# The Effects of Pedagogical Agent Voice and Animation on Learning, Motivation and Perceived Persona

Amy L. Baylor, Ph.D
Department of Educational Psychology and Learning Systems
Director, PALS (Pedagogical Agent Learning Systems) Research Laboratory http://pals.fsu.edu
Florida State University
baylor@coe.fsu.edu

Jeeheon Ryu
Department of Educational Psychology and Learning Systems
Florida State University
jjr7148@garnet.fsu.edu

E Shen
Department of Educational Psychology and Learning Systems
Florida State University
ess0086@garnet.fsu.edu

**Abstract**: This 2x2 factor experimental study investigated the effects of pedagogical agent voice (human, machine-generated) and animation (present, absent) on learning, motivation, and perceived pedagogical agent persona. 80 participants were randomly assigned into four conditions. A main effect for animation indicated that participants learned significantly more when the agent was animated even though they also reported that the agent was significantly *less* facilitative of learning when it was animated. Similarly, a main effect for animation indicated that the participants were significantly less motivated about the topic when the agent was animated. In addition, a significant interaction for motivation revealed that participants were more motivated if the agent was either animated with machine-generated voice or not animated with a human voice. A main effect for voice indicated that participants rated the agent persona as more engaging and human-like when it had a human voice.

## Introduction

Research with lifelike pedagogical agents has indicated positive effects on learners' attitude toward learning and performance (Baylor, 2002a, 2002b; Baylor & Ryu, 2003a; Moreno, Mayer, Spires, & Lester, 2001). The assumption underlying the advantages of using pedagogical agents is that they can facilitate learners' interaction with the computer by rendering the system more communicative. Among computer science researchers and developers there has been increasing attention toward developing more lifelike pedagogical agents. However, it is challenging to create lifelike pedagogical with natural communication and behaviors resembling human-to-human conversation. Since human communication involves synchronized speech and gesture (Wachsmuth, 1999), for a pedagogical agent to be an effective communicator it should be able to communicate through artfully combined aural and expressive communication (Lester, Towns, & FitzGerald, 1999; Lester, Voerman, Towns, & Callaway, 1999).

Prior research has indicated that voice has a superiority effect to visual appearance for communication in computer-based media (Mayer & Moreno, 1998; Moreno & Mayer, 1999), and pedagogical agents in particular (Atkinson, 2002; Moreno et al., 2001). According to communications research, voice provides a powerful indicator of meaning. Mayer and Moreno have suggested through results of their research that the presence of a voice promotes deeper processing (Mayer, Sobko, Fennell, & Mautone, 2002; Moreno & Mayer, 2002; Moreno, Mayer, & Lester, 2000; Moreno et al., 2001). However, many studies examining the effect of voice have focused on a presence/absence of voice rather than the *nature* of the voice (e.g., machine generated vs. human). While there have been anecdotal reports that a human voice is preferable for agents (Atkinson, 2002) there have been no systematic controlled studies comparing machine-generated and human voice for agents.

Another key characteristic for a lifelike pedagogical agent is its capacity for animated behaviors, including emotional expression and deictic gestures. Several studies have reported that an animated agent leads to better interaction between the agent and learner (Cassell et al., 1994; Doyle, 1999; FitzGerald, 2000). Further, agents' gaze and/or gesture can guide the learners' attention (Rickel, 2001), and provide clarity of meaning for a spoken utterance. Lester and colleagues have suggested that if lifelike pedagogical agents can provide contextually appropriate facial expressions and expressive gestures, they can promote learning and motivation (Lester, Towns et al., 1999; Towns, FitzGerald, & Lester, 1998). These results support Atkinson and colleagues (2002) findings that participants' learning was improved when agents were animated. They concluded that animation enabled the learners to dedicate their limited cognitive resources to the task of understanding the underlying conceptual segments of worked-out examples.

Yet, such animations also work *together with* the agent's verbal communication to promote interaction. For example, according to Link (2001), participants relied on both the linguistic expression (from the verbal modality) and the mouth curve (from the nonverbal modality). This finding is consistent that information from various modalities are integrated during perception. That is, emotional identification will be most accurate when information is given from both verbal and nonverbal communication. Although vocal feature is superior to nonverbal communication, if information from verbal communication is ambiguous, learner will rely on nonverbal communication to identify the emotion. Although voice may be superior in comparison to visual features of agents for communication, it seems that learners also do significantly rely on the visual features in order to more clearly interpret the meaning of the communication or emotion (Lester, Towns, Callaway, Voerman, & FitzGerald, 2000).

