Virtual Rapport 2.0

Lixing Huang[1], Louis-Philippe Morency[1], Jonathan Gratch[1]

[1]Institute for Creative Technologies, University of Southern California,
12015 Waterfront Drive, Playa Vista, CA 90094, USA,
{lhuang, morency, gratch}@ict.usc.edu

Abstract. Rapport, the feeling of being "in sync" with your conversational partners, is argued to underlie many desirable social effects. By generating proper verbal and nonverbal behaviors, virtual humans have been seen to create rapport during interactions with human users. In this paper, we introduce our approach to creating rapport following Tickle-Degnen and Rosenberg's three-factor (positivity, mutual attention and coordination) theory of rapport. By comparing with a previously published virtual agent, the Rapport Agent, we show that our virtual human predicts the timing of backchannel feedback and end-of-turn more precisely, performs more natural behaviors and, thereby creates much stronger feelings of rapport between users and virtual agents.

Keywords: Rapport, Virtual human, Positivity, Mutual attention, Coordination

1 Introduction

You feel the connection and harmony with your partner when you are engaged in a good conversation. This phenomenon, formally known as *rapport*, has been studied extensively in social psychology. Rapport is argued to underlie successful negotiation [1], improved quality of child care [2], social engagement [3], and success in teacher-student interactions [4]. Tickle-Degnen and Rosenthal [5] argued that rapport is created through behaviors indicating positive emotions (such as head nods or smiles), mutual attention (such as mutual gaze), and coordination (such as postural mimicry or synchronized movements). They further claimed that as the friendship between two conversants deepens, the importance of positivity decreases, while the importance of coordination increases. Along these lines, Cassell et al. [6] divided rapport into short-term and long-term, where short-term rapport focuses on building instant rapport, while long-term rapport models the unfolding of both verbal and nonverbal behaviors over the course of a relationship. Here we consider the former.

The power of rapport in social interactions has inspired researchers in human-computer interaction and a number of virtual agents have been motivated by these findings. For example, Bailenson et al. [7] showed that a virtual agent was more persuasive and better liked if it mimicked a human speaker's head movement. Bickmore et al. [8] developed an animated agent with text-based dialogue generation that performed nonverbal behaviors such as hand gestures, head nods, eye gaze movements and facial displays of emotion. Their pilot evaluation study showed that the agent promotes antipsychotic medication adherence among patients with schizophrenia. The recent SEMAINE project built the Sensitive Artificial Listener [9]. By exhibiting different styles of audiovisual listener feedback, the listener is able to express four different personalities.

In our previous work "Virtual Rapport" [12], we introduced an intelligent virtual agent, the Rapport Agent. In a series of subsequent studies by us and outside collaborators, we demonstrated that the Rapport Agent could induce the subjective feeling and many of the behavioral benefits of the psychological concept of rapport. The Rapport Agent has proved a valuable tool for advancing Intelligent Virtual Agent (IVA) research, both by demonstrating that virtual agents have important social effects on human users [10,15,16], and by illuminating the factors that contribute towards or sometimes undermine these social consequences [12,15-17].

Although the Rapport Agent clearly influences human users in important ways, it is less clear how well it is performing at this task, and there is growing evidence in our subsequent subjective experiments that it falls well-short of the potential IVA's hold for shaping human behavior. Several lines of evidence highlight shortcomings to this system: participants give the system mediocre ratings with respect to subjective measures of rapport and social presence [11]; it generally underperforms human users in terms of subjective and behavioral measures [29][1]; and our research on data-driven methods for behavior generation suggests the Rapport Agent's hand-crafted algorithms and animations could be considerably improved [31].

