Effects of Virtual Human Presence on Task Performance

Catherine Zanbaka, Amy Ulinski, Paula Goolkasian, Larry F. Hodges
University of North Carolina at Charlotte
{czanbaka, aculinsk, pagoolka, lfhodges}@uncc.edu

Abstract

Do people respond to a virtual human in the same way they would to a real human? To answer this question, we designed a study which replicates a classical test of human-human social interaction. Particularly, we chose to replicate the social facilitation/social inhibition effects. Social facilitation/inhibition theory simply states that when in the presence of others, people perform learned tasks better and novel tasks worse. Participants first learned a task and were then randomly assigned to perform the same task or a novel task either alone, in the presence of a real human, or in the presence of a virtual human. Our results showed that people reacted to the virtual human similarly to the way they reacted to the real human. In particular, female participants performing in the presence of the virtual human demonstrated the social inhibition effect. We also found that more women learned the novel task when alone than when being observed by either a human or a virtual human.

Keywords: virtual humans, social facilitation, social inhibition, social influence, human computer interaction, avatars

1. Motivation

There are a growing number of virtual environments in which it appears that virtual humans would be useful in clinical or training applications. Case studies have indicated that exposure to a virtual audience can be helpful in the treatment of the fear of public speaking [1]. Other studies have shown that a virtual audience induces social anxiety and that the degree of anxiety is related to the type of feedback the audience provides [2]. In addition, virtual humans are currently being used in the development of virtual classroom scenarios for the assessment and rehabilitation of Attention Deficit Hyperactivity Disorder (ADHD) [3]. In all of these applications there is an underlying assumption that there exists a measurable similarity between a person's response to a virtual human and that person's response to a real human. In social psychology literature, one of the classical tests of how the presence of others affects task performance is social facilitation/inhibition [4, 5, 6]. Social facilitation/inhibition refers to performance enhancement of a simple or well learned task, and performance impairment of a complex or novel task, when done in the presence of others. Comparing the results of human-human interaction with human-virtual human interaction in the context of social facilitation/inhibition may provide evidence of whether learning to interact with virtual humans leads to better social interactions with real humans.

2. Related Work

Tripplett's first investigation of social influence in 1898 led to the development of many social facilitation theories and studies [6]. These social facilitation theories include Zajonc's *drive theory*, Cottrell, Wack, Sekerak, and Rittle's *socialization theory,* Sanders, Baron and Moore's *attentional conflict theory,* and Guerin and Innes' *social monitoring theory* [4, 7, 8, 9]. Numerous studies have been conducted to test the effect of the presence of others on task performance; Bond and Titus performed a meta-analysis of 241 social facilitation studies and summarized the results of these studies [5].

In a more recent study, Blascovich et al. conducted an experiment on social facilitation based on the biopsychosocial model of challenge and threat [10]. This model states that in goal-relevant situations involving affective and cognitive processes, challenge occurs when the resources from individual experiences meet demands of the situation, whereas for threat, these resources are insufficient to meet demands. In the experiment, the authors measured cardiovascular responses of participants while they either performed a novel or well-learned task alone or in the presence of others. They found that participants performing the well-learned task in the presence of others had an increased cardiac response and decreased vascular resistance, whereas participants performing a novel-learned task in the presence of others had an increased cardiac response and increased vascular resistance. Both of which fit the challenge and threat model. Participants performing the task alone, learned or unlearned, demonstrated no appreciable reactivity from baseline.

Work has also been done with avatars and virtual environments; Hoyt et al. assessed the utility of using immersive virtual environment technology for social psychological research [11]. They replicated Blascovich's social facilitation/inhibition study within an immersive virtual environment [10]. In their study, participants mastered one of two tasks and subsequently performed the mastered or non-mastered task either alone or in the presence of a virtual human audience whom they were led to believe were either computer-controlled *agents* or human-controlled *avatars.* The authors found that participants performing in the presence of avatars demonstrated classic social inhibition performance impairment effects relative to those performing alone or in the presence of agents. However, this study introduced a possible confound by having the research assistants physically present in the experimental room in the avatar audience condition. Additionally, the research data did not strongly indicate the effect of audience.

