Life-Like Animated Virtual Pedagogical Agent Enhanced Learning

Jie Yan, Ruth Agada

Bowie State University
jyan@bowiestate.edu
doi: 10.4156/jnit.vol1.issue2.1

Abstract

Research has shown that learning programs with well designed animated virtual pedagogical agents engage and motivate students, produce greater reported satisfaction and enjoyment by students, and produce greater learning gains than programs without these agents. How to design an animated virtual pedagogical agent behaves much like a sensitive and effective human tutor is a very challenge work to do. In this paper, we developed a procedure for producing head and face movements during speech by a virtual pedagogical agent by combining different voice recordings, different facial expressions and different head movement patterns. We then developed three experimental conditions to evaluate how the facial expressions and head movements contribute to learning experiences and outcomes. Results showed that facial expressions and head movements have great impact on student’s impressions and engagement with the virtual pedagogical agent. Experimental results suggest that virtual pedagogical agent that produce natural head movements and appropriate facial expressions while narrating a story produce much more positive user experiences than virtual pedagogical agent that lack these behaviors.

Keywords: Virtual pedagogical agent, Facial expressions, Computer animation

1. Introduction

In recent years, converging evidence indicates that learning gains can be achieved by designing computer programs that use pedagogical agents that foster social agency [1][2][3][4][5][6]. An animated pedagogical agent can have a positive outcome on the learning experience by personification whereby the student tries to impress the agent; the agent collaborates with the student, and provides several types of feedback including causal, congratulatory, deleterious, assistive, background, and motivational responses; and by engagement [7][8][9][10][11][12][13][14]. Research has shown that learning programs with well designed animated pedagogical agents engage and motivate students, produce greater reported satisfaction and enjoyment by students, and produce greater learning gains than programs without these agents. How to design an animated virtual pedagogical agent behaves much like a sensitive and effective human tutor is a very challenge work to do.

In this paper, we have 1) developed procedures for producing head and face movements during speech by a virtual Pedagogical agent by combining different voice recordings, different facial expressions and different head movement patterns, and 2) we have conducted investigations that suggest virtual animated tutor that produce natural head movements and appropriate facial expressions while narrating a story produce much more positive user experiences than virtual tutors that lack these behaviors.

2. Related work

Much prior research has focused on understanding the nature of gestures during speech communication. This research has shown that gestures are universal, interpretable and idiosyncratic. For example, at the McNeil lab at the University of Chicago, David McNeil and his colleagues have spent over two decades analyzing the relationship between speech and gesture through careful analysis of videos of individuals communicating with each other. They have concluded that spontaneous gestures occur universally, are generated almost exclusively during speech, that gestures and speech are synchronous and have a constant relationship in time, and that gesture and speech are semantically and pragmatically coexpressive—that is, they are aspects of a single underlying process. “My argument,
in a nutshell, is that gestures are an integral part of language as much as are words, phrases, and sentences—gesture and language are one system.” [15].

Given that gestures are an integral component of speech, “tightly intertwined with spoken language in time, meaning and function,” but are also idiosyncratic, it is problematical and perhaps infeasible to develop a set of rules for predicting or generating gestures from text or speech. However, given that gestures, including head movements and facial expression, are the physical manifestation of ideas that are also represented in the discourse and acoustic structure of speech, it should be possible to learn statistical relationships between these linguistic structures and the head and facial behaviors that accompany speech production.

During daily conversations, meaning is communicated by both speakers and listeners in parallel both within and across auditory and visual channels. In the auditory speech signal, meaning is communicated through movements of the articulators to produce an acoustic phonetic representation of words that carry meaning, and through changes in fundamental frequency and amplitude that communicate emphasis, emotional states (excitement, anger) and other shades of meaning (e.g., sarcasm). Speakers also produce visual information through movements of the lips, tongue and jaw that complements and enhances the acoustic phonetic information in the auditory signal, especially in noisy environments. In addition, rigid head movements (e.g., vertical and horizontal head nods indicating agreement and disagreement, moving the head sideways and up to indicate thinking, etc.) are combined with a wide variety of facial expressions to communicate emotions and other communicative states during speech [16]. We note that listeners in conversations also produce a range of auditory behaviors, head movements and facial expressions which are processed by the speaker and often affect the content and flow of the conversation. The tight and meaningful integration of communicative gestures during speech production is a remarkable feat of human communication that seems even more remarkable when it is considered that speech production, head and face movements and movements of the torso, arms and hands are controlled by different cortical mechanisms and muscle groups, each with their corresponding temporal constraints.