In this article, we will return to the original motivation for the Rapport Agent – Tickle-Degnen and Rosenberg's three-factor theory of rapport – and illustrate how subsequent research has illustrated ways to enhance the positivity, mutual attention and coordination of systems like the Rapport Agent. In particular, we will emphasize the importance of data-driven methods for behavior generation. After reviewing the benefits and limitations of the Rapport Agent, we will introduce a new system designed to enhance the subjective and objective measures of rapport. In a head-to-head comparison, 90% of participants prefer this new system and rate it almost twice as good as the original Rapport Agent along a number of measures of rapport. Our hope is that this new approach will be even more useful as a tool for demonstrating the important benefits of intelligent virtual agents.

2 Virtual Rapport 1.0

The Rapport agent was designed to establish rapport with human participants by providing contingent nonverbal feedback while a participant speaks. The initial system focused on a "quasi-monolog" paradigm, where a human speaker (the narrator) retells some previously observed series of events (e.g., the events in a recently-watched video) to a non-speaking but nonverbally attentive agent [11,12]. More recently, we have extended the system to engage in more interactive dialogs, such as acting as an interviewer [10].

[1] although some subgroups – e.g. shy users – seem to prefer the animated agent [29]

Table 1. Rapport Agent Behavior Mapping Rules

Human Speaker Behavior	Rapport Agent Response
Lowering of pitch	Head nod
Raised loudness	Head nod
Speech disfluency	Posture/gaze shift
Shift posture	Mimic
Gaze away	Mimic
Head nod/Head shake	Mimic

In designing the Rapport Agent, we extracted a small number of simple rules (as shown in Table 1) from social science literature. To produce listening feedback, the agent first collects and analyzes the speaker's upper-body movement and voice. To detect features from the participants' movement, it uses Waston [13] to track the head position and orientation. With the head tracking data, it can detect head gestures, posture shifts and gaze direction. Acoustic features are derived from properties of the pitch and intensity of the speech signal, using a signal processing package, LAUN. The recognized speaker's features are then mapped to reactions through a set of authorable mapping rules. These reaction animations are passed to the SmartBody [14] animation system using Behavior Markup Language (BML); and finally, the animations are rendered by a commercial game engine and displayed to human users. The animations that the Rapport Agent can perform are relatively simple, such as two continuous nods with equal amplitude and posture shifts. When used in an interview setting, the Rapport Agent steps through a series of predefined questions, taking its turn either when indicated by a human controller [10], or more recently after waiting for a 1.5s pause - the timing based on an analysis of data collected in previous studies [24].

2.1 Benefits

The Rapport Agent has been applied in a series of empirical studies to investigate how people are influenced by such computer-generated behaviors. In these studies, human participants sit in front of the Rapport Agent and are prompted to either retell some previously experienced situation (monologue) or interviewed by the agent to answer some predefined questions (interview). After the interaction, participant rapport is assessed by a variety of subjective and behavioral measures. These studies showed that by interacting with the Rapport Agent, people have: greater feelings of self-efficacy [10], less tension [15] and less embarrassment [10], greater feelings of rapport [15], a greater sense of mutual awareness [16], and greater feelings of trustworthiness [10].The contingent nonverbal feedback of the Rapport Agent also changes participants' behavior. Behavioral effects include: more disclosure of information including longer interaction time and more words elicited [11,12,15,16], more fluent speech [11,12,15,16], more mutual gaze [15] and fewer negative facial expression [17].

2.2 Limitations

Although it has been demonstrated effective in many studies, the current models and behaviors of the Rapport Agent have limitations with regard to the three-factor theory of rapport.

Mutual Attention and Coordination: Tickle-Degnen and Rosenthal emphasize that, with rapport, participants fall into a cohesive, unified pattern of behavior arising through close attention to and tight-coordination of nonverbal signals. The Rapport Agent attempts to realize these two factors by attending to human nonverbal cues (e.g., gestures and prosodic signals) and utilizing them to coordinate its responses (such as backchannel feedback and turn-taking). However, there are reasons to suspect the agent's attention and coordination could be significantly improved. Like many virtual agents, the Rapport Agent's behavior is driven by general rules derived from social science literature, in contrast to more recent approaches [13,22] that attempt to learn behaviors directly from large datasets. Although based on human-behavioral studies, such "literature-based" rules are often intended to make general theoretical points rather than to drive behaviors. Further, such rules are often generated in a variety of social contexts that may differ considerably from the situations to which the Rapport Agent has been applied. Consequently, such rules are unlikely to capture the subtlety in both timing and realizations of nonverbal behaviors.

