

Scripted Animation Towards Scalable Content Creation for eLearning—A Quality Analysis

Nicoletta Adamo-Villani¹(✉), Jian Cui², and Voicu Popescu²

¹ Department of Computer Graphics Technology, Purdue University,
401 N. Grant Street, West Lafayette, IN 47907, USA
nadamovi@purdue.edu

² Department of Computer Science, Purdue University, 305 N. University Street,
West Lafayette, IN 47907, USA
{cui9, popescu}@purdue.edu

Abstract. The success of eLearning depends on the broad availability of educational materials that provide a high-quality delivery of high-quality content. One approach for high-quality delivery is to rely on a computer animated instructor avatar that not only speaks, but that also gestures to elucidate novel concepts and to convey an engaging personality that captures and maintains the learners' focus. The traditional approach of manual key frame animation does not scale, as it requires a substantial time investment as well as artistic talent. We have developed a system that allows animating an instructor avatar quickly and without the prerequisite of artistic talent through a text script. In this paper we quantify the speed/quality tradeoff made by our scripted animation by comparison to manual animation.

Keywords: Instructor avatar · Instructor gesture · Scripted animation · Manual key frame animation · Effective online learning materials

1 Introduction

The proliferation of inexpensive yet powerful internet-connected computing platforms such as laptops, tablets or even smartphones creates opportunities for eLearning to supplement, and, in some cases to supplant, traditional classroom education. The success of eLearning depends not only on the quality of the content of the online learning materials, but also on the quality of the delivery. This is even more important for young learners whose language skills are still developing. For such learners, text is not enough, nor are verbal explanations from an invisible narrator. The lesson is best delivered by an online instructor, which evokes teacher-student and parent-child interactions that are known to work with young learners. One solution is to deliver the lessons through videos. A skilled instructor is videotaped giving the lesson to the camera and the video is placed online to be accessed by learners from school and from home. The skilled instructor gives a good lesson that is captured faithfully by the video camera yielding great results. However, the approach has important limitations.

Lack of interactivity. Instructors cannot ask questions, they cannot provide feedback to students, and they cannot adapt the lesson pace and content to each particular student or group of students.

Constrained delivery. Asking the instructor to follow a script and to give the lesson in front of the camera can result in an unnatural delivery, with the instructor being worried about following the script and about staying in the field of view of the camera as opposed to simply being a teacher; moreover, in a studio setting there is no audience to connect with, and the delivery can become an unenthusiastic monologue.

Lack of scalability. Making videos is a tedious process and covering all the ways in which a concept can be explained, all concepts, for all disciplines, for all student age groups requires a huge investment.

A promising alternative is computer animation, which has long shown that it can tell stories convincingly. Computer animation characters could serve as believable and effective instructor avatars, alleviating the challenges enumerated above.

Interactivity. Computer animation is more amenable to *interactivity* than video. The instructor avatar can request input from the learner, analyze the correctness of the answer, and react accordingly. The instructor avatar has perfect memory and infinite energy, which, paradoxically, could result in a more natural *delivery*.

Scalability. What is needed is a fast and accessible method for creating e-lessons delivered by instructor avatars; the entertainment industry uses two main approaches for animating characters—manual animation and motion capture; in manual animation the character pose is defined by a digital artist through a graphical user interface for each key frame; for complex animations one needs multiple key frames per second; key frame animation is slow and requires artistic talent; it is simply not feasible to manually animate the delivery of the world's ever-expanding knowledge base. Motion capture requires expensive specialized hardware and talent to perform the animation to be captured and thus it does not scale to our context either.

Scripted animation—a promising solution. We have developed a system that provides a computer animation instructor avatar that is animated quickly and effectively based on a text script [21]. The script is created by the eLearning content creator with a conventional text editor. The script specifies what and when the avatar does and says. The script is executed automatically to obtain the desired animation. The animation is obtained quickly, and without the requirements of artistic talent, of familiarity with complex animation software, and of programming expertise.

