Title
Animation Stimuli System for Research on Instructor Gestures in Education

Journal
IEEE Computer Graphics and Applications, 37(4)

ISSN
0272-1716

Authors
Cui, Jian
Popescu, Voicu
Adamo-Villani, Nicoletta
et al.

Publication Date
2017

DOI
10.1109/mcg.2017.3271471

Peer reviewed

Animation Stimuli System for Research on Instructor Gestures in Education

Jian Cui, Voicu Popescu, and Nicoletta Adamo-Villani ■ *Purdue University*

Susan Wagner Cook ■ *University of Iowa*

Katherine A. Duggan ■ *University of Pittsburgh*

Howard S. Friedman ■ *University of California, Riverside*

Education research has shown that instructor gestures can help capture, maintain, and direct the student's attention during a lecture as well as enhance learning. The proposed system allows users to efficiently create accurate and effective stimuli for complex studies on gesture, without the need for computer animation expertise or artist talent.

Gestures play an important role in education. Gestures such as pointing, circling, and underlining can help capture, maintain, and direct the student's attention during a lecture. Instructor gestures serve as a complementary communication channel that helps students parse and grasp the information presented to them verbally and visually.[1,2] For example, in the context of learning mathematical equivalence, a balance gesture elicits the student's first-hand experience with the concept of physical equilibrium, which is used to scaffold the novel concept of balancing the two sides of an equation. Gestures can also convey an instructor's appealing and engaging personality, which enhances learning. As important as gestures seem to be, however, much remains to be discovered through education research regarding whether, when, and how specific gestures benefit learning.

Traditional education research on instructor gestures relies on video stimuli (Figure 1).[3] With this approach, student participants are shown prerecorded videos of instructor actors giving lessons in various gesture conditions. To create the video stimuli, the education researcher first composes a detailed script of the lesson for each condition, which includes the gestures that must occur and when. Then the instructor actor learns and performs the script, and the performance is recorded. An obvious advantage of video stimuli is that they are by definition photorealistic: students have no difficulty in associating what they see in the video with a real-world instructor.

However, video stimuli have important disadvantages. First, learning long scripts for multiple conditions is challenging. Any delivery error requires an additional take to ensure the correctness and fluidity of the final stimuli. The instructor actor may have difficulty repeating, from condition to condition, the same level of energy, enthusiasm, and voice intensity, as well as the same secondary motions (such as head, body, and nongesturing arm and hand motions). As a consequence, producing video stimuli is time consuming, especially when gesture precision and consistency across conditions are strictly enforced. This effectively limits the complexity of the studies that can be conducted and forces researchers to rely on simple, short learning activities so they can more easily

 Published by the IEEE Computer Society 0272-1716/17/$33.00

control the delivery parameters. Moreover, video stimuli often come across as constrained and unnatural. Instructor actors have to focus on remembering the script and the exact secondary parameters of the delivery while reciting the script, so they cannot simply be teachers. To ensure the exact same speech from condition to condition, researchers can have the instructor actor lip-sync to prerecorded audio, but that takes the stimulus even further from a realistic in-class lesson delivered by a skilled teacher. Furthermore, not all differences in secondary parameters between conditions can be eliminated, which can confound the analysis of the effect on which the study focuses. Consequently, although there is an initial body of evidence that gesture increases learning, fundamental methodological issues in studying gesture using video stimuli limit our progress toward understanding what, when, and for which learners gestures lead to more learning. (See the "Related Work in Gesture Research and Instructor Avatars" sidebar for more details.)

Advances in computer graphics software and hardware allow us to use computer animation characters as instructor avatars. With perfect memory and infinite energy, instructor avatars are well suited for precisely executing complex scripts. Repeatability is not a problem—all secondary parameters can be perfectly controlled. Thus, we can use the exact same audio and secondary motion from condition to condition. The computer animation characters' precision and consistency during gesture performance makes such instructor avatars well suited for studying this line of research.

Because the avatar is not a "real" instructor, we must ask several questions. Do students learn from such avatars? Does what is discovered about avatar instructor gestures apply to human instructors? First, education research has shown that students do learn from instructor avatars populating comic books, cartoons, and video games. Consequently, improving instructor avatar effectiveness is a worthy goal in and of itself. Second, emerging research shows that what is learned about gesture in the context of instructor avatars does apply to human instructors.[4,5] Third, finding out how to deliver learning activities effectively using an instructor avatar will also pay off in terms of achieving scalability of the currently expensive process of authoring digital-learning activities. Fourth, an eloquent instructor avatar can open the door to effective learning personalization based on specific affective and cognitive characteristics of individual learners. Such personalization is inconceivable in the context of conventional classroom education with one teacher to tens of students. (See the "Related Work in Gesture Research and Instructor Avatars" sidebar for more details.)

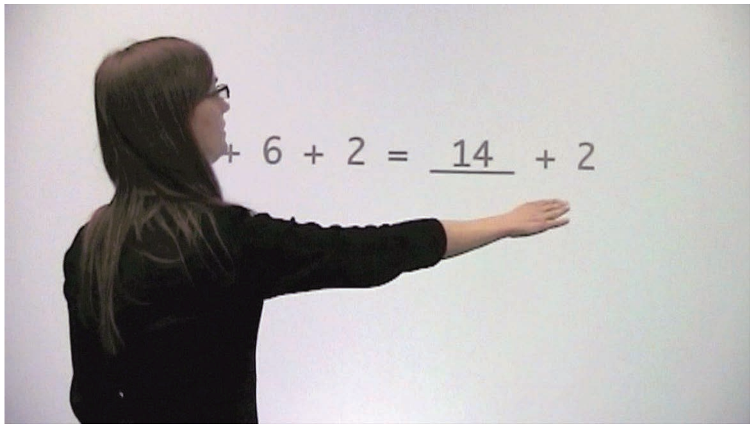


Figure 1. Sample frame from traditional video stimuli used in research on instructor gesture in education. The stimuli are created by video recording instructors giving lessons scripted for various gesture conditions.

