Building Rapport between Human and ECA: A Pilot Study

David Novick and Iván Gris

Department of Computer Science, The University of Texas at El Paso,
500 West University Avenue, El Paso, TX 79968-0518 USA
novick@utep.edu, ivangris4@gmail.com

Abstract. This study is part of a longer-term project to provide embodied conversational agents (ECAs) with behaviors that enable them to build and maintain rapport with their human partners. We focus on paralinguistic behaviors, and especially nonverbal behaviors, and their role in communicating rapport. Using an ECA that guides its players through a speech-controlled game, we attempt to measure the familiarity built between humans and ECAs across several interactions based on paralinguistic behaviors. In particular, we studied the effect of differences in the amplitude of nonverbal behaviors by an ECA interacting with a human across two conversational sessions. Our results suggest that increasing amplitude of nonverbal paralinguistic behaviors may lead to an increased perception of physical connectedness between humans and ECAs.

Keywords: Embodied conversational agent, familiarity, rapport, paralinguistic, nonverbal communication.

1 Introduction

An embodied conversational agent is a computer program that produces an intelligent agent that lives in a virtual environment and communicates through an elaborate user interface. Graphically, an embodied agent can take almost any form, often human-like, and aims to unite gesture, facial expression and speech to enable face-to-face communication with users, providing a powerful means of human-computer interaction (Cassell, 2000). We are interested in exploring how ECAs can build rapport with humans.

Face-to-face conversation is an ongoing collaborative process in which conversants coordinate their verbal and paralinguistic actions (Cassell et al., 2007). However, human-agent communication cannot yet achieve the naturalistic and spontaneous communication that humans do unconsciously; familiarity-enabled ECAs are a step towards a more naturalistic human-agent conversation. By increasing an ECA's extraversion as the relationship progresses across time, participants should experience a stronger sense of physical connection. Thus the question we address in this paper is how ECAs can build rapport with humans through behaviors linked to familiarity—the sense of knowing someone built in more than one conversation.

M. Kurosu (Ed.): Human-Computer Interaction, Part II, HCII 2014, LNCS 8511, pp. 472–480, 2014.

In this paper we describe several definitions and measures of rapport and familiarity. We then classify and merge these into a more comprehensive model. We test this model by having users play a variation of a speech enabled text-based game. During the game, the paralinguistic behaviors of the ECA can change over time to simulate an increase in familiarity between ECA and human. Finally, we analyze the results of the experiment, review the study's limitations, and discuss the future work.

2 Rapport Models

Previous research has identified multiple constituent factors for rapport, including positivity, attention and coordination (Tickle-Degnan & Rosenthal, 1987), and sense of connection, sense of understanding, and what and how things are said (Gratch et al., 2007). In this section, we analyze multiple approaches to rapport with a view toward creating a unified, comprehensive model that can then be implemented in an ECA as a way to test the model.

Prior research on interaction between humans and embodied conversational agents (ECAs) has studied differences in ECAs' nonverbal behaviors as expressions of extraversion or attention. Some studies were based on analysis of a single recording of a human-ECA interaction rather than through a between-subjects comparison of responses to differences in ECA behaviors (e.g., Neff et al., 2010; Huang, Morency & Gratch, 2011). Other studies, though they did use a between-subjects design, looked at rapport-building behaviors through single-session experiments. That is, they compared subjects' responses across conditions based on a single encounter with the ECA (e.g., Cafaro et al., 2012). Several studies have examined how ECAs and humans build rapport over time. In these studies, however, the agents used a multiple-choice text interface and sprite-based characters, which limited nonverbal interactions and dialog flow, particularly during turn taking (e.g., Bickmore & Cassell, 2001; Bickmore & Picard, 2004). Ideally, the way ECAs interact with humans would change as a function of prior interactions between individuals, analogous to how humans interact differently with friends than with strangers.

2.1 Natural Rapport Model

Tickle-Degnen and Rosenthal (1987) described rapport in terms of three dimensions:

- *Attentiveness:* The conversants focus is directed toward the other. They experience a sense of mutual interest in what the other is saying or doing.
- *Positivity:* The conversants feel mutual friendliness and caring.
- *Coordination:* Balance and harmony, where the conversants are "in sync." In addition to its positive valence, coordination conveys an impression of equilibrium, regularity and predictability between the conversants.

This model assumes that positivity becomes less necessary over time, while coordination increases in frequency and importance. This is one of the simpler yet robust models of rapport, and it was not developed with ECAs in mind. Due to its

high level of abstraction, the model does not specifically address many of the major nonverbal rapport-building behaviors within each dimension, and these paralinguistic behaviors are particularly important when implementing ECAs.