We have used the scripted animation system in two studies on research instructor gesture. The first study investigates which instructor gestures make the instructor avatar appear to students as having a more engaging personality [21]. The second study investigates whether deictic and embodied cognition gestures improve student learning [21]. The system of scripted instructor avatars enabled the efficient creation of tens of high-quality and precise stimuli for these studies. Creating the stimuli for these two studies relying on manual animation would have been prohibitively slow.

In this paper we examine the questions of whether the efficiency of scripted animation compared to manual animation comes at the cost of a loss of animation quality, and, if yes, of how much this cost is. We have chosen a 1 min mathematical equivalence lesson sequence, and we have animated the sequence with both the scripted animation and the manual animation methods. The scripted animation was created with our system in 1 h and it is available via YouTube at the following

URL: <https://www.youtube.com/watch?v=rgSq5lm7yY0>. The manual animation was created by a computer animator in 23 h and is available via YouTube at the following URL: <https://www.youtube.com/watch?v=s-a0FUytpNQ>.

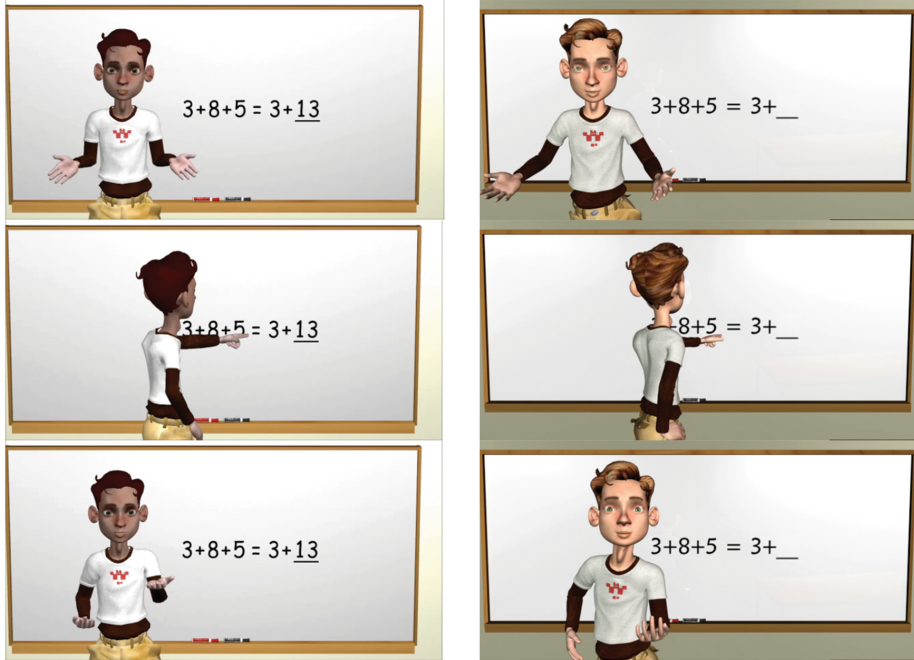


Fig. 1. Frames from the scripted animation (*left*) and from the manual animation (*right*).

As can be seen in Fig. 1, the two animations are similar but not identical. For example, in the scripted animation the pointing to the equal sign is more precise, in the manual animation the right hand rests on the hip as the left hand makes the pointing gesture, and the balance gesture indicating equality is more evocative of physical equilibrium in the scripted animation compared to the manual animation. The animator decided that the answer “13” should appear at the end of the lesson.

The two animations were then shown to computer animators, computer science researchers working in graphics and visualization (and not in animation), and psychology researchers working on gesture. After each animation, the viewer was asked three questions regarding the quality of the animation, regarding the quality of the synchronization of gestures with speech, and regarding the perceived personality of the instructor avatar. Computer animators were asked seven more questions regarding the quality of the motion, the quality of the poses, and the degree to which the animation adheres to each of five principles of animation. The overall score for the scripted vs the manual animation was 3.0 vs 4.1 on a 1 to 5 scale. Psychologists, for whom the animations are intended, liked both animations (4.2 vs 4.6).

