedback is crucial when it comes to creating rapport. The primary goal in this study was evaluative: can an agent generate behavior that engenders feelings of rapport in human speakers and how does this compare to human generated feedback? A secondary goal was to answer the question: Is contingency (as opposed to frequency) of agent feedback crucial when it comes to creating feelings of rapport? Results suggest that contingency matters when it comes to creating rapport and that agent generated behavior was as good as human listeners in creating rapport. A "virtual human listener" condition performed worse than other conditions.

Keywords: rapport, virtual agents, evaluation

1 Introduction

You know that harmony, fluidity, synchrony, flow one feels when engaged in a good conversation with someone? Known formally as *rapport*, these features are prototypical characteristics of many successful interactions. Speakers seem tightly enmeshed in something like a dance. They rapidly detect and respond to each other's movements. Tickle-Degnen and Rosenthal [1] equate rapport with behaviors indicating positive emotions (e.g. head nods or smiles), mutual attentiveness (e.g. mutual gaze), and coordination (e.g. postural mimicry or synchronized movements). Numerous studies have demonstrated that, when established, rapport facilitates a wide range of social interactions including negotiations [2], management [3], psychotherapy [4], teaching [5] and caregiving [6].

Several research groups are currently exploring the potential of embodied agents to establish rapport with humans through simple contingent nonverbal behavior. Such systems, for example, can generate positive feedback (e.g., nods) by recognizing and responding to vocal or behavioral cues of a human speaker [7-13]. Further, there is growing empirical evidence that such simple contingent behaviors can make agents more engaging [9, 14, 15] and persuasive [13], promote fluent speech [9, 14] and reduce user frustration [12]. These effects can be subtle; many studies indicate the benefits of such feedback fall outside of conscious awareness in that people often show measurable impacts on their observable behavior without reporting significant differences when introspecting upon their experience.

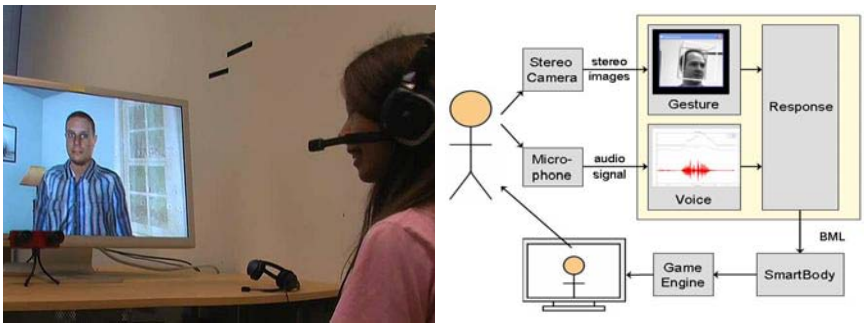

Fig 1: A speaker interacting with the Rapport Agent (left) and the system architecture (right)

Our research on the Rapport Agent [9] investigates how virtual characters can elicit the harmony, fluidity, synchrony, flow one feels when achieving rapport. Among human dyads, rapport can be conceptualized as a phenomenon occurring on three levels: the emotional, the behavioral, and the cognitive. Emotionally, rapport is an inherently rewarding experience; we feel a harmony, a flow. Cognitively, we share an understanding with our conversation partner; there is a convergence of beliefs or views, a bridging of ideas or perspectives. Behaviorally (or interactionally), there is a convergence of movements with our conversational partner; observers report increased synchrony, fluidity and coordination in partners' movements. Are virtual characters capable of establishing rapport with us, on each of these levels, when we are their conversation partners?

Although our primary goal in this study was evaluative (i.e., can agent-generated behavior engender feelings of rapport in human speakers comparable to that of real human listeners?), a secondary goal of ours was to attempt to answer the question: Is contingency, not just the frequency of positive feedback in agents, crucial when it comes to creating feelings of rapport? We define *contingent feedback* as nonverbal movements by a listener (e.g, nods or posture shifts) that are tightly coupled to what the speaker is doing in the moment and *non-contingent feedback* as listener movements that share the same frequency and characteristics of contingent feedback divorced from what the speaker is doing in the moment. In other words, does feedback have to be tightly coupled to what the speaker is doing *in the moment* (imposing fairly challenging computational requirements) or would random positive feedback suffice (greatly simplifying the task of embodied agent design)? This article describes our current progress on addressing these questions.

Rapport Agent and Prior findings

The Rapport Agent is designed to elicit rapport from human participants within the confines of a dyadic narrative task. In this setting, a speaker (the narrator) retells some previously observed series of events (i.e., the events in a sexual harassment awareness and prevention video) to a graphical character. The speaker is led to believe that the character accurately reflects the nonverbal feedback of a human listener. In fact, these movements are generated by the Rapport Agent (see Figure 1).