Positive Emotion Communication: A third component of rapport relates to sense of emotional alignment and positivity that participants experience in the course of rapportful interactions. Nonverbally, this feeling arises from the positive and empathetic expression of emotion. Other research on rapport has emphasized the equal importance of verbal expressions of emotion, for example, through the reciprocal self-disclosure of hopes and fears [25]. Thus, a clear limitation of the Rapport Agent is its inability to engage in emotional communication, both verbally and nonverbally; a point highlighted in some of the evaluations of the system [17].

3 Virtual Rapport 2.0

Virtual Rapport 2.0 improves over the previous work by directly addressing the main limitations of the Rapport Agent. We enhance mutual attention and coordination by applying data-driven approach to build context-specific (i.e. the same context where the virtual human is deployed) response models, which better model the subtlety of timing and realization of nonverbal behaviors. By integrating affective information and strengthening reciprocity, we also enable the virtual human to communicate positive emotions both verbally and nonverbally.

3.1 Enhanced Mutual Attention and Coordination: Data-driven Approach

To enhance mutual attention and coordination, we learn models to predict backchannel and turn-taking opportunity points from the human behavior observed in the same dyadic conversation settings in which the Rapport Agent is intended to be used. By using such "contextually-appropriate" data, and employing more sophisticated techniques than are typically used in the social sciences, we expect to better model the subtlety and variability in both timing and realizations of the nonverbal behaviors.

To collect human behavior data for the response models, we adopt the method of Huang and colleagues [20]. Participants are guided to interact with media representations of people parasocially so that it is possible to gather multiple different views on the same interaction, which are later combined to build the consensus view of how a typical response would be. It is showed in [20] that the resulting *parasocial consensus data* generates better virtual human behavior than actual human behavior does.

Backchannel Prediction Model.

A backchannel is a kind of feedback within face-to-face interactions that signals a person's attention and interest. It is usually expressed via head nods or paraverbals like "um-huh". By giving backchannel feedback appropriately, the virtual human creates the feeling of mutual attention and coordination. Our backchannel prediction model predicts both when to give feedback and how to display it (i.e. the realization of the head nod).

We build a probabilistic backchannel prediction model based on the parasocial consensus data, which is collected using the Rapport 06-07 data set[2]. Inspired by previous work [21], we used pause and eye-gaze (i.e. looking at the listener) in our model. The model fuses multimodal features at the early stage and captures the long-range dependency between them by applying the encoding dictionary technique [21]. A Conditional Random Field (CRF) model [27] is learned to find the mappings between speaker's nonverbal behaviors and the feedback time. Forty-five video sequences (from the Rapport 06-07 data set) are used in the training stage and the best regularization factors of the CRF model are found by applying 4-fold cross-validation. Conventional CRF uses the forward-backward inference engine that requires the full sequence (i.e. offline processing), which is not applicable for real-time predictions. Instead, we implement real-time CRF using the forward-only inference [30] so that it can make predictions in real-time. The output of CRF indicates the likelihood of giving backchannel feedback. By setting a threshold (based on a preliminary study), we predict the time of feedback by comparing the output with the threshold.

The CRF model predicts when to give feedback and how to give such feedback is learned from actual listeners' behavior. We found the typical styles of head nods from the listeners' behavior in the Rapport 06-07 data set. First, the listeners' head positions were tracked by Watson [13] and converted to frequency domain by Fast Fourier Transform. Then K-means (k=3) was applied to cluster all head nods to find typical styles, which are implemented in Behavior Markup Language (BML):

- *Small and continuous head nod*: four continuous small nods with decayed amplitudes and speed;

[2] Datasets are available for research purpose at rapport.ict.usc.edu

- *Normal nod*: two continuous head nods with decayed amplitudes and equal speed;
- *Single nod*: one slow head nod.