Pertaub et al. conducted a study to assess the extent to which social anxiety, especially fear of public speaking, was induced by a virtual audience and the extent of influence of degree of immersion [2]. In this study, subjects were randomly assigned to one of three groups: a negative (hostile, unappreciative) audience followed by a positive (friendly, appreciative) audience, a positive audience followed by a negative audience, or a static audience followed by static audience. The virtual audience consisted of eight formally dressed male avatars seated in a semi-circle around a table. The avatars continuously exhibited small twitching movements, blinking and shifting about in their chairs in order to make them look lifelike. Measures taken for this study included a personal report of confidence, an anxiety checklist, and the participants' own assessment of their performance on a scale from 1 to 100. The authors found that social anxiety was induced by the virtual audience and that the degree of anxiety experienced was directly related to the type of feedback the speaker received from the audience.

Rickenberg and Reeves ran an experiment to test the effects of different animated character presentation on user anxiety, task performance, and subjective evaluations of two commerce Web sites [12]. They found that the effects of monitoring and individual differences in the way a person thinks about control worked as they do in real life. Users felt more anxious when characters monitored their Web site work and this effect was strongest for users with an external control orientation. Monitoring characters also decreased task performance, but increased trust in Web site content.

3. Experimental Study

The intent of this study was to replicate classical tests of human social interaction, substituting a virtual human into a role usually occupied by a real person. We hypothesized that the presence of a virtual human will have a similar effect as the presence of a real human on task performance. Social facilitation/inhibition theory simply states that people perform simple tasks better when in the presence of others, and complex tasks worse.

3.1 Participants

The participants were 89 students from the University of North Carolina at Charlotte. Volunteers were recruited from the psychology department subject pool, through fliers, and by word of mouth. Volunteers from the psychology pool received extra credit points towards their psychology class grade. Data from five participants who did not learn to criterion in the learning phase were thrown out, leaving data from 84 participants, 44 males and 40 females, to be used in analysis. The learning criterion was to perform above chance, namely better than 50% correct.

3.2 Tasks

The tasks were adapted from Blascovich and Hoyt's pattern recognition and categorization tasks; the latter of which was an adapted version of Maddox and Ashby's perceptual boundary exercise [10, 11, 13]. These tasks were designed to be equivalent in difficulty, but unique visually and in their methods of operation.

For the categorization task, participants were presented with two numbers on a computer screen and were required to determine whether the numbers belonged to one of two categories, *Group 1* or *Group 2*. The numbers in *Group 1* ranged from 25 to 68, had a mean = 46.5, and SD = 12.8. Group 2 numbers ranged from 69 to 112, had a mean = 90.5, and SD = 12.8.

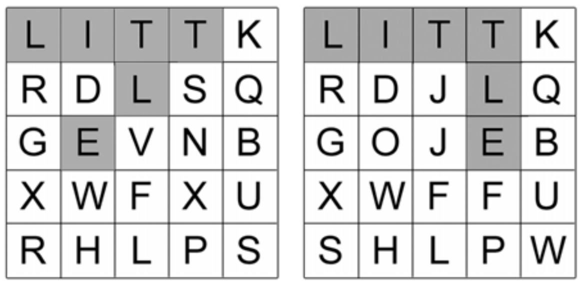

Fig. 1 Left: Incorrect pattern in pattern recognition task. Right: Correct pattern in pattern recognition task.

In the pattern recognition task, participants viewed a 5x5 letter matrix on the screen and a word in the matrix was highlighted one letter at a time. The participants' task was to determine whether the word was highlighted in the correct pattern or not. A correct pattern was one that depicted right angles only, whereas an incorrect pattern had obtuse or acute angles (Figure 1).