The central challenge for the rapport agent is to provide the nonverbal listening feedback associated with rapportful interactions. Such feedback includes the use of

backchannel continuers [16] (nods, elicited by speaker prosodic cues, that signify the communication is working), postural mirroring, and mimicry of certain head gestures (e.g., gaze shifts and head nods). The Rapport Agent generates such feedback by real-time analysis of acoustic properties of speech (detecting backchannel opportunity points, disfluencies, questions, and loudness) and speaker gestures (detecting head nods, shakes, gaze shifts and posture shifts).

Prior evaluations have demonstrated the Rapport Agent's social impact at the emotional and behavioral levels when contrasted with either an "unresponsive agent" that produced random, neutral (as opposed to positive) behaviors or with visible human listeners [9, 14]. The Rapport Agent produced benefits at the emotional level through increased speaker engagement (as indexed by duration of the interaction and the number of meaningful words produced). It produced improvements at the behavioral or interactional level when compared with the unresponsive agent (as indexed by the number and rate of disfluencies). We have not addressed the question of influences at the cognitive level.

Although these prior studies have demonstrated a social impact, it is less clear what aspects of agent behavior are critical and where improvements can be made. One relevant fact is the *form* of the feedback. People utilize a variety of behavioral movements, posture shifts, and facial expressions, and some research has shown that subtle features of how these behaviors are expressed can influence interpretation. For example, Krumhuber et al. showed that variation in the onset and offset rates of facial expressions would influence interpretations of trust and sincerity [17]. One way to gain insight into such factors is to capture the actual nonverbal feedback displayed by human listeners and use this to drive the behavior of virtual characters.

Another relevant factor in the establishment of rapport is the *contingency* of feedback: does listener feedback have to be contingent on speaker behavior? Few empirical studies of embodied agents have specifically controlled for the contingency of behavior. For example, studies of the Rapport agent [9, 14] did not control separately for the contingency and the distribution of feedback, leaving open the possibility that *frequency of feedback* is the crucial variable when it comes to creating rapport and that non-contingent (i.e., randomly timed) head nods and posture shifts could be just as effective as well-timed feedback when it comes to creating feelings of rapport, obviating the need for complex techniques for sensing user behavior.

When contingency has been carefully controlled, empirical findings are mixed. For example, Bailenson and Yee found a significant increase in the persuasiveness of virtual characters if they mirrored a human listener's head motion with a four second delay [13]. On the other hand, Burleson found no significant difference on a number of dependent variables from a similar mirroring intervention [18], though a recent post-hoc analysis suggests an interaction dependent on gender [12]. Research on delays in human-to-human interaction also has shown a mixed relationship with rapport. [19]. Collectively, such findings point to a need to further investigate the role of contingency in the context of listener feedback.

The present study seeks to deepen and generalize our prior findings on the cognitive, emotional and behavioral impact of rapport and to specifically investigate the role of contingency.

