Can Virtual Human Build Rapport and Promote Learning?

Ning WANG[a,1] and Jonathan GRATCH[a]
[a]*Institute for Creative Technologies*
University of Southern California

Abstract. Research show that teacher's nonverbal immediacy can have a positive impact on student's cognitive learning and affect [31]. This paper investigates the effectiveness of nonverbal immediacy using a virtual human. The virtual human attempts to use immediacy feedback to create rapport with the learner. Results show that the virtual human established rapport with learners but did not help them achieve better learning results. The results also suggest that creating rapport is related to higher self-efficacy, and self-efficacy is related to better learning results.

Keywords. Pedagogical agent, virtual human, rapport

Introduction

In recent years, there has been significant progress in the development of advanced computer-based learning environments. These systems have demonstrated the potential to promote learning gains that far exceed what is typical of classroom instruction. Some of the systems are successfully making the transition out of the research laboratory and into widespread instructional use, for example the math tutor developed by Carnegie Learning and foreign language and culture course developed by Alelo Inc. Despite these impressive achievements, computer-based instruction falls short of the effectiveness of human tutors [4][5], and much research remains both in terms of identifying why certain techniques are effective and how to exploit these techniques to promote efficient learning.

Researchers have investigated the potential of pedagogical agents to promote learning for years. Few formal studies have demonstrated that pedagogical agents can improve learning gains [8][1][2]. What constitutes an effective teacher is a fundamental question to pedagogical agents design. Past research has shown that teacher's verbal and non-verbal immediacy can have a positive impact on student's cognitive learning and affect [32, 31]. Much of the research on teacher immediacy has focused on nonverbal cues and seems to indicate that immediacy does increase teaching effectiveness. Anderson [32] defines nonverbal cues of immediacy as eye contact, gestures, relaxed body position, direct body position toward students, smiling, vocal expressiveness, movement and proximity.

We have designed a virtual human that attempts to use nonverbal immediacy to produce a sense of rapport from a human speaker. The virtual human, called the Rapport Agent, tracks the human speaker's prosody, head movements and posture in

[1] Corresponding Author.

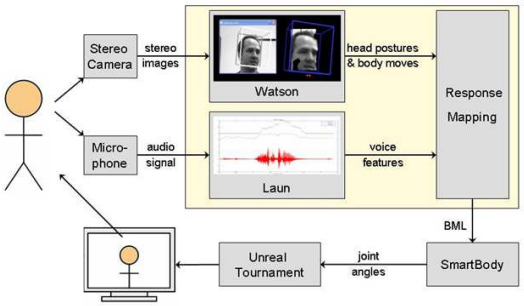
real time, and rapidly producing contingent feedback (head nods, postural mirroring)[11]. Rapport is argued to underlie success in negotiations and conflict resolution [12][13], improving worker compliance [14], psychotherapeutic effectiveness [15], improved quality of child care [16] and, improved test performance in classrooms [17]. Cappella [9] 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." In this paper, we investigate whether the virtual human who exhibit immediacy behavior can build rapport with learners and help them learn. We hypothesize that immediacy feedback will be perceived as helpful and can help student learn better.

Rapport Agent

The Rapport Agent was designed to establish a sense of rapport with a human participant in "face-to-face monologs" where a human participant 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. Creation of the system is inspired by findings that feelings of rapport are correlated with simple contingent behaviors between speaker and listener, including behavioral mimicry [18] and back-channeling (e.g., nods [3]). Rapport Agent uses a vision based tracking system and signal processing of the speech signal to detect features of the speaker and then uses a set of reactive rules to drive the listening mapping displayed in Figure 1. The architecture of the system is also 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 [19] uses stereo video to track the participants' head position and orientation and incorporates learned motion classifiers that detect head nods and shakes from a vector of head velocities. Other features are derived from the tracking data. 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, using a signal processing package, LAUN, developed by Mathieu Morales. Speaker pitch is approximated with the cepstrum of the speech signal [20] and processed every 20ms. Audio artifacts introduced by the motion of the Speaker's head are minimized by filtering out 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 from [21]).