One challenge to generating animation stimuli with instructor avatars is that current computer animation software systems require artistic talent and technical expertise to model, animate, and control characters. Presenting stimuli involves 3D avatar and environment modeling, animation creation (such as condition gestures, necessary pose changes, and secondary gestures), lip-syncing, and final delivery including refining iterations. Such a workflow can only support one stimulus at a time; for a similar stimulus such as postponing all gestures by 1 second, a similar amount of work is needed. Many gesture studies have common pieces that can be used across studies, and many components can be automated, such as lip-syncing and deistic gestures, but no such system exists that exploits these common assets effectively and easily.

In this article, we present our work toward overcoming this challenge. Specifically, we describe an approach that enables the creation of animation stimuli for research on instructor gesture in education and describe a system that implements the approach. The system does not require artistic talent or programming expertise, as it provides an animation character that serves as an instructor avatar. The avatar is controlled with a text script that specifies when and what it says and does. The script is executed automatically to create the animation stimuli. The system has been used so far to create stimuli in five studies that demonstrate it can be used efficiently to create accurate and effective stimuli for complex studies on gesture.

System Overview

Our approach is based on two observations. First, we cannot expect education researchers to model

Virtual Reality's Advantages over Video Playback

For over a hundred years, ethnology and linguistics studies have shown that we respond to gestures with extreme alertness and in myriad ways.[1] Researchers have conducted broad studies regarding how instructor gestures influence the learning process. Math and science teachers frequently use their hands to depict information.[2] Gesture in instruction has been shown to increase learning of a variety of mathematical concepts, including mathematical equivalence.[3] For example, a study of Piagetian conservation[4] compared instruction with gesture to instruction without gesture and found that the first condition led nearly two times as many students to demonstrate deep learning.

In addition to facilitating initial learning, gesturing during mathematics instruction also makes it more likely that children will retain their new learning over time, above and beyond the amount of retention shown by children who learn the same amount without gesture, and it increases the transfer of knowledge to novel contexts.[5] Gesture is a spontaneous behavior that is routinely incorporated into human communication, so it is perhaps not surprising that it is helpful for listeners.

Studies have also shown that an appealing and engaging nonverbal style and charismatic personality are rated more favorably.[6] The general finding of these studies was that an enthusiastic, expressive lecturer is more highly valued.[7] All the studies mentioned so far were conducted using videos of actors.

Animated avatars have been incorporated into general information systems such as Jack[8] and PPP Persona.[9] Early examples of instructor avatars are Cosmo,[10] a cosmonaut who explains how the Internet works; Herman,[11] a bug-like creature that teaches children about biology; and Steve,[12] who trains users to operate complex machinery using speech, pointing gestures, and gaze behavior. In the Virtual Human Project,[13] a teacher gives astronomy lessons by following different pedagogical paradigms and uses various nonverbal behaviors that reflect different parameterized personality settings. Many studies confirm the intended positive influences on education by systems using avatars.[14,15] Studies also suggest that teaching avatars could be employed in e-learning environments to enhance the users' attitude toward online courses.[16] Moreover, animated avatars help create a positive learning environment[17] and can be used to create immersive environments, such as in virtual classrooms where human teachers can practice.[18] Whereas these studies confirm the potential of avatars in education, the avatar animations are one of kind, commissioned by the authors from professional digital artists, and cannot serve as a flexible testbed for education research.

To facilitate the development of animated avatars applications, software toolkits have been created for producing and adding animated characters to e-content. For example, the Character Builder tool (www.mediasemantics.com) helps nonanimators create e-presentations that include talking and gesturing characters. It uses a nonlinear animation engine to generate animation from high-level commands such as look, point, and say. Noah virtual instructor technology allows users to add 2D animated talking avatars to slideshow presentations and websites. Codebaby (codebaby.com/products/codebaby-studio/support/studio-knowledge-base/) enables nonexpert users to create and integrate 3D animated characters into e-learning courseware. Gesture Builder (www.vcom3d.com/index.php?id=gesturebuilder) is a commercial software product for producing animated 3D characters that gesture and sign.

Although the characters produced using these 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 ver-

and animate their own characters. Instead, a database of premodeled characters should be available from which the user can select an instructor avatar. Whereas simple animations should be computed on the fly using inverse kinematics animation algorithms, complex animations should be predefined and made available through the database. Second, we cannot expect researchers to program the generation of the stimuli. Instead, much like they are doing now when creating video stimuli, education researchers should have the ability to control the avatar through a high-level English-like scripting language.

The system we developed implements this approach. Our instructor avatar corresponds to a casually dressed teenager (see Figure 2). The avatar can speak, write on a nearby white board, and point to, circle, and underline any location on the board. The avatar can also make embodied cognition gestures and charisma gestures (see Figure 2). *Embodied cognition gestures* aim to leverage the learner's experience with motor and perceptual interaction with the environment to improve learning. *Charisma gestures* aim to portray the instructor as an engaging, enthusiastic personality, a style thought conducive to increased learning. The avatar is controlled with a script written by the education researcher using a text editor. The scripting language lets the researcher specify what the avatar says and does and when. The script is executed automatically to create the animation stimuli. The accompanying video web extra (see bal and nonverbal behavior. Systems that automate the production of animated nonverbal behavior have begun to emerge,[19] but the majority of psychology and education research investigating nonverbal communication still relies on traditional video stimuli.