2.2 Relational Models

Other researchers provided different approaches to modeling rapport, although these other approaches are not, by themselves, as comprehensive as that of Tickle-Degnen and Rosenthal. We look at four relational models proposed by different researchers, whose findings have not yet been unified. Each of these relations by itself only explains a part of rapport as an overall relationship. Moreover, these interaction traits relay heavily on context and verbal disclosure, which can be difficulty to implement in ECAs. These four relational models are

- *Affinity:* The process through which people try to induce others to have positive feelings towards them; this has also been described as a sense of connection (Bell & Daly, 1984).
- *Reciprocity:* A preference of similarity, often expressed as the Golden Rule: One should treat others as one would like others to treat oneself (Cole & Teboul, 2004).
- *Intimacy:* An interpersonal process, where a person expresses personally revealing feelings or information to another. The process continues when the listener responds supportively and empathically. For an interaction to become intimate, the discloser must feel understood, validated, and cared for (Reis & Shaver, 1988).
- *Continuity:* A progressive pattern of interactions, where each conversation ends with the possibility of continuing the interaction at a later time (Fisher & Drecksel (1983).

Each of these relations can be viewed through a lens of nonverbal behaviors based on extraversion across time, creating animations and expressions on behalf of the agents that are easily observable and controlled (e.g., increased) across time to create the familiarity effect.

2.3 Virtual Rapport Model

Gratch et al. (2007) proposed a model of rapport specifically for ECAs. Indeed, this model defines virtual rapport as rapport generated for human-ECA interactions. In this model, rapport comprises three dimensions:

- *Emotional Rapport:* The sense of connection with the user
- *Cognitive Rapport:* The sense of mutual understanding
- *Behavioral Rapport*: Verbal properties, such as speech duration, pitch, etc

This model, while clearly more useful for implementing ECAs, does not provide details on some of nonverbal behaviors that trigger these dimensions of rapport, especially full-body gesture and interaction.

2.4 Paralinguistic Rapport Model

With a view toward providing a model of rapport that (a) accounts for the factors identified in the natural, relational, and virtual models and (b) provides a basis for supporting full-body ECA paralinguistics, we examined the common elements of the three approaches. We suggest that these elements can be described as encompassing three dimensions: a sense of emotional connection, a sense of mutual understanding, and a sense of physical connection. Figure 1 presents our "paralinguistic" model of rapport, showing the correspondence of the model's dimensions to the dimensions or factors of the antecedent models. Two of the dimensions in the paralinguistic model—emotional connection and mutual understanding—arise from a combination of verbal and nonverbal behaviors. However, the third dimension—physical connection—arises solely from paralinguistic behaviors. Our model is broader than that of Tickle-Degnen and Rosenthal with respect to physical behaviors, in that the physical collaboration and cooperation of familiar conversants can go beyond mimicry to include many other kinds of paralinguistic behaviors, such as ways of expressing continuers and ways of indicating attitudes.

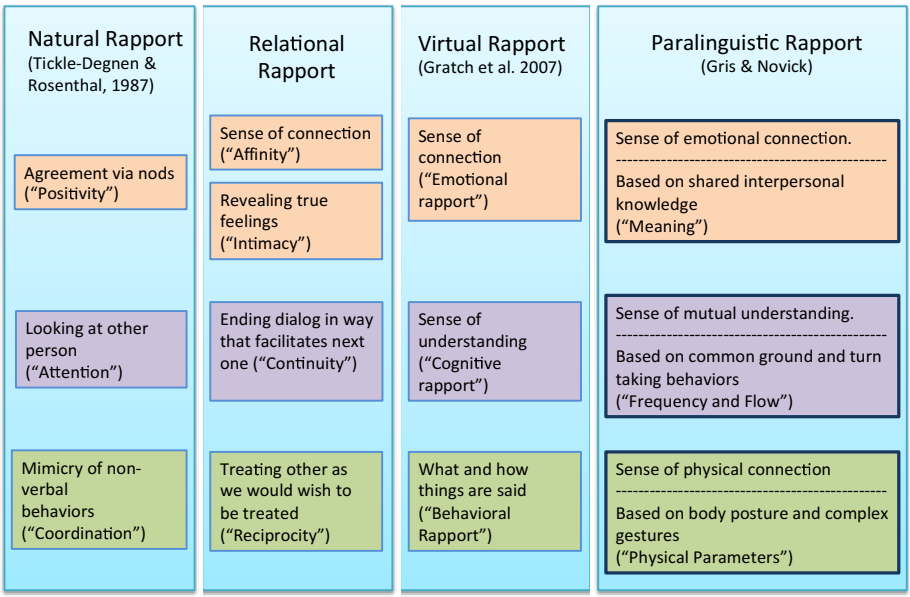

Fig. 1. Paralinguistic rapport model and its relation to the natural rapport mode, the relational rapport model (affinity: Bell & Daly, 1984; reciprocity Cole & Teboul, 2004; intimacy: Reis & Shaver, 1988; and continuity: Fisher & Drecksel, 1983), and the virtual rapport model