2 Background

Computer animated characters have been used in e-learning environments to teach and supervise. Early examples of pedagogical avatars are Cosmo [1], a cosmonaut who explains how the internet works, Herman [2], a bug-like creature that teaches children about biology, and STEVE [3] who trains users in operating complex machinery using speech, pointing gestures, and gaze behavior. PETA is a 3-D computer animated human head that speaks by synthesizing sounds and conveys different facial articulations [4]. PETA allows children to acquire a new language in a spontaneous, unconscious manner. A similar example is the “Thinking Head” [5], a virtual anthropomorphic software agent able to speak and to display emotion through complex facial expressions, vocal prosody, and gestures. Gesturing avatars have also been used to teach sign language, mathematics, and science to young deaf children using sign language, e.g. Mathsigner and SMILE [6]. The ASL software system [7] allows educators to create and add animated signing avatars to e-learning materials.

Rigorous empirical testing was used to assess the contributions of pedagogical agents to learning, and the affective impact on students. Many studies confirm the intended positive influences on education by systems using these agents [8, 9]. Studies also suggest that teaching avatars could be employed in e-learning environments to enhance users’ attitude towards online courses [10]. Agents interacting using multiple modalities appear to lead to greater learning than agents that interact only in a single channel [11]. A comparative study of three e-learning interfaces suggests that e-learning materials incorporating full-body teaching agents that speak and gesticulate are the most efficient, effective and engaging [12].

Animating a 3D character is a challenging task that has been approached from various directions. In manual 3D animation, a skilled animator uses a 3D animation software package (e.g. Maya) and a variety of techniques, such as keyframe animation, to craft the character poses and motions by hand. Manual animation is time consuming, has a steep learning curve and requires artistic talent. In data-driven animation (e.g., motion capture), live motion is recorded directly from an actor, digitized, and then mapped onto a 3D character. Motion capture animation requires highly expensive equipment and the recorded data often needs to be manually refined by a skilled animator.

In automated (or scripted) animation the character’s speech and gestures are automatically generated from input text. BEAT [13] is an example of fully automated character animation system which takes plain text as input, runs a linguistic analysis and generates speech intonation, facial expressions and gestures. GESTYLE [14] annotates text with hand/head/face gestures based on “style” definitions. “Style” determines the gesture repertoire and the gesturing manner of the animated character. Virtual Presenter [15] is an animation system in which gestures can be added to input text manually, or can be automatically generated with keyword-triggered rules.

In addition to fully automated systems, to facilitate the development of embodied agent applications, software toolkits have been created that allow people with no animation expertise to produce and add animated characters to e-content (we call these systems “partially automated”). While these tools do not generate the animations automatically from text, they provide an easy-to-use interface and do not require any

training in animation. Examples include Character Builder [16], NOAH virtual instructor technology [17], Codebaby [18] and Gesture Builder [19]. Although the characters produced using existing fully or partially automated systems speak and gesticulate, their gesture repertoire is limited and generic and the occurrence of facial and manual gestures in concurrence with speech is not driven by research-based rules on the relationship between verbal and non-verbal behavior.

3 Scripted Animation

We have developed a system of computer animation avatars that are controlled with a text script [21]. The input to the system is pre-recorded audio of what the avatar has to say and a text script. The system bypasses the need for digital artistic talent by animating the avatar either automatically or using pre-generated animation stored in a database. The automatic animation relies on lip-syncing and inverse kinematics algorithms to have the avatar utter words and to perform deictic gestures, which include pointing, circling, and underlining at any location on the whiteboard. More complicated gestures, such as the balance gesture used to indicate the equality of the left and right sides of a correctly solved mathematical equivalence problem, are pre-animated by a digital artist, stored in an animation database, and invoked using the script. The script specifies which gestures have to occur and when in relation with the audio file. The script is executed automatically, creating the avatar animation.