References

1. E. Sapir, "The Unconscious Patterning of Behavior in Society," *Language, Culture, and Society: A Book of Readings*, B. Blount, ed. Waveband Press, 1927, pp. 29–42.
2. S. Goldin-Meadow, S. Kim, and M. Singer, "What the Teacher's Hands Tell the Student's Mind about Math," *J. Educational Psychology*, vol. 91, no. 4, 1999, pp. 720–730.
3. S. Goldin-Meadow and M. Alibali, "Gesture's Role in Speaking, learning, and Creating Language," *Ann. Rev. Psychology*, vol. 64, 2013, pp. 257–283; doi:10.1146/annurevpsych-113011-143802.
4. R.B. Church, S. Ayman-Nolley, and S. Mahootian, "The Role of Gesture in Bilingual Education: Does Gesture Enhance Learning?" *Int'l J. Bilingual Education and Bilingualism*, vol. 7, no. 4, 2004, pp. 303–319.
5. S.W. Cook, R.G. Duffy, and K.M. Fenn, "Consolidation and Transfer of Learning after Observing Hand Gesture," *Child development*, vol. 84, no. 6, 2013, pp. 1863–1871.
6. H.S. Friedman, R.E. Riggio, and D.F. Casella, "Nonverbal Skill, Personal Charisma, and Initial Attraction," *Personality and Social Psychology Bulletin*, vol. 14, no. 1, 1988, pp. 203–211.
7. H.W. Marsh and J.E. Ware, "Effects of Expressiveness, Content Coverage, and Incentive on Multidimensional Student Rating Scales: New Interpretations of the Dr. Fox Effect," *J. Educational Psychology*, vol. 74, no. 1, 1982, pp. 126–134.
8. T. Noma, L. Zhao, and N.I. Badler, "Design of a Virtual Human Presenter," *IEEE Computer Graphics and Applications*, vol. 20, no. 4, 2000, pp. 79–85.
9. E. André, J. Müller, and T. Rist, "The PPP Persona: A Multipurpose Animated Presentation Agent," *Proc. ACM Workshop on Advanced Visual Interfaces*, 1996, pp. 245–247.
10. J.C. Lester et al., "Cosmo: A Life-Like Animated Pedagogical Agent with Deictic Believability," *Proc. IJCAI Workshop on Animated Pedagogical Agents*, 1997, pp. 61–69.
11. J.C. Lester, B.A. Stone, and G.D. Stelling, "Lifelike Pedagogical Agents for Mixed-Initiative Problem Solving in Constructivist Learning Environments," *User Modeling and User-Adapted Interaction*, vol. 9, nos. 1–2, 1999, pp. 1–44.
12. J. Rickel and W.L. Johnson, "Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control," *Applied Artificial Intelligence*, vol. 13, nos. 4–5, pp. 343–382, 1999.
13. N. Reithinger et al., "Virtual Human: Dialogic and Affective Interaction with Virtual Characters," *Proc. 8th ACM Int'l Conf. Multimodal Interfaces*, 2006, pp. 51–58.
14. J.C. Lester et al., "The Persona Effect: Affective Impact of Animated Pedagogical Agents," *Proc. ACM SIGCHI Conf. Human Factors in Computing Systems*, 1997, pp. 359–366.
15. A.L. Baylor et al., "The Impact of Three Pedagogical Agent Roles," *Proc. 2nd ACM Int'l Joint Conf. Autonomous Agents and Multiagent Systems*, 2003, pp. 928–929.
16. L.A. Annetta and S. Holmes, "Creating Presence and Community in a Synchronous Virtual Learning Environment Using Avatars," I*nt'l J. Instructional Technology and Distance Learning*, vol. 3, Aug. 2006, pp. 27–43.
17. S. Barlow et al., "Animated Pedagogical Agents and Problem Solving Effectiveness: A Large Scale Empirical Evaluation," *Proc. 8th World Conf. Artificial Intelligence in Education* (AI-ED), 1997, pp. 23–30.
18. L. Dieker et al., *Evolving Environments in Learning: The TLETeachLivE™ Experience*, Am. Assoc. Colleges for Teacher Education, 2012.
19. P. Petta, C. Pelachaud, and R. Cowie, eds., *Emotion-Oriented Systems: The Humaine Handbook*, Springer, 2011.

https://youtu.be/xnYsjw6vv_4) illustrates the process for creating stimuli with our system as well as the range of charisma gestures that our instructor avatar can make.

As we mentioned earlier, our system has been used so far in five studies: three studies investigate the cognitive benefits and two the affective benefits of instructor gesture in the context of learning mathematical equivalence. A detailed description of the learning science underlying the hypothesis investigated by each study, the procedure followed by each study, and the analysis of the data collected by each study is beyond the scope of this article, which focuses on describing the system of instructor avatars. However, we briefly summarize the studies here and in the "Results" section to convey the complexity of the stimuli that these studies call for, stimuli that were successfully generated with our system.

Producing accurate stimuli for these five complex studies using the conventional approach of video recording an instructor actor would have been challenging, if not impossible. For example, in one study, the gesture had to precede or succeed speech by exactly one second, and another study required $3 \times 3 \times 2 = 18$ stimuli to cover all combinations of parallelism, direction, and amplitude.

Our work contributes a software system that allows generating stimuli for complex research studies on instructor gesture. The system was designed from the ground up with iterative feedback

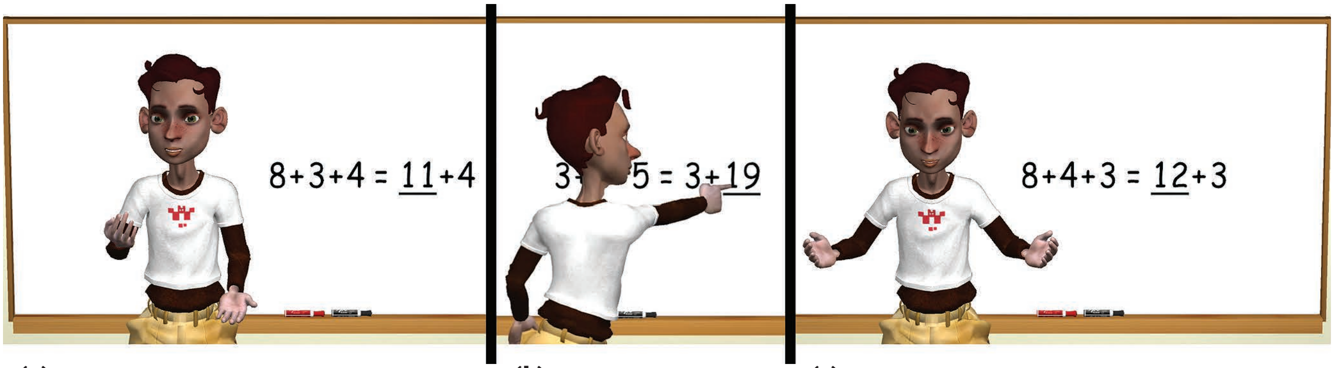


Figure 2. Frames from stimuli created with our system. In the context of learning mathematical equivalence, the stimuli were used to study the effect of (a) embodied cognition gestures, such as this balance gesture; (b) reference gestures, such as this pointing gesture; and (c) instructor charisma gestures, such this parallel outward-focused gesture.

