Adaptive Virtual Rapport for Embodied Conversational Agents

Ivan Gris
Department of Computer Science
The University of Texas at El Paso
500 West University Avenue, El Paso, TX
79968-0518 USA
igris@miners.utep.edu

ABSTRACT

In this paper I describe my research goals and hypotheses regarding human-computer relationships with embodied conversational agents (ECAs). I include important studies of related research that inform and direct my own efforts. I explain the current state and some technical aspects of the ECAs I have contributed to create, and past experiments regarding human-ECA familiarity, ECA design and analysis, and multiparty ECA interaction, including our semi-automated corpora collection techniques, analysis methodology, and their respective results to date. Finally, I conclude with an overall presentation of all current studies I have worked on, and future possibilities for my final dissertation and post-dissertation research related to virtual human-ECA rapport.

Keywords: Virtual rapport, embodied conversational agents.

1. INTRODUCTION

The use of ECAs can improve the user's experience in certain applications. For example, when interacting with ECA enabled systems, users prefer a non-verbal visual indication of an embodied system's internal state to a verbal indication [1]. In addition, there are several advantages in human-ECA interactive systems such as communication parallelism (e.g. the user can communicate non-verbally, verbally and can perform other tasks at the same time) [2], face-to-face interaction, and increased recall. Users are able to perform multiple tasks and remember the interaction history due to the social component and face-to-face affordances that ECAs provide [3].

ECAs are part of a multi-billion dollar industry, with applications ranging from entertainment to complex training systems. By creating ECAs that are able to enact realistic and natural paralinguistic behaviors, we can simulate relational processes and traits, such as rapport and familiarity, and enhance the user experience across long-term human-ECA interactions. This increase in naturalness will allow human-ECA relationships to escalate into long-term, non-context-based, verbal and non-verbal interactions, which can result in ECAs performing convincingly a larger, more complex set of actions. Given the importance and emerging adaptation and possibilities in ECA design, my research question aims to solve the problem of deciding what non-verbal behaviors should ECAs present, what should trigger those behaviors, and if those behaviors should evolve or change during prolonged interactions.

In the following sections I explain and classify ECAs according to their functionality and characteristics. I then proceed to explain rapport and the current models that attempt to explain this trait. Next I explain studies I have conducted in human-ECA familiarity and the methodology for both, current and future experiments. I explain the results found to date, and how our ECA systems work for both, automatic annotation and agent's behaviors. Finally, I conclude with my expected dissertation work and post-degree research directions.

2. BACKGROUND

The term Embodied Conversational Agent (ECA) refers to a form of human-computer interaction, represented by intelligent agents that live in a virtual environment and communicate through elaborate user interfaces. Graphically embodied agents can take almost any form, often human-like, and aim to unite gesture, facial expression and speech to enable face-to-face communication with users, providing a powerful means of human-computer interaction [4].

2.1 Embodied Conversational Agents

To facilitate the study of ECAs and generalize some of the particular features of different agents, I classified them according to their representation, features and purpose into four categories: commercial, mediator, pedagogical, and specialized agents.

Commercial ECAs are used to improve the customer service experience either by presenting the company's information in a more attractive manner, or to provide automated customer support. The main advantage of this approach is the possibility of uninterrupted service and a reduced workload for the human operators. These ECAs do not need realistic behavior; conversations are usually scripted and take the form of "frequently asked questions" instead of dynamic, unscripted conversation. Their representations are usually human-like 2D images.

Another thriving area that focuses on improving virtual agents is the $34.7 billion market [5] of the videogame industry, where ECAs are the key element in enhancing storytelling and creating more immersive player interactions. In videogames, ECAs usually represent characters that the player encounters across the flow of the game. Since the player is represented as an avatar and ECAs interact with the player' indirectly by responding to the avatar's actions and not the player itself, these interactions are mediated.

ICMI'13, December 9–13, 2013, Sydney, Australia.
ACM 978-1-4503-2129-7/13/12.
http://dx.doi.org/10.1145/2522848.2532190

In other words, the player controls a virtual character, and all interactions usually occur with that virtual character. This mediated type of interaction poses usability and playability challenges [6], and interaction mechanics are limited to the affordances of the avatar that the player controls.