In the current implementation, we randomly choose one of the three styles when it is proper to give backchannel feedback.

End-of-Turn Prediction.

Within a dyadic conversation, the roles of speaker and listener are regulated seamlessly by a negotiation process of turn-taking. A smooth turn-taking strategy without long mutual silence and interruption increases the feeling of coordination.

Previous work [19] suggests that pause in speech is an important cue for when to take the turn, but that the amount of time people wait before jumping in can vary considerably depending on the speaker's nonverbal signals. The Rapport Agent, when applied to interview settings, uses a fixed turn-taking strategy: it takes the turn whenever a speaker pauses for more than one and a half seconds. Instead, we construct a multimodal end-of-turn prediction model which takes advantage of visual information such as eye-gaze and nod. The model is based on the consensus data from the Self-Disclosure data set[2]. We analyze the co-occurrence pattern between the turn-taking behavior and human speakers' nonverbal features and how these nonverbal features influence the pause duration before taking a turn, and build the model as follows:

(1) When the pause duration is longer than 1.5s and the speaker has been looking at the virtual human for more than 1.0s, it is a turn-taking place;

(2) When the speaker is looking away, the virtual human will wait until the speaker looks back and then go to (1);

(3) When head nod co-occurs with a longer-than-1.5s pause, the virtual human will take the turn 200ms after the end of the head nod.

After the virtual human takes a turn, the human speaker is allowed to interrupt him. The virtual human will stop and yield his turn to the human speaker by saying "I'm sorry, keep going" with a regretful facial expression.

3.2 Enhanced Positive Emotion Communication: Affective Response and Reciprocity

The feeling of positivity, which is important in establishing rapport in initial encounters, can be enhanced by communicating positive emotions both nonverbally and verbally.

Facial expression is an important channel to convey positive emotion nonverbally. To generate proper affective responses, the visual feature detector (a confidential commercial product) of the virtual human tracks the facial feature points of the human speaker in real-time, from which it infers the level of smiling (continuous value from 0 to 100). By setting the threshold to 50, we can reliably determine whether the human speaker is smiling or not. When there is a backchannel opportunity and the human speaker happens to smile at the same time, the virtual human will display backchannel feedback with a smiling face.

Recent research by Kang et al. [28] has emphasized some simple strategies for conveying positive feelings verbally. In her study, the interviewee discloses more intimate information if the interviewer (virtual human) discloses itself first. The mutual self-disclosure, or reciprocity, positively affects the human user's social attraction to the virtual human. In our system, we follow the same strategy of strengthening reciprocity. Before the virtual human asks its human partner questions, he will first disclose the information about himself; that is, sharing some of his autobiographical back story. For example, instead of simply asking "how old are you?", the virtual human says "I was created about three years ago. How old are you?".

3.3 System Architecture

The system (as shown in Figure 1) consists of three main parts: (1) perception, which detects the audiovisual features of human speakers in real-time; (2) response models, which predict the timing of backchannel feedback and end-of-turn and the affective state; and (3) generation, which animates the virtual human's behaviors such as head movements and facial expression.

Perception: The four main audiovisual features extracted in real-time are silence, head nods, eye-gaze (looking at listener or not) and smile. The audio feature detector extracts intensity from the raw signal every 100ms using the signal processing package, Praat [18]. With intensity information, it outputs a binary feature, speech or silence, every 100ms. The visual feature detector (a confidential commercial product) tracks the position of face and facial feature points, the direction of eye-gaze and the smile level. With this information, it outputs visual features indicating the human is nodding or not, looking away or not and smiling or not.