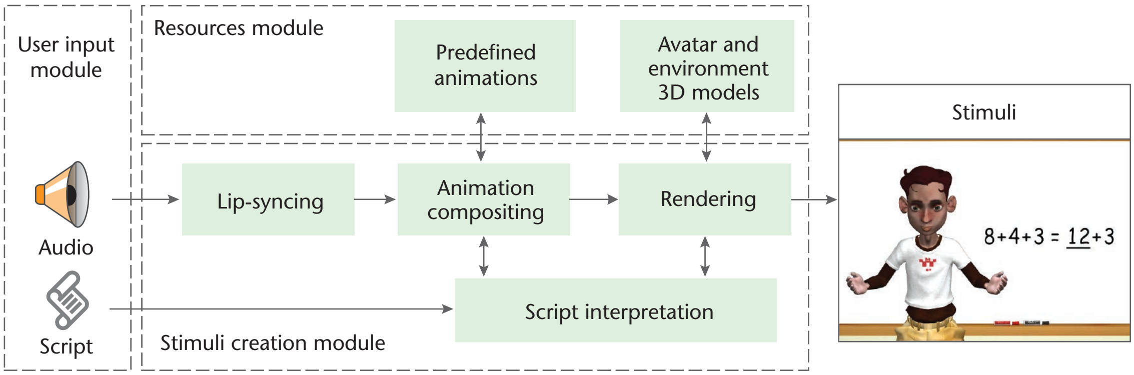


Figure 3. System architecture. The system's user input, resources, and stimuli creation modules enable the efficient creation of effective computer animation stimuli.

from our team members who are education and psychology researchers. The system has proven to be useful for studies that simply would not have been possible with the prior approaches for generating stimuli.

System Architecture

We have developed a system that enables the efficient creation of effective computer animation stimuli for research on gestures in education. Figure 3 shows the system architecture's three main modules: user input, resources, and stimuli creation.

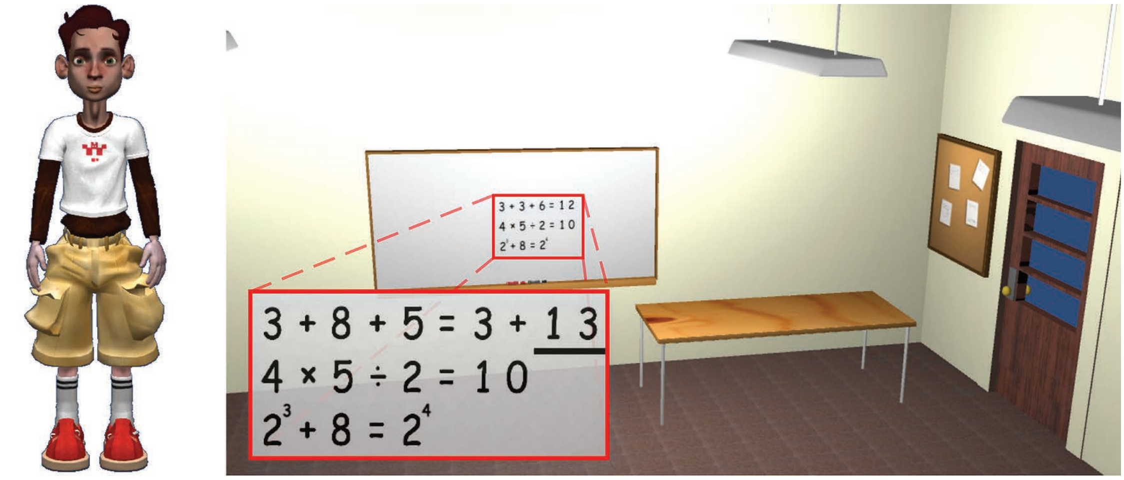
The *user input module* allows the user to create the script file that controls the stimuli generation and record the audio file that defines what the avatar says. The user edits the script file with an external text editor of their choice. The text spoken by the avatar can be read by anyone with a matching voice, and the audio is recorded using a microphone connected to a computer using an external audio recording application. The script text file and voice audio files are loaded into the stimuli creation module.

The *resources module* provides 3D models and complex animations, created by digital artists and stored in a database of resources to be used by the stimuli creation module. The resources include the avatar and the classroom environment models, avatar visemes, poses, transitions between poses, and charisma and embodied cognition gestures. This module relieves education researchers from having to create these resources themselves.

The *stimuli creation module* creates the animation stimuli leveraging the user input and the 3D model and animation resources. The avatar speaks, moves, and gestures according to the script. The avatar is lip-synced automatically to the audio. Complex animation (such as complex gestures or transitions between poses) is retrieved from the resources database, whereas simple animation is computed on the fly. Secondary motion is added according to pseudorandom patterns. The avatar, the environment, and the audio are rendered together to create the stimuli.

Resources

One major component of the resources module is the 3D models of the instructor avatars and background environment. Our long-term goal is for our system to provide multiple instructor avatars of different ages, genders, ethnicities, and personalities, covering multiple topics and participants. The work



(a) **(b)**

Figure 4. Default system resources for mathematical equivalence learning: (a) instructor avatar Jason and (b) a classroom environment with scripted equations on the whiteboard.

we report here focuses on mathematical equivalence learning, which occurs in late elementary school or in early middle school. Research indicates that children tend to be influenced by children slightly older than them.[6] Consequently, we developed an instructor avatar called Jason, which corresponds to a 15-year-old boy (Figure 4a).

The environment is a classroom with a whiteboard that can be scripted to display mathematical equivalence exercises (Figure 4b). The avatar is a partially segmented character with 12 polygon meshes and a total polygon count of 107,000. The avatar is rigged with a skeleton deformation system that includes 65 joints for the body and 24 joints for the face. The skin is attached to the skeleton via smooth binding with a maximum of five influences.

Another important component is a collection of predefined animations of the avatars. The first studies supported by our system investigated the benefits of instructor gestures in the contexts of making the instructor more engaging and improving mathematical equivalence learning. Animations too complex to be computed on the fly were created by animators in advance. These include charisma gestures, embodied cognition gestures, poses and transitions, secondary motion, and visemes.

