Don't Just Stare at Me!

Ning Wang and Jonathan Gratch
Institute for Creative Technologies
University of Southern California
13274 Fiji Way, Marina del Rey, CA 90292 USA
{nwang, Gratch}@ict.usc.edu

ABSTRACT
Communication is more effective and persuasive when participants establish rapport. Tickle-Degnen and Rosenthal [57] argue rapport arises when participants exhibit mutual attentiveness, positivity and coordination. In this paper, we investigate how these factors relate to perceptions of rapport when users interact via avatars in virtual worlds. In this study, participants told a story to what they believed was the avatar of another participant. In fact, the avatar was a computer program that systematically manipulated levels of attentiveness, positivity and coordination. In contrast to Tickel-Degnen and Rosenthal's findings, high-levels of mutual attentiveness alone can dramatically lower perceptions of rapport in avatar communication. Indeed, an agent that attempted to maximize mutual attention performed as poorly as an agent that was designed to convey boredom. Adding positivity and coordination to mutual attentiveness, on the other hand, greatly improved rapport. This work unveils the dependencies between components of rapport and informs the design of agents and avatars in computer mediated communication.

Author Keywords
Virtual human, rapport, back-channel, gaze, head nod, posture mirroring.

ACM Classification Keywords
H.1.2. User/Machine Systems: Human Factors.

General Terms
Design, Experimentation, Human Factors

INTRODUCTION
When we interact, our embodied behaviors of speech prosody, gesture, gaze, posture, and facial expression contribute to the establishment and maintenance of interpersonal *rapport* that serves to scaffold effective social interaction. Rapport has drawn intense interest in the social sciences for its impact in a wide range of interpersonal domains including social engagement [52], classroom learning [22], success in negotiations [20], improving worker compliance [18], psychotherapeutic effectiveness [59], and improved quality of child care [11]. Recent research in virtual environments has demonstrated the possibility of translating these findings into computer-mediated (CMC) and human-computer interactions (HCI) where embodied communicated behaviors can not only be reproduced but altered in novel ways to perhaps amplify their interpersonal consequences [26] [5].

Tickle Degnan and Rosenthal [57] define rapport as a subjective feeling of connectedness and argue that it arises in face-to-face interaction from the expression of these essential components of nonverbal behavior: mutual attention, positivity and coordination. The positivity correlates of rapport are behaviors, such as smiling and head nodding, that indicate participant liking and approval of one another. The coordination correlates, on the other hand, are behaviors that signal that the participants are "with" one another, functioning as a coordinated unit, such as postural mirroring and interactional synchrony. Forward lean and orienting body towards one another are behaviors indicate mutual attention. However, one of the most important indicators of mutual attention is gaze. As we grew up, we were taught by our parents to "look someone in the eye" when we speak. During initial interaction, mutual gaze signals interest, a precondition to the continuation of the interaction. Later, gaze signals the unity of the dyad members, both in terms of the unity of their experience and the mutuality of their relationship goals.

Although rapport was developed to explain properties of face-to-face interaction, recent work has examined Tickle-Degnan and Rosenthal's theoretical framework [57] within the context of CMC [53] and HCI [13]. Indeed, rapport bears close similarity to the CMC notion of *social presence* [50] and subjective measures of rapport index many of the same concepts as social presence scales. In our own work, we have examined how the nonverbal behaviors of virtual characters can influence subjective and behavioral correlates of rapport in interactions within virtual worlds [26]. This work emphasized the positivity and coordination components of the model. In this paper, we examine the impact of mutual attention component by manipulating the gaze behavior of a virtual agent. This is part of a large scale in-

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

vestigation to systemically assessing the nonverbal behavior components of rapport and their validity for informing the design of computer-mediated and human-agent interaction in virtual environments. The results of this article not only deepen understanding of the role of rapport in interactions in the virtual environments but also raise important cautions in directly applying theories of face-to-face communication to computer-based interaction.