In the context of this research, the key point is that auditory and visual components of speech production by the head and face provide critically important information about the message being produced. In the context of speech communication by a lifelike computer character in learning contexts, communication should be more interesting, believable and meaningful when it is accompanied by the full range of auditory and visual behaviors used in daily interactions.

We also believe that comprehension of speech produced by a lifelike computer character with accurate visual speech and simultaneous expressive head and face movements will produce better comprehension and learning than listening to the auditory message alone. The ultimate goal is to develop a fully automatic approach for generating these believable and contextually appropriate head and face behaviors and demonstrate that children and adults find these agents more engaging, believable and are more effective tutors.

3. Methods

The main objective of this research is to investigate the role of facial expressions and head movements produced by a lifelike animated character to indicate important information in a story or the emotions of the characters in a story. The experiments we designed will manipulate head movements and facial emotions of a virtual storyteller corresponding to intervals of speech that were either emphasized or provided with emotional expression by the storyteller’s voice. We will analyze videotapes of subjects who are listening to and looking at the storyteller narrating a story to measure how attentive the subject is to the storyteller, and use questionnaires to measure the subjects’ impressions of how believable, credible and human-like the subjects believe the virtual storyteller to be. We will also have subjects summarize the story and analyze the summaries for both comprehension and spontaneous recollection of emotional content. We hypothesize that stories in which the virtual storyteller emulates natural head movements and emotions will produce increased engagement by subjects, produce more positive ratings of the storyteller and lead to better comprehension of the stories.
3.1. Story Development

We first developed a story in which one or more of the characters in the story experience all or a subset of the emotions happiness, surprise, fear, anger, sadness and disgust. Then the story was rehearsed and recorded by an expert human storyteller with instructions to communicate the indicated emotions while narrating the story or producing dialogs between the characters. We videotaped the storyteller recording the story in order to analyze the head movements and facial expressions; these analyses will inform the head movements and facial expressions of the virtual tutor in the experiments.

3.2. Manipulating Voice, Facial Expressions and Head Movements of the Virtual Agent

**Voice:** The human voice is a remarkable instrument. The words we produce in everyday conversations are often enhanced by features of voice that communicate emotions (e.g., joy, surprise, fear, anger, disgust and sadness), moods (excitement, impatience, boredom) and other communicative functions (e.g., sarcasm). In our studies, we manipulated the voice of the virtual storyteller by having a professional voice talent rehearse and then produce two narrations of the story, one in a normal “conversational mode” and the other in a more theatrical “dramatic mode.” We recorded each of these narrations for use in different stories. We note that the visual speech produced by the virtual storyteller, that is, the movements of the lips, tongue and jaw was animated independently for the two recordings. Each recording was transcribed phonetically and the resulting time-aligned phonetic transcription was used to generate the visual speech for each story, synchronized with the auditory signal.

**Facial Expressions:** In the story we wrote for our initial investigation, a series of events caused the main character, Jessica, to experience six basic emotions: sadness, joy, anger, fear, surprise and disgust. In our investigation, the virtual storyteller, Julie, either did or did not produce the six of the emotions shown in Figure 1 at appropriate times when narrating the story. The emotions were designed in collaboration with Erika Rosenberg, an expert in the field of facial display emotions [17][18].

**Figure 1.** Six basic emotions of the virtual storyteller (Julie): sadness, joy, anger, fear, surprise and disgust.

**Head Movements:** In our investigation, we compared three head movement patterns produced by the virtual storyteller: (a) No head movements during speech, (b) having the virtual tutor move her head gracefully up and down while saying each sentence, with the movements adjusted algorithmically to “fit” the duration of the utterance, and (c) having the head movements of the virtual tutor mimic the head movements of the human storyteller in the dramatic condition. This was done using a video annotation system that produced parameters that were used to program the movements of the virtual tutor.

3.3. Experimental Design

We presented separate groups of college students with the same story produced with different combinations of voice, face and head movements by the virtual storyteller. In the **Conversational**
Control condition (Condition 1), Julie, the virtual storyteller, presented the story in a conversational voice with no head movements or facial expressions. In the Conversational Head Movement condition (Condition 2), Julie presented the story in a conversational voice with essentially the same up and down head movement pattern in each sentence. In the Dramatic condition (Condition 3), Julie read the story in a dramatic voice, her head movements were based on those of the human storyteller, and appropriate facial emotions were displayed during an appropriate interval of speech corresponding to the expression of the emotion by the dramatic voice.