Psychology research indicates that public speakers who gesture with their hands and arms during discourse are perceived as more charismatic and more engaging.[7] As such, our avatar can make vertical, inward, or outward gestures with the left hand, right hand, both hands in parallel, or both hands unsynchronized and with full or half amplitude. This results in 3 × 4 × 2 = 24 *charisma gestures* (Figure 5).

Embodied cognition gestures are one of a kind, and they depend on the learning context. Research on

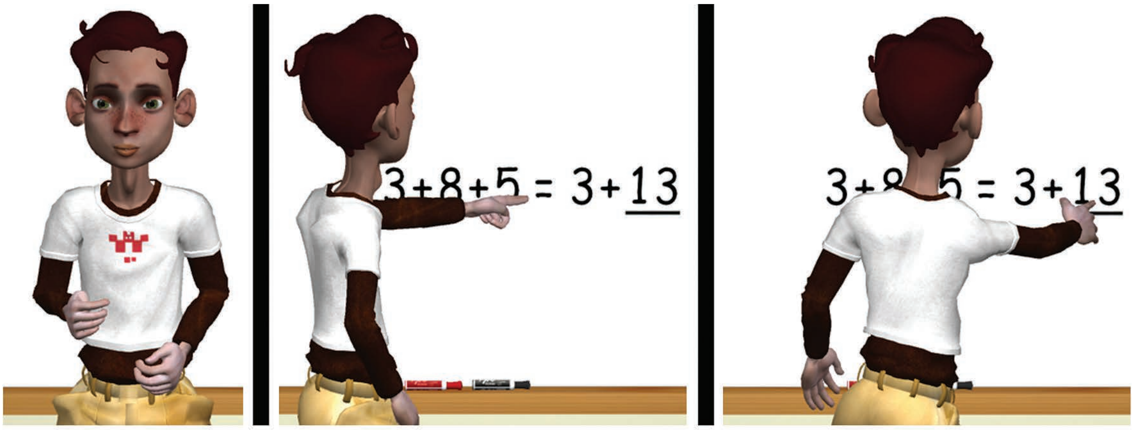
(a) **(b)** **(c)**

Figure 5. Charisma gestures: (a) vertical, (b) inward, and (c) outward two-handed parallel charisma gestures with full amplitude. Two extremal states are shown for each gesture.

mathematical equivalence learning suggests that the balance gesture is important (Figure 2a). In addition, our database contains a beat gesture (Figure 6a) that can be invoked for the avatar to segment and emphasize verbal explanations.

The avatar must be able to face the students and point anywhere on the whiteboard. These requirements are satisfied with *poses and transitions*. Specifically, we developed three poses: one frontal pose (Figure 6a), one profile pose for pointing nearby (Figure 6b), and one profile extended pose for pointing at the far end of the board (Figure 6c).

Secondary motion includes blinking, gaze shifts, head turns, and minor torso motions. Secondary



(a) (b) (c)

Figure 6. Avatar poses: (a) frontal, (b) profile, and (c) profile extended avatar poses.

motion can be switched on or off throughout the script. When on, secondary motion is added in a pseudorandom fashion: the occurrences are randomly distributed within one stimulus, but the occurrences are repeated identically for the same mathematical equivalence problem, across conditions and participants.

Lastly, the avatar mouths the words it speaks using *visemes* invoked by the lip-syncing module according to the input audio. The avatar uses a set of eight standard visemes: four for consonant sounds and four for vowel sounds.

Script Language and Script Interpretation

Education researchers are used to creating video stimuli by writing textual scripts. To change this process as little as possible, we developed a script-based interface to our avatar system. The scripting language lets users control what is written on the whiteboard, what the avatar says, the gestures the avatar makes, and how speech and gestures are synchronized.

Whiteboard commands control what is displayed on the whiteboard and when. They also let the user define interaction targets, which are used by reference gestures. This script snippet defines the whiteboard used in Figures 6b and 6c:

```
board 1
write 3+8+5{=}3+{13}
underline T2
```

In this example, the board is defined as having a single row. The mathematical equivalence is written next. Two targets are defined using braces, one for the equal sign and one for the number 13, which are automatically named T1 and T2. The number 13, referred to as T2, is underlined.

Speech commands control what the avatar says. The audio files are prerecorded and lip-sync animation is precomputed. The audio file is connected to the lip-sync animation automatically and transparently. The avatar can speak all or part of an audio recording. In the following script snippet, the second command makes the avatar speak the 3 seconds starting at second 22.

```
speak singleSentenceAudio
speak severalSentencesAudio 22.0 25.0
```

It is up to the user whether an audio file corresponds to an individual sentence, a paragraph, or the entire script. Longer audio files achieve better uniformity, but they come at the cost of multiple takes during recording to remove errors, less flexibility in combining sentences, and more complex time offsets for referencing.

Gesture synchronization commands let the user attach avatar gestures to speech. A gesture synchronization command is a subcommand of a speak command with the following general structure: `<syncType> <timeOffset> <command> <parameters>`.

The time offset is specified in seconds and is defined relative to the beginning of the current speech command, after the end of the previous synchronization command, or before the end of the previous synchronization command, using a value of the `<syncType>` field of `@`, `+`, or `–`, respectively. The `point`, `underline`, and `circle` commands let the user refer to a target defined by a `write` command with a pointing, underlining, and circling gesture. The `gesture` command invokes an animation database gesture. The `move` command transitions the avatar from the current to the new pose specified. The `pause` command suspends speech for the duration specified.

The following script snippet has the avatar speak the words in the "singleSentenceAudio" resource.

```
speak singleSentenceAudio
    @ 0.5 move B
    @ 1.2 point left T1
    + 1.1 move C
    + 0.0 point right T2
    + 0.3 move A
    - 0.1 gesture balance speed 1.8
```

After 0.5 sec from the beginning of the speech sequence, the avatar moves from the initial pose A to pose B. At 1.2 sec into the sequence, the avatar points with its left hand to target T1, attached to the equal sign in the script snippet. Then the avatar moves to pose C, which allows it to reach target "13," to which it points with the right hand. Finally, the avatar moves back to pose A and makes the balance gesture. The gesture begins 0.1 sec before actually reaching pose A, and the gesture is performed accelerated by a factor of 1.8.