These studies overall suggest that communication between agent and learner will be most accurate and meaningful when information is provided both verbally (e.g., through voice) and visually (e.g., through animation). Given the evidence for the importance of providing voice and animation for lifelike pedagogical agents, it seems that providing both voice and animation would be critical; however, these two factors –particularly with respect to the *nature* of the agent's voice -- have not yet been systematically examined together within the same study. Therefore, the purpose of this study was to examine the interaction of voice (machine-generated vs. human voice) and animation (presence/absence) on learning, motivation, and the perception of agent persona.

## Methods

### Participants and Experimental Design

80 undergraduate students (17.2% male and 82.8% female) enrolled in an educational technology course in a public southeast university participated in this study. The average age of the participants was 20.32 (SD=3.77). This study employed a 2X2 factorial design with agent voice (machine-generated, human) and animation (presence, absence) as the two factors. The participants were randomly assigned to one of 4 conditions: human voice with animation, human voice without animation, machine-generated voice with animation, and machine-generated voice without animation.

### Agent Format

Four three-dimensional pedagogical agents were developed in Poser to represent the four experimental conditions, varying voice and animation. Each agent had identical scripts, but was narrated by either the Microsoft Agent machine-generated voice, or a recorded human voice. Given that prior research has suggested that the optimal combination is for voice narration to be presented together with the corresponding text as a way to reduce cognitive load and improve learning (Moreno & Mayer, 2000; Mousavi, Low, & Sweller, 1995), a text bubble was presented together with the agent's spoken narration. In all conditions, the agent spoke with lip-synchronization. In terms of animation, the agent either was presented as a static image (as shown below) or as animated, with gestures and emotional expression. The factorial design is shown in Figure 1 with the sample sizes for each condition.

| | | Voice | |
|---|---|---|---|
| | *Sample Size* | Machine-generated | Human |
| **Animation** | Present | 20 | 21 |
| | Absent | 20 | 19 |

**Figure 1: Agent image and Factorial Combinations of Voice and Animation with Sample Sizes**

**Procedure**

During the experiment, the students participated the MIMIC (Multiple Intelligent Mentors Instructing Collaboratively) agent-based research environment, designed to help students learn how to develop instructional plan based on a case study of a 13-year old girl trying to learn the economic concepts of supply and demand. The participants were able to move among instructional planning phases (Case Study, Blueprints, Plan, Assessment) by clicking navigation buttons. When the participant entered each phase, the agent, referred to as "Rick," provided instructional information about the phase, and the participant could request additional information from the agent on his/her own initiative at any time. Once participants completed the four phases, they answered post-test questions in the dependent measures. It took approximately 75 minutes on average for students to complete the task.

**Measures**

Measures for the four sets of dependent variables (learning, facilitating leaning, motivation, and perceived agent persona) are described in this section.

***Learning***

The dependent variable of learning was measured in terms of recall and transfer. To assess *recall*, participants were asked to "List all of the information that you can recall from using the program. List it in the order that you recall it. List as much information as possible." Each recall answer was decomposed into idea units (a procedure implemented by Mayer & Gallini, 1990). Credit (one point) was given for each idea in the student's answer that conveys the same meaning as an idea unit from the program. Incomplete ideas were acceptable. Three researchers coded a sample of the data until a criterion of $r$>.90 was reached to establish inter-rater reliability. Once there was agreement in the coding methods, one researcher performed the coding.

To assess *transfer*, participants were provided with the following question:

> Applying what you've learned, develop an instructional plan for the following scenario: Imagine that you are a sixth grade teacher of a mathematics class. Your principal informs you that a member of the president's advisory committee will be visiting next week and wants to see an example of your instructional about multiple of fractions.

Each instructional plan was scored according to a scale (where 1=poor and 5=excellent) that evaluated the overall plan in terms of how well the participant applied his/her knowledge of instructional planning to this particular situation. Three researchers met and together discussed what characterized a score of 1 through 5 while evaluating sample plans. Disagreements were resolved through discussion. Following that, each researcher independently scored 10 instructional plans. Inter-rater reliability between the two researchers was established at $r > .90$ for the ten instructional plans. One of the researchers then scored the remainder of the instructional plans using the same scale. In scoring each instructional plan, the researchers were blind as to which tool was used by the participant.