RELATED WORK

Eye contact is an invitation to communicate. It clearly signals a person's "availability" for communication and usually produces positive perceptions in receivers. Goldberg, Kiesler, and Collins [25] found that people who spent more time gazing at an interviewer received higher socio-emotional evaluations. Increased eye contact was also associated with greater perceived dynamism, likability, and believability [4]. Burgoon [10] studied differential gaze behavior (e.g. nearly constant, normal or nearly constant aversion) and found that they resulted in varied impressions of attraction, credibility, and relational communication, with gaze aversion producing consistently negative effects. Gaze also serves to facilitate the learning process and enhance task performance. During instruction, gaze helps learning, in that college students had higher performance on a learning task when the instructor gazed at them than when the instructor did not [21] [4] also showed that students learned better when looked at more by a virtual teacher. Mutlu [42] found that increased gaze from a storytelling robot facilitated greater recall of story events.

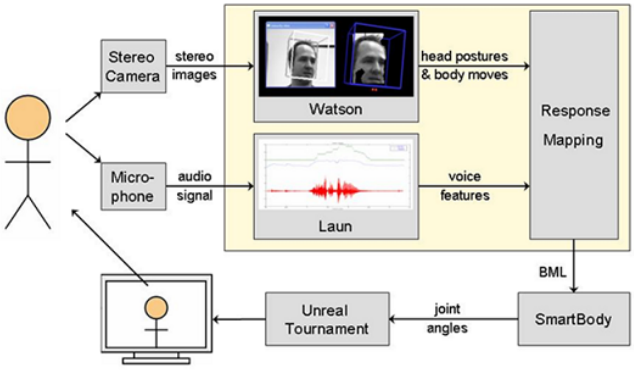
The amount of eye contact in a human-human encounter varies widely. Argyle [1] found that in dyadic conversations, the listener spent an average of about 75% of the time gazing at the speaker. Kendon [33] reported that a typical pattern of interaction when two people converse with each other consists of the listener maintaining fairly long gazes at the speaker, interrupted only by short glances away.

Gaze modeling for the virtual agent has been researched extensively in the area of turn management [8] [12] [14] [56] [47] and indicating object of interest [8] [37] [55]. Evaluations of the effects of gaze on the quality of interactions in mediated conversation have shown that improving the gaze behavior of agents or avatars in human-agent or human-avatar communication has noticeable effects on the way communication proceeds [54] [60] [23] [19] [28]. However, most of the evaluation of gaze in mediated communication had been with human-human conversations in video-conferencing and not, to any great extent, with conversations between human and autonomous embodied conversational agents [51].

Previous studies evaluating a virtual Rapport Agent using contingent feedback to establish rapport with human speaker showed that people who interacted with the Rapport Agent felt stronger feelings of rapport, increased engagement and improved speech fluency compared to people who interacted with agents that provided non-contingent feedback [26]. Other studies also show that people high in trait-anxiety are more engaged speaking with the Rapport Agent than they are speaking with a stranger face-to-face [31]. A more recent study showed that people who are more agreeable established more rapport with the Rapport Agent and suffered less speech disfluency [32].

VIRTUAL RAPPORT AGENT

In this paper, we continue our investigation of mutual attention using the virtual 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 [15] and back-channeling, e.g., nods [62]. 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.



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

Figure 1: Rapport Agent architecture and behavior mapping table.

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 [41] 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 [46] 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 [61]).

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 [30] using a standardized API [34]. 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.

HYPOTHESIS

Based on the literature review, we hypothesize that:

H1: Self-reported rapport will be the highest when the Rapport Agent provides feedback that reflects all three components of rapport. When only mutual attention is expressed, the level of rapport level would be lower. When none of the three components is included in the feedback, rapport would be the lowest.

H2: Human Speaker will speak most fluently when the Rapport Agent feedback reflected all three components of rapport. The speaker will speak less fluently when only mutual attention is expressed by the Rapport agent. The speaker will suffer the most speech disfluency when none of the three components is provided in the agent feedback.

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).

Design

The study adapts a common paradigm used for studying the impact of listener behavior on speech production [6] [40]. In this "quasi-monolog" elicitation, one participant, the Speaker, has previously observed some incident, and describes it to another participant, the Listener. In this study, the Listener corresponds to some experimental manipulation of the Rapport Agent.

It should be noted that restricting the study to quasi-monologs potentially limits the generality of our results but we adopt this paradigm for several reasons. First, it allows us to assess the effectiveness of our Rapport Agent which was designed to give feedback in such storytelling contexts. Although limited, several potential applications of virtual human technology correspond to quasi-monologs including survey interviewing [13], story elicitation or psychotherapeutic applications. Second, free natural language dialogue is beyond the capability of current dialogue systems but we wanted to avoid the common use of confederates or "wizard of Oz" designs. Such designs should be avoided as the confederate/wizard can unconsciously recognize the experimental manipulation and introduces biases into their own behavior [48].