From a pedagogical perspective, ECAs are employed as guides or teachers for very specific tasks. Although they are far from replacing human instructors, this type of agent is an appropriate alternative to areas or topics with low availability of real instructors. Public museums have employed ECAs as guides, where they engage with visitors in natural face-to-face communication while providing information about the exhibits. [7]. In classroom settings, ECAs with a variety of gestures and facial expressions have been implemented in small groups, and proven to increase attitudinal and procedural learning [8]. Even when agents for pedagogical purposes show great potential, the long-term interaction component, along with the hardware and technical skills necessary to implement ECAs in classrooms limits their use.

Finally, ECAs have been developed to serve causes with a high social impact. Examples include military training, computer-assisted speech and language tutors for hard-of-hearing and autistic children [9] and many other applications. These ECAs are realistic lifelike characters that use speech recognition, natural language, non-verbal behavior and realistic scenarios.

During the last few decades, state-of-the-art technology has overcome some of the basic technical obstacles of rich human-agent interaction, including speech recognition, text-to-speech, and character three-dimensionality and animation. Immersive applications, such as ECAs, are not only perceptual, but highly interactive and require user action.

Interfaces with rich behavior, such as ECAs, present additional complexities for both, design and development. In particular, researchers argue that although the functionality provided by speech-enabled and kinetic aware interfaces is impressive by itself, the interfaces used to interact with these agents and access their functionality are often inconsistent (e.g. depending on the agents domain, each one recognizes different words), imprecise [10] (e.g. motion trackers jitter or misrecognized speech), and by emphasizing naturalness in communication and expression in their designs, user interaction metaphors are confusing and almost non-existent.

An ECA should designed to interact without the need of most of the traditional interface elements, that is, humans should interact with ECAs as natural as possible and preferably without explicit traditional interface elements such as buttons, text boxes, or point-and-click items (unless they are part of the agent's tasks and goals, for example, a training agent for a new software). Instead, ECAs need to detect and simulate complex behaviors that encompass a combination of non-verbal cues such as gaze, gestures and mimicry, and verbal feedback such as backchannels, in conjunction to context specific content to create a shared state of understanding between human and the ECA. This Multimodal interaction in everyday life seems so effortless. However, a closer look reveals that such interaction is indeed complex and comprises multiple levels of coordination, from high-level linguistic exchanges to low-level couplings of momentary bodily movements [11].

One of the main goals of ECA development and research is to raise the believability and perceived trustworthiness of agents, and increases the user's engagement with the system; in other words, to create ECAs that follow social conventions, similar to those in natural interaction [12].

One way to increase the non-verbal naturalness of human-ECA communication is the use of rapport. Several models for rapport are explained in the following section.

2.2 Existing Rapport Models

To create and apply virtual rapport, one must first understand inter-human rapport. Rapport is not an individual trait but rather a collective combination of qualities that emerge from each individual during interaction [13]. One generalized definition of rapport is the feeling of mutual understanding; the connection and harmony experienced when two people are engaged in conversation [14], or as it is often informally described, the feeling of being in "sync".

In this study I analyze several definitions and measures of rapport to create a unified, comprehensive model, which can then be implemented on an ECA.

2.2.1 Rapport Measures

Tickle-Degnen and Rosdenthal divided rapport into 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, and are "in sync". Where in addition to its positive valence, in an interpersonal context coordination conveys an image of equilibrium, regularity and predictability between the interactants.

This model assumes that positivity becomes less necessary over time while coordination increases in frequency and importance. One problem with this model may be in the definition of rapport itself. Since it is possible to have both, a mutual understanding and a disagreement, positivity may not be as important.

2.2.2 Relational Models

There are four relational models that when combined provide another definition for rapport.

- Affinity: The process in which people try to make others have positive feelings towards them. Also described as a sense of connection [15].
- Reciprocity: It's the preference of similarity, in other words, the golden rule: One should treat others as one would like others to treat oneself [16].
- Intimacy: Intimacy is an interpersonal process. One person expresses personally revealing feelings of information to another. It continues when the listener responds supportively and empathically. For an interaction to become intimate the discloser must feel understood, validated, and cared for [17].