Silence → gaze up/straight
Raised loudness → head nod
Backchannel → head nod
Ask question → head nod
Speaker shifts posture → mimic
Speaker gazes away → mimic
Speaker nods or shakes → mimic

Figure 1. Rapport Agent architecture and behavior mapping table.

Recognized speaker features are mapped into listening animations through a set of authorable mapping language. This language supports several advanced features. Authors can specify contextual constraints on listening behavior, for example, triggering different behaviors depending on the state of the speaker (e.g., the speaker is silent), the state of the agent (e.g., the agent is looking away), or other arbitrary features (e.g., the speaker's gender). One can also specify temporal constraints on listening behavior: For example, one can constrain the number of behaviors produced within some interval of time. Finally, the author can specify variability in behavioral responses through a probability distribution of different animated responses.

These animation commands are passed to the SmartBody animation system [22] using a standardized API [23]. SmartBody is designed to seamlessly blend animations and procedural behaviors, particularly conversational behavior. These animations are rendered in the Unreal Tournament™ game engine and displayed to the Speaker.

Method

One-hundred forty-four people (62.5% women, 37.5% 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 $30 for one and half hour of their participation. On average, the participants were 39.5 years old (*min* = 19, *max* = 60, *std* = 11.6) with 15.8 years of education (*min* = 12, *max* = 20, *std* = 1.6).

1. Design

To investigate the impact of rapportful feedback on learning, we conducted the study using three kinds of virtual agents. The first virtual agent is a "good virtual listener" (the "Responsive" condition). The agent provides rapportful feedback by synthesizing head gestures and posture shifts in response to features of the learner's speech and movements. The second virtual agent, a "not responsive listener" (the "Non-responsive" condition), is one that does not provide rapportful feedback but still tries to be attentive. The agent gazes at the learner but does not provide any feedback in response to the learner. The last agent is an "ignoring listener" (the "Ignore" condition), who does not pay attention to the learner. This agent does not maintain gaze with the learner nor respond to the learner. The manipulation of gaze was inspired by Bailenson's study [24], which illustrated that students learned better when looked at

more by a virtual teacher. Mutlu et al. [25] also found that increased gaze from a storytelling robot facilitated greater recall of story events.

The study design was a between-subjects experiment with three conditions: Responsive (n = 51), Non-responsive (n = 47), and Ignore (n = 46), to which participants were randomly assigned.

2. Procedure

The participant first signed the consent form and completed the pre-questionnaire. Then the participant was assigned the role of the speaker and the confederate was assigned to the role of the listener. Next, the speaker was led to the computer room while the listener waited in a separate side room. The speaker viewed one of two videos. One of the videos was a Tweety and Sylvester cartoon. The other video is taken from the Edge Training Systems, Inc. Sexual Harassment Awareness video. The video clip, "CyberStalker," is about a woman at work who receives unwanted instant messages from a colleague at work. Which one of the videos was shown was randomly decided.

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 sat in front of a computer monitor and sat approximately 8 feet apart from the listener, who sat in front of a TV. They could not see each other, being separated by a screen. The speaker saw the virtual agent displayed on the computer monitor. The Speaker was told that the virtual agent 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 TV. Next, the experimenter led the speaker to a separate side room. The speaker completed a questionnaire about the contents of the video he/she just saw. During this time, the listener (the confederate) remained in the computer room and spoke to the camera what he/she had been told by the speaker.

Later, the speaker was led back to the computer room and watched remaining of the two videos. The speaker then retold the stories portrayed in the clips to the listener. After that, the speaker filled out another questionnaire about the contents of the video while the listener (the confederate) remained in the computer room and spoke to the camera what he/she had been told by the speaker. Then the speaker completed the post-questionnaire. Finally, participants were debriefed individually. No participants indicated that they believed the listener was a confederate in the study.

