Design of a Virtual Peer Tutor

Rebecca Lai

Introduction

This report concerns the design and front end development of a virtual peer tutor for Carnegie Mellon's ArticuLab.

I. Overview

Rapport can be described as "a close and harmonious relationship in which the people or groups concerned understand each other's feelings or ideas and communicate well." In colloquial terms, the presence of rapport determines whether two people "click." One question the ArticuLab asks is: *How* do people build rapport? That is, in the process of building harmonious relationships, what verbal and nonverbal behaviors do people evince? The answer to this question has important implications: after all, knowing how to build rapport benefits virtually (no pun intended) any field that puts a large emphasis on interpersonal relationships.

Virtual agents play a large role in answering the aforementioned question. Because the behavior of virtual agents is linked to controllable parameters, virtual agents can first be used to test rapport: by switching certain parameters on and off, we can determine which behaviors *augment* rapport and which behaviors *diminish* rapport. Once a basic understanding of rapport has been established, virtual agents can be used to create rapport, enhance interpersonal interactions, and help achieve specific goals. For instance, a virtual teacher that can easily build rapport with its students will be a far more effective educator than teachers (virtual or living) that cannot.

It is with this understanding that the ArticuLab seeks to develop a virtual agent capable of building rapport. This summer, I designed and developed the front end of such an agent, a process which I will detail in this final report.

II. Design Considerations: Age, Gender, Race

A natural first question one might ask is: What does the front end development of a virtual agent entail? The term "virtual agents" is a broad one. A virtual agent can be a static 2D image that "interacts" with individuals through text, or it can be a 3D model that talks and moves as if it were a real person. The ArticuLab seeks to build the latter. As such, the front end development of our virtual agent encompassed the following: 1. designing the appearance of the agent, 2. creating 3D models of the agent, and 3. animating the agent. Once these assets have been created, back end development would ensure that the agent reacts appropriately to stimuli.

Front end development is an important step in the creation of a virtual agent; the appearance of an agent, after all, can significantly affect how people interact with it. As Baylor (2009) notes, "the *visual presence* and *appearance* of…agents can have a major impact on motivation and affect regardless of the

underlying technical sophistication." Agent characteristics that often merit consideration include age, gender, and race.

Fortunately, age was not a field that required much contesting for the agents we developed. Our agent has a very specific purpose: it will be working as a peer tutor for students in the 7th to 10th grades, teaching math to them and learning math from them. As such, the agent should appear to be the same age as the students with whom it will interact.

Gender was an area that required more careful deliberation: which gender would be most conducive for rapport? Baylor (2003) notes that male agents are perceived as more interesting, intelligent, useful, and generally leading to greater satisfaction, although it may be significant that the participants of Baylor's study were predominantly female (70%). Indeed, one might initially think that the female majority in Baylor's study corroborates the results of Karacora et. al (2012), which observes that participants' performance and effort were greatly enhanced when they worked with agents of the opposite gender. However, both of these studies are contradicted by Baylor et. al (2009), which reports that in general, observers tend to be more influenced by agents of the *same* gender. Furthermore, the relevance of *any* of the aforementioned conclusions is brought into question when one considers that the studies involved participants 18 years and older. The study most relevant to our particular needs was Silvervarg et. al (2012) who worked with children ages 14-15 and notes that 1. female agents are abused more than male agents and 2. that gender-ambiguous agents are abused less than female agents but more than male agents. However, while these results are certainly informative, I ultimately decided that it would be best to make two agents – a male agent, and a female agent – so that ArticuLab researchers could choose which agent to use based on the specific research context. In general, it seems best to agree with Baylor (2009) when she notes that "generalizations are context-dependent."

I next considered the agent's race. Here, existing studies (though somewhat scarce in number) suggest that in general, participants work better with agents who share their race and ethnicity (Baylor, 2003; Baylor, 2009). Of course, as is the case with gender, the effects of an agent's race/ethnicity is often also context-dependent; for instance, if the agent is intended to act as an expert in a particular area of knowledge, Baylor (2004) has shown that young adults prefer that the agent appear African-American and male. Based on these observations, I proposed that we first create agents that appear Caucasian. Then, we could modify the features and skin tone of the agents until we settled on an appearance that appeared racially ambiguous. That is, African-American students would perceive the agents as being African-American, and Caucasian students would perceive them as being Caucasian.