The pattern recognition task consisted of 10 trials per block and the number categorization task consisted of 20 trials per block. The stimuli for both tasks were randomly presented.

3.3 Apparatus

A Pentium IV 2.4 GHz Dell PC with an nVidia GeForce4 Ti 4200 graphics card served as the graphics generator for the virtual human. The graphics were rendered with OpenGL then projected using a Sony VPL-CX5 data projector.

Stimulus presentation and data collection were controlled by SuperLab Pro running on a Sony Vaio Pentium IV 2.0 GHz laptop. The laptop was attached to a 17 inch flatscreen monitor (Figures 3 and 4).

Fig. 2 Screenshot of virtual audience

We used one of Haptek Corporation's interactive 3-D characters for our virtual human [14] (Figure 2). Haptek also has a library which allowed us to create our own realistic animations and behaviors.

The nature of this experiment requires that the audience, both human and virtual, exhibit only non-verbal gestures and behaviors. In order to make our virtual human, Diana, human-like, we modeled her actions based on the non-verbal behaviors of the human audience in this experiment and executed them at random. These behaviors included coughing, sniffling, yawning, looking around, clearing throat, and shifting in her chair. In addition, Diana displays life-like behaviors such as breathing, blinking, and other subtle gestures. Two speakers at the bottom of the projection screen were used to output the various sounds from Diana (coughing, sniffing, etc.).

3.4 Experiment Design and Procedures

A 2x3x2 mixed design was used. The study manipulated three independent variables: task type, audience type, and phase (Learning vs. Testing). The first two variables were between subjects and the last was within subjects. There was a total of six experimental groups: 2 (Task Type: unlearned or well-learned) x 3 (Audience Type: Human Audience, Virtual Human Audience, or Alone). The dependent variables were task performance and reaction time.

Participants were randomly assigned to one of six experimental conditions, performing:

1. A *learned* task *alone*
2. A *learned* task in the presence of a *virtual human*
3. A *learned* task in the presence of a *real human*
4. A *novel* task *alone*
5. A *novel* task in the presence of a *virtual human*
6. A *novel* task in the presence of a *real human*

The experiment consisted of two phases: a learning phase and a testing phase. The experiment took approximately half an hour to complete.

Fig. 3 Participant with virtual audience

Fig. 4 Participant with real audience

Learning Phase. In a separate room from the testing room, participants filled out an informed consent form. They were then given instructions regarding the experimental procedures. The participants were taken to the testing room where they sat in a chair at a desk, facing the stimuli screen, with a clearly marked two-button mouse in front of them. They were then instructed on how to perform both tasks. Specifically, they were told the objectives of the tasks, shown how to provide their responses on the mouse, and led through a sample trial for each task. This familiarized them with the procedures and the tones: a high-pitched tone following a correct response and a low-pitched tone following an incorrect response. Participants were instructed to complete five blocks of the pattern recognition task, and to return to the main area of the lab after they are done. The experimenter left the room. There was a five minute rest period between the learning and the testing phase of the experiment.

Testing Phase. The participant returned to the testing room along with the experimenter. Participants were randomly assigned to perform the learned pattern recognition task or the unlearned categorization task either alone, in the presence of a human audience, Amy, or in the presence of a virtual human audience, Diana. Participants completed five blocks of the assigned task. For all the conditions requiring a virtual human, the participants were told that Diana was able to "see" them via a camera mounted in the testing room. In the virtual human condition, Diana was projected on a screen to the right side of the participant and at a 135 degree angle from the stimuli screen such that she could "observe" the participant and the testing screen. In the human audience condition, Amy was seated in a position equivalent to the position of Diana, such that she too can observe the participant and the testing screen (Figures 3 and 4). Diana exhibits nonverbal responses that indicate some genuine interest in what the participant is doing. Finally, in the alone condition, the setting of the testing room was the same as in the learning phase. Upon completion of the testing phase, participants returned to the main area of the lab, where they filled out a short questionnaire, and were debriefed and thanked for their participation.