3. Equipment

Two Videre Design Small Vision System stereo cameras were placed in front of the speaker and listener to capture their movements. 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 DELL desktop computers were used in the experiment. 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 30-inch TV to the listener.

4. Measures

Learning Scale. We constructed a learning questionnaire for each video. There are 14 questions regarding the content of the Tweety and Sylvester cartoon video and 15 questions about the Edge Training Systems, Inc. Sexual Harassment Awareness video. Each correct answer gets 1 point. Sample questions include: "When Sylvester first saw Tweety, what was Tweety doing?" and "What was the woman's response to her co-worker when he suggested that she report the harassment?" We constructed 3 learning scales: Total Score, First Video Score and Second Video Score. The Total Score is the sum of points from questionnaires regarding both videos. Since which video was shown first had no significant interaction with experiment condition on all the scales, we constructed a First Video Score scale that's the sum of points from questionnaires regarding the video that's shown first and a Second Video Score scale for questionnaires regarding the video that's shown second.

Rapport scale. We constructed a 10-item rapport scale (coefficient alpha = .89), presented to speakers in the post-questionnaire. This scale was measured with an 8 point metric (1 = 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."

Self-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.

Helpfulness, distraction, agent naturalness. For helpfulness and distraction scale, we constructed 2 items for each scale, with Cronbach's alpha coefficient of .64 and .49, respectively. These scales were measured with an 8 point metric (1 = 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 8 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.

Pre-questionnaire packet. In addition to the scales listed above, the pre-questionnaire packet also contained questions about one's demographic background, personality [26], self-monitoring [27], self-consciousness [28] and shyness [29]. Scales ranged from 1 (disagree strongly) to 5 (agree strongly).

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, speaker's goals while explaining the video and listener's traits [30]. Scales from [30] range from 0 (not at all) to 7 (very). Other scales ranged from 1 (disagree strongly) to 8 (agree strongly).

Result

Data from 11 participants were excluded due to technical difficulties and missing data. As a result, data from 133 sessions were included in the analysis, 48 in the Responsive condition, 41 in the Non-responsive condition and 44 in the Ignore condition.

We performed a pairwise means analysis on means of scales across 3 conditions using the Tukey test (see Table 1). On the overall duration, participants interacted with the Responsive agent talked longer than those interacted with the Ignore agent. The subjects from the Responsive condition also talked longer when retelling the second video.

Table 1. Comparison of learning results, rapport and other self-report measures. Columns share the same subscripts connote a significant difference at an alpha level of .05 between them.

Measures	Responsive	Non-responsive	Ignore
Duration	249.83_{a}	241.24	212.57_{a}
First Video Duration	114.13	114.29	99.34
Second Video Duration	133.52_{a}	126.95	113.23_{a}
Total Score	20.61	19.90	21.18
First Video Score	10.07	9.46	10.39
Second Video Score	10.54	10.44	10.80
Rapport	$4.47_{a\ b}$	3.70_{b}	3.66_{a}
Self-performance	5.19	5.06	5.20
Helpfulness	5.43_{a}	4.51	4.34_{a}
Distraction	3.76	4.85	4.59
Agent naturalness	4.39	4.42	4.27

Overall, there was no significant difference between the three conditions on the Total Score, First Video Score and Second Video Score. Participants from the Responsive condition reported higher rapport than those from the Non-responsive condition and Ignore condition. The subjects interacted with the Responsive agent also felt that the feedback provided by the agent was more helpful than those who interacted with the Ignore agent. There was no significant difference on how distractive the agent feedback was and how natural the agent's behavior was. On average, subjects from all conditions evaluated their own performance about the same.

Table 2. Correlation between Rapport, Total Score and other variables. Columns with a * indicating the correlation is significant at an alpha level of .05.