Response Models: Based on the perceived audiovisual features, the backchannel, end-of-turn and affective models decide in real-time the most appropriate responses. All three models take advantage of the data-driven approach described in Section 3. These responses also take into account the agent state (e.g. whether the virtual human is holding the turn or not). For example, if the virtual human is holding the turn, the output from backchannel model is ignored. The backchannel model takes silence and eye-gaze as input, the end-of-turn prediction model uses features such as silence, eye-gaze, and head nod and the affective model takes smile as input.

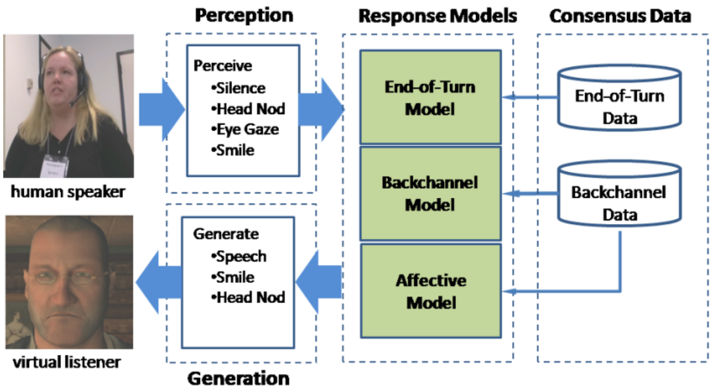

Fig. 1. System Architecture of Virtual Rapport 2.0: The perception module detects human behavior (e.g. silence in speech, nod, gaze aversion, and smile) in real-time; then the data-driven based response models take these feature as input and predict the timing of backchannel feedback and turn-taking, and the affective response; finally, the generation module generates speech and animations (e.g. smile and nod) to display to the human speaker.

Generation: The output from the response models drives the virtual human behaviors. For example, if the human speaker smiles, the virtual human will smile as well when giving the backchannel feedback. These animations are first converted to BML and then sent to an action scheduler module, which keeps track of the duration of each animation. If the current animation has not completed yet, new animations will be ignored. The BMLs are passed to the animation system, Smartbody [14], which is a virtual human animation system designed to seamlessly blend animations and procedural behaviors. Finally, animations are rendered by a commercial game engine, Gamebryo, and displayed to users.

4 Subjective Evaluation

To evaluate the performance of our virtual human (*Virtual Rapport 2.0*), we conducted a subjective evaluation to compare it with *Rapport Agent* along four dimensions: rapport, overall naturalness, backchannel feedback and end-of-turn prediction.

4.1 Experiment Design

We guided human participants to interact with both virtual humans one after the other, where the virtual human acts as an interviewer and steps through a series of questions one by one, and the human participant acts as the interviewee. For each interaction, we used different question sets derived from [25]. The order of virtual humans and question sets were randomized in the experiment. After each interaction, the human participant was asked to assess the virtual human's performance.

In a within-subject design, 21 participants were recruited to evaluate both virtual humans. Before the experiment started, the participant was required to read the instructions and ask questions about anything they do not understand. They were told "Your partner will ask you several questions and your task is to answer as best as you can. For each question, please try to answer in at least one or two sentences. You partner will listen when you answer. Please do not ask your partner questions. Your partner does not know who you are, your behavior will not be recorded and your identity will be kept anonymous". When the experiment was done, the participant was forced to choose the one s/he likes better.

The virtual human is evaluated along the four dimensions:

Rapport:

The rapport is measured by using the 5-item social presence scales suggested in [26], which ask several questions such as "I perceive that I am in the presence of another person in the room with me (1(strongly disagree) - 7(strongly agree))".

Overall Naturalness:

Do you think the virtual agent's overall behavior is natural? (1(not natural at all) - 7(absolutely natural))

Backchannel Feedback:

- *Precision*: How often do you think the virtual human generated feedback at inappropriate time? (1(all the time) - 7(never inappropriate))
- *Recall*: How often do you think the virtual human missed feedback opportunities? (1(always miss) - 7(never miss))

End-of-Turn prediction:

- *Correct time*: How often do you think the virtual human ask the next question too early? (1(always) - 7(never))
- *In time*: How often do you think the virtual human ask the next question too late? (1(always) - 7(never))

4.2 Results

The results are summarized from Figure 3(a) to 3(d). In each figure, the left bar is for Rapport Agent and the right bar is for Virtual Rapport 2.0. The star (*) means there is significant difference between the versions under the bracket.