Following the standard setup adopted by McNeill [40], we designed the study as having a human participant watch a short cartoon and then describe it to a listening agent. We designed three different virtual agents to play the listener role for the study. Behaviors of these three virtual agents are listed in Table 1.

Responsive	Continuous gaze, Head nod, Posture Mimicking, idle-time behavior
Staring	Continuous gaze, idle-time behavior
Ignoring	Gaze occasionally at speaker (otherwise gaze randomly about the room), idle-time behavior

Table 1: Agent behavior in three experimental conditions.

The first virtual agent is a "good virtual listener" (the "Responsive" condition). The agent exhibits attentive listening behaviors including head nods and posture mimicking that have previously been demonstrated to create self-reported feelings of rapport [26]. Agent's posture mimicking includes posture shifts (left, right, front and back) and head nods. Between these attentive listening behaviors, the agent does idle-time behaviors including blinking and random posture shifts. This agent gazes continuously at the speaker except when he is blinking and nodding. The second virtual agent, a "gaze only listener" (the "Staring" condition), gazes continuously at the speaker 100% of the time and exhibit random idle-time behaviors. Finally, the "ignoring listener"

(the "Ignoring" condition), does not maintain continuous gaze with the speaker (it gazes randomly about the room and occasionally gaze at the speaker) and exhibit random idle-time behaviors. This agent's gaze behavior is shown at the bottom of Figure 2.

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

Figure 2: The gaze behavior of the Rapport Agent. On the top is the gaze behavior of the agent in the Responsive and Staring condition. At the bottom is the gaze behavior of the agent in the Ignoring condition.

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. The speaker was also told that the listener will later retell the story to the camera. 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. The size of the agent is approximately the same size of the human listener sitting 8 feet away. 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 saw before interacting with the virtual agent. During this time, the listener (the confederate) remained in the computer room and pretended to speak to the camera what he/she had been told by the speaker so that the participant would not suspect that the listener is a confederate.

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.

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.

Measures

Rapport Scale

We constructed a 10-item rapport scale (coefficient alpha = .89), presented to speakers in the post-questionnaire packet. 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."

Listener Focus, Distraction, Agent Naturalness

For listener focus and distraction scale, we constructed 2 items for each scale, with Cronbach's alpha coefficient of .70 and .71, respectively. We also constructed a 6-item agent naturalness scale, with Cronbach's alpha coefficient of .94. All the scales were measured with an 8 point metric (1 = Disagree Strongly; 8 = Agree Strongly). These three

scales were issued to speakers in the post-questionnaire packet. These scales indexed how much the speaker paid attention to the listener, how distracting the listener's feedback was, and how natural the agent appeared to be, respectively.

Pre-questionnaire Packet

Participants also completed a pre-questionnaire packet that contains questions about one's demographic background, personality [27], self-monitoring [34], self-consciousness [46] and shyness [14]. 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 embarrassment, speaker's goals while explaining the video and listener traits. Listener's traits were measured using items such as "likeable", "tense" and "trustworthy" taken from the dependent measure from Krumhuber study [33]. Scales from Krumhuber [33] range from 0 (not at all) to 8 (very). Other scales ranged from 1 (disagree strongly) to 8 (agree strongly).

RESULT

Data from 11 participants were excluded due to equipment failure during the experiment and missing data. As a result, data from 133 sessions were included in the analysis, 48 in the Responsive condition, 41 in the Staring condition and 44 in the Ignoring condition.

Self-report Analysis

We first conducted Univariate Analysis of Variance tests on the self-report items. The tests showed that the between subjects effect of the experiment condition is statistically significant ($p<.05$) for the variables in Table 1.

Variable	df	F	p
Rapport	2	12.71	<.001
Agent Naturalness	2	1.74	.180
Distraction	2	5.53	.005
Listener Focus	2	3.25	.042
Tense	2	3.26	.042
Disfluency Freq (per minute)	2	9.06	<.001

Table 2: Main effect results of Univariate Analysis of Variance tests on the self-report items and speech disfluency.