To generate animations of these three experimental conditions, we used an animation authoring tool embodied in CU Animate system [19] that previously developed by the author this paper and colleagues.

The CU Animate system provides several important functions. First, the system will automatically generate accurate visual speech synchronized with a spoken utterance, given as input the speech waveform of an utterance and its associated text string. The quality of the visual speech is excellent as it is generated from a database of motion capture data from human lips producing phrases containing all sequences of English phonemes. Second, the system contains an easy to use yet powerful authoring environment for controlling head movements and facial expressions. Using this authoring tool, it is possible to generate animation sequences in which the 3D models closely mimics the movements of a person captured on videotape. The CU Animate system also enables the six emotions produced by 3D models to be generated across any sequence of phonemes within any utterance. On top of it, we have extended the functionality of the CU Animate system by developing the algorithm that produces natural head movements within the system automatically. The CU Animate system provides a complete research environment for designing experiments that control the behaviors of a virtual storyteller.

To generate the three experimental conditions above, First of all, we aligned the recordings at the phonetic level automatically using automatic alignment procedures, then by using the authoring tools provided in CU Animate, we were able to generate three experimental conditions respectively.

Specifically, in experimental condition 3, we generated fully animated agent with emotions, head movements with word prominence. Figure 2 shows the procedure to produce the animation sequences in condition 3. Firstly, we videotaped the human storytellers while she narrating the story, then we use an annotation tool to find the facial expressions, hand movements patterns to inform the virtual storyteller. We used the authoring tools provided in CU Animate system to manually add corresponding expression and head gesture tags in the story utterances to produce dramatic animations sequences. The facial expressions and head movements were synchronized with the human video who told the story. Figure 3 is the Annotation tool used to mark the facial expressions and head movement patterns.

Figure 4, 5 and 6 show that the animated agent says the same utterance in three experimental conditions respectively. Figure 7 shows a sentence produced by the human storyteller and the corresponding animation produced by Julie in the Dramatic condition. They are saying the same utterance “Jessica was so sad that she started to cry”.

![Figure 2. Procedure to produce the animation sequences in condition 3 (Dramatic condition)](image-url)
The above combinations of variables were chosen in order to determine whether differences in user experience and/or comprehension would occur for conditions that differ greatly (condition 3, relative to conditions 1 and 2). The conversational control condition provides a baseline condition to gain insights about potential benefits of voice, head movements and facial expressions in the other two conditions. Subsequent research will focus on the relative contributions of these conditions in controlled experiments.
3.4. Assessment

Thirty college students were participated in the testing stage. Each student was presented one of the three versions of the story. Each subject was randomly assigned to one of the three experimental conditions.

Each subject was seated in front of a computer console. The animated storyteller was centered in the screen and told the story to the subject. Students listened to the story through high quality headphones while watching the screen.

Following the story, the subject was given a survey with questions designed to assess the subject’s perception of the storyteller in terms of expressiveness, believability, storytelling ability and effectiveness. The survey was followed by a comprehension test which included writing a brief summary of the story (which was scored for key points) and answering both multiple choice and short answer questions.

To assess students’ impressions of the storyteller, we used 12 survey questions. Students indicated their response to each question on a 5 point Likert Scale. The verbal anchors corresponding to the rating scale for each question are indicated following the questions below:

1. How would you rate the overall quality of storytelling by the agent? (Terrible storyteller – Just ok – Excellent storyteller) Storytelling ability
2. Overall how would you rate the agent? (Awful & Unbelievable – So-So – Wonderful) Overall rating
3. The agent stimulated interest in the story. (Bored me to death – So-So – Totally held my interests) Interest
4. How believable was the animated agent? (Not at all – Somewhat – Completely believable) Believable
5. How effectively did the animated agent deliver the story? (Totally ineffective – Moderately effective – Highly effective) Effective delivery
6. How convincing was the animated agent? (Not convincing – Very convincing) Convincing
7. How well does the storyteller’s voice match her image? (Mismatch – Just ok – Perfectly) Voice face match
8. How well did the agent hold your attention? (Not at all – Completely) Hold attention
9. How would you rate the contribution of the animated storyteller? (Complete Distraction – Major benefit) Contribution

10. If you were buying a software program to teach your child to read, would you want to have the animated storyteller? (No – Wouldn’t matter – Absolutely) Desirability

11. Do you think that having the story read to you by animated storyteller helped you remember the story better? (Not at all – Absolutely) Comprehension

12. How emotional and dramatic was the animated storyteller? (Not at all – Very much) How emotional

Each of the questions was rated with a score range from 1-5 with 1 as the most negative and 5 as the most positive response. Subjects were able to place their response anywhere along a line between the negative and positive anchors.