Correlation	Rapport	Total Score
Duration	-0.118	.075
Rapport	N/A	-.069
Self-performance	.336 *	.251*
Helpfulness	.484 *	-.024
Distraction	-.539 *	.051
Agent naturalness	.312 *	-.057

We then conducted a correlation analysis to test the correlation between rapport and the learning measures. From Table 2 we can see that rapport is positively correlated with Self-performance, feedback helpfulness and agent naturalness and negatively correlated with how distractive the feedback was. However, there was no significant correlation between self-reported rapport and total score. Duration was also not correlated with either rapport or Total Score. Feedback helpfulness, feedback distraction and agent naturalness did not correlate with Total score. Interestingly, self-performance is positively correlated with the total score.

Discussion

In this paper, we presented our investigation of feedback immediacy using a virtual human. The results showed that immediacy feedback induced higher sense of rapport but did not help the learner perform better on the recall test. However, we found a "ceiling effect" on the recall test probably because the learning materials (the videos) were too easy. Some studies [1, 8] show that pedagogical agents may not make a significant impact on learning easy concepts but do so on learning difficult concepts.

One factor that had a significant impact on the learning results is age. Age had a significant negative correlation with learning results (r=-.222). We conducted a hierarchical multiple regression relating experiment condition to learning measures, controlling for the potential effect of age. The result showed that the model as a whole is significant ($F(2,127)=3.829$, p=.024) but age is still the significant predictor of the learning result ($beta_{age}$=-.232, p=.008) compare to the experiment manipulation.

The results showed that self assessment of performance was positively correlated with both rapport and learning gain. Self-performance can be considered as an index of learner's self-efficacy. This result suggested that feedback immediacy can be related to higher self-efficacy. And higher self-efficacy can be related to better learning results.

There are several limitations of the current study. The learning materials (contents of the videos) were relatively straightforward. The relatively easy learning material created a "ceiling effect" on the recall test. For example, even on the most difficult question about the Tweety and Sylvester cartoon, half of the people answered the question correctly. Over 65% of the learning questions had over 70% of correct answer rate. This means that over 70% of the subjects answered those questions correctly.

The analysis presented here was mostly based on self-report measures. In the future, the analysis can be extended to behavior measures such as speech disfluency, explaining style (e.g. summarizing or reasoning) and the quality of the explanation the speaker produced (e.g. how much correct information was conveyed). Previous studies also showed that some individual differences such as shyness can influence how one react to virtual human's rapport building behavior [6]. Similar analysis of individual differences can be conducted on the learning measure. Some studies on pedagogical agents [2][8] indicated that pedagogical agents may not make a significant impact on learning easy concepts but do so on learning difficult concepts. Better learning material can be used to avoid the "ceiling effect" encountered here and facilitate this investigation in future studies.

References

[1] Wang, N., Johnson, W.L. The Politeness Effect in Intelligent Foreign Language Tutoring System. In Proc. of the 9th International Conference on Intelligent Tutoring Systems, 2008

[2] Wang, N., Johnson, W.L., Mayer, R.E., Rizzo, R., Shaw, E., Collins, H. The politeness effect: Pedagogical agents and learning gains. The 12th International Conference on Artificial Intelligence in Education, 2005

[3] Yngve, V. H. (1970). On getting a word in edgewise. Paper presented at the Sixth regional Meeting of the Chicago Linguistic Society.

[4] Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems (Chapter 37). M. G. Helander, T. K. Landauer, & P. Prabhu, (Eds.) Handbook of Human-Computer Interaction, 2nd edition. Amsterdam, The Netherlands: Elsevier Science.

[5] Lane, H.C. (2006). Intelligent Tutoring Systems: Prospects for Guided Practice and Efficient Learning. Whitepaper for the Army's Science of Learning Workshop, Hampton, VA. Aug 1-3, 2006.