- Continuity: A progressive pattern of interactions. End each conversation with the possibility of continuing the interaction at a later time.

One problem with these models is that they have not been unified in previous research, and each dimension by itself only explains a small fragment of rapport. In addition, some of the dimensions relay on the context of verbal disclosure, which is irrelevant to our attempt to produce and explain this behavior in terms of paralinguistics.

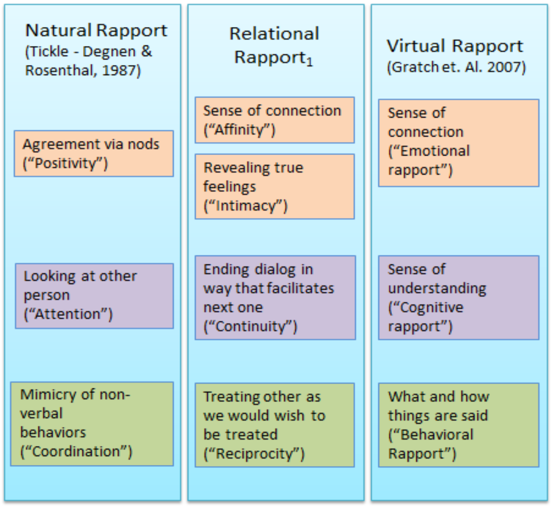
2.2.3 Virtual Rapport

One of the latest models describes virtual rapport for human-ECA interaction [14]. According to this model, rapport can be divided in three dimensions:

1. Emotional Rapport: The sense of connection with the user.
2. Cognitive Rapport: The sense of mutual understanding.
3. Behavioral Rapport: Verbal properties, such as speech duration, pitch, etc.

This model, however, lacks the specifics for the non-verbal behaviors that trigger these dimensions of rapport.

Figure 1 compares and categorizes different models of rapport in three main dimensions, each one represented by a different color.



Natural Rapport (Tickle - Degnen & Rosenthal, 1987)	Relational Rapport[1]	Virtual Rapport (Gratch et. Al. 2007)
Agreement via nods ("Positivity")	Sense of connection ("Affinity") Revealing true feelings ("Intimacy")	Sense of connection ("Emotional rapport")
Looking at other person ("Attention")	Ending dialog in way that facilitates next one ("Continuity")	Sense of understanding ("Cognitive rapport")
Mimicry of non-verbal behaviors ("Coordination")	Treating other as we would wish to be treated ("Reciprocity")	What and how things are said ("Behavioral Rapport")

1. Affinity (Bell & Dalay, 1984), Reciprocity (Cole & Teboul, 2004), Intimacy (Reis & Shaver, 1988), Continuity (Fisher & Drecksel, 1983)

Figure 1. Comparison across different models of rapport

2.3 New Model for Rapport

The combination of all previous models renders our own interpretation of rapport as shown in Figure 2.

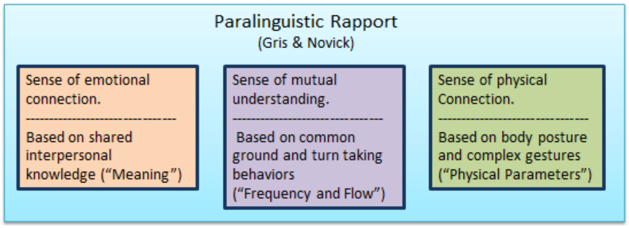

Paralinguistic Rapport
(Gris & Novick)
Sense of emotional connection.
Based on shared interpersonal knowledge ("Meaning")
Sense of mutual understanding.
Based on common ground and turn taking behaviors ("Frequency and Flow")
Sense of physical Connection.
Based on body posture and complex gestures ("Physical Parameters")

Figure 2. Gris-Novick rapport model

My research aims to define and enact through an ECA non-verbal behaviors at the appropriate moment to enhance rapport.