The script for the sequence used in the comparison consists of 66 lines, organized according to 17 audio sentences. The partition of the audio into sentences facilitates script writing by allowing quick previews of the part currently edited and by simplifying time references. The script from Fig. 2 plays the audio sentence Lesson9.

At 1.2 s in, a pause is inserted to allow for the completion of the gestures up to that point. The ability to insert pauses greatly simplifies the audio recording process, which can proceed without concerns for allowing enough time for gestures to complete. At 0.2 s, the avatar is instructed to make a right hand underlining gesture, spanning characters 6 through 9 on the whiteboard (i.e. the right side of the equation in Fig. 1, left). As soon as the underlining gesture finishes, the avatar is instructed to move to position B. We use three positions: facing the students (A), profile, looking at the board (B), and extended profile, reaching for the right edge of the board (C). After reaching B, the avatar underlines the left side of the equation, at a slower speed, and then turns to face the students again.

Writing the initial script takes about 60 min. Changing gestures to obtain a different experimental condition takes 10 min, with most of the time spent to rework the synchronization. Switching from gesture to control (i.e. no gesture) condition takes less than a minute, as does creating an exercise for a different mathematical equivalence. The script is executed in real time using interactive rendering techniques, therefore the animation is available as soon as the script is written.

```

PlayAudio Lesson9
  @ 1.2 Pause 1.5
  @ 0.2 Deictic RightUnderline 6 9
  + 0.0 Move B
  + 0.0 Deictic LeftUnderline 0 4 SPEED=0.8
  + 0.0 Move A

```

Fig. 2. Script for one audio sentence.

4 Manual Animation

The manual animation was created by a digital artist with 4 years of experience in 3D animation. The artist was given a video of the scripted animation sequence, the computer animation character of the instructor, and the lip synch animation used to generate the scripted animation. He was asked to reproduce the sequence in a professional grade computer animation software system (Maya) using traditional animation techniques. He employed key frame animation to set the character's main body poses and used various interpolation types provided by the software to generate the in-between frames. Then he manipulated the animation curves by hand to attain realistic timing and fluid motions. Because of the limitations of the character's facial rig, facial articulations could not be animated.

The artist took 9 h to complete the animation. The sequence was rendered using a high-quality offline rendering engine (Mental Ray); the rendering process took 14 h. Whereas removing gestures as needed to transform a gesture stimulus into a control (no-gesture) stimulus is straightforward, changing the animation for a different mathematical equivalence problem takes approximately 2 h and changing the type of charisma gestures takes approximately 4 h.

5 Results and Discussion

We conducted a survey to compare the two animations. We drew respondents from three groups of experts: psychologists working in gesture research (3), computer scientists working in graphics and visualization research (7), and computer animators (16). Each respondent was shown both animations. The order in which the animations were shown was randomized. The respondents were asked the same questions after each animation. All questions were answered on a five point scale: strongly disagree (score of 1), disagree, neutral, agree, and strongly agree (score of 5). We first present the survey questions and answers and we then discuss the results.

The survey had three parts. The first part had three questions addressed to all respondents (Table 1). The second part contained questions that depended on expertise. The psychologists were asked whether they would use the animation in their research on gesture. The mean scores were 4.33 for both animations. The computer animators were asked about specific aspects of animation (Table 2), including whether the motion is fluid and realistic, whether the quality of the animation poses is high, and whether the animation adheres to five fundamental principles of animation (i.e. a subset of the 12 Disney principles of animation [20]). The third part of the survey consisted of an essay question posed to all respondents. The respondents were asked to comment on the animation they had just seen, and to point out and explain the subsequences they liked/disliked the most.

Two of the computer scientists liked the scripted animation for its rendition of the balance gesture, two liked the animation overall; two computer scientists complained of the quality of the audio file, including the high background noise during the speech compared to the perfect silence (of the inserted pauses). The computer scientist who

Table 1. General questions about the scripted (S) and manual (M) animations addressed to computer scientists (CS), psychologists (Psych), and computer animators (CA).