Upon finishing, the script can be simply treated as any text file and can be published easily via online download or can be simply attached to an email. Then the script can be played back at any time in the desired resolution locally for user studies or can be recorded into videos for noncompatible operation systems.

In a pilot user study with seven individuals (two had a computer science background, one an animation background, and four a psychology background), all participants were able to produce a meaningful result within 1 hour after watching the online tutorial, and all gave positive feedback on using only text.

Results

We implemented our system at Purdue University using the third-party game engine Unity 3D (unity3d.com). Unity 3D scripting was done in C#. The 3D model and animation resources were developed in the commercial animation system Maya (www.autodesk.com/products/autodesk-maya/overview) and were imported directly into Unity. The stimuli are delivered either as a standalone executable for the Microsoft Windows and Apple Mac OS platforms or as video files. The audio files were recorded by the education researchers. Lipsyncing was done using the Maya plug-in Voice-O-Matic (www.di-o-matic.com/products/plugins/VoiceOMatic/). The system was developed in the Computer Science Department at Purdue University, with formative feedback from psychology and education psychology researchers from the University of California, Riverside and the University of Iowa.

The system has since been used to generate stimuli for five different studies on instructor gestures in the context of mathematical equivalence. Three studies conducted at the University of Iowa investigated the cognitive benefit of instructor gestures. Two studies conducted at the University of California, Riverside investigated the affective benefit of instructor gestures. See the accompanying video web extra for sample stimuli and the gesture vocabulary.

Cognitive Studies

Historically, it has been difficult to empirically probe the effect of gesture on learners because it is difficult for human actors to produce controlled, yet natural gestures. In any experiment using live actors, there are a number of potential confounds across experimental gesture conditions. For example, prosody, eye gaze, speech rate, and other nonverbal behaviors generally vary with gesture.

To address this, our studies used the gesturing avatar to determine whether adding gesture to instruction increases learning. Using the avatar, we were able to control for prosody, eye gaze, speech rate, and other nonverbal behaviors in three studies with multiple conditions, while ensuring that the gestures produced were naturalistic.

Study 1: Do instructor gestures promote learning? In the first study, third and fourth grade children (N = 65) participated in a tutorial on mathematical equivalence delivered by an avatar that either did or did not use gestures. Here we only give a brief overview of the study, which is described in detail in a separate publication.[8] All stimuli were presented on computers using the PsyScope system.[9] Prior to the stimuli, children took a pretest of six matching addends mathematical equivalence problems on the computer. Each problem had three single-digit numbers (2–9) on the left side of the equal sign and one number and an answer blank on the right side. The answer blank occurred in either the left or right positions on the right-hand side. The addend on the right side was always the same as one addend from the left side, and it was located in a matching position across the equal sign—for example, 5 + 9 + 7 = 5 + _ or 7 + 8 + 4 = _ + 4.

After this pretest, the children (N = 38) who had scored correctly on fewer than three of the six problems participated in a video tutorial. Those children first viewed an introduction from the avatar about the principle of mathematical equivalence and then viewed six explanations of novel, equal addends mathematical equivalence problems presented by the avatar in a fixed order.

The children were randomly assigned to an experimental condition: a *gesture condition* where the introduction and the six videos contained accompanying hand gestures or a *no gesture condition* where the videos did not contain any hand gestures. After each avatar explanation, the children solved an equal addends mathematical equivalence problem of their own using the computer. After the lesson, children completed a posttest on the computer. Posttest problems were presented in a fixed order: six problems of novel equal addends equivalence problems identical to form to the pretest, then four transfer problems where two with equal addends not located in a matching position and two with no equal addends, and finally six conceptual true or false questions about equality adapted from prior research.[10] The children's responses and reaction times were recorded for analysis.

Posttest data suggests that avatar gesture contributes to learning and that gestures are particularly helpful when they anticipate or synchronize with speech.

We used a logistic mixed-effects model to predict the log odds of answering each problem correctly as a function of the interaction of condition and test, with a random intercept and random slopes for all terms. We found that the children who saw the gesturing avatar performed much better than those who saw the nongesturing avatar (posttest: 76 versus 61 percent, transfer: 82 versus 63 percent, and conceptual: 94 versus 86 percent), which was reflected in a significant effect of condition ($\beta = -2.05$, $p = 0.024$) and no interaction with test.

These findings were expected given similar findings using live and prerecorded stimuli.[3] However, these findings go beyond prior work in that they provide more convincing evidence that the effect observed is due to hand gestures rather than to some associated confound. Although of course the gestures produced by the avatar are somewhat contrived, children in our study found the avatar appealing and engaging, and the learning rates observed were high when compared with prior work. These findings suggest that the avatar can provide a useful tool for more nuanced studies of the role of gesture in learning. Our future work will vary the nature of the gestures produced by the avatar to help us better understand the learning process.

Study 2: Does introducing ideas in gesture first promote learning? We used our system to develop stimuli for a second study that investigated whether introducing ideas with gestures before they are introduced in speech leads to greater learning than introducing ideas using speech and gestures simultaneously.

Third and fourth graders ($N = 48$) were randomly assigned to two conditions. In the gesture-first condition, representations of mathematical equivalence were produced with gestures prior to being produced via speech. In the gesture-first condition, gestural representations of mathematical equivalence were produced with gestures on instructional trials prior to these explanations being produced in speech. In the gesture with speech condition, gestural and spoken representations of equivalence were produced on the same trial. The study results shows that introducing ideas concurrently in gesture and speech led to significantly more learning.

Study 3: Which timing of speech-gesture coordination promotes learning? We next investigated the importance of gesture and speech timing in a third study where we manipulated the in-the-moment timing of gesture as it synchronizes with speech. Specifically, we used our system to develop stimuli to investigate the cognitive benefits of instructor gesture in the context of learning mathematical equivalence. Again, third and fourth graders ($N = 78$) participated in a tutorial lesson on mathematical equivalence on the computer. The participants were randomly assigned to one of three conditions presented by a gesturing avatar. The animations contained gestures timed to anticipate speech by 1 second, synchronize with speech as in the first cognitive study, or follow speech by 1 second.

