2014

An Animation Stimuli System for Research on Instructor Gestures in Education

Jian Cui
Purdue University, cui9@purdue.edu

Voicu Popescu
Purdue University, popescu@purdue.edu

Nicoletta Adamo-Villan
Purdue University

Susan Cook
University of Iowa

Katherine Duggan
University of California - Riverside

See next page for additional authors

Report Number:

14-001

An Animation Stimuli System for Research on Instructor Gestures in Education

Jian Cui, Voicu Popescu, Nicoletta Adamo-Villan, Susan Cook, Katherine Duggan, Howard Friedman

***Abstract*—We present a system for creating animation stimuli for the study of instructor gestures in education. The system does not require artistic talent or programming expertise. The system provides an animation character that serves as an instructor avatar. The avatar can speak, write on a nearby white board, and make pointing, embodied cognition, and charisma gestures. The avatar is controlled with a text script that specifies when and what the avatar says and does. The script is executed automatically to create the animation stimuli. The system has been used so far in two studies. The first study investigated which type of charisma gestures makes the instructor more appealing by testing 18 gesture conditions. The second study compared the learning of mathematical equivalence after a lesson given by the avatar, with and without gestures. This shows that the system can be used efficiently to create accurate and effective stimuli for complex studies.**

***Index Terms*—Computers and Education, Computer-assisted instruction (CAI)**

I. Introduction

Gestures play an important role in education. Gestures such as pointing, circling, or underlining can help capture, maintain, and direct the student's attention during lecture. Instructor gestures serve as a complementary communication channel that helps students parse and grasp the information presented to them verbally and visually. 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 help convey an instructor's appealing and engaging personality, which enhances learning. Important as gestures seem to be, much remains to be discovered through education research regarding what, when, and how instructor gestures benefit learning.

Traditional education research on instructor gestures relies on video stimuli. 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 what gestures have to occur and at what time; then the instructor actor learns and performs the script, and the performance is recorded. An obvious and important advantage of video stimuli is that they are by definition photo-realistic: 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 and detailed scripts for multiple conditions is challenging. Any error in delivery requires an additional take, from the beginning, to ensure the correctness and fluidity of the final stimuli. The instructor actor has difficulty repeating, from condition to condition, the same level of energy, enthusiasm, and voice intensity, as well as the same secondary motion (e.g. head, body and non-gesturing arm/hand motions). As a consequence, producing video stimuli is very time consuming, which limits the complexity of the studies that can be conducted. Moreover, video stimuli often appear constrained and unnatural. Instructor actors have to focus on remembering the script and the exact secondary parameters of the delivery, and they cannot simply be teachers. Not all differences in secondary parameters between conditions can be eliminated, which can confound the analysis of the effect on which the study focuses.

Advances in computer graphics software and hardware present the opportunity to enroll computer animation characters as instructor avatars to be used in education research. Instructor avatars have perfect memory and infinite energy thus they are well suited for precisely executing complex scripts. Repeatability is not a problem–all secondary parameters can be perfectly controlled. One can use the exact same audio and secondary motion from condition to condition. However, generating animation stimuli with instructor avatars presents challenges of its own.

The first potential challenge is that the avatar is not a "real" instructor. Do students learn from such avatars? Does what is discovered about avatar instructor gestures apply to human instructors? 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 by itself. Moreover, emerging research shows that what is learned about gesture in the context of instructor avatars does apply to human instructors. The second challenge is that current computer animation software systems require artistic talent and technical expertise to model, animate, and control characters.

In this paper we present our work towards overcoming this second challenge: we describe an approach that enables the creation of animation stimuli for research on instructor gesture in education, without the prerequisites of artistic talent or programming expertise; we describe a system that implements the approach; and we report on using our system to investigate the benefits of instructor gesture in the context of mathematical equivalence learning.