After defining my new model of rapport based on inter-human interactions, we must then answer the questions:

1. What non-verbal behaviors, such as grounding and turn-taking, are necessary to represent each of the three dimensions of rapport in the composite model?
2. When are the appropriate times for displaying non-verbal rapport behaviors in human-ECA conversations?
3. How should this behaviors evolve through time (for example, should a dimension be preferred over another after a history of events, or if a dimension decreases in importance after a longer period of interaction)?

I hypothesize that users will not only be able to notice the difference between our agent and a non-rapport agent, but that the interaction with our rapport enabled agent will be preferred.

In the next section I describe the latest implementation of the agent and our expected test cases.

3. METHODOLOGY

The experiments, both, current and past, follow a bottom-up approach, where the observations made on natural (human-human) and virtual (human-ECA) dyadic conversations are used to develop the paralinguistic representations of our new models, in this case, the Gris-Novick rapport model. Observations focus on when do paralinguistic signals such as, turn taking, control acts, mutuality confirmation signals appear, and how they differ across time. Gaze, nods, and upper body gestures are annotated for both, the natural interactions and the resulting virtual interactions.

To aid with the annotation tasks, I have developed a flexible system using a Kinect™ that detects a set of poses and gestures and automatically annotates them and their time-stamp.

3.1 Validation

The new models are validated by comparing the observations from the corpus against the pilot experiments with the ECA.

One pilot experiment considers human-ECA familiarity. In this experiment, the participant is exposed to an ECA for two half an hour sessions at least one day apart. During the first session, the agent exhibits non-familiar behavior, which attempts to mimic the paralinguistic actions of a person when they make acquaintance for the first time. Throughout the second session, the agent changes behavior for a more extroverted, fast paced behavior, thus assuming the user's familiarity with it.

The agent's behavior is controlled and limited by its grammar and a pre-defined set of movements where it chooses its reaction from. The users' behavior is recorded and analyzed for different non-verbal reactions between the familiar and non-familiar conditions.

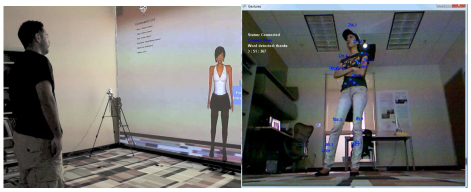


Figure 3. [Right] Pose and speech recognizer. [Left] Interaction with our agent, Mia, during the human-ECA familiarity pilot study

These initial observations will be contrasted with the new proposed model to validate the most distinguishable behaviors.

3.2 ECA Implementation

The ECAs are designed using a three-tiered architecture. The top layer is the Kinect sensor, which supports RGB video, audio provided by an array of seven microphones, and a depth field based on infrared sensor information. The middle layer contains scripts in Unity3D Pro Game Engine, which are used to render and animate the ECAs and contains the virtual environment. The bottom layer contains all the logic of the agent's behavior and sensor interpretation, including speech recognition, text-to-speech, and gesture recognition.

The agent is represented with her environment in the Interactive Systems Group Immersion Lab. She is projected at actual human scale on a wall.

So far two versions of the agent have been developed. The first version was used in a previous study, where we examine familiar and non-familiar embodied conversational agent (ECA) behavior, and how it affects user interaction. We examined the effects of user's perception of the agent's familiarity levels based on the agent's extroversion and analyzed the effects of the user's experience and the user's behavioral changes with respect to the agent's current state, and explore automated annotation methods for both, agent and user verbal and non-verbal behaviors. The interactions occur while playing a verbal version of a text based game.

The latest version of the agent will lead the user through a series of activities and conversations while playing a game. The game simulates a survival scenario, where the user has to collaborate, cooperate, and build a relationship with the agent to survive. This simulation is built with the intention to maximize rapport building opportunities, as well as to take advantage of the non-verbal behaviors in a more immersive environment, where both, the user and the agent can interact with the same objects in virtual space. The storyline allows the necessary flexibility and decision making, without creating a completely open environment where tasks are difficult to set up and evaluate.

4. RESULTS

The results of the familiarity study have found a difference between the perception of the familiar and unfamiliar behaviors, however, there is not a clear connection yet as to what particular behaviors lead to the different perceptions and if they reflect the participant's reactions, as the analysis is still an ongoing process.