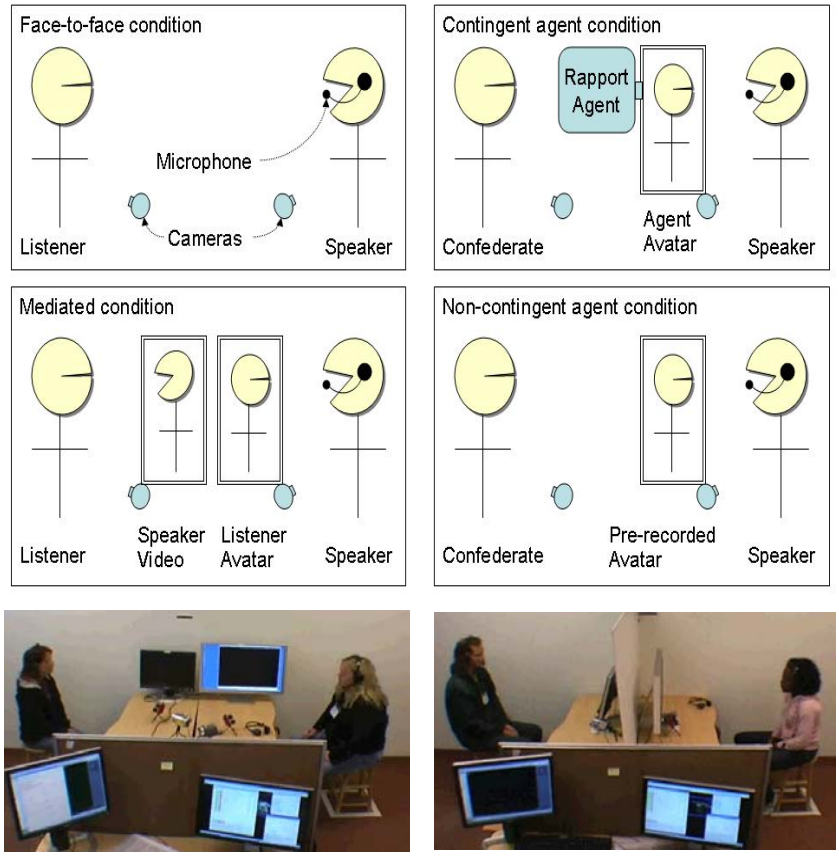

Fig 2. Graphical depiction of the four conditions. The actual face-to-face condition is illustrated on the lower left and the setup for the other three conditions on the lower right.

2 Method

One-hundred thirty-one people (61% women, 39% men) from the general Los Angeles area participated in this study. They were recruited by responding to recruitment posters posted on Craigslist.com and were compensated $20 for one hour of their participation. On average, the participants were 37.5 years old (*min* = 18, *max* = 60, *std* = 11.3) with 15.6 years of education (*min* = 5, *max* = 24, *std* = 3.0). For female subjects, the average age is 38.7 (*min* = 20, *max* = 60, *std* = 11.4), and the average years of education is 15.7 (*min* = 5, *max* = 24, *std* = 3.3). For male subjects, the average age is 35.7 (*min* = 18, *max* = 59, *std* = 11.1), and the average years of education is 13.7 (*min* = 10, *max* = 20, *std* = 2.5).

Design

To investigate the importance of feedback form and contingency, we studied two kinds of virtual characters: one, a "good virtual listener" (the "Responsive" condition) using the Rapport Agent to synthesizes head gestures and posture shifts in response to features of a real human speaker's speech and movements, and the other, a "virtual representation of a real listener" (the "Mediated" condition), which reproduces the actual head movements and posture shifts of a real human listener. To investigate whether these two characters could engender feelings of rapport in human speakers comparable to that of real human listeners, we added a "face-to-face" condition, in which speakers spoke directly to real human listeners, for comparison. In a fourth condition, we created "a non-contingent response virtual listener" that provided positive feedback that was unsynchronized with the speaker's movements and speech. Equivalence in feedback frequency across conditions was created by experimental design.

The study design was a between-subjects experiment with four conditions: Face-to-face (n = 40: 20 speakers, 20 listeners), Mediated (n = 40: 20 speakers, 20 listeners), Responsive (n = 24), and Non-contingent (n = 24), to which participants were randomly assigned using a coin flip. A confederate listener was used in the Responsive and Non-Contingent conditions.

Face to Face. In the Face-to-face condition, the participant talked to a human listener face-to-face.

Mediated. In the Mediated condition, the participant interacted with a virtual character whose head movements and posture were copied from the movements of a real human listener. Through the use of stereo camera and image-based tracking software, the head position and orientation of the listener were captured and displayed by a virtual human character to the speaker. Facial expression feedback was not recognized or displayed.

Responsive. In the Responsive condition, the participant interacted with a virtual character displaying proper listening behaviors. These behaviors were contingent on the recognition of features of the participant's speech (acquired by microphone) and head movements (acquired by a stereo camera) and driven according to predefined behavior-mapping rules (see [9]). For example, certain prosodic contours in the speaker's voice would cause the character to nod and mirror posture shifts. Facial expressions were not generated.

Non-contingent. Finally, in the Non-contingent condition, the participant interacted with a virtual character whose behaviors are identical to the responsive condition in terms of their frequency and dynamics, but not contingent on the behaviors of the speaker. Each subject is presented with a pre-recorded behavior sequence taken from the responsive condition. Equivalence in feedback frequency across conditions was created by experimental design: Following the "yoking" design of Bailenson and Yee [13], the behavior corresponded to what was seen by the previous speaker in the Responsive condition (i.e., each Non-contingent speaker was paired with a Responsive speaker, and saw their feedback).[1]

[1] In the case where duration of the Non-contingent session is longer than the last Responsive session, the system would loop to the beginning of the recording.

Procedure

Participants in groups of two entered the laboratory and were told they were participating in a study to evaluate a communicative technology. The experimenter informed participants:

The study we are doing here today is to evaluate a communicative technology that's developed here. An example of the communicative technology is a web-camera used to chat with your friends and family.