4. Measures

Task Performance Data. Accuracy of responses and reaction times were automatically logged for each participant. Task performance data were computed by summing the number of correct responses across the five blocks and converting these to percentages. Reaction times were measured starting from the viewing of the stimulus and terminated by the participant's mouse press. There was no limit set for response time. The reaction time data were the means of the trimmed response times.

Task Novelty, Task Anxiety, and Copresence Scales. A post experiment questionnaire was used to measure the participants' degree of task novelty, task anxiety, and co-presence with the virtual human. This questionnaire was developed by Hoyt and Blascovich [11]. Participants indicated on a seven-point Likert-type scale the extent to which the task they performed in the testing phase was novel, the extent to which they felt anxious about performing the task and, in the human audience and virtual audience conditions, the extent to which they felt like there was another person co-present with them. The questionnaire also included questions about participant demographic characteristics.

Debriefing Data. After the experiment, participants were asked subjective open-ended questions about their experience. They were asked what they thought about the task. Those with a real human or virtual human condition were asked if they felt that they were being watched by Diana or Amy. Participants with a virtual human condition were also asked about Diana's appearance and animations.

5. Results

Learning Phase. For the learning phase data of the experiment, the main effects of audience and task type were not significant. There were also no interaction effects for audience and task type.

Task Novelty. Participants who performed the novel task during the testing phase reported that the task was significantly more novel (M=3.60, SD=1.74) than those who performed the learned task in the testing phase (M=2.36, SD=1.29), p<0.001. Figure 5 shows a boxplot of participants' mean scores on the task novelty questionnaire.

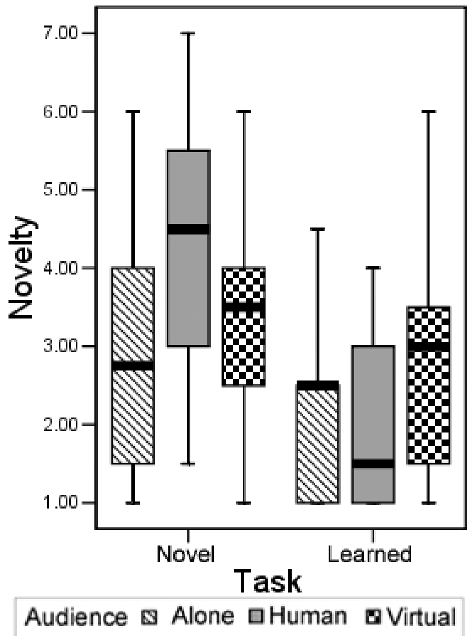

Fig. 5 Boxplot of participants' scores on task novelty questionnaire (1- most novel -> 7-least novel). The "box" area of the boxplot contains the 50% of the values that fall between the 25th and 75th percentiles. The heavy black line through the box is the median. The "whiskers," or lines that extend from the box, show the spread of scores from highest to lowest.

Task Anxiety. Male participants who performed the novel task during the testing phase reported significantly higher levels of task anxiety (M=2.83, SD=1.42) compared to those in the learned condition (M=1.79, SD=0.93), p<0.01. The effect including males and females was near significance with p=0.054, and the effect with females alone was not significant p=0.812 (Figure 6).

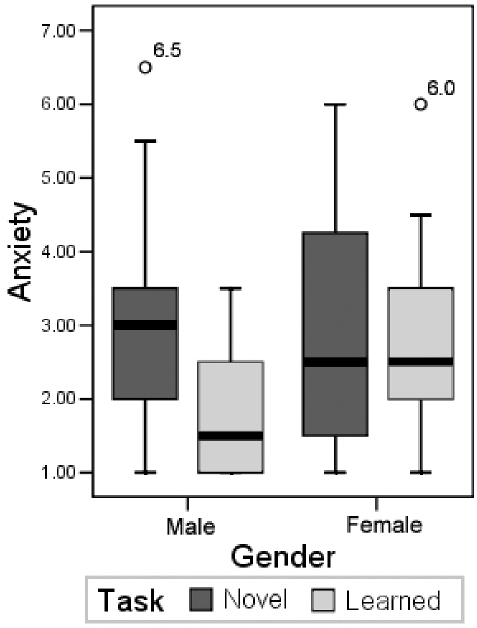

Fig. 6 Boxplot of participants' scores on task anxiety questionnaire (1- most anxious -> 7-least anxious). See Fig. 5 for an explanation of boxplots. Circles represent extreme values and outliers.