***Agent facilitation of learning***

Seven Likert-scale items, ranging from 1 for "strongly disagree", 2 for "disagree", 3 for "neutral", 4 for "agree", and 5 for "strongly agree" were used to measure how the agent (referred to as "Rick") facilitated the learning process. These questions were as follows:

1. *It was easy to learn from Rick.*
2. *Rick did not interfere with my understanding of the content. (*scored in reverse)*

3. *Rick believed I was knowledgeable.*
4. *Rick was useful.*
5. *Rick was effective on convey ideas.*
6. *Rick was precise in providing advice.*
7. *Rick helped me concentrate.*

The Cronbach's alpha for the overall reliability of the seven items was assessed at 0.79.

***Motivation***

To assess motivation, questions were asked regarding self-efficacy, disposition, and satisfaction. A single item was used to measure students' *self-efficacy toward instructional planning* based on Bandura and Schunk's (1981) guidelines for specificity, given that self-efficacy is the degree to which one feels capable of performing a particular task at certain designated levels (Bandura, 1986). All participants were asked, "How sure are you that you can write a lesson plan?" on a scale from 1 being not sure to 9 being very sure.

To assess *disposition regarding instructional planning,* all participants were asked to write two adjectives to "Describe what you think about instructional planning." This method was employed to obtain the participants' personal affect regarding instructional planning as opposed to the response set that could bias them to choose more favorable adjectives if presented in a list. These adjectives were coded according to three levels: as -1 if both were negative, as 0 if 1 was negative and the other positive, and as +1 if both were positive. Two raters coded the items independently and interrater reliability was established at .96. There were only two disagreements about two sets of adjectives, which were resolved through discussion. Two adjective pairs were discarded because they could not be classified. The concurrent validity of this measure was supported in Kitsantas and Baylor (2001) by a significant positive correlation between initial disposition and initial self-efficacy scores. Prior research has shown that self-efficacious students generally have positive affect (Bandura, 1986).

To assess *satisfaction regarding instructional planning performance*, following the development of each instructional plan, participants were asked to rate their satisfaction with their performance on the instructional plan on a Likert scale of 1-5 where 1=Not at all satisfied and 5=Extremely satisfied.

***Perceived agent persona***

Three sub-scales from the API (Agent Persona Instrument) were used, to assess the agent in terms of how much it was *credible*, *engaging*, and *human-like*. The Cronbach's alpha for the overall reliability of the instrument was assessed at 0.97. See Baylor & Ryu (2003b) for more information regarding this instrument.

## Results

Four two-way MANOVAs were conducted to examine how agent voice and animation affected learning, agent facilitation of learning, motivation, and perceived agent persona. The alpha level was set at 0.05 to determine statistical significance.

**Learning**

Learning was analyzed through a two-factor MANOVA, with recall and transfer as the two dependent measures, and with animation (present, absent) and voice (human, machine-generated) as the two between-subject factors. The two-factor MANOVA indicated that there was an overall positive effect of the presence of animation on learning, Wilk's Lambda = .93, $F(2,80) = 3.07$, $p=.05$. Follow-up univariate analyses (ANOVA) indicated that differences approaching statistical significance occurred in both dependent measures indicating the positive affect of animation: for transfer, $F(1,81)=2.89$, $p=.09$ and for recall, $F(1,81)=3.26$, $p=.07$.

**Agent facilitation of learning**

Agent facilitation of learning was analyzed through a two-factor MANOVA, with seven items of facilitating learning as dependent measures, and with animation (present, absent) and voice (human, machine-generated) as the two between-subject factors. The two-factor MANOVA indicated that there was an overall *negative* effect of the presence of animation on agent facilitation of learning, Wilk's Lambda = .83, $F(2,80) = 2.16$, $p=.05$.

**Motivation**

Motivation was analyzed through a two-factor MANOVA, with self-efficacy, disposition, and satisfaction as

the three dependent measures, and with animation (present, absent) and voice (human, machine-generated) as the two between-subject factors. The two-factor MANOVA indicated that there was an overall *negative* effect of the presence of animation on motivation, Wilk's Lambda = .91, $F(2,82) = 4.06$, $p<.05$. Follow-up univariate analyses (ANOVA) indicated that there was a significant difference for satisfaction with performance $F(1,83) = 7.35$, $p<.01$, and one approaching significance for self-efficacy, $F(1,83) = 2.86$, $p=.09$. In both cases, the presence of animation negatively affected satisfaction and self-efficacy. In addition there was a significant interaction Wilk's Lambda = .90, $F(1,82) = 4.33$, $p<.05$ between the voice and animation factors on motivation. The interaction indicated that learners were more motivated when the agent had no animation with human voice or with animation and machine-generated voice.