We then conducted a Post Hoc analysis on each of these variables to how the conditions different from one other. Table 3 summarizes the means of various self-report and behavior measures. Items sharing the same superscripts are significantly different from each other in the Post Hoc tests. For example, in each row, items sharing superscript "a" are statistically significantly different from each other and items sharing superscript "b" are statistically significant from each other.

From Table 3 we can see that, Hypothesis 1 is partially supported. Participants interacted with the Responsive agent reported significantly higher level of rapport than those interacted with the Staring and Ignoring agent. However, there was no significant difference between Staring and Ignoring condition.

Conditions	Responsive	Staring	Ignoring
Rapport	$^{a, b}$4.61	^{a}3.49	^{b}3.34
Agent Naturalness	4.39	4.42	4.27
Distraction	$^{a, b}$3.76	^{a}4.85	^{b}4.59
Listener Focus	^{a}5.23	^{a}4.52	4.69
Tense	^{a}2.02	^{a,b}2.95	^{b}2.18
Disfluency Freq (per minute)	$^{a, b}$22.44	^{a}36.75	^{b}30.27

Table 3: Results of Post Hoc analysis of self-report and speech disfluency. In each row, items sharing the same superscripts (e.g. a, b) are significantly different from each other. For example, in each row, items sharing superscript "a" are statistically significantly different from each other and items sharing superscript "b" are statistically significant from each other.

Participants from all conditions did not differ significantly on their evaluation of the naturalness of the agent's appearance and behavior. However, participants found the agent in the Staring and Ignoring condition more distracting than the one in the Responsive condition. Participants found the Staring agent as distracting as the Ignoring agent.

We asked participants how much do they think they focused on the listener (the agent) when they interacted with him. Participants focused significantly more to the Responsive agent than the Staring agent, according to their self-report. Interestingly, participants that interacted with the Staring agent rated the human listener more tense than those interacted with the Responsive and Ignoring agent.

Speech Fluency Analysis

We annotated three types of speech disfluencies from the interaction: pause fillers (e.g. "um" and "er"), prolonged words (e.g. "I li::ke it", where ":" signifies lengthened vow "I") and incomplete words (e.g. univers-). Since the interactions are of various length (M=3.91, *min*=1.27, *max*=8.72, *std*=1.43, in minutes), we divided the sum of the three types of disfluencies by duration and defined it as the *disfluency frequency* scale. From Table 3, we can see that Hypothesis 2 is partially supported. Participants in the Responsive condition spoke with less disfluencies than participants from the Staring and Ignoring condition. Again, there was no significant difference between the Staring and Ignoring condition.

Gaze Analysis

In addition to speech disfluency, speaker's gaze pattern could also be an important behavior measure of the agent's feedback. For example, long duration of speaker gaze may indicate interest in the agent's feedback. Change of gaze duration may reveal how interpretation of agent's behavior

evolves over time. Thus, we annotated the instances when participants gazed at the agent during their interaction. On the overall percentage of time the participants gazed at the agent, ANOVA test show that there was no significant difference between the Responsive, Staring and Ignoring conditions ($M_{Responsive}$=.346, $M_{Staring}$=.305, $M_{Ignoring}$=.30, p=.599). However, the means shows a trend that participants in the Responsive condition spent more percentage of time gazing back at the agent than the ones in the Staring and Ignoring condition. And the gaze percentage from the Staring condition and Ignoring condition are almost identical. An earlier study of human-human interaction in similar experiment setup showed that human speaker spent about 65% of the time gazing at a human listener ($M_{Face\text{-}to\text{-}Face}$=.645).

We divided each interaction into 5 sections evenly to analyze how gaze duration changes over time. Since each participant interacted with the agent twice, once explaining the Tweety and Sylvester cartoon and once explaining the Sexual Harassment Prevention Training video, we divided each interaction into 5 sections. For each section, we then calculated the percentage of time the participant gazed at the agent. Figure 3 shows how the average of gaze duration changes from the first interaction to the second interaction. Note that first (or second) interaction is defined as the first (or second) time the participant interacted with the agent, regardless which video they are describing.