The analysis of the posttest data suggests that avatar gesture contributes to learning and that gestures are particularly helpful when they anticipate or synchronize with speech. The data suggests that students do poorly when they view gestures that occur after the accompanying speech, as students in this condition performed particularly poorly on the posttest (57 percent correct versus 82 percent correct when gestures came early and 70 percent correct when gestures were synchronized with speech). Moreover, an examination of eye-tracking data suggests that eye movements to appropriate problem elements are delayed in this condition, suggesting a mechanism for this effect.

The first of its kind, this study using our system allowed for the precise control of the timing of gesture and speech coordination. Generating stimuli for a study such as this using the traditional approach of video recording a teacher actor would simply not be possible. Using our system, modifying the stimuli of the first cognitive study to obtain these new stimuli took less than 5 minutes; it only required shifting the gesture timeline 1 second earlier or later. This important result regarding the benefit of gesture anticipating speech will be applied to all future developments of instructor gesture stimuli.

Affective Studies

We know that teachers who gesture can increase student engagement, but research in nonverbal communication has struggled to identify the best-suited gestures because individual gestures have not yet been able to be manipulated and tested independently of secondary motion. Human actors asked to perform different charisma gestures for comparison involuntarily include additional, ever-changing motions that confound the studies.[11,12]

Study 4: Which gestures make the instructor more appealing? Stimuli for our fourth study were developed iteratively. First, we located real-world examples of relevant gestures in online video clips of charismatic speakers. A digital artist then replicated these examples for the instructor avatar. The avatar gestures were refined several times with feedback from psychology researchers. Once the gestures were deemed acceptable, animation stimuli were created using a script that controls the gestures' type, timing, arc, and speed in the context of mathematical equivalence instruction. A total of 18 stimuli were created based on the level of parallelism (single arm, two arms unsynchronized, or two arms in parallel), direction of movement (vertical, inward, or outward), and movement amplitude (full or partial). Once one script is created, the other scripts can be created quickly via minor modifications to the first script.

These stimuli were shown to undergraduate students ($N = 56$). Each student saw all 18 stimuli in one of 14 preselected counterbalanced orders in which the amplitude alternated and no more than two levels of parallelism or direction of movement appeared next to each other. After watching each video, the participants answered nine questions about avatar characteristics such as attractiveness, dominance, likeability, confidence, and clarity. During the debriefing, some participants commented that they noticed changes in the avatar's voice from segment to segment, although in actuality, the avatar's voice was exactly the same for all segments. This indicates the importance of gestures in influencing perceptions of the avatar.

The participants also answered questions about their own charisma, personality, mood, and demographics. A factor analysis isolated two dimensions in the ratings: charisma and attractiveness. Then contrast analyses showed that parallel outward focused gestures were particularly charismatic. This was confirmed via a mixed-modeling procedure, which showed that parallelism did not significantly predict charisma ratings (compared with a null model, $\Delta \chi^2(1) = 2.9$, $p = 0.09$). However, adding the direction of movement ($\Delta \chi^2(1) = 5.7$, $p = 0.02$) and the interaction ($\Delta \chi^2(1) = 8.8$, $p = 0.003$) significantly decreases deviance, confirming the results of the focused contrast analyses. This is the first time gestural charisma has been isolated in a true experimental research design.

Compared with previous studies using human instructor actors, the avatar system enables us to test the effects of individual channels of communication while efficiently and conveniently holding other channels constant, which has never been done before. The most appealing charisma gestures can be used by human and animated instructors to increase engagement and learning in various levels of education.

Study 5: Do charisma gestures promote embodied cognition? The findings from our fourth study—namely, that parallel outward-focused gestures are particularly charismatic—were used in a subsequent study. Undergraduate students ($N = 72$) were randomly assigned to view one of four avatar animations, with versions of the same elementary mathematics lesson content, presented by a gesturing avatar (in a 2 × 2 between-subjects design). The animated lessons contained either no gestures; charismatic, parallel outward-focused gestures; explanation-related comprehensive movements and beats; or both types of gestures. The students were then filmed giving a mathematics lesson and were told that the recordings would later be shown to elementary students. The participants were not primed to note gestures when viewing the avatar videos. The participants did not know the true goal of the study was to investigate whether those exposed to particular gestures would give better lessons.

Six judges later rated the participant videos for charisma and attractiveness. Results were reliable ($\alpha = 0.87 - 0.91$) and in line with theoretical predictions, showing that the participants exposed to charisma gestures were later rated more favorably in their own lessons by judges than were the participants who were not exposed to charisma gestures. The participants who used more gestures ($r = 0.32$, $p = 0.007$), produced a higher amount of outward-focused gestures ($r = 0.39$, $p = 0.0009$), and produced more balance gestures ($r = 0.25$, $p = 0.04$) were rated by judges as more charismatic.

The participants were less likely to produce nervous gestures (such as fidgeting and clenched fists) if they viewed videos with charismatic or comprehensive gestures. Finally, relevant to embodied cognition, the participants giving the lessons were more likely to produce comprehensive, explanation-related mathematics gestures (V, sweep, and balance gestures) if they were exposed to them via the animated avatar. This demonstrates a possible core mechanism by which an avatar's gestures can help facilitate learning, outside of awareness.

Our system has the potential to change the way stimuli are produced for the study of gesture in education and to become the standard tool for doing so.

Lastly, the participants exposed to charismatic videos were rated more favorably by naive judges, but they did not exactly imitate the charisma gestures they were exposed to. Rather the students who viewed more comprehensive, explanation-related gestures picked up these gestures themselves in a comparable situation, in line with theories of embodied cognition.

Our system relies on animation that is either computed on the fly or precomputed and stored in a database. This allows users to generate stimuli without the need for computer animation expertise or artistic talent. However, this comes at the cost of a loss of animation quality compared with a one-of-a-kind animation produced manually by a digital artist using conventional keyframe animation or motion capture. We have conducted a comparison of the animation produced by our system with one-of-a-kind animation produced manually by a skilled digital artist. Although the one-of-a-kind animation was judged to be more realistic, the gesture researchers found the animation produced by our system more appropriate for formal user studies because the avatar in the one-of-a-kind animation exhibited nonverbal expression variability, which could be a confounding factor.[13] Furthermore, manual animation does not scale and generating complex stimuli for multiple conditions and multiple studies is prohibitively expensive. Finally, manual animation does not have the potential of scripted animation to remove the authoring bottleneck in e-learning.