Questions	Mean scores							
	CS		Psych		CA		Avg	
	S	M	S	M	S	M	S	M
Animation is of high quality	2.4	3.0	4.3	4.7	1.9	4.3	2.9	4.0
Gestures well synchronized with speech	3.2	3.8	4.3	4.7	2.5	4.5	3.3	4.3
Avatar has engaging personality	2.8	3.0	4.0	4.3	1.7	4.4	2.8	2.9

liked the scripted animation the least complained about the appearance of the character, about the unenthusiastic voice, and about the simplicity of the mathematical problem. The psychologists liked the scripted animation overall, lauding the balance gesture and the body movements. The only complaint mentioned was that the avatar says “11” and points to “8”, which is the choice of the psychologist who designed the lesson—11 is the total so far, up to and including the addend 8. Whereas psychologists occasionally disagree, their disagreement should not blemish the record of the system. The computer animators were much more critical of the scripted animation. The main complaints were about the lack of adherence to the principles of animation and about the poor quality of the rendering.

Regarding the manual animation, the comments of CS respondents were slightly more positive. The CS respondent who was most disapproving of the earlier animation sees progress, but only marginal, limited to improved lighting that is “less gloomy”; the character was still perceived as “ugly”, and with a monotonous voice. Two CS respondents liked that the answer is only revealed later, and one respondent thought that the reflection of the character in the whiteboard was distracting. The psychologists thought that the avatar was engaging, but reported “superfluous” arm gestures at the beginning, and uncertainty about the “body movement toward [the viewer]”. The computer animators strongly preferred the manual animation noting the adherence to the principles of animation and superior rendering quality. The negatives noted include the lack of facial expressions, an occasional “stiffness”.

Table 1 shows that, on average, the scripted animation was perceived of lower quality, with scores roughly one point below the scores for the manual animation. For the psychologist respondents, who are the users for whom these animations are intended, approve of the scripted animation and give it virtually identical scores to the manual animation. The computer animation respondents were the most critical of the scripted animation, noting the lack of adherence to animation principles (Table 2) and the lower rendering quality. As discussed in the earlier sections, the scripted animation was completed in a fraction of the time it took to put together the manual animation, and real time rendering was used, which are facts not disclosed to the animators. Moreover, some of the features added during manual animation and that increased the score from the computer animation respondents were judged by the psychologists as harmful to the experiments (e.g. reflections, superfluous gesture).

Table 2. Animation-specific questions addressed to computer animators.

Motion quality high		Pose quality high		Anticipation		Arcs		Slow in & slow out		Secondary action		Stretch and squash	
S	M	S	M	S	M	S	M	S	M	S	M	S	M
2.0	4.3	1.9	4.7	2.0	4.1	2.2	4.1	2.2	4.3	1.8	4.1	1.8	4.0

6 Conclusions and Future Work

Although falling short of the highest quality manual animation, scripted animation is of sufficiently high quality to provide a scalable option in support of research on education and eLearning. Another important conclusion of our work is that the animation quality is application and user dependent—whereas computer animators consider animation principles and highest quality rendering non-negotiable, education researchers and eLearning applications might be willing to tradeoff in favor of authoring efficiency.

We have addressed the issue of eLearning scalability by simplifying the task of animation. As future work we will pursue scalability by adding more gestures to our animation database, and by adding support for more concepts, disciplines, and student age groups (e.g. more avatars, more whiteboard drawing capabilities, more types of math problems). Finally, we will investigate extending the system in two divergent directions: to bridge the gap between scripted and manual animation by adding adherence to animation principles, and to further reduce the animation authoring time by scripting. The latter effort will first focus on developing a graphical user interface for editing the script, which promises to lower the script language learning curve and to avoid the possibility of syntax errors. Then, we will investigate automating the animation based on instructor gesture rules, which eliminates scripting altogether.