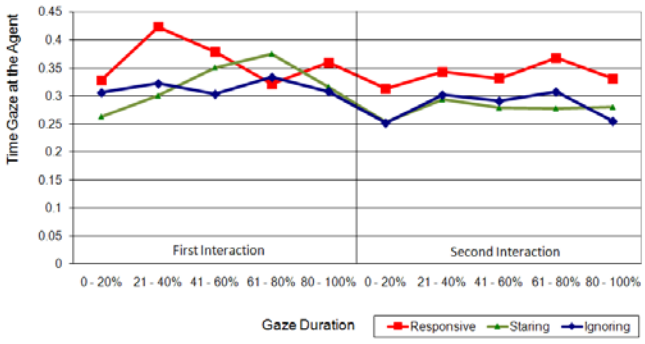


Figure 3: Percentage of time participants gaze at the agent changes during the interaction.

We then conducted a General Linear Model (GLM) Repeated measure test to analyze the change over time. The result shows that the main effect of condition (Responsive, Staring and Ignoring) is not significant (df=2, F=.674, p=.512). This means that the overall, the average percentage of gaze duration does not differ significantly between the three conditions. However, the ANOVA test of gaze duration does show trend that participants in the Responsive condition spent more time gazing back at the agent than the ones in the Staring and Ignoring condition, while the Staring and Ignoring condition did not differ. The overall gaze duration change over time is significant (df=9, F=2.55, p=.007). And the interaction between time and condition is also not significant (df=18, F=.868, p=.013). This means that gaze duration changed differently over time in the three conditions. From Figure 3 we can see that, gaze duration in the Responsive condition remains relatively high throughout the two interactions, while gaze durations in the Ignoring condition remains low. In the Staring condition, gaze duration is high toward the beginning and slowly decreases over time to the level similar to the Ignoring condition.

DISCUSSION AND CONCLUSION

In this paper, we explored the psychological construct of rapport as an explanatory construct for guiding the design of virtual agents and avatars. Consistent with the predictions of Tickel-Degnen and Rosenthal [57], participants experienced more rapport when the virtual representation of their conversation partner showed more attention, positivity and coordination: participants interacting with the Rapport Agent had greater subjective experience of rapport and exhibited more fluent speech when compared to an agent that only exhibited attention (Staring Agent) or an agent that exhibited none of the constituents of rapport (Ignoring Agent). This is consistent with Hypothesis 1 and 2. Unexpectedly, gazing alone (without positivity or coordination) had surprisingly strong negative impact on user performance: an agent that simply stares is just as bad as an agent that conveys disinterest in terms of creating distractions, reducing rapport and disrupting speech production, despite the fact that all agents were perceived as equally natural. This finding is significant in that many avatars and virtual agents convey attention to the user by staring.

Although inconsistent with the model of Tickle-Degnen and Rosenthal, the negative effects of staring are consistent with some theoretical perspectives on human interaction. Several theories of nonverbal interaction emphasize that, although gaze communicates interest and intimacy, this intimacy may not be desired and its expression can have negative consequences in certain circumstances. For example, according to discrepancy-arousal theory, the expression of nonverbal intimacy in face-to-face conversations can reduce participant gaze and create feelings of discomfort or embarrassment if the rate or magnitude of these nonverbal cues differ from what is expected (see O'Conner and Gifford [45] for a review of this and related models). To assess if negative arousal played some role, we asked the participants to evaluate how uncomfortable they were when interacting with the agent (the embarrassment scale) in the post-questionnaire packet. ANOVA test shows no significant difference between the three conditions. However, the means show a trend that participants in the Staring condition felt more uncomfortable than participants in the Responsive and Ignoring condition ($M_{Responsive}$=3.35, $M_{Staring}$=3.87, $M_{Ignoring}$=3.30).

A completely different explanation is offered by theories of conversational grounding [17]. According to this research, speakers in a conversation expect frequent and incremental feedback from listeners that their communication is understood. When listeners provide grounding feedback, speech can proceed fluently and presumably with a greater sense of

rapport. Such feedback can take the form of nods (corresponding to Tickle-Degnen's positivity dimensions) but also complex patterns of making and breaking gaze [6]. Indeed, Nakano [43] found that listener staring can be interpreted by speakers a failure to establish mutual ground. Thus, the poor performance of speakers in the Ignore and Staring conditions, according to grounding theory, can be explained by the failure of these agents to produce grounding cues. Further research will attempt to distinguish between these theoretical perspectives.