Copresence. Participants in the virtual human condition (M=2.89, SD=1.12, N=28) had a slightly higher copresence rating than those in the human condition (M=2.66, SD=1.00, N=29). However, the copresence rating for both conditions was relatively low, which was probably due to lack of interaction between the participants and the human or virtual human audience.

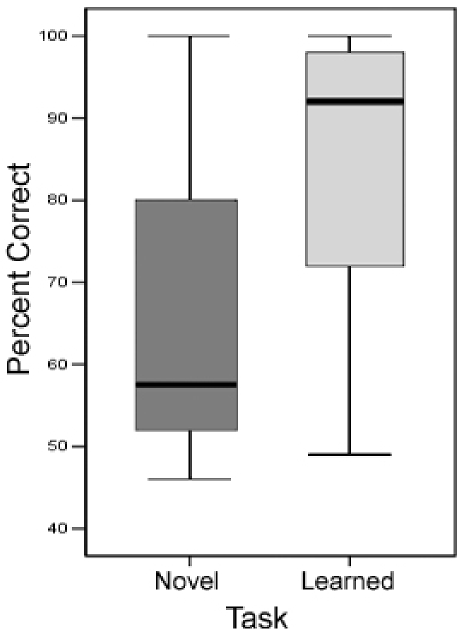

Fig. 7 Participant performance on novel and learned tasks. There was a significant main effect of task type. See Fig. 5 for an explanation of boxplots.

5.1 Task Performance Data

Percent Correct Data. As expected, there was a significant main effect for task type. Participants in the learned task condition (M=0.85, SD=0.14, N=42) performed significantly better than those in the novel task condition (M=0.66, SD=0.18, N=42), $p<0.001$ (Figure 7).

Although the main effect of audience was not significant, the performance data results for the human audience (M=0.75, SD=0.19, N=29) and virtual audience (M=0.73, SD=0.18, N=28) conditions were similar (Figure 8). Furthermore, a block by block analysis of the task performance data revealed a significant main effect of audience type in the novel task condition in blocks 2 and 3 for female participants, $F(2, 16) = 3.80$, $p<0.05$. All blocks considered, the effect was not significant, we believe this is due to a learning effect in blocks 4 and 5 and initial adaptation in block 1.

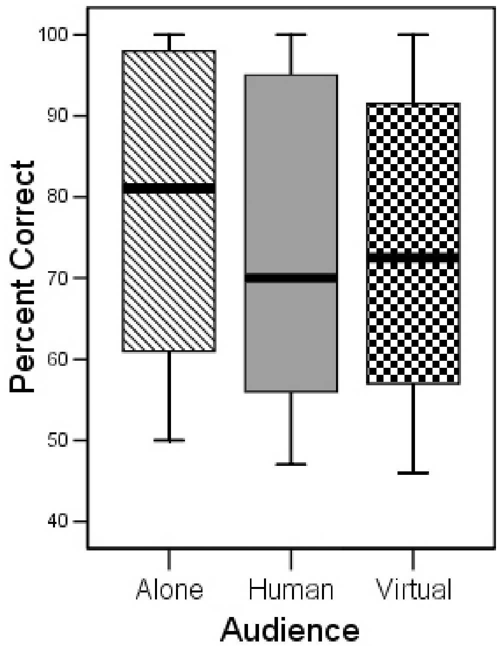

Fig. 8 Participant performance in the three audience conditions. See Fig. 5 for an explanation of boxplots.