Given our paralinguistic rapport model, our longer-term goal involves assessing the model's validity and usefulness for implementing ECAs that can build rapport with humans. As a first step toward this goal, we focus on the rapport-building effects of physical paralinguistic behaviors. The harmony and engagement of rapport can be seen as relatively weak during initial interactions and developing strength over time;

we refer to this development across time as familiarity. Conversants signal increased familiarity by, among other things, increasing the amplitude of nonverbal communicative behaviors such as hand gestures and head nods (Neff et al., 2010; Cafaro et al., 2012; Clausen-Bruun, Ek, & Haake, 2013). In other words, the ECA's gestures and their degree of extraversion, as expressed through greater amplitude, can build the physical-connection dimension of rapport over time.

Thus our specific question in this study is whether subjects interacting with an ECA over two sessions, where the ECA uses higher-amplitude gestures in the second session, would feel an increase in rapport in the second session.

3 Methodology

To test whether subjects would perceive more rapport with an ECA in the increased-familiarity condition, this study piloted an investigation of how to signal increased familiarity over repeated interactions as a component of rapport. In particular, we studied the effect of differences in the amplitude of nonverbal behaviors by an ECA interacting with a human across two conversational sessions. In the first session, the ECA used nonverbal behaviors with a lower-amplitude baseline. Our independent variable was whether, in the second session, the ECA used same baseline amplitude nonverbals, indicating no increase in familiarity, or used higher-amplitude nonverbals to convey an increase in familiarity.

Our experimental protocol had 20 subjects interact for a 20-minute session with an ECA and then interact for a second 20-minute session with the ECA at least one day later. The sessions involved a conversation with a life-sized, front-projected ECA in UTEP's Immersion Laboratory, where the ECA served as the narrator for an adventure game developed specifically for this study.

The game, "Escape from the Castle of the Vampire King," was inspired by early text-based adventure games such as Zork (Anderson & Galley, 1985) and Colossal Cave. Subjects can move from room to room, pick up, drop, and use objects, and kill vampires. We chose this application because such games are known to be engaging, and we wanted our human subjects to want to interact with the ECA. A text-based game was also helpful from a practical standpoint, in that it limited the amount of speech that needed to be recognized; the subject's possible utterances were both simple and highly constrained by the game's context. Based on our own experiences with text-based games, we instructed subjects to draw a map as they explored.

A trial run of the experiment strongly suggested that asking subjects to draw a map as they explored the castle meant that the subjects usually had their gaze directed at the map rather than at the ECA. This was problematic for our experiment because if subjects kept their gaze focused on their map they would not be watching the ECA and thus would not see the paralinguistic behaviors we were changing as the independent variable. As a result, we modified the game so that the map was drawn automatically on the wall behind the ECA as the subject explored the castle. We observed that subjects now directed their gaze toward the ECA to a much greater extent. Figure 2 shows a person interacting with the ECA. Figure 3 shows the game mid-way through a typical session. The game was extensive enough to support easily the two 20-minute sessions.

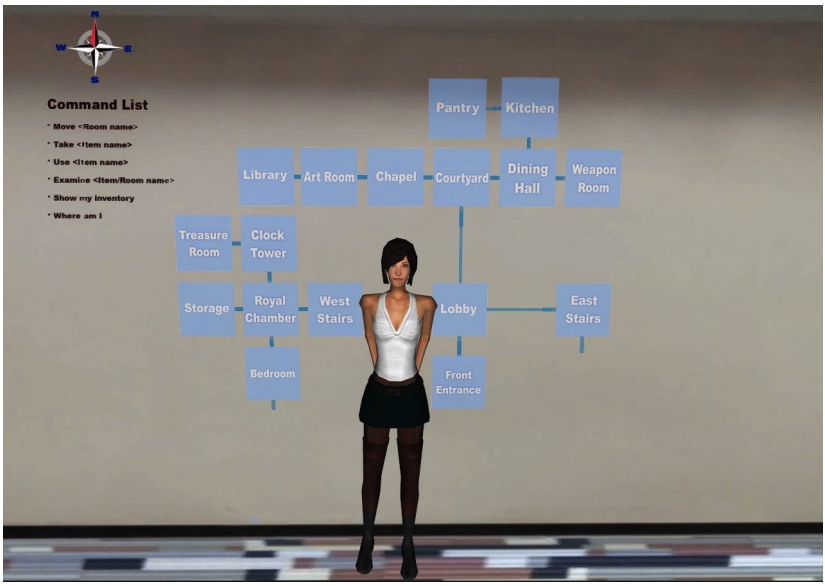

Fig. 2. "Escape from the Castle of the Vampire King" game, with the game at about the mid-point of play