Males and females can respond differently to nonverbal behaviors, particularly in the case of eye gaze cues [2]. Studies in Computer Mediated Communication have shown gender differences on interpretation of gaze and presence [7]. In this study, we did not find any significant main or interaction effect of participant's gender on self-reported and behavior variables. This could partly due to the mismatch of the confederate (the human listener) and the virtual agent's gender. The participants were led to believe that the virtual agent's behavior was controlled by the human listener (the confederate). But the confederate of the study was female and the agent was male. However, in our prior studies, when human listener and the virtual agent's gender were matched, we didn't find any significant effect of gender.

Harrigan [27] studied the nonverbal behavior of high/low rapport doctors and found that high rapport doctors engaged in less extensive eye-contact than low rapport doctors. In their study, the low rapport doctors maintained mutual gaze with the patient throughout 85% of interaction. The high rapport doctors maintained mutual gaze 70% of the time. However, Tickle-Degnan and Rosenthal [57] later pointed out that, in a non-helping context (e.g. interacting with an interviewer, new acquaintances), as oppose to the helping context (e.g. meeting with doctors), directed gaze is positively correlated with participant's evaluative impression. Gaze behavior can have different social implications in different social context. The findings presented here are observed in a monolog setting. Further studies need to be conducted to better understand how they generalize to contexts where agents/avatars can provide other forms of feedback.

The work presented here has intriguing implications to the design of agents and avatars in the collaborative virtual environments and virtual worlds in general. Thanks to the recent technology advancements, collaborative virtual environments (CVE) where user can interact and collaborate via avatars in 3D worlds have become more and more common. CVEs are already used in a variety of different fields: gaming (e.g. World of Warcraft) [9], business (e.g. Second Life) [38], education (e.g. MOVE, Whyville) [24] [44], social communication (e.g. There) [39], and cooperative development (Workspace 3D) [58]. In these virtual environments, people gather online, design virtual avatars to represent themselves and interact with other virtual avatars in the virtual world. State of the art of the social behavior of the avatar is to be desired. For example, even though in some of the virtual worlds, user could choose the nonverbal behavior accompanies the verbal signals, the default listening behavior of the avatar is gazing directly at the user or accompanied by random posture shifts. Results from this paper show that during social interactions, including the ones in the virtual worlds, simply use directed gaze to establish mutual attention is not only not enough to create positive impression but could have negative social effect. There are other nonverbal behaviors that can build mutual attention. For example, forward lean and orienting body towards one another. However, regardless which nonverbal behavior is used to express mutual attention, it is not that agents and avatars should not maintain eye contact with the user, but mutual attention should be accompanied by behaviors that indicate positivity and/or coordination to create positive interaction.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 0713603. This work was also 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. Argyle, M. *Bodily Communication (Second Edition)*. Routledge, Oxford, 1988.
2. Argyle, M. and Cook, M. *Gaze and Mutual Gaze*. Cambridge University Press, Cambridge, 1976.
3. Bailenson, J.N., Yee, N., Blascovich, J., Beall, A.C., Lundblad, N., and Jin, M. The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences*, (2008) 17, 102-141.
4. Beebe, S. A. Effects of eye contact, posture and vocal inflection upon credibility and comprehension. *Australian SCAN: Jounal of Human Communication*, (1980) 7-8, 57-70
5. Bevacqua, E., Mancini, M. and Pelachaud, C. A listening agent exhibiting variable behaviour, in *8th International Conference on Intelligent Virtual Agents*, (Tokyo, 2008), Springer, 262-269.
6. Bavelas, J. B., Coates, L. and Johnson, T. Listener responses as a collaborative process: The role of gaze. *Journal of Communication*, (2002) 52, 566–580
7. Bente, G., Eschenburg, F. and Aelker, L. Effects of simulated gaze on social presence, person perception and personality attribution in avatar-mediated communication, *in 10th Annual International Workshop on Presence*, (Barcelona, 2007)