After subjects signed the consent form and completed the pre-questionnaire, the experimenter asked the question "what's your favorite animal?" The subject whose answer came first alphabetically was assigned the speaker role and the other subject was assigned the listener role. In the Responsive and Non-contingent conditions, the confederate always gave the answer "zebra" to ensure their being assigned to the listener role.

Next, subjects were led to two separate side rooms to fill out the pre-questionnaire.

After both subjects completed the pre-questionnaire, subjects were led into the computer room. The experimenter then explained the procedure and introduced participants to the equipment used in the experiment.

Next, the speaker remained in the computer room while the listener was led to a separate side room to wait. The speaker then viewed a short segment of a video clip taken from the Edge Training Systems, Inc. Sexual Harassment Awareness video. The video clip was merged from two clips: The first, "CyberStalker," is about a woman at work who receives unwanted instant messages from a colleague at work (CLIP 1), and the second, "That's an Order!", is about a man at work who is confronted by a female business associate, who asks him for a foot massage in return for her business (CLIP 2).

After the speaker finished viewing the video, the listener was led back into the computer room, where the speaker was instructed to retell the stories portrayed in the clips to the listener.

Speakers in all conditions (except the face-to-face condition) sat in front of a 30-inch computer monitor and sat approximately 8 feet apart from the listener, who sat in front of a 19-inch computer monitor. They could not see each other, being separated by a screen. The speaker saw an animated character displayed on the 30-inch computer monitor. Speakers in all conditions (but the face-to-face condition) were told that the avatar on the screen represents the human listener. While the speaker spoke, the listener could see a real time video image of the speaker retelling the story displayed on the 19-inch computer monitor. The monitor was fitted with a stereo camera system and a camcorder. For capturing high-quality audio, the subject wore a lightweight close-talking microphone and spoke into a microphone headset.

Next, the experimenter led the speaker to a separate side room. The speaker completed the post-questionnaire while the listener remained in the computer room and spoke to the camera what he/she had been told by the speaker.

Finally, participants were debriefed individually and probed for suspicion using the protocol from Aronson, Ellsworth, Carlsmith, and Gonzales [20]. No participants indicated that they believed the listener was a confederate in the study.

Equipment

To produce listening behaviors, the Rapport agent first collects and analyzes the features from the speaker's voice and upper-body movements. Two Videre Design

Small Vision System stereo cameras were placed in front of the speaker and listener to capture their movements. Watson, an image-based tracking library developed by Louis-Phillipe Morency, uses images captured by the stereo cameras to track the subjects' head position and orientation [21]. Watson also incorporates learned motion classifiers that detect head nods and shakes from a vector of head velocities. Both the speaker and listener wore a headset with microphone. Acoustic features are derived from properties of the pitch and intensity of the speech signal using a signal processing package, LAUN, developed by Mathieu Morales [9].

Three Panasonic PV-GS180 camcorders were used to videotape the experiment: one was placed in front the speaker, one in front of the listener, and one was attached to the ceiling to record both speaker and listener. The camcorder that was in front of the speaker was connected to the computer monitor in front of the listener, in order to display video images of the speaker to the listener.

Four desktop computers were used in the experiment: two DELL Precision 670 computers, one with Intel Xeon 3.2 GHz CPU and 2 GB of RAM (for speaker) and another one with Intel Xeon 3.80 GHz CPU and 2 GB of RAM (for listener), run Watson and record stereo camera images, one DELL Precision 690 (Intel Xeon 3.73 GHz CPU with 3 GB of RAM) runs the experiment system and one DELL Precision 530 (Intel Xeon 1.7 GHz with 1 GB of RAM) stores logs.

The animated agent was displayed on a 30-inch Apple display to approximate the size of a real life listener sitting 8 feet away. The video of the speaker was displayed on a 19-inch Dell monitor to the listener.

Measures

Rapport scale. We constructed a 10-item rapport scale (coefficient alpha = .89), presented to speakers in the post-questionnaire. This scale was measured with a 9 point metric (0 = Disagree Strongly; 8 = Agree Strongly). Sample items include: "I think the listener and I established a rapport" and "I felt I was able to engage the listener with my story."

Emotional rapport. We indexed the emotional component of rapport using the item "I felt I had a connection with the listener." This is taken from the Rapport scale listed above.

Cognitive rapport. We indexed the cognitive component of rapport using the item "I think that the listener and I understood each other." This is taken from the Rapport scale listed above.

Behavioral or interactional rapport. Behavioral or interactional measures of rapport included duration or speech, word count, number of pausefillers, number of prolonged words, number of incomplete words, number of disfluencies (pausefillers + incomplete words), number of meaningful words (wordcount-pausefillers-incomplete words), and variations thereof (i.e., calculations per word and per minute).

