Life-Like Animated Virtual Tutor Embedded Learning Module

Jie Yan, Ruth Agada
Department of Computer Science, Bowie State University, Bowie, Maryland, USA

Abstract - In order to engage students in their learning experience, a life-like animated virtual tutor enhanced learning module that control learning interactions with students was presented in this paper. This module is designed to present scientific lectures to students in optimal ways through narrated animations and to provide interactive support to students as they learn science expositions, and to assess and train comprehension through question-answer dialogs with a life-like animated virtual tutor. Results show that animated virtual tutor with life-like facial expressions and head movements has great impact on student’s impressions and engagement in the learning process.

Keywords: Life-Like, Animated virtual Tutor, Computer Animation, Learning Module

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]. Research has shown that learning programs with well designed pedagogical agents engage and motivate students, produce greater reported satisfaction and enjoyment by students, and produce greater learning gains than programs without these agents.

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 scientific topic 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.” [7].

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 [8]. 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 scientific topic or the emotions of the characters in a topic. The experiments we designed will manipulate head movements and facial emotions of a virtual tutor corresponding to intervals of speech that were either emphasized or provided with emotional expression by the tutor’s voice. We will analyze videotapes of subjects who are listening to and looking at the tutor narrating a scientific topic to measure how attentive the subject is to the tutor, and use questionnaires to measure the subjects’ impressions of how believable, credible and human-like the subjects believe the virtual tutor to be. We will also have subjects take quiz afterward to test both the comprehension and spontaneous recollection of the content. We hypothesize that scientific lectures in which the virtual tutor emulates natural head movements and emotions will produce increased engagement by subjects, produce more positive ratings of the animated virtual tutor and lead to better comprehension of the lectures.

3.1 Scientific Lectures Development

We first developed a series of scientific lectures for the animated virtual tutor to present. Then the scientific lectures were rehearsed and recorded by an expert human tutor with instructions to communicate the indicated emotions while narrating the lecture or producing dialogs between the characters. We videotaped the real human tutor recording the lecture 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 tutor by having a professional voice talent rehearse and then produce two narrations of the scientific lectures, one in a normal “reading mode” and the other in a more theatrical “conversational mode.” We recorded each of these narrations for use in different scientific lectures. We note that the visual speech produced by the virtual tutor, 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 scientific lecture, synchronized with the auditory signal.

Facial Expressions: In the scientific lectures we prepared for our initial investigation, the six basic emotions of the virtual tutor known as sadness, joy, anger, fear, surprise and disgust were developed. In our investigation, the virtual tutor either did or did not produce the six of the emotions shown in Figure 1 at appropriate times when narrating the science lectures. The emotions were designed in collaboration with Erika Rosenberg, an expert in the field of facial display emotions [9][10].

Figure 1: Six basic emotions of the animated virtual tutor: sadness, joy, anger, fear, surprise and disgust
Head Movements: In our experiment, we compared two head movement conditions produced by the animated virtual tutor: one is no head movements during speech; the other one is having the head movements of the virtual tutor mimic the head movements of the real human tutor. 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 Conditions

To study how facial expressions and head movements influence the student learning experience and learning outcomes with the animated virtual tutor, three animation conditions of the animated tutor were created and used to instruct the same scientific lecture (learning material). Each condition differs from each other with respect to the expression and animation levels:

1. Fully Animated: the animated tutor is fully animated and realistically expressive. We used the annotated videos of the real expert tutor to create the facial animations so the animated tutor behaves just like the real expert tutor.
2. Partially Animated: the animated tutor is limited to only audio expression and head movements, the facial expressions are muted.
3. Muted: the animated tutor has its head and voice expression completely muted; a static version of the virtual tutor.

Figure 2: Animated virtual tutor interface (lecture mode)

3.4 Subject Testing Results

90 College students were participated in the evaluation of the animated virtual tutor. In the subject testing, the student is presented with a lecture specified by the animated tutor, as well topics to be covered in that lecture. Topics covered in the different lecture vary based on the type of instruction specified, but may include introduction to the lecture, explanation of terms related to the lecture topic, various examples to further explain concepts. Some of these examples contain different media like static images to illustrate a concept or provide a visual cue of an object or video media to demonstrate how tasks are to be completed. Based on the subtopic selected by the student, only information pertinent to that subtopic is displayed. Students are allowed to view and review material based on the subtopic covered so as to allow them the opportunity to re-explore certain aspects of the subtopic not fully understood.

Following each lecture, quizzes are made available to test the students’ comprehension and retention of the lecture (Figure 3). The quiz mode test students on information covered in each lecture. The quizzes are cumulative.

Figure 3: Animated virtual tutor interface (quiz mode)

These 90 students were separated into three groups (30 each) with the same scientific lectures presented by the virtual tutor but with different combinations of voice, face and head movements (Fully animated condition, partially animated condition and Muted condition).

After the test, students were given a questionnaire to rate the animated virtual tutor in terms of 12 related aspects including 1) Story telling ability 2) Overall rating 3) Interest 4) Believable 5) Effective delivery 6) Convincing 7) Voice face match 8) Hold attention 9) Contribution 10) Desirability 11) Comprehension 12) How emotional of the tutor. Results showed (See Figure 4) that facial expressions and head movements had great impact on student’s impressions of and engagement with the virtual tutor, who was given the highest ratings on all dimensions. Unfortunately, no differences in comprehension were obtained across conditions because the quiz questions we designed were too simple, so subjects produced correct answers to all of the comprehension questions. We plan to replicate this study with more complex
learning materials and quiz questions for college students in different comprehension level.

4 Discussion

Progress in our exploratory research has been hindered by the amount of time and effort required to create animation sequences for the virtual tutor that accurately modeled the behaviors of the real human tutor. One of the great potential directions of this research will be the development and demonstration of a research methodology and tools that facilitate design and testing of animated virtual tutors’ behaviors in learning tasks.

5 Conclusions

In Summary, We developed an animated virtual tutor to evaluate how the facial expressions and head movements contribute to learning experiences and learning outcomes. Results showed that facial expressions and head movements have great impact on student’s impressions and engagement with the virtual animated tutor.

This research has significant impact on individuals and society by providing more efficient and effective computer interfaces. The development of animated virtual tutor that behave like sensitive and effective teachers will have untold impact on quality of life for many individuals. Computer-based learning and tutoring systems with life-like animated tutor provide unprecedented opportunities for millions of individuals to benefit from accessible and inexpensive tutoring, distance learning systems.

6 References