III. Modeling in Pixologic ZBrush and Autodesk Maya

I chose to use Pixologic ZBrush as the 3D modeling software for the agents due to my familiarity with it. ZBrush is largely used only for modeling; for animation purposes, models created in ZBrush are typically imported into animation software like Autodesk Maya. There are a number of differences between modeling in ZBrush and modeling in Maya. In general, ZBrush is perhaps a less meticulous process – it is easier to "sketch out" 3D models in ZBrush because its modeling process is more akin to traditional sculpting (as opposed to Maya, in which models must be built up polygon by polygon).[1]

Creating basic body and facial forms for the models took about three weeks. Following that, the models were revised based on feedback from lab members. Most of the changes concerned the degree of realism in the models: to avoid the "uncanny valley effect," the models were made to look more cartoonish in style (i.e. simplified features, fewer facial details). Both male and female models were dressed in standard school clothing: polo shirts, khaki pants, and sneakers.

Once the models were deemed acceptable, a few steps were necessary for the models to be "animation-ready." First, the number of polygons in the models needed to be reduced so that the models could be smoothly animated. Second, the polygons that constituted the models (called their "topology") needed to follow the general flow of the models' bodies Both of these steps could be achieved through a process called re-topologizing, in which the polygons of a model can be re-drawn manually. Re-topologizing was completed in ZBrush. Then, the models were brought into Maya where their topologies were further refined.

IV. Evaluating Agent Appeal through Surveys

The initial design of the model was based largely on intuition. That is, what did we, as researchers, think would appeal most to kids in 7th to 10th grade? However, precisely because these design decisions were based on intuition, it was eventually necessary to test the validity of our assumptions. As such, I designed a survey that would be distributed to middle school students. This survey was intended to test students' perception of the following: 1. the agents' age and gender, 2. the agents' personalities and general disposition (as derived from just their appearance), and 3. the agents' capabilities as potential peer tutors (i.e. their ability to learn, their ability to teach).

The students who take the survey rate both models by answering questions referring to Character A before those referring to Character B. If the student is given the Red Survey, Character A refers to a picture of the male model and Character B refers to a picture of the female model. If the student is given the Green Survey, the student rates the female model (now Character A) before the male model (now Character B).

Although the survey questions have undergone several iterations, both of our current survey colors (Red and Green) contain the same 34 questions. The first seventeen questions are about Character A, the latter seventeen about Character B. The two sets of seventeen questions are identical except for the character to which they refer. Of each seventeen questions, nine are Likert scale questions, three are multiple choice questions without significance in order, and five are free response questions.

The survey has gone through three iterations and all together, these iterations have been distributed to thirty students ranging from ages 8 to 15. The second version of the survey is worded slightly differently from the first version. The third version of the survey is the same as the second, except for the addition of two questions. Four students rated the first version of the survey, ten students rated the second version, and sixteen students rated the third version. The surveys thus provide 59 ratings for both the male and female surveys (due to a slight error in survey distribution, two participants only rated one model). To analyze the survey results, I created a coding key that assigns numerical values to the Likert scale questions. The results of questions that appear in all three versions of the survey were compiled (e.g. participant's perception of agent age, participant's perception of agent gender, etc.); otherwise, only answers to the third version of the survey were used.

Regarding agent gender, the majority of participants correctly perceived the male agent as being male; indeed, most participants (23 out of the 59) answered that the male agent was "definitely a boy," as opposed to "probably a boy" or "could be either." Although a plurality of participants (7) correctly identified the gender of the female agent, answers for the female agent were more ambiguous, with many participants perceiving the female agent as being gender ambiguous (2) or even a male (4). We posit that this observed result could be due to the reference pictures used in the survey. The female agent's ponytail cannot be seen clearly in the reference picture (because the picture is a front-facing portrait). Given the cartoonish style of the models, it is likely that participants use hairstyle (i.e. whether a model's hair is long or short) to pinpoint gender and are thus unable to correctly pinpoint the gender of the female agent.