Our approach is based on two observations. First, one cannot expect education researchers to model and animate their own characters. Instead, a database of pre-modeled characters should be available from where the instructor avatar is selected. Whereas simple animations should be computed

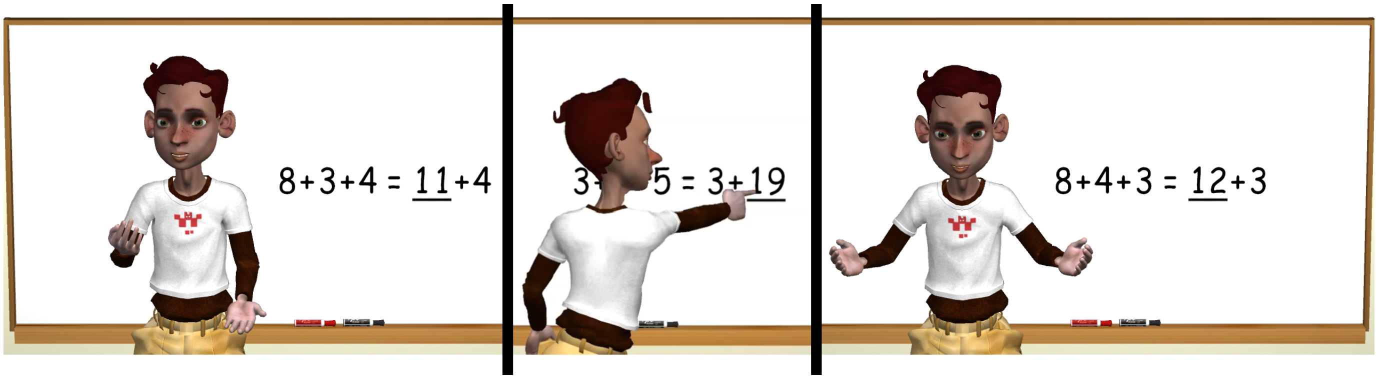

Fig. 1: Frames from animation stimuli created with our system. The stimuli were used to study the effect of embodied cognition gestures, e.g. balance gesture in left frame, of reference gestures, e.g. pointing gesture in middle frame, and of instructor charisma gestures, e.g. parallel outward-focused gesture in right frame, in the context of learning mathematical equivalence.

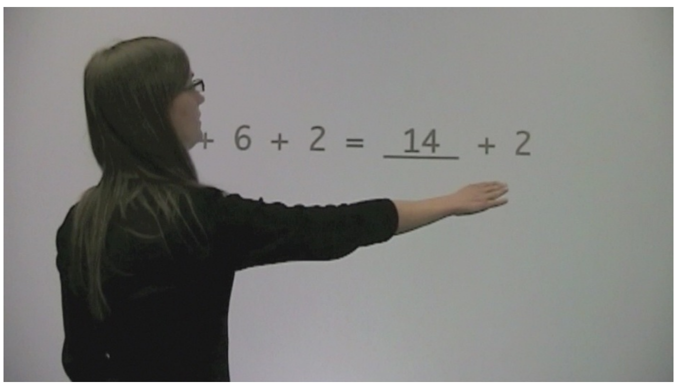

Fig. 2: Sample frame from traditional video stimuli used in research on instructor gesture in education. The stimuli are created by video recording instructors giving lessons scripted for various gesture conditions.

on the fly, using inverse kinematics animation algorithms, complex animations should be predefined and made available through the database. Second, one 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 that is executed automatically.

We have developed a system that implements this approach. The system provides an animation character that serves as an instructor avatar. The avatar can speak, write on a nearby white board, and point to, circle, and underline any location on the board. The avatar can make embodied cognition gestures, such as the balance gesture, and charisma gestures, such as parallel outward-focused gestures (Figure 1). The avatar is controlled with a script written by the education researcher using a text editor. The scripting language allows specifying what and when the avatar says and does. The script is executed automatically to create the animation stimuli. We also refer the reader to the accompanying video that illustrates the process for creating stimuli with our system, as well as the range of charisma gestures that our instructor avatar can make.