Our system was designed with extensibility in mind. We are currently extending it with a second avatar that corresponds to a professionally dressed young adult and to additional math topics, such as linear dependency and polynomial multiplication. We are currently using the system to generate stimuli for four more studies on the use of gesture to link various representations of concepts from these additional topics. Our system has the potential to change the way stimuli are produced for the study of gesture in education and to become the standard tool for doing so. Three education research groups have contacted us to adopt our system for the production of stimuli for their research studies.

One direction of future work is to reduce the scripting learning curve by providing a way to synchronize gestures to speech based on text and not on time stamps and by providing a GUI to help avoid scripting language syntax errors. Another area of future work is to research whether the optimal gesture rules inferred from using the avatar translate to actual teachers. The system has potential beyond research on gesture such as achieving scalability in the authoring of high-quality digital-learning materials, professional development for education students and teachers, and beyond education, in all contexts where effective public speaking is important.

Acknowledgments

We thank Thomas Haley-Hermiz and Mary Spalla for their help with 3D modeling and animation. This work was supported by the United States National Science Foundation through grants 1217215, 1217137, and 1216984.

References

1. E. Sapir, "The Unconscious Patterning of Behavior in Society," *Language, Culture, and Society: A Book of Readings*, B. Blount, ed. Waveband Press, 1927, pp. 29–42.
2. S. Goldin-Meadow, S. Kim, and M. Singer, "What the Teacher's Hands Tell the Student's Mind about Math," *J. Educational Psychology*, vol. 91, no. 4, 1999, pp. 720–730.

3. S.W. Cook, R.G. Duffy, and K.M. Fenn, "Consolidation and Transfer of Learning after Observing Hand Gesture," *Child Development*, vol. 84, no. 6, 2013, pp. 1863–1871.
4. L.A. Annetta and S. Holmes, "Creating Presence and Community in a Synchronous Virtual Learning Environment Using Avatars," *Int'l J. Instructional Technology and Distance Learning*, vol. 3, Aug. 2006, pp. 27–43.
5. J. Holmes, "Designing Agents to Support Learning by Explaining," *Computers & Education*, vol. 48, no. 4, 2007, pp. 523–547.
6. S. Ellis and B. Rogoff, "The Strategies and Efficacy of Child Versus Adult Teachers," *Child Development*, vol. 53, no. 3, 1982, pp. 730–735.
7. R.B. Church, S. Ayman-Nolley, and S. Mahootian, "The Role of Gesture in Bilingual Education: Does Gesture Enhance Learning?" *Int'l J. Bilingual Education and Bilingualism*, vol. 7, no. 4, 2004, pp. 303–319.
8. S. Cook et al., "Hand Gesture and Mathematics Learning: Lessons From an Avatar," *Cognitive Science*, vol. 41, no. 2, 2017, pp. 518–535.
9. J.D. Cohen et al., "Psyscope: A New Graphic Interactive Environment for Designing Psychology Experiments," *Behavioral Research Methods, Instruments, and Computers*, vol. 25, no. 2, 1993, pp. 257–271.
10. P. Matthews et al., "Measure for Measure: What Combining Diverse Measures Reveals about Children's Understanding of the Equal Sign as an Indicator of Mathematical Equality," *J. Research in Mathematics Education*, vol. 43, no. 3, 2012, pp. 316–350.
11. H.S. Friedman, R.E. Riggio, and D.F. Casella, "Nonverbal Skill, Personal Charisma, and Initial Attraction," *Personality and Social Psychology Bulletin*, vol. 14, no. 1, 1988, pp. 203–211.
12. H.W. Marsh and J.E. Ware, "Effects of Expressiveness, Content Coverage, and Incentive on Multidimensional Student Rating Scales: New Interpretations of the Dr. Fox Effect," *J. Educational Psychology*, vol. 74, no. 1, 1982, pp. 126–134.
13. N. Adamo-Villani, J. Cui, and V. Popescu, "Scripted Animation towards Scalable Content Creation for E-Learning: A Quality Analysis," *E-Learning, E-Education, and Online Training*, Springer, 2014, pp. 1–9.

Jian Cui *is a software engineer at Google. During this research, he was a PhD candidate and research assistant in the Department of Computer Science at Purdue University. His research interests include computer vision and computer graphics in education and psychological research. Cui has a PhD in computer science from Purdue University. Contact him at cui9@purdue.edu.*

Voicu Popescu *is an associate professor with the Department of Computer Science at Purdue University. His research interests include camera model design, remote visualization, aggressive and exact visibility computation, and applications of computer graphics in education. Popescu has a PhD in computer science from the University of North Carolina at Chapel Hill. Contact him at popescu@purdue.edu.*

Nicoletta Adamo-Villani *is a professor in the Department of Computer Graphics Technology and a faculty scholar at Purdue University. She is also the cofounder and director of the IDEAlaboratory. Her research interests include character animation, character design, 3D animation technology to education, human-computer communication, and visualization. Contact her at nadamovi@purdue.edu.*

Susan Wagner Cook *is an assistant professor in the Department of Psychological and Brain Sciences and a member of the DeLTA Center at the University of Iowa. Her research interests include communication and learning, with a focus on the nature and function of hand gesture in reflecting and supporting ongoing cognitive processes. Cook has a PhD in psychology from the University of Chicago. Contact her at susan-cook@uiowa.edu.*

Katherine A. Duggan *is a postdoctoral fellow in the Department of Psychiatry at the University of Pittsburgh. Her research interests include associations between personality, sleep, health throughout the lifespan, and applied work in charisma and nonverbal communication. Duggan has a PhD in social and personality psychology from the University of California, Riverside. Contact her at dugganka@upmc.edu.*

Howard S. Friedman *is distinguished professor in the Department of Psychology at the University of California, Riverside. His research focuses on nonverbal expressiveness and charisma. He is the editor in chief of the Journal of Nonverbal Behavior and coauthor of The Longevity Project (Plume, 2012). He received the James McKeen Cattell Award from the Association for Psychological Science (APS) and is a fellow of the American Association for the Advancement of Science (AAAS). Friedman has a PhD in psychology from Harvard University. Contact him at howard.friedman@ucr.edu.*