The interaction effect of task type by audience type was not significant overall. There was a decline in performance on the novel task as a function of audience type, whereas performance on the learned task was unaffected by audience type. Further analysis looking at men and women participants separately did show a significant interaction effect of task type by audience type for female participants in blocks 2 and 3 [$F(2, 34) = 3.49$, $p<0.05$]. Figure 9 highlights the differences in male and female participants' performance on the novel task.

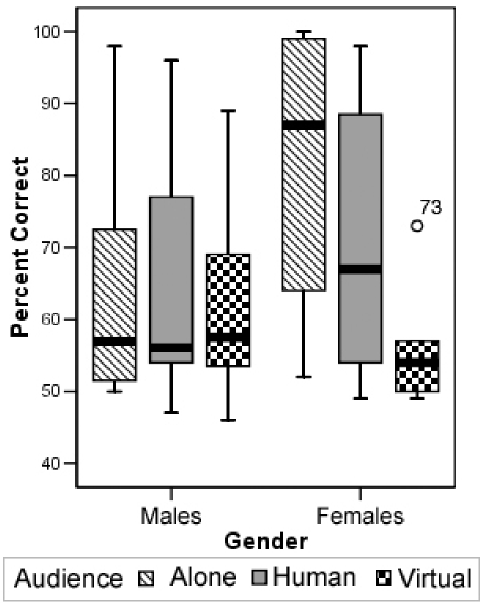

Fig. 9 Male and female participant performance on novel task. See Fig. 5 for an explanation of boxplots. Circles represent extreme values and outliers.

Reaction Time Data. The main effects of audience type and task type were not significant for the reaction time data. However, Figure 10 shows that there was a strong trend [$F(2, 16) = 3.22$, $p = 0.067$] for women in the novel task condition to respond faster when alone (M=1362 ms, SD=518, N=6) than in the presence of a human (M=1955 ms, SD=605, N=7) or a virtual human (M=2150 ms, SD=551, N=6).

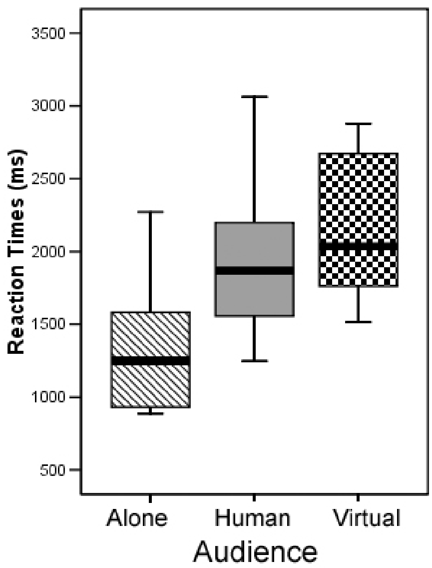

Fig. 10 Reaction time data for women on the novel task. See Fig. 5 for an explanation of boxplots.

Learning effect. In order to emphasize the interaction between task type and audience type, we used a more sensitive measure for participants who learned the task well (performed above average, M = 75.8, SD=18.7, N=84) and those who did not. The percentage correct scores were broken into two groups: those that learned the task well (performed > 76%) and those that did not learn the task well (performed <= 76%).

The results of a task type by audience type analysis showed a significant interaction effect for female participants who learned the task versus those who did not [$F(2, 34) = 3.49$ $p<0.05$].