5. FUTURE WORK

My current research, which includes building human-ECA familiarity relationships, human-ECA turn-taking based on non-verbal behaviors, building realistic ECAs, multimodal and multiparty ECAs specific behaviors, and studying interaction in cross-functional teams has made possible several follow-up topics leading to new possibilities for my dissertation research.

I plan to pursue paralinguistic behavior, including nods, gaze-shifts, full body gestures, and pose for grounding, turn taking, and misunderstanding detection and recovery in mid-term human-ECA interactions. Post-degree research may extend this to multiparty-multi-agent settings, cross-cultural settings, and human-ECA artifact interaction with objects on virtual environments.

6. REFERENCES

[1] Marsi, Erwin; van Rooden, Ferdi (2007), "Expressing uncertainty with a talking head in a multimodal question-answering system", Proceedings of the Workshop on Multimodal Output Generation (MOG 2007), Aberdeen, UK, pp. 105–116

[2] Kipp, Michael; Kipp, Kerstin H.; Ndiaye, Alassane; Patrick (2006), "Evaluating the tangible interface and virtual characters in the interactive COHIBIT exhibit.", Proceedings of the International Conference on Intelligent Virtual Agents (IVA'06)

[3] Beun, Robbert-Jan; de Vos, Eveliene; Witteman, Cilia (2003), "Embodied conversational agents: Effects on memory performance and anthropomorphisation.", Proceedings of the International Conference on Intelligent

[4] Cassell, J. (Ed.). (2000). Embodied conversational agents. The MIT Press.

[5] Bond, Paul (18 June 2008). "Video game sales on winning streak, study projects". Reuters. Retrieved 2011-03-12.

[6] Desurvire, H., Caplan, M., & Toth, J. A. (2004, April). Using heuristics to evaluate the playability of games. In CHI'04 extended abstracts on Human factors in computing systems (pp. 1509-1512). ACM.

[7] Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., ... & White, K. (2010, January). Ada and Grace: Toward realistic and engaging virtual museum guides. In Intelligent Virtual Agents (pp. 286-300). Springer Berlin Heidelberg.

[8] Baylor, A. L., & Kim, S. (2008, January). The effects of agent nonverbal communication on procedural and attitudinal learning outcomes. In Intelligent virtual agents (pp. 208-214). Springer Berlin Heidelberg.

[9] Bosseler, A., & Massaro, D. W. (2003). Development and evaluation of a computer-animated tutor for vocabulary and language learning in children with autism. Journal of autism and developmental disorders, 33(6), 653-672.

[10] Wooldridge, M., & Jennings, N. R. (1998, May). Pitfalls of agent-oriented development. In Proceedings of the second international conference on Autonomous agents (pp. 385-391). ACM.

[11] Zhang, H., Fricker, D., & Yu, C. Modeling Real-time Multimodal Interaction with Virtual Humans.

[12] Bickmore, T., & Cassell, J. (2001, March). 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.

[13] Linda Tickle-Degnen and Robert Rosenthal. Group rapport and nonverbal behavior. In Clyde Hendrick, editor, Group processes and intergroup relations, volume 9 of Review of Personality and Social Psychology, pages 113-136. Sage, Newbury Park, CA, 1987.

[14] Lixing Huang, Louis-Philippe Morency, and Jonathan Gratch. Virtual rapport 2.0. In Hannes H. Vilhjalmsson, Stefan Kopp, Stacy Marsella, and Kristinn R. Thorisson, editors, Intelligent Virtual Agents, volume 6895 of Lecture Notes in Computer Science, chapter 8, pages 68-79. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2011.

[15] Bell, R. A., & Daly, J. A. (1984). The affinity-seeking function of communication.*Communications Monographs*, *51*(2), 91-115.

[16] Cole, T., & Teboul, B. (2004). Non-zero-sum collaboration, reciprocity, and the preference for similarity: Developing an adaptive model of close relational functioning. *Personal Relationships*, *11*(2), 135-160.

[17] Reis, H. T., & Shaver, P. (1988). Intimacy as an interpersonal process.