Helpfulness, distraction, agent naturalness. For helpfulness and distraction scla, we constructed 2 items for each scale, with Cronbach's alpha coefficient of .64 and .49, respectively. These scales were measured with a 9 point metric (0 = Disagree Strongly; 8 = Agree Strongly). We also constructed a 6-item agent naturalness scale, with Cronbach's alpha coefficient of .77. This scale was measured with a 9 point metric (0 = Disagree Strongly; 8 = Agree Strongly). These three scales were issued to speakers in the post-questionnaire. These scales indexed how helpful the listener's

feedback was, how distracting the listener's feedback was, and how natural the agent appeared to be, respectively.

Performance. Speakers' self-assessed performance in the speaking task was measured using this scale we constructed (coefficient alpha = .85). Sample items include: "I think I did a good job telling the story" and "I had difficulty explaining the story" (reverse coded). This scale was issued in the post-questionnaire.

Trustworthiness and Likableness. Speakers from all conditions were asked to evaluate the listener on these traits, using the items 'likeable' and 'trustworthy' taken from the dependent measure used in the Krumhuber, Manstead, Cosker, Marshall, and Rosin study [17]. This scale was measured with an 8 point metric (0 = Not At All; 7 = Very). These items were issued in the post-questionnaire packet.

Pre-questionnaire packet. In addition to the scales listed above, the pre-questionnaire packet also contained questions about one's demographic background, personality [22], self-monitoring [23], self-consciousness [24] and shyness [25]. Scales ranged from 1 (disagree strongly) to 5 (agree strongly). Speakers and listeners from all conditions filled out the pre-questionnaire.

Post-questionnaire packet. In addition to the scales listed above, the post-questionnaire packet also contained questions to examine speaker self-focus, other-focus, embarrassment, and speaker's goals while explaining the video. Scales from [17] range from 0 (not at all) to 7 (very). Other scales ranged from 0 (disagree strongly to 8 (agree strongly). It was completed by speakers across all conditions.

3 Results

Although our primary goal in this study was evaluative (i.e., do the agents we created engender feelings of rapport in human speakers comparable to that of real human listeners?), a secondary goal of ours was to attempt to answer the question: Is contingency, not just the frequency of feedback in agents, crucial when it comes to creating feelings of rapport?

To investigate whether embodied agents could engender feelings of rapport in human speakers comparable to that of real human listeners, we performed a pairwise means analysis on rapport means across these 4 conditions using the Tukey test (see Table 1). Results indicate that the responsive agent was as good as human listeners in creating rapport, but that the mediated avatar was not, with the mediated avatar eliciting less rapport, as captured by our self report scale, and more pause fillers (each in terms of raw counts, rate, and per word) than a real human listener. The mediated avatar also elicited more prolonged words (each in terms of raw counts, rate, and per word) than in the responsive condition. As compared with real human listeners in the face-to-face condition, the non-contingent agent elicited more speaker pause fillers. Although the mediated avatar was as likeable and trustworthy, it was found to be less helpful and more distracting than either a real human listener or the responsive agent, and to be less natural than the responsive agent. The non-contingent agent was also found to be more distracting than a real human listener. The responsive agent, however, was found to be less trustworthy than human listeners. Speakers rated themselves as performing equally well across all conditions.

To answer the question as to whether contingency of feedback is crucial when it comes to creating feelings of rapport, participants across the Non-contingent agent and Responsive agent conditions were paired by feedback frequency, and a dependent samples t-test was conducted, comparing the rapport means across conditions. This analysis was chosen because equivalency in feedback frequency across conditions was created by experimental design: A feedback recording taken from the Agents in

Table 1: Tukey Table of Means for Speakers

Measures	Non-Contingent	Responsive	Mediated	F-to-F
Helpful	5.44	5.73 b	4.43 b, c	5.85 c
Distracting	3.64 a	2.62 b	4.15 b, c	2.45 a, c
Trustworthy	4.56	3.65 d	4.40	4.89 d
Likeable	4.67	4.15	4.55	4.63
Natural Agent	3.12	3.79 b	2.63 b	NA
Performance	5.20	5.22	5.50	5.52
Rapport Scale	4.79	5.04	4.46 c	5.53 c
Emo Rapport	4.56	4.65	4.20	5.60
Cog Rapport	4.88	5.35	4.80	5.55
Beh Rapport:				
Duration	144	131	139	115
Word Count	403	345	353	319
Pausefiller Count	14.00 a	13.08	15.40 c	6.75 a, c
Prolonged Count	3.96	3.00 b	7.00 b	4.60
Incomplete Word Count	5.56	3.92	3.70	4.55
Disfluency Count	19.56	17.00	19.10	11.30
Meaningful Word Count	384	328	334	308
Word Rate	166	158	152	169
Pausefiller Rate	6.08	5.66	6.89 c	3.60 c
Prolonged Rate	1.56	1.40 b	2.82 b	2.12
Incomplete Word Rate	2.26	1.75	1.53	2.26
Disfluency Rate	8.33	7.41	8.42	5.86
Meaningful Rate	157	151	144	163
Pausefiller per Word	.04	.04	.05 c	.02 c
Prolonged per Word	.01	.01 b	.02 b	.01
Incomplete Wd per Word	.01	.01	.01	.01
Disfluency per Word	.05	.05	.06	.03
Meaningful per Word	.95	.95	.94	.97