Rapport: The answers of the five-item social presence scales are highly correlated with each other (the Cronbach's alpha is 0.9). Therefore, we average them into one scale. It is 2.6 for Rapport Agent and 3.84 for Virtual Rapport 2.0, and the difference is significant ($p<0.01$).

Overall Naturalness: The overall naturalness for Rapport Agent is 2.55, while it is 4.5 for Virtual Rapport 2.0, and the difference is significant ($p<0.01$).

Backchannel Feedback: For the precision question, the mean value of Rapport Agent is 3.6 while it is 5.25 for Virtual Rapport 2.0; for the recall question, the mean value of Rapport Agent is 3.7, while it is 4.6 for Virtual Rapport 2.0. Virtual Rapport 2.0 is significantly better ($p<0.05$) than Rapport Agent in both questions.

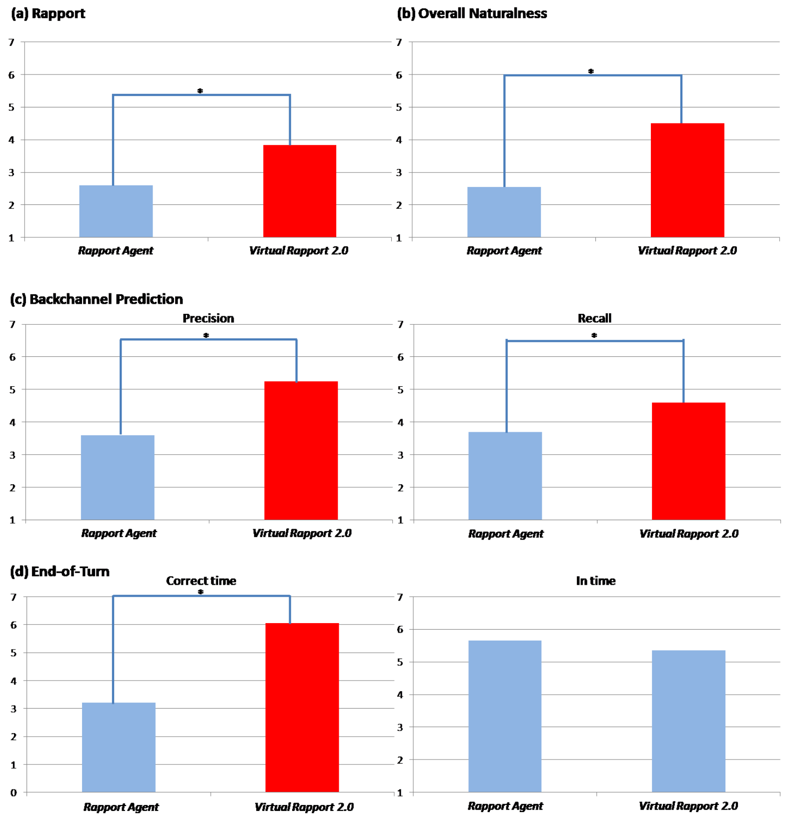

Fig. 2. The comparison of subjective evaluation results between Rapport Agent and Virtual Rapport 2.0. Virtual Rapport 2.0 is significantly better than Rapport Agent in predicting the timing of backchannel feedback (c) and end-of-turn (d); it is also significantly better than Rapport Agent in overall naturalness (b). Therefore, Virtual Rapport 2.0 creates much stronger feeling of rapport (a) than the Rapport Agent does.