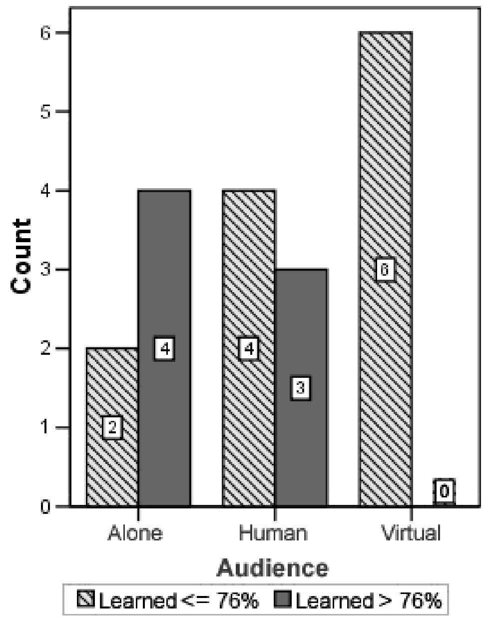

Fig. 11 Proportion of women who learned the novel task varied significantly with audience type. Numbers on the bars indicate count.

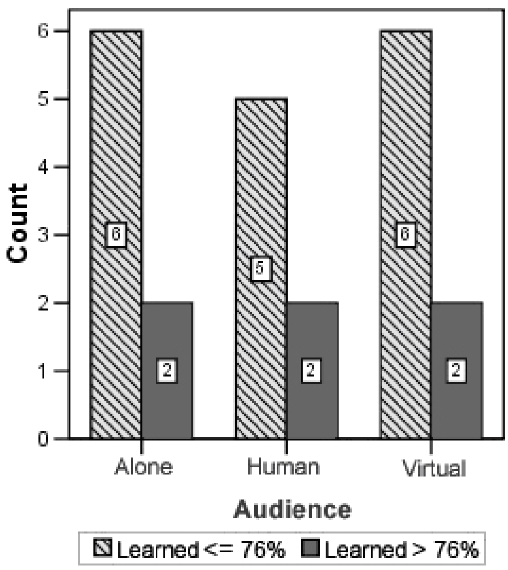

Fig. 12 Proportion of men who learned the novel task did not vary with audience type. Numbers on the bars indicate count.

Figure 11 shows that the proportion of women who learned versus those who did not learn varied by audience type. More women learned in the alone condition than in the human or virtual human conditions, this difference was significant, χ^2 (2) = 5.90, p=0.05.

Figure 12 shows that the same number of men learned the novel task across the three audience types [χ^2 (2) = 0.03, p=0.98].

Debriefing Trends. Analysis of the post-experiment interviews resulted in the following trends:

- 69.2% of the participants in the novel-human condition, 61.5% of the participants in the novel-virtual human and learned-human conditions, and 71.4% of the participants in the learned-virtual human condition felt that they were being watched by Diana/Amy.
- When asked: "What percentage of the time did you feel Diana/Amy was watching you?" The mean response of the participants in the novel-human condition was 68.1% (SD = 42.2), 45.4% in the novel-virtual condition (SD = 40.2), 54.9% in the learned-human condition (SD = 40.0), and 47.7% in the learned-virtual human condition (SD = 35.9). The main effect for task type was near significance with p=0.058.

6. Discussion

In this study, we were able to replicate the social inhibition effect with our female participants using a virtual human. Women performed worse on the novel task in the presence of our virtual human, Diana, than when they performed the task alone. The presence of Diana had more influence on the female participants while they completed the novel task than the learned task. Female participants did not seem to be affected by the presence of Diana during the learned task, whereas during the novel task, Diana hindered participants' performance. Male participants, however, were neither affected by the presence of Diana during the learned nor the novel task. The human observer, Amy, affected the participants' performance in a similar manner as Diana.

Furthermore, we found that the proportion of women who learned the novel task varied by audience type. More women learned the novel task when alone than when being observed by either a human or a virtual human. Men, on the other hand, were not affected by the audience type. The same number of men learned the novel task when alone or while observed by a human or a virtual human.

We were unable to successfully replicate the social facilitation effect. Most likely this was due to a ceiling effect which is a common problem in social facilitation research [5]. Having learned the correct pattern in the learning phase left little room for improvement and resulted in inadequate variability in the participants' scores in the testing phase.

7. Conclusions and future work

Overall, our data shows that participants responded to Diana similarly to the way they responded to the female human observer. The participants' reaction to Diana in the virtual audience condition, shown by performance on the novel task, was similar to that of their reaction to a real human for a novel task.