Note, columns share the same subscripts connote a significant difference at an alpha level of .05 between them.

the Responsive condition was replayed to speakers in the Non-contingent condition.[2] However, because the same feedback tape from the Responsive condition was sometimes played to more than one participant in the Non-Contingent condition, to obtain one-to-one correspondence for this analysis, a randomly selected subsample of par-

[2] As individual speakers vary in the length of their narrative, the frequency of feedback is not strictly identical as the non-contingent behavior may be either a shortened or looped display of the Responsive behavior.

ticipants was drawn from those participants who shared the same feedback tape. Using this procedure, in effect, controlling for feedback frequency, all rapport variables were examined. Figure 3 shows that the total number of words spoken was significantly greater for speakers interacting with non-contingent agent (M = 458.83) than with the responsive agent (M = 299.17, t (11) = 2.17, p = .05), the raw number of pause fillers was significantly greater for speakers interacting with the non-contingent agent (M = 15.75) than with the responsive agent (M = 8.00, t (11) = 3.06, p = .01), the pause filler rate was significantly greater for speakers interacting with the non-contingent agent (M = 6.11) than with the responsive agent (M = 3.80, t (11) = 2.44, p = .03), and the raw number of disfluencies was significantly greater for speakers interacting with the non-contingent agent (M = 22.17) than with the responsive agent (M = 11.75, t (11) = 2.30, p = .04), suggesting that contingency matters.

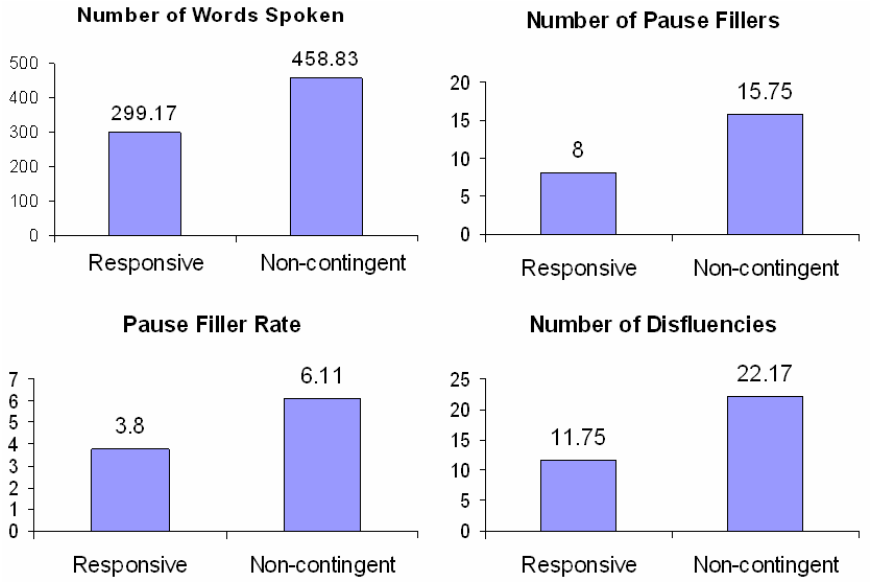

Fig 3. Significant differences were found in number of words spoken, number of pause fillers, pause filler rate and number of disfluencies between speakers interacting with Responsive agent and with the Non-contingent agent.

4 Discussion and Future Work

The primary goal in this study was evaluative: Do the two agents we created engender feelings of rapport in human speakers comparable to that of real human listeners? A secondary goal was to answer the question: Is contingency (as opposed to frequency) of agent feedback crucial when it comes to creating feelings of rapport?