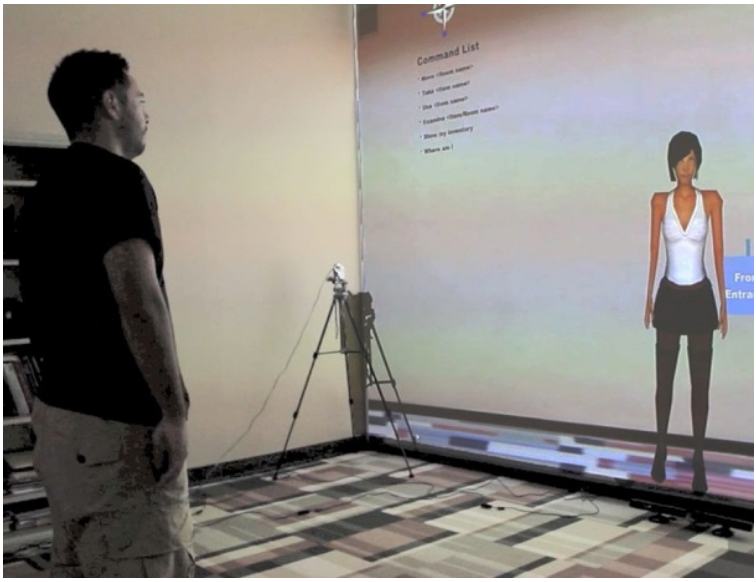

Fig. 3. Interaction between a human and the ECA during testing of the game

In the subjects' first session, the ECA used the baseline amplitude for nonverbals. In the second session, the subjects were randomly assigned to one of two conditions:

(1) the ECA continued to use the baseline nonverbal or (2) the ECA used nonverbal with increased amplitude. The subjects completed a rapport instrument after each session. We adapted and extended the survey of Acosta and Ward (2011) into an instrument of twelve Likert-scale questions, balanced for positive and negative responses that covered the three rapport factors in our model. Table 1 lists the questions in the rapport instrument.

Table 1. Rapport instrument. Subjects indicated agreement or disagreement on a five-point scale.

The agent understood me
The agent seemed unengaged
The agent was excited
The agent's movements were not natural
The agent was friendly
The agent was not paying attention to me
The agent and I worked towards a common goal
The agent and I did not seem to connect
I sensed a physical connection with the agent
The agent's gestures were not lively
I feel the agent trusts me
I didn't understand the agent

4 Results

Our analysis compared the second-session responses across the two conditions. A one-tailed t-test indicated that there was no significant main effect ($p=0.37$). Similarly, there was no significant effect for each of the three rapport factors composing the instrument: emotional connection ($p=0.40$), sense of mutual understanding ($p=0.29$), physical connection ($p=0.17$). However, a post-hoc power analysis suggests that conducting the study with 60 subjects would likely produce a significant result for the physical-connection factor.

These results suggest that increasing the amplitude of nonverbal paralinguistic behaviors may not by itself be sufficient to induce a perception of increased rapport throughout all three major dimensions. Nevertheless, the results also suggest that increasing the amplitude of nonverbal paralinguistic behaviors may lead to increased perception of physical connectedness between humans and ECAs. We suspect that the low emotional connection observed in our study was likely due to the lack of emotion in the speech synthesizer; varied emotion would have increased engagement for critical parts of the game, such as the player fighting a vampire or dying. In our follow-on work, we are building a new game with greater interactivity that uses recorded rather than synthesized speech.

An open question at the end of the survey asked subjects to comment on their overall perception of the agent. This question was not taken into account for the analysis of the results, but it provided some interesting information. Some subjects went

as far as believing that the agent was a vampire trying to kill them, and they expected a major plot twist near the end of the game. This is arguably a sense of mutual understanding on the part of the player, even if it was unintended and a complete misconception. Another reason for this perception may be a mismatch between the ECA's nonverbal and verbal behaviors, due in large part to the speech synthesizer. That is, when the player encountered a vampire, the ECA would physically show exaggerated expressions alerting the player to the presence of the vampire in the room while verbally explaining the situation in a calm, synthesized voice.