[6] Sin-Hwa Kang, Jonathan Gratch, Ning Wang, James Watt. Does Contingency of Agents' Nonverbal Feedback Affect Users' Social Anxiety? 7th International Conference on Autonomous Agents and Multiagent Systems. Estoril, Portugal. May 2008
[7] Reeves, B., Nass, C. (1996). The media equation. New York: Cambridge University Press.
[8] Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., Bhogal, R. S. (1997). The persona effect: Affective impact of animated pedagogical agents. In CHI '97. 359-366
[9] Capella, J. N. (1990). On defining conversational coordination and rapport. Psychological Inquiry, 1(4), 303-305.
[10] Bickmore, T. and Cassell, J. "Relational Agents: A Model and Implementation of Building User Trust." ACM CHI 2001 Conference Proceedings, Seattle, Washington, 2001.
[11] Jonathan Gratch, Anya Okhmatovskaia, Francois Lamothe, Stacy Marsella, Mathieu Morales, R. J. van der Werf and Louis-Philippe Morency. Virtual Rapport. 6th International Conference on Intelligent Virtual Agents, Marina del Rey, CA, 2006
[12] Drolet, A. L., & Morris, M. W. (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.
[13] Goldberg, S.B., The secrets of successful mediators. Negotiation Journal, 2005. 21(3): p. 365-376.
[14] Cogger, J.W., Are you a skilled interviewer? Personnel Journal, 1982. 61: p. 840-843.
[15] 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.
[16] Burns, M., Rapport and relationships: The basis of child care. Journal of Child Care, 1984. 2: p. 47-57.
[17] Fuchs, D., Examiner familiarity effects on test performance: implications for training and practice. Topics in Early Childhood Special Education, 1987. 7: p. 90-104.
[18] Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. Journal of Personality and Social Psychology, 76, 893-910.
[19] Morency, L.-P., Sidner, C., Lee, C., & Darrell, T. (2005). Contextual Recognition of Head Gestures. Paper presented at the 7th International Conference on Multimodal Interactions, Torento, Italy.
[20] Oppenheim, A. V., & Schafer, R. W. (2004). From Frequency to Quefrency: A History of the Cepstrum. IEEE Signal Processing Magazine, September, 95-106.
[21] Ward, N., & Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. Journal of Pragmatics, 23, 1177-1207.
[22] Kallmann, M., & Marsella, S. (2005). Hierarchical Motion Controllers for Real-Time Autonomous Virtual Humans. Paper presented at the 5th International Working Conference on Intelligent Vir-tual Agents, Kos, Greece.
[23] Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., et al. (2006). Towards a common framework for multimodal generation in ECAs: The behavior markup language. Paper presented at the Intelligent Virtual Agents, Marina del Rey, CA.
[24] Bailenson, J.N., Yee, N., Blascovich, J., Beall, A.C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. The Journal of the Learning Sciences, 17, 102-141.
[25] Mutlu, B., Hodgins, J.K., and Forlizzi, J., (2006). A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior. In Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids'06), December 2006, Genova, Italy.
[26] John, O.P. and S. Srivastava, The Big-Five trait taxonomy: History, meas-urement, and theoretical perspectives. Handbook of personality: Theory and research, 1999. 2: p. 102–138.
[27] Lennox, R.D. and R.N. Wolfe, Revision of the Self-Monitoring Scale. Journal of Personality and Social psychology, 1984. 46: p. 1349-1364.
[28] Scheier, M.F. and C.S. Carver, The Self-Consciousness Scale: A revised version for use with general populations. Journal of Applied Social Psychol-ogy, 1985. 15: p. 687-699.
[29] Cheek, J.M., The Revised Cheek and Buss Shyness Scale (RCBS). 1983, Wellesley College: Wellesley MA.
[30] Krumhuber, E., et al. Temporal aspects of smiles influence employment deci-sions: A comparison of human and synthetic faces. in 11th European Con-ference Facial Expressions: Measurement and Meaning. 2005. Durham, United Kingdom.
[31] Gorham, J. (1988). The relationship between verbal teaching immediacy behaviors and student learning. Communication Education, 17, 40-53.
[32] Anderson, J. F. (1979) Teacher immediacy as a predictor of teaching effectiveness, in: D. Nimmo (Ed.) Communication Yearbook 3 Transaction Books, New Brunswick, N.J, pp. 543-559.