Results indicate that the responsive agent was as effective as human listeners in creating rapport, but that the mediated avatar was not as effective, with the mediated

avatar eliciting less rapport, as captured by our self report scale, and by some of our behavioral/interactional indices. Although the mediated avatar was as likeable and trustworthy, it was found to be less helpful and more distracting than either a real human listener or the responsive agent we created. Several factors could have contributed to relatively poor performance of the mediated condition. It is possible that listeners in the mediated condition gave less visible feedback. For example, the Rapport Agent always generates bodily feedback (nods, posture shifts) in response to speaker cues; however human listeners often responded with facial feedback that would be seen in the face-to-face condition but was not recognized or displayed in the mediated condition. Listeners may have also felt less engaged from watching a video than listeners in the face-to-face condition and therefore exhibited less feedback. Finally, there may have been subtle errors introduced by the video processing equipment that disrupted rapport. A direction for future studies is to understand the factors that contributed to the lower measures for the mediated condition.

Controlling for the frequency of feedback, some behavioral indices of rapport were significantly greater for speakers interacting with the non-contingent agent than with the responsive agent, suggesting that contingency of agent feedback matters when it comes to creating virtual rapport. This is the first experimental evidence supporting this (often unspoken) assumption of much embodied agent research.

Several open questions remain for future work. The current study reveals interesting differences in the impact of virtual character behavior on rapport-related variables when compared with our prior experimental findings. For example, a previous study [14] found that subjects spoke significantly longer in the responsive condition when compared to face-to-face, whereas the current study found no significant difference on this dimension. One key difference in the current study is the use of more provocative narrative content – a sexual harassment video as opposed to a funny cartoon. One may expect subjects to be less comfortable and more concerned with impression management (e.g., using "politically correct" terminology in their narratives). These factors would be expected to negatively impact rapport, though their differential impact across conditions is unclear.

Another important question is how to provide more semantically meaningful feedback to the speaker. The Rapport Agent responds without attending to the content of the speaker's narrative. Such feedback has been called *envelope feedback* [26] or *generic feedback* [27] and, despite being non-specific to the meaning of speech, plays an important interaction function. It seems to signal "everything is ok, please continue," or "I'm paying attention", and can contribute to a sense of mutual understanding and liking; factors associated with rapport. Several studies have demonstrated that envelope feedback can be woefully inadequate in certain contexts if not bolstered by *specific,* or *content feedback,* that makes reference to the content of the speech. For example, Bavelas et al. [27] found when speakers were telling personally emotional stories, storytellers expected emotional feedback to key events in the story and found it hard to construct effective narratives without it. A major challenge is how to recognize and respond meaningfully to a speaker with the rapidity seen in human dyads. See [28, 29] for some initial explorations in this direction.

Finally, within the virtual human's community, rapport has been conceptualized as short-term construct that arises in a single interaction, as discussed here, or as a deepening sense of interdependence that arises over time [30, 31]. Both approaches, how-

ever, demand greater attention to multi-modal recognition, a greater understanding of the functional role nonverbal behavior plays in co-construction of meaning and deeper models of the social cognitions that underlie the generation and interpretation such reciprocal behaviors. As such, rapport can serve as a productive theoretical construct to propel the advancement of virtual human research.

Overall, the current study and related findings add further evidence that the nonverbal behavior of virtual characters influence the behavior of the humans that interact with them. This gives confidence that embodied agents can facilitate social interaction between humans and computers, with a host of implications for application and social psychological research.

Acknowledgments. We are grateful for the substantive contributions of a number of individuals. Wendy Treynor played an indispensable role in the experimental design and data analysis, and contributed to the draft. Jeremy Bailenson, Anya Okhmatovskaia and Alison Wiener also gave valuable input to the experimental design. Sue Duncan, Nicole Kraemer and Nigel Ward informed the theoretical underpinnings of the work. We also thank Edge Training Systems, Inc., 9710 Farrar Court, Suite P, Richmond, VA 23236, for granting us the right to use their vignettes in our research. 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

1. Tickle-Degnen, L. and R. Rosenthal, *The Nature of Rapport and its Nonverbal Correlates.* Psychological Inquiry, 1990. **1**(4): p. 285-293.
2. Drolet, A.L. and M.W. Morris, *Rapport in conflict resolution: accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts.* Experimental Social Psychology, 2000. **36**: p. 26-50.
3. Cogger, J.W., *Are you a skilled interviewer?* Personnel Journal, 1982. **61**: p. 840-843.
4. Tsui, P. and G.L. Schultz, *Failure of Rapport: Why psychotheraputic engagement fails in the treatment of Asian clients.* American Journal of Orthopsychiatry, 1985. **55**: p. 561-569.
5. Fuchs, D., *Examiner familiarity effects on test performance: implications for training and practice.* Topics in Early Childhood Special Education, 1987. **7**: p. 90-104.
6. Burns, M., *Rapport and relationships: The basis of child care.* Journal of Child Care, 1984. **2**: p. 47-57.
7. Tosa, N., *Neurobaby.* ACM SIGGRAPH, 1993: p. 212-213.
8. Breazeal, C. and L. Aryananda, *Recognition of Affective Communicative Intent in Robot-Directed Speech.* Autonomous Robots, 2002. **12**: p. 83-104.
9. Gratch, J., et al. *Virtual Rapport*. in *6th International Conference on Intelligent Virtual Agents*. 2006. Marina del Rey, CA: Springer.