5 Conclusion

Based on our observation of the 20 second-session human-ECA interactions discussed in Section 4, we now return to the question that motivated our study: Do subjects perceive the agent in the increased-familiarity condition in the second session as having higher rapport? Although subjects did notice a behavioral change, and they scored the agent higher on the physical connection dimension, there was not enough information to affirm that this by itself leads to an overall higher sense of rapport.

Our future work will address many of the limitations in this study. In particular, we plan to have a pool of 60 rather than 20 subjects, a recorded human voice rather than a synthesized voice, a more immersive game experience to maintain the participants' attention for a longer period of time, and visual aids to accompany the verbal descriptions of the game state that the agent describes. In the longer term, we plan to address the additional two dimensions of rapport by expressing them in terms of perceivable nonverbal behaviors, to truly create a greater sense of rapport that grows as a function of additional interaction.

Acknowledgments. The authors thank Guillaume Adoneth and David Manuel for their contributions to the design of this study, Jonathan Daggerhart for permission to adapt his original text-based adventure game into "Escape from the Castle of the Vampire King," and Diego Rivera, Alex Rayon, Adriana Camacho, Baltazar Santaella, Juan Vicario, Joel Quintana and Anuar Jauregui for their help in developing the game.

References

1. Acosta, J.C., Ward, N.G.: Achieving rapport with turn-by-turn, user-responsive emotional coloring. Speech Communication 53(9), 1137–1148 (2011)
2. Anderson, T., Galley, S.: The history of Zork. The New Zork Times 4(1-3) (1985)
3. Bell, R.A., Daly, J.A.: The affinity-seeking function of communication. Communications Monographs 51(2), 91–115 (1984)
4. Bickmore, T., Cassell, J.: Relational agents: A model and implementation of building user trust. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 396–403. ACM (March 2001)

5. Bickmore, T.W., Picard, R.W.: Establishing and maintaining long-term human-computer relationships. ACM Transactions on Computer-Human Interaction (TOCHI) 12(2), 293–327 (2005)
6. Cafaro, A., Vilhjálmsson, H.H., Bickmore, T., Heylen, D., Jóhannsdóttir, K.R., Valgarðsson, G.S.: First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters. In: Nakano, Y., Neff, M., Paiva, A., Walker, M. (eds.) IVA 2012. LNCS, vol. 7502, pp. 67–80. Springer, Heidelberg (2012)
7. Cassell, J. (ed.): Embodied conversational agents. The MIT Press (2000)
8. Cassell, J., Gill, A.J., Tepper, P.A.: Coordination in conversation and rapport. In: Proceedings of the Workshop on Embodied Language Processing, pp. 41–50. Association for Computational Linguistics (June 2007)
9. Clausen-Bruun, M., Ek, T., Haake, M.: Size certainly matters–at least if you are a gesticulating digital character: The impact of gesture amplitude on addressees' information uptake. In: Intelligent Virtual Agents, pp. 446–447. Springer, Heidelberg (2013)
10. Cole, T., Teboul, B.: Non-zero-sum collaboration, reciprocity, and the preference for similarity: Developing an adaptive model of close relational functioning. Personal Relationships 11(2), 135–160 (2004)
11. Crowther, W., Woods, D., Black, K.: Colossal cave adventure. Computer Game (1976)
12. Fisher, B.A., Drecksel, G.L.: A cyclical model of developing relationships: A study of relational control interaction. Communications Monographs 50(1), 66–78 (1983)
13. Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy, R.: Creating rapport with virtual agents. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 125–138. Springer, Heidelberg (2007)
14. Huang, L., Morency, L.-P., Gratch, J.: Virtual rapport 2.0. In: Vilhjálmsson, H.H., Kopp, S., Marsella, S., Thórisson, K.R. (eds.) IVA 2011. LNCS, vol. 6895, pp. 68–79. Springer, Heidelberg (2011)
15. Neff, M., Wang, Y., Abbott, R., Walker, M.: Evaluating the effect of gesture and language on personality perception in conversational agents. In: Safonova, A. (ed.) IVA 2010. LNCS, vol. 6356, pp. 222–235. Springer, Heidelberg (2010)
16. Reis, H.T., Shaver, P.: Intimacy as an interpersonal process. In: Duck, S.W. (ed.) Handbook of Personal Relationships, pp. 367–389. John Wiley, NY (1988)
17. Tickle-Degnen, L., Rosenthal, R.: Group rapport and nonverbal behavior. Review of Personality and Social Psychology 9, 113–136 (1987)