End-of-Turn: For the correct time question, the mean value of Rapport Agent is 3.2, while it is 6.05 for Virtual Rapport 2.0, and there is significant difference ($p<0.01$) between the two; for the in time question, the mean value of Rapport Agent is 5.65 and it is 5.35 for Virtual Rapport 2.0, and the difference is not significant.

In the force-choice task, among all 21 participants, 19 (90%) participants preferred our virtual human to Rapport Agent.

4.3 Discussion

Rapport: Our virtual human is significantly better than the Rapport Agent in creating rapport. One of the main advantages of our virtual human is that it is based on models learned from human behavior data. This innovative approach is reflected in all the response models. The data-driven approach promotes the feeling of mutual attention and coordination. Besides, the strengthened reciprocity and affective response communicate positive emotions both verbally and nonverbally.

Timing: Our backchannel prediction model significantly outperforms the Rapport Agent's in precision and recall, which indicates that our virtual human is more "in sync" with the human speaker during the interaction. Rapport Agent tends to take a turn (ask the next question) too quickly. Such turn-taking strategy is most likely associated with negative and strong personality [23], which is opposite to the goal of creating rapport.

Behavior: Compared to Rapport Agent, our virtual human has a richer set of behaviors that is correlated with creating rapport. For example, the virtual human mimics the human speaker's smiles, it performs more natural head gestures and strengthens reciprocity by self-disclosure. All these improvements may explain the significant difference on the overall naturalness between our virtual human and the Rapport Agent.

5 Conclusion and Future work

In this paper, we introduced our effort towards building a virtual human whose goal is to create rapport during interactions with human users. Our design follows the three-factor theory of rapport by focusing on creating feelings of positivity, mutual attention and coordination. By comparing with Rapport Agent, we found that our virtual human predicts the timing of backchannel feedback and end-of-turn more precisely, performs more natural behaviors and thereby creates much stronger feelings of rapport between users and virtual agents. As future work, we plan to deploy our virtual human in various scenarios to assess how it will influence the human partner in different situations.

Acknowledgement

This study was funded by the National Science Foundation under Grant # IIS-0916858 and the U.S. Army Research, Development, and Engineering Command and. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

1. 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, 36, 26--50 (2000).
2. Burns, M. Rapport and relationships: the basis of child care. J. Child Care, 2, 47--57 (1984).
3. Tatar, D. Social and personal consequences of a preoccupied listener. Stanford University, Stanford, CA (1997).
4. Bernieri, F.J. and Rosenthal, R. Interpersonal coordination: Behavior matching and interactional synchrony. In: Feldman, R.S. and Rime, B. (Eds.), Fundamentals of nonverbal