10. Cassell, J. and K.R. Thórisson, *The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents.* International Journal of Applied Artificial Intelligence, 1999. **13**(4-5): p. 519-538.
11. Brand, M. *Voice puppetry*. in *ACM SIGGRAPH*. 1999: ACM Press/Addison-Wesley Publishing Co.
12. Burleson, W. and R.W. Picard, *Evidence for Gender Specific Approaches to the Development of Emotionally Intelligent Learning Companions.* IEEE Intelligent Systems, Special issue on Intelligent Educational Systems, 2007. **Jul/Aug**.
13. Bailenson, J.N. and N. Yee, *Digital Chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments.* Psychological Science, 2005. **16**: p. 814-819.
14. Gratch, J., et al. *Can virtual humans be more engaging than real ones?* in *12th International Conference on Human-Computer Interaction*. 2007. Beijing, China.
15. Smith, J., *GrandChair: Conversational Collection of Family Stories*. 2000, Media Lab, MIT: Cambridge, MA.
16. Ward, N. and W. Tsukahara, *Prosodic features which cue back-channel responses in English and Japanese.* Journal of Pragmatics, 2000. **23**: p. 1177-1207.
17. Krumhuber, E., et al. *Temporal aspects of smiles influence employment decisions: A comparison of human and synthetic faces.* in *11th European Conference Facial Expressions: Measurement and Meaning*. 2005. Durham, United Kingdom.
18. Burleson, W., *Affective Learning Companions: Strategies for Empathetic Agents with Real-Time Multimodal Affective Sensing to Foster Meta-Cognitive and Meta-Affective Approaches to Learning, Motivation, and Perseverance*. 2006, Unpublished PhD Thesis, MIT Media Lab: Boston.
19. Manning, T.R., E.T. Goetz, and R.L. Street, *Signal delay effects on rapport in telepsychiatry.* CyberPsychology and Behavior, 2000. **3**(2): p. 119-127.
20. Aronson, E., et al., *Methods of Research in Social Psychology*. 2nd Edition ed. 1990, New York: McGraw-Hill.
21. Morency, L.-P., et al. *Contextual Recognition of Head Gestures*. in *7th International Conference on Multimodal Interactions*. 2005. Torento, Italy.
22. John, O.P. and S. Srivastava, *The Big-Five trait taxonomy: History, measurement, and theoretical perspectives.* Handbook of personality: Theory and research, 1999. **2**: p. 102–138.
23. Lennox, R.D. and R.N. Wolfe, *Revision of the Self-Monitoring Scale.* Journal of Personality and Social psychology, 1984. **46**: p. 1349-1364.
24. Scheier, M.F. and C.S. Carver, *The Self-Consciousness Scale: A revised version for use with general populations.* Journal of Applied Social Psychology, 1985. **15**: p. 687-699.
25. Cheek, J.M., *The Revised Cheek and Buss Shyness Scale (RCBS)*. 1983, Wellesley College: Wellesley MA.
26. Thórisson, K.R., *Communicative Humanoids: A Computational Model of Psycho-Social Dialogue Skills*, in *The Media Lab*. 1996, Massachusetts Institute of Technology.

27. Bavelas, J.B., L. Coates, and T. Johnson, *Listeners as Co-narrators.* Journal of Personality and Social Psychology, 2000. **79**(6): p. 941-952.
28. Jondottir, G.R., et al. *Fluid Semantic Back-Channel Feedback in Dialogue: Challenges and Progress*. in *Intelligent Virtual Agents*. 2007. Paris, France: Springer.
29. Heylen, D. *Challenges Ahead. Head Movements and other social acts in conversation*. in *AISB*. 2005. Hertfordshire, UK.
30. Cassell, J. and T. Bickmore, *Negotiated Collusion: Modeling social language and its relationship effects in intelligent agents.* User Modeling and Adaptive Interfaces, 2002. **12**: p. 1-44.
31. Cassell, J., A. Gill, and P. Tepper. *Conversational Coordination and Rapport*. in *Proceedings of Workshop on Embodied Language Processing at ACL 2007*. 2007. Prague, CZ.