To assess comprehension of the narrated stories in each condition, we asked students to write a summary of the story. We scored each summary for a set of key points. At the end of the test, we asked the subjects to comment on the virtual storyteller.

4. Results

The results shown in Figure 8 indicate clear differences between the three conditions. In general, subjects felt that the dramatic storyteller was more believable, convincing, effective, desirable, emotional, etc. For all 12 questions, this condition (Dramatic condition) received the highest average rating on each question. The virtual storyteller was considered least believable, convincing etc., when the storyteller had a normal conversational voice with no head movements or emotions, and this condition (Conversational Control condition) produced the lowest scores on each question. Interestingly, adding head movements to the conversational voice produced a relatively large change in opinions about and perceptions of the virtual storyteller except for the question on how emotional she was. It appears that simply adding head movements greatly improves opinions about the naturalness and believability of the agent, even when the movements are not synchronized in any way with the prosodic structure of the speech being produced. We discuss possible reasons for this below.

Over 90% participants in the dramatic voice condition provided highly positive feedback about the storyteller. Their comments included: “The story was very well told and I understood it very well”. “Expressions were very good, I really like those”. “I believe the agent showed great emotional gestures when telling the story. It was pretty dynamic which grabs the viewers’ attention.” “It can help an individual remember the story since it can relate to the gestures of the agent.”

About 70% of the participants in experimental condition 1 and 2 commented on the lack of emotions expressed by the agent. Their comments include “I liked the agents’ appearance and the sound of the storyteller’s voice. But I didn’t like the lack of emotions on the agent”. “Unfortunately, I didn’t like anything about the agent. I felt that the words were just being read off the paper with no emotion.” “She could use more facial expressions such as the eyes.”

Measurement of comprehension did not produce differences between groups because subjects in each condition had essentially perfect comprehension. The stories were designed for testing with children, but initial testing was conducted with college students, who found the stories easy to comprehend. In subsequent experiments, we will revise the stories to incorporate more inferences, test children as well as adults, and test comprehension both immediately after the initial storytelling and following a delay of two weeks.

We hypothesize that comprehension of stories (or science explanations) with emotional expression will improve comprehension, for at least two reasons. First, research by Mayer and colleagues (summarized in [20]) on multimedia learning, demonstrates improved recall, comprehension and knowledge transfer when speech and visual information are presented simultaneously, as in narrated animations. Other research has demonstrated improved learning when explanations are narrated by pedagogical agents [2][3]. We therefore hypothesize that presentation of stories or narrations of science animations by a pedagogical agent with an emotional voice and accurate facial emotions will result in an enriched mental representation of the content produced by the storyteller, leading to improved comprehension. Second, based on subjects’ ratings, we expect that listeners are more engaged and attentive when listening to a dramatic voice accompanied by facial emotions. Listeners who are more engaged in tasks are likely to benefit from improved attention and motivation. In the future, we will
test this hypothesis using both stories and narrated animations that present science concepts. We predict that narrated animations in which the voice and facial expressions of the storyteller are more emotional and believable will lead to improved learning of science concepts.

![Subject Test Results Comparisons](image)

**Figure 8.** Subject Test Results Comparisons

5. Conclusion

In Summary, We developed three experimental conditions of animated virtual pedagogical agent to evaluate how the facial expressions and head movements contribute to learning experiences and outcomes. Results showed that facial expressions and head movements have great impact on student’s impressions and engagement with the virtual pedagogical Agent.

The development of animated virtual pedagogical agent that behave like sensitive and effective teachers will have untold impact on quality of life for many individuals. E-learning and tutoring systems with life-like animated pedagogical agent provide unprecedented opportunities for millions of individuals to benefit from accessible and inexpensive tutoring, distance learning.

6. References