**Perceived Agent Persona**

Perceived agent persona was analyzed through a two-factor MANOVA, with the four sub-scales as the dependent measures, and with animation (present, absent) and voice (human, machine-generated) as the two between-subject factors. The two-factor MANOVA indicated that there was an overall positive effect of the human voice on perceived agent persona, Wilk's Lambda = .86, $F(4,73) = 2.91$, $p<.05$. Follow-up univariate analyses (ANOVA) indicated that significant differences occurred in two of the three sub-scales: *Engaging*, and *Human-like*. Specifically, those with the human voice ($M=3.64$, $SD=0.72$) reported the agent as more Engaging than those in Machine-generated Voice ($M=3.12$, $SD=0.86$) with $F(1,76)=8.58$, $p<0.05$. Also, students in the human voice condition ($M=3.34$, $SD=0.70$) reported that the agent was more Human-like than those in the machine-generated voice condition ($M=2.80$, $SD=0.97$) with $F(1, 76)=8.18$, $p<0.05$.

## Discussion

Overall, results suggest that that the presence of basic animation for pedagogical agents may help promote learning but also may negatively impact motivation toward the content. Interestingly, while the presence of animation positively contributed to the participants' learning, they did not perceive it as such -- the significant main effect of animation on agent facilitation of learning revealed that the presence of animation led the agent to be perceived as significantly *less* facilitative of learning.

As expected, a *human voice* was found to be particularly valuable as compared to machine-generated voice in promoting outcomes related to perceived pedagogical agent persona. This finding adds to prior research which has substantiated the value of having an agent voice present (as opposed to absent) in order to promote anthropomorphism (Nass & Steuer, 1993; Reeves & Nass, 1996), and learning (Atkinson, 2002; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Moreno et al., 2001). Here, it was found that the human-like voice positively impacted how the learner perceived the agent as engaging and human-like. These outcomes are important as developers strive to develop pedagogical agents that are perceived as more like real human mentors.

Regarding learning, animation may have facilitated learning because the visibly-displayed emotional expressions and gestures relating to the content focused the learner on the meaning and context of the information, as suggested by Lester and others (Lester, Towns et al., 1999; Towns et al., 1998). Another explanation is that the presence of animation forced the learner to pay more attention to what the agent was saying (because it was moving), thus suggesting that a little distraction may be beneficial to focus the learner to pay attention. Along this line, learners reported that the agent was significantly *less* facilitative of learning when it was animated (even though they were learning more), suggesting that the agent animation increased the learners' cognitive effort toward the learning process, leading them to be less satisfied (i.e., it didn't make it "easier" for them). Future research should try to better ascertain learners' metacognitive awareness of their learning as it relates to agent animation by having them quantitatively predict how much they believe they have learned and compare it to their actual learning.

In terms of motivation, it was found that the presence of animation (as compared to a static agent image) provided a negative effect. For motivation, particularly in terms of learner self-efficacy and satisfaction regarding performance, perhaps the agent animation distracted learners, as found earlier by Koda & Maes (1996), making them feel less confident and satisfied. In addition, a significant interaction for motivation revealed that participants were more motivated if the agent was either animated with machine-generated voice or not animated with a human voice. This interaction suggests the importance for the agent to be human-like, but not *too* human-like in order to promote motivation. As suggested by Norman (1997), if the learner's expectation toward the agent's capabilities and human-likeness are too high (e.g., human voice with animation) or too low (e.g., machine-generated voice and no animation), this could negatively impact motivation.

It is important to note that more research is necessary to delineate the *nature* of the animation for promoting

learning and motivation. Here, only basic emotional expression and deictic gestures were provided from only the waist and above. Along this line, Baylor & Ryu (2003a), using full-bodied agents, found that the presence of animation significantly contributed to perceived pedagogical agent persona, which was not replicated here. This could be due to the more limited animation in the current study as well as the different content as presented by the agents. Future research should also consider the effects of more elaborate and full-body animations.

## References


Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology, 94(2)*.

Atkinson, R. K., Merrill, M. M., & Patterson, P. E. (2002). *In Search Of an Embodied Agent Effect.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Bandura, A. (1986). *Social foundations of thought and action: A social-cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology, 41*, 586-598.

Baylor, A. L. (2002a). Agent-based learning environments for investigating teaching and learning. *Journal of Educational Computing Research, 26*(3), 249-270.

Baylor, A. L. (2002b). Expanding preservice teachers' metacognitive awareness of instructional planning through pedagogical agents. *Educational Technology Research & Development, 50*(2), 5-22.