The results from this study suggest that virtual humans may in fact be useful in clinical or training applications where exposure to a virtual audience is necessary. The fact that participants' reaction to the virtual audience was comparable to their reaction to the real audience indicates that there is a measurable similarity between a person's response to a virtual human and that person's response to a real human. This in turn implies that learning to interact with virtual humans may lead to improved social interactions with real humans.

The following comments from the participants during the debriefing session illustrate that participants related to Diana as though she was a real human:

- "The sneezing did make me relate a little to when I had a cold and I felt a little sorry for her, even though she was a virtual human."
- "I felt like she was sighing when I got a problem wrong, like 'oh man, you almost got it!'".
- "Whenever she leaned forward, it seemed like she was getting impatient with me."
- "Diana looked pretty real, she reminded me of a friend I know."
- "Diana actually made me more relaxed or maybe it was just the fact that it felt like there was more than just me in the room. I didn't have to think as hard with her in the room… she took my mind off the numbers."
- "Diana is a beautiful person. I would like to get to know her better, had she been a real person. It's been quite a realistic representation. Great Job."
- "Diana was quite lifelike in her movements. Having her presence there was like having a person in the room, although I did understand that she was computer-generated. Pretty cool."
- "It was like taking a test; it's comforting knowing there is someone else in the room, but when you're the only one left you stress. So Diana made it more relaxed, but at the same time distracted because you're looking around trying to see what she is doing."

There are a number of remaining questions that could be answered with further research. Future experiments, varying the gender of the human and virtual human observer, could investigate why women in this study exhibited the social inhibition effect, whereas men did not. Additionally, potential studies could show the effect of a more immersive environment, such as one utilizing a head mounted display, on social facilitation/inhibition. Possible experiments may also examine how performance on a more interactive task is affected by the presence of a virtual human. Finally, other aspects of social interaction, such as persuasion and obedience effects, can also be studied with virtual humans.

References

[1] Anderson, P., Rothbaum, B.O., & Hodges, L. F. (2003). Virtual reality exposure in the treatment of social anxiety. *Cognitive and Behavioral Practice*, 10, 240-247.

[2] Pertaub, D. P., Slater, M., & Barker, C. (2002). An experiment on pubic speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and Virtual Environments*, 11, 68-78.

[3] Rizzo, A. et. al. (2000). Virtual environment applications in clinical neuropsychology. *Proceedings IEEE VR*, pp.63-70.

[4] Zajonc, R.B. (1965). Social Facilitation. *Science 149,* 269.

[5] Bond, C.F., & Titus, L.J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin, 94,* 265-292.

[6] Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *Journal of Psychology, 9,* 507-533.

[7] Cottrell, N. B., Wack, D.L., Sekerak, G.J., & Rittle, R.H. (1968). Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of Personality and Social Psychology, 9,* 245-250.

[8] Sanders, G. S., Baron, R. s., & Moore, D. L. Distraction and social comparison as mediators of social facilitation effects. *Journal of Experimental Social Psychology,* 1978, 14, 291 – 303.

[9] Guerin, B., & J. M. Innes. (1982). Social facilitation and social monitoring: A new look at Zajonc's mere presence hypothesis. *British Journal of Social Psychology, 21,* 7-18.

[10] Blascovich, J., Mendes, W. B., Hunter, S. B., & Salomon, K. (1999). Social “facilitation” as challenge and threat. *Journal of Personality and Social Psychology*, 31, 422-429.

[11] Hoyt, C., Blascovich, J., Swinth, K. Social inhibition in immersive virtual environments. *Presence: Teleoperators and Virtual Environments*, 12, 2, p.183-195.

[12] Rickenberg R., Reeves B. The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems*, p.49-56.

[13] Maddox, W. T. & Ashby, F.G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception & Performance*, 22, 795-817.

[14] Haptek Corporation: http://www.haptek.com/