8. Beskow, J. Animation of talking agents., in *Proceedings of the ESCA Workshop on Audio-Visual Speech Processing*, (1997) 149—152.
9. Blizzard: http://www.wow.com
10. Burgoon, J. K., Coker, D. A., and Coker, R. A. Communicative effects of gaze behavior: A test of two contrasting explanations. *Human Communication Research*, 1986 (12), 495-524.
11. Burns, M., Rapport and relationships: The basis of child care. *Journal of Child Care*, 1984. 2, 47-57.
12. Cassell, J., Bickmore, T.W., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H.H., and Yan, H. Embodiment in Conversational Interfaces: Rea. In *Proceedings of Computer Human Interaction*, 1999, 520—527.
13. Cassell, J. & Miller, P. Is it self-administration if the computer gives you encouraging looks? In Conrad, F.G. & Schober, M.F. Eds. *Envisioning the Survey Interview of the Future*, Wiley, Hoboken, NJ, 2008, 161-178.
14. Cassell, J., Torres, O. and Prevost, S. Turn Taking vs. Discourse Structure: How Best to Model Multimodal Conversation. In Wilks, I., editor, *Machine Conversations*. Kluwer, The Hague, 1999.
15. Chartrand, T. L., and Bargh, J. A. The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, (1999) 76, 893-910.
16. Cheek, J.M., *The Revised Cheek and Buss Shyness Scale (RCBS)*. Wellesley College, Wellesley MA, 1983.
17. Clark, H.H. and E.F. Schaefer, Contributing to discourse. *Cognitive Science*, (1989) 13, 259-294.
18. Cogger, J. W. Are you a skilled interviewer? *Personnel Journal*, (1982) 61, 840-843.
19. Colburn, R. A., Cohen, M. F., and Drucker, S. M. Avatar Mediated Conversational Interfaces. *Technical Report MSR-TR-2000-81*, Microsoft, 2000.
20. Drolet, A. L., and Morris, M. W. Rapport in conflict resolution: accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. *Experimental Social Psychology*, (2000) 36, 26-50.
21. Fry, R. and Smith, G. F. The effects of feedback and eye contact on performance of a digit-encoding task. *Journal of Social Psychology*, (1975) 96, 145- 146.
22. Fuchs, D. Examiner familiarity effects on test performance: implications for training and practice. *Topics in Early Childhood Special Education*, (1987) 7, 90-104.
23. Garau, M., Slater, M., Bee, S., and Sasse, M.A. The Impact of Eye Gaze on Communication Using Humanoid Avatars. In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, 2001, 309–316.
24. Garcia, P. Move: Component groupware foundations for collaborative virtual environments, in *4th International Conference on Collaborative Virtual Environments*, 2000, 55–62.
25. Goldberg, G. N., Kiesler, C. A., and Collins, B. E. Visual behavior and face-to-face distance during interaction. *Sociometry*, (1969) 32, 43-53.
26. Gratch, J., Wang, N., Gerten, J., Fast, E. and Duffy, R. Creating Rapport with Virtual Agents. In *7th International Conference on Intelligent Virtual Agents*, (Paris, 2007), Springer, 125-138.
27. Harrigan, J. A., Oxman, T. E., Rosenthal, R. Rapport expressed through nonverbal behavior. *Journal of Nonverbal Behavior* (1985) 9(2).
28. Heylen, D.K.J. and van Es, I. and Nijholt, A. and van Dijk, E.M.A.G. Controlling the Gaze of Conversational Agents, in *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*. Kluwer Academic Publishers, 2005, 245-262.
29. John, O.P. and S. Srivastava, The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, (1999), 2, 102–138.
30. Kallmann, M., & Marsella, S. Hierarchical Motion Controllers for Real-Time Autonomous Virtual Humans. In *5th International Conference on Intelligent Virtual Agents*, (Kos, 2005), Springer, 253-265.
31. Kang, S-H, Gratch, J., Wang, N., Watt, J. Does Contingency of Agents' Nonverbal Feedback Affect Users' Social Anxiety? in *7th International Conference on Autonomous Agents and Multiagent Systems*. (Estoril, 2008), 120-127.
32. Kang, S-H, Gratch, J., Wang, N., Watt, J. Agreeable People Like Agreeable Virtual Humans, *in 8th International Conference on Intelligent Virtual Agents*, (Tokyo, 2008), 253-261.
33. Kendon, A. Some functions of gaze direction in social interaction. *ACTA PSYCHOLOGICA*, (1967), 26, 22-63.
34. Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., et al. Towards a common framework for multimodal generation in ECAs: The behavior markup language. In *6th International Conference on Intelligent Virtual Agents*, (Marina del Rey, 2006), 28-41.
35. Krumhuber, E., Cosker, D., Manstead, A. S. R., Marshall, D., & Rosin, P. L. Temporal aspects of smiles influence employment decisions: A comparison of human and synthetic faces, in *11th European Conference Facial Expressions: Measurement and Meaning*, (Durham, United Kingdom, 2005).
36. Lennox, R.D. and R.N. Wolfe, Revision of the Self-Monitoring Scale. *Journal of Personality and Social psychology*, (1984). 46: p. 1349-1364.
37. Lester, J.C., Stuart, S.G., Callaway, C.B., Voerman, J.L., and Fitzgeral, P.J. Deictic and emotive communica-

tion in animated pedagogical agents. In S. Prevost, J. Cassell, J. Sullivan, and E. Churcill, eds, *Embodied Conversational Characters*. MIT Press, Cambridge, MA, 2000.