References

1. Lester, J., Voerman, J., Towns, S., Callaway, C.: Cosmo: a life-like animated pedagogical agent with deictic believability. In: Notes of the IJCAI '97 Workshop on Animated Interface Agents: Making Them Intelligent, Nagoya, Japan, pp. 61–70 (1997)
2. Lester, J., Stone, B., Stelling, G.: Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *User Model. User-Adap. Inter.* **9**(1–2), 1–44 (1999)
3. Johnson, W.L., Rickel, J., Stiles, R., Munro, A.: Integrating pedagogical agents into virtual environments. *Presence: Teleoper. Virtual Environ.* **7**, 523–546 (1998)
4. PETA – a pedagogical embodied teaching agent. In: Proceedings of PETRA08: 1st International Conference on Pervasive Technologies Related to Assistive Environments, ACM Digital Library, Athens (2008)
5. Davis, C., Kim, J., Kuratate, T., Burnham, D.: Making a thinking-talking head. In: Proceedings of the International Conference on Auditory-Visual Speech Processing (AVSP 2007), Hilvarenbeek, The Netherlands (2007)
6. Adamo-Villani, N., Wilbur, R., Eccarius, P., Abe-Harris, L.: Effects of character geometric model on the perception of sign language animation. In: IEEE Proceedings of IV09 -13th International Conference on Information Visualization, pp. 72–75 (2009)

7. Hayward, K., Adamo-Villani, N., Lestina, J.: A computer animation system for creating deaf-accessible math and science curriculum materials. In: Proceedings of Eurographics 2010 – Education Papers, EG Digital Library, Norrköping, Sweden, May 2010
8. Lester, J., Converse, S., Kahler, S., Barlow, T., Stone, B., Bhogal, R.: The persona effect: affective impact of animated pedagogical agents. In: Proceedings of CHI '97, pp. 359–366 (1997)
9. Lester, J., Converse, S., Stone, B., Kahler, S., Barlow, T.: Animated pedagogical agents and problem-solving effectiveness: a large-scale empirical evaluation. In: Proceedings of the Eighth World Conference on Artificial Intelligence in Education, pp. 23–30 (1997)
10. Annetta, L.A., Holmes, S.: Creating presence and community in a synchronous virtual learning environment using avatars. *Int. J. Instr. Technol. Distance Learn.* **3**, 27–43 (2006)
11. Lusk, M.M., Atkinson, R.K.: Varying a pedagogical agent's degree of embodiment under two visual search conditions. *Appl. Cogn. Psychol.* **21**, 747–764 (2007)
12. Alseid, M., Rigas, D.: Three different modes of avatars as virtual lecturers in elearning Interfaces: a comparative usability study. *Open Virtual Real. J.* **2**, 8–17 (2010)
13. Cassell, J., Vilhjalmsson, H., Bickmore, T.: BEAT: the behavior expression animation toolkit. In: Proceedings of SIGGRAPH, pp. 477–486 (2001)
14. Noot, H., Ruttkay, Z.: Gesture in style. In: Camurri, A., Volpe, G., (eds.) GW 2003. LNCS (LNAI), vol. 2915, pp. 324–337. Springer, Heidelberg (2004)
15. Noma, T., Zhao, L., Badler, N.: Design of a virtual human presenter. *IEEE Comput. Graph. Appl.* **20**(4), 49–85 (2000)
16. Character Builder (2011). <http://www.mediasemantics.com/>
17. NOAH Animated Avatar Technology (2011). noahx.com/index.asp
18. Codebaby (2010). www.codebaby.com/products/elearning-solutions
19. Gesture Builder (2010). www.vcom3d.com/index.php?id=gesturebuilder
20. Johnston, O., Thomas, F.: *The Illusion of Life: Disney Animation*. Disney Editions, New York (1995)
21. Cui, J., Popescu, V., Adamo-Villan, N., Cook, S., Duggan, K., Friedman, H.: An animation stimuli system for research on instructor gestures in education. Technical reports, Department of Computer Science, Purdue University. doi:<http://docs.lib.purdue.edu/cstech/1771/>