behaviors. Studies in emotion and social interaction. pp. 401--432. Cambridge University Press (1991).
5. Tickle-Degnen, L. and Rosenthal, R. The Nature of Rapport and its Nonverbal Correlates. Psychological Inquiry, 1(4), 285--293 (1990).
6. Cassell, J., Gill, A.J. and Tepper, P.A. Coordination in conversation and rapport. In: Proceedings of the workshop on Embodied Language Processing, pp. 41--50 (2007).
7. Bailenson, J.N., Yee, N. Digital Chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. Psychological Science, 16, 814--819 (2005).
8. Bickmore, T., Puskar, K., Schlenk, E., Pfeifer, L., and Sereika, S. Maintaining Reality: Relational Agents for Antipsychotic Medication Adherence. J. Interacting with Computers special issue on Mental Health (2010).
9. Bevacqua, E., Mancini, M., and Pelachaud, C. A Listening Agent Exhibiting Variable Behavior. In: Proceedings of 8th International Conference on Intelligent Virtual Agents (2008).
10. Kang, S.-H., Gratch, J., and Watts, J. The Effect of Affective Iconic Realism on Anonymous Interactants' Self-Disclosure. In: Proceedings of Interaction Conference for Human-Computer Interaction (2009).
11. Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy, R. Creating Rapport with Virtual Agents. In: Proceedings of 7th International Conference on Intelligent Virtual Agents (2007).
12. Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S., Morales, M., Werf, R.J., and Morency, L.-P. Virtual Rapport. In: Proceedings of 6th International Conference on Intelligent Virtual Agents (2006).
13. Morency, L.-P., et al. Contextual Recognition of Head Gestures. In: Proceedings of 7th International Conference on Multimodal Interactions (2005).
14. Thiebaux, M., Marshall, A., Marsella, S., & Kallmann, M. SmartBody: Behavior Realization for Embodied Conversational Agents. In: Proceedings of 7th International Conference on Autonomous Agents and Multi-Agent Systems (2008).
15. Wang, N., and Gratch, J. Don't just stare at me. In: Proceedings of 28th Annual CHI Conference on Human Factors in Computing Systems (2010).
16. von der Pütten, A., Krämer, N., and Gratch, J. Who´s there? Can a Virtual Agent Really Elicit Social Presence. In: Proceedings of 12th Annual International Workshop on Presence (2009).
17. Wang, N., and Gratch, J. Rapport and Facial Expression. In: Proceedings of the International Conference on Affective Computing and Intelligent Interaction (2009).
18. Praat: http://www.fon.hum.uva.nl/praat/
19. Jonsdottir, G.R., Thorisson, K.R., Nivel, E. Learning Smooth, Human-Like Turntaking in Realtime Dialogue. In: Proceedings of International Conference on Intelligent Virtual Agents (2008)
20. Huang, L., Morency, L.-P., Gratch, J. Parasocial Consensus Sampling: Combining Multiple Perspectives to Learn Virtual Human Behavior. In: Proceedings of 9th International Conference on Autonomous Agents and Multiagent Systems (2010).
21. Morency, L.-P., Kok I. de, Gratch, J. Predicting Listener Backchannels: A Probabilistic Multimodal Approach. In: Proceedings of 8th International Conference on Intelligent Virtual Agents (2008).
22. Lee, J. and Marsella, S. Learning a Model of Speaker Head Nods using Gesture Corpora. In: Proceedings of 8th International Conference on Autonomous Agent and Multiagent System, 189--296 (2009).
23. ter Maat, M., Truong K. and Heylen, D. How turn-taking strategies influence users' impressions of an agent. In: Proceedings of 10th International Conference on Intelligent Virtual Agents (2010).

24. Huang, L., Morency, L.-P., Gratch, J. A Multimodal End-of-Turn Prediction Model: Learning from Parasocial Consensus Sampling. In: Proceedings of 10th International Conference on Autonomous Agents and Multiagent Systems (2011).
25. Moon, Y. Intimate exchanges: Using computers to elicit self-disclosure from consumers. J. Consumer Research, 26(4), 323--339 (2000).
26. Bailenson, J.N., Blascovich, J., Beall, A.C. and Loomis, J.M. Equilibrium revisited: Mutual gaze and personal space in virtual environment. PRESENCE: Teleoperators and Virtual Environments, 10. 583--598 (2001).
27. Lafferty, J., McCallum, A., Pereira, F. Conditional Random Field: Probabilistic Models for Segmenting and Labeling Sequence Data. In: Proceedings of 18th International Conference on Machine Learning (2001).
28. Kang, S.-H., Gratch, J. People Like Virtual Counselors that Highly-Disclose About Themselves. The Annual Review of Cybertherapy and Telemedicine (2011).
29. Kang, S.-H., Watts, J., Gratch, J. Associations between interactants' personality traits and their feelings of rapport in interactions with virtual humans. In: 59th Annual Conference of the International Communication Association (2009).
30. Murphy, K. Dynamic Bayesian Networks: Representation, Inference and Learning. Ph.D. Thesis, UC Berkeley, Computer Science Division, July 2002.
31. Huang, L., Morency, L.-P., Gratch, J. Learning Backchannel Prediction Model from Parasocial Consensus Sampling: A Subjective Evaluation. In: Proceedings of 10th International Conference on Intelligent Virtual Agent (2010).