We observed interesting results regarding participants' perception of agent age. In the Age of Participant v. Perceived Age of Male Agent graph shown below, there is a correlation of 0.5 and a slope of 0.668. In the Age of Participant v. Perceived Age of Female Agent graph, there is a correlation of 0.5 and a slope of 0.94. Thus, it appears that for both male and female agent, participants view the agents as generally being around the same age as them (the participants). That is, an eight-year old participant will perceive the agents as being 8-10 years old, just as a fifteen-year old participant will perceive the agents as being 14-15 years old. Although this effect was not intentional, it is certainly a development with exciting implications, as it means we have developed agents

that are age-flexible. Age-flexible agents could possibly expand the range of students with whom the agents can work.

Regarding agent ability to teach math, there was no significant difference between the ratings for the male agent and those for the female agent. Seventeen participants rated the male agent with a 0 or higher (where 0 corresponded to the answer "Ok, I guess" and 2, the maximum, corresponded to the answer "The best"). Thirteen participants rated the female agent with a 0 or higher, but overall there were fewer participants who answered this question for the female agent, as some participants had skipped over it.
Regarding the agents' test scores on their last math exams, participants seemed to think quite highly of both agents. There were 14 ratings of "4" for the female agent (corresponding to a grade of A) and 11 for the male agent. However, when asked whether the agents were good at learning, participants seemed to think that the female agent was better than the male agent: 9 participants scored the female agent with a 1 or higher (where 1 corresponded to the answer "Kind of" and 2 corresponded to the answer "Yes, he/she is"), whereas only 5 scored the male agent with a 1 or higher. Similarly, when asked whether the agents *liked* learning, more participants rated the female agent higher than the male agent – although the difference is not enough to be significant.

Regarding the popularity of the agents, participants seemed to believe that the male agent would be slightly more popular than the female agent. Nine participants scored the male agent with a 1 or 2 (where 1 corresponded to the answer "A little popular" and 2 corresponded to the answer "As one of the most popular kids), whereas only 5 participants scored the female agent with a 1 or 2. Indeed, whereas no one scored the male agent with a -2 (corresponding to the answer "As one of the least popular kids"), the female agent received two scores of -2. However, when questioned about agent trustworthiness, more participants scored the female as being "completely" trustworthy (11 as opposed to 5 for the male agent).

Overall, the results of the surveys are certainly interesting. However, I believe that more surveys will need to be administered to obtain a significantly large number of datum, particularly for questions that have (up until now) produced seemingly ambiguous ratings.

V. Conclusion

Much work remains to be done for the peer tutor agents. Although much time has been spent designing and creating the current models, they made need to be redesigned depending on additional survey results. Then, further work will need to be done to rig and animate the models – before the back end development of the agents

can begin. However, I believe my work this summer has established a good starting point and will expedite the development that is yet to come.

Works Cited

Baylor, A. L. (2009). Promoting motivation with virtual agents and avatars: role of visual presence and appearance . *Philosophical Transactions of the Royal Society* .

Baylor, A. L., & Kim, Y. (2004). Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role. *Intelligent Tutoring Systems* , 5-22.

Baylor, A., Shen, E., & Huang, X. (2003). Which Pedagogical Agent do Learners Choose? The Effects of Gender and Ethnicity . *Proceedings of World Conference on E-Learning in Corporate Government, Healthcare, and Higher Education 2003* , 1507-1510 .

Karacora, B., Dehghani, M., Kramer-Mertens, N., & Gratch, J. (2012). The Influence of Virtual Agents' Gender and Rapport on Enhancing Math Performance. *In Proceedings of the 34th Annual Meeting of the Cognitive Science Society* .

Silvervarg, A., Raukola, K., Haake, M., & Gulz, A. (2012). The Effect of Visual Gender on Abuse in Conversation with ECAs. *Intelligent Virtual Agents* , 153-160.