38. Linden Lab: http://secondlife.com
39. Makena Technologies: http://www.there.com
40. McNeill, D. Hand and mind: *What gestures reveal about thought*. The University of Chicago Press, Chicago, 1992.
41. Morency, L.-P., Sidner, C., Lee, C., and Darrell, T. Contextual Recognition of Head Gestures. In *7th International Conference on Multimodal Interactions*, (Toreno, 2005), 18-24.
42. Mutlu, B., Hodgins, J.K., and Forlizzi, J. A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, (Genova, 2006), 518-523.
43. Nakano, Y. I., Reinstein, G., Stocky, T., and Cassell, J. 2003. Towards a model of face-to-face grounding. In *41st Annual Meeting on Association For Computational Linguistics - Volume 1* (Sapporo, 2003). Association for Computational Linguistics, Morristown, NJ, 553-561.
44. Numedeon: http://www.whyville.net
45. O'Connor, B., and Gifford, R. A test among models of nonverbal intimacy reactions: Arousal-labeling, discrepancy-arousal, and social cognition. *Journal of Nonverbal Behavior*, (1988), 12, 6-33.
46. Oppenheim, A. V., and Schafer, R. W. From Frequency to Quefrency: A History of the Cep-strum. *IEEE Signal Processing Magazine*, (2004) 9, 95-106.
47. Pelachaud, C. and Bilvi, M. Modelling gaze behavior for conversational agents. In *4th International Conference on Intelligent Virtual Agents*, (Kloster Irsee, 2003), Springer.
48. Rosenthal, R. Interpersonal expectancy effects: A 30-year perspective. *Current Directions in Psychological Science*, (1994) 3, 176-179
49. 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, 687-699.
50. Short, J.A., Williams, E., & Christie, B. *The social psychology of telecommunications*. New York: John Wiley & Sons, 1976.
51. Swaab, R. and Swaab, D. Sex differences in the effects of visual contact and eye contact in negotiations, *Journal of Experimental Social Psychology* (2009) 45 (1) 129-136.
52. Tatar, D. Social and personal consequences of a preoccupied listener. Department of Psychology. Stanford, CA, Stanford University: Unpublished doctoral dissertation (1997).
53. Thompson, L., & Nadler, J. Negotiating via information technology: Theory and application. *Journal of Social Issues*, (2002), 58, (1), 109-124.
54. Thórisson, K. R. and Cassell, J. Why Put an Agent in a Body: The Importance of Communicative Feedback in Human-Humanoid Dialogue. *In Proceedings of Lifelike Computer Characters*, (1996) 44–45.
55. Thórisson, K. R. Layered modular action control for communicative humanoids. In *Computer Animation*. IEEE Computer Society Press, Geneva, Switzerland, 1997.
56. Thórisson, K. R. Natural turn-taking needs no manual. In I. Karlsson, B. Granstr¨om, and D. House, eds, *Multimodality in Language and speech systems*, . Kluwer Academic Publishers, 2002, 173-207.
57. Tickle-Degnen, L., Rosenthal, R. The nature of rapport and its nonverbal correlates, *Psychological Inquiry*, (1990) 1, 285-93.
58. Tixeo Soft: http://www.workspace3d.com
59. Tsui, P., & Schultz, G. L. Failure of Rapport: Why psychotheraputic engagement fails in the treatment of Asian clients. *American Journal of Orthopsychiatry*, (1985) 55, 561-569.
60. Vertegaal, R. The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration, in *Proceedings of Conference on Human Factors in Computing Systems*, (Pittsburgh, 1999), ACM, 294–301,
61. Ward, N., & Tsukahara, W. Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics*, (2000) 23, 1177-1207.
62. Yngve, V. H. On getting a word in edgewise, in *6th regional Meeting of the Chicago Linguistic Society*, 1970.