Baylor, A. L., & Ryu, J. (2003a). Does the presence of image and animation enhance pedagogical agent persona? *Journal of Educational Computing Research, 28*(4), 373-395.

Baylor, A. L., & Ryu, J. (2003b). *The API (Agent Persona Instrument) for Assessing Pedagogical Agent Persona.* Paper presented at the ED-MEDIA, Hawaii.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., et al. (1994). *Animated Conversation: Rule-based Generation of Facial Expression, Gesture & Spoken Intoation for Multiple Conversational Agents.* Paper presented at the Siggraph'94, Orlando, FL.

Doyle, P. (1999, May). *When is a Communicative Agent a Good Idea?* Paper presented at the Third International Conference on Autonomous Agents, Seattle WA.

FitzGerald, P. (2000, June). *Design Considerations for Animated Intelligent Agents.* Paper presented at the The Fourth International Conference on Autonomous Agents, Barcelona, Spain.

Kitsantas, A., & Baylor, A. L. (2001). The Impact of the IPSRT (Instructional Planning Self-Reflective Tool) on Preservice Teachers' Performance, Disposition, and Self-Efficacy Beliefs Regarding Systematic Instructional Planning. *Educational Technology Research & Development, 49*(4), 101-110.

Koda, T., & Maes, P. (1996). *Agents with face: the effects of personification of agents.* Paper presented at the HCI'96, London, UK.

Lester, J. C., Towns, S. G., Callaway, C. B., Voerman, J. L., & FitzGerald, P. J. (2000). Deictic amd Emotive Communication in Animated Pedagogical Agents. In J. Sullivant (Ed.), *Embodied Conversational Agents*. Boston: MIT Press.

Lester, J. C., Towns, S. G., & FitzGerald, P. J. (1999). Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents. *The International Journal of Artificial Intelligence in Education, 10*(3-4), 278-291.

Lester, J. C., Voerman, J. L., Towns, S. G., & Callaway, C. B. (1999). Deictic Believability: Coordinated Gesture, Locomotion, and Speech in Lifelike Pedagogical Agents. *User Modeling and User-Adapted Interaction, 9*(1-2), 1-44.

Link, K. E., Kreuz, R. J., Graesser, A. C., & Tutoring Research Group. (2001). Factors that Influence the Perception of Feedback Delivered by a Pedagogical Agent. *International Journal of Speech Technology, 4*, 145-153.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312-320.

Mayer, R. E., Sobko, K., Fennell, S., & Mautone, T. (2002). *Social Cues in Multimedia Learning: Role of Speaker's Voice and Conversational Style.* Paper presented at the AERA.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity effects. *Journal of Educational Psychology, 91*, 358-368.

Moreno, R., & Mayer, R. E. (2002). Verbal Redundancy in Multimedia Learning: When Reading Helps Listening. *Journal of Educational Psychology, 94*(1), 156-163.

Moreno, R., Mayer, R. E., & Lester, J. C. (2000). *Life-Like Pedagogical Agents in Constructivist Multimedia Environments: Cognitive Consequences of their Interaction.* Paper presented at the World Conference on Educational Multimedia, Hypermedia, and Telecommunication (ED-MEDIA), Montreal.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction, 19*(2), 177-213.

Nass, C., & Steuer, J. (1993). Anthropomorphism, agency, and thopoeia: Computers as social actors. *Human Communication Research, 19*(4), 504-527.

Norman, D. A. (1997). How might people interact with agents. In J. M. Bradshaw (Ed.), *Software Agents* (pp. 49-56). Menlo Park, CA: MIT Press.

Reeves, B., & Nass, C. (1996). *The Media Equation*. Stanford, CA: CSLI Publications.

Rickel, J. (2001, September 10-11). *Intelligent Virtual Agents for Education and Training: Opportunities and Challenges.* Paper presented at the Intelligent Virtual Agent, Madrid, Spain.

Towns, S. G., FitzGerald, P. J., & Lester, J. C. (1998). *Visual Emotive Communication in Lifelike Pedagogical Agents.* Paper presented at the International Conference on Intelligent Tutoring Systems, San Antonio, TX.

Wachsmuth, I. (1999, March 17-19). *Communicative Rythm in Gesture and Speech.* Paper presented at the International Gesture Workshop, Gif-sur-Yvette, France.

## Acknowledgements

This work was sponsored by National Science Foundation (NSF) Grant # IIS-0218692.