Our system has been used so far in two studies on instructor gestures. The first study investigated which type of charisma gesture makes the instructor more appealing. The study involved 56 college students. The gestures tested differed based on whether they were executed with one hand, with two hands unsynchronized, or with two hands in parallel, whether they were inward, vertical, or outward, and whether they were of small or large amplitude (i.e. $3 \times 3 \times 2 = 18$ stimuli). The study reveals that parallel outward gestures make the instructor most appealing. Our system allowed creating accurate stimuli for the 18 conditions with little effort. Doing the same with conventional video stimuli would be prohibitively challenging.

The second study measured learning of mathematical equivalence before and after exposure to a lesson given by the avatar, consisting of 6 examples. The study involved 51 third and fourth grade students. In one condition the avatar made parallel outward-focused, balance, beat, pointing and underlining gestures, whereas in the control condition the avatar made no overt gestures but had comparable body, head, and arm movement. The study reveals that students exposed to the gesture conditions learned substantially more. This study demonstrates that students do learn from computer animation instructor avatars, and that gestures help in the case of instructor avatars. Moreover, the measured benefit of gestures is comparable to that measured in similar studies involving human instructors.

In summary, our paper makes the following contributions:

- a script-based system for creating animation stimuli to be used in educational psychology studies;
- a database seeded with a first instructor avatar and multiple gesture animations;
- a validation of the system in two studies conducted by psychology and educational psychology researchers.

II. Prior work

Studies of ethnology and linguistics have shown for over a hundred years that we respond to gestures with extreme alertness and in an unknown way [1]. Broad studies have been conducted regarding how instructor gestures influence the learning process. Math and science teachers frequently use their hands to depict information [2], [3], [4], [5]. Gesture in instruction has been shown to increase learning of a wide variety of mathematical concepts, including mathematical equivalence [6], [7], [8]. For example, a study of Piagetian conservation [9] compared instruction with gesture to instruction without gesture and found that instruction with gesture 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 increases transfer of knowledge to novel contexts [10]. Gesture is a spontaneous behavior that is routinely incorporated into human communication, and so it is perhaps not surprising that it is generally helpful for listeners.

Studies also show that an appealing and engaging instructor with a charismatic personality can also promote learning [11]. The general finding of these studies was that an enthusiastic, expressive lecturer is more highly valued [12], [13], [14]. All the studies mentioned so far were conducted using videos of instructors or instructor actors, and generating accurate video stimuli is challenging as described in the introduction.

Animated avatars have been incorporated into general information systems such as Jack [15] and PPP Persona [16]. Early examples of instructor avatars are Cosmo [17], a cosmonaut who explains how the Internet works, Herman [18], [19], a bug-like creature that teaches children about biology, and STEVE [20], [21], who trains users in operating complex machinery using speech, pointing gestures, and gaze behavior. In the Virtual Human project [22] a teacher gives astronomy lessons by following different pedagogical paradigms and shows a variety of nonverbal behaviors that reflect different parameterized personality settings. Many studies confirm the intended positive influences on education by systems using avatars [23], [24], [25], [26], [27], [28]. Studies also suggest that teaching avatars could be employed in e-learning environments to enhance the users' attitude towards on-line courses [29]. Moreover, animated avatars help create a positive learning environment [30]. 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 test bed 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. Character Builder [31] is a tool that lets non-animators 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 [32] allows users to add 2-dimensional animated talking avatars to slide show presentations and web sites. Codebaby [33] enables non-expert users to create and integrate 3D animated characters into e-learning course-ware. Gesture Builder [34] is a commercial software product for producing animated 3D characters that gesture and sign. Although the characters produced using these products 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.

III. System Overview

We have developed a system which enables the efficient creation of effective computer animation stimuli for research on gestures in education. The system architecture comprises three main modules as shown in Figure 3: *user input module*, *resources module*, and *stimuli creation module*.

User input module. This module allows creating the script file that controls the stimuli generation, and recording 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 is 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 the voice audio files are loaded into the stimuli creation module.

Resources module. This module provides 3-D 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, transition between poses, and charisma and embodied cognition gestures, as described in Section IV. This module relieves the education researchers from having to create these resources themselves.

Stimuli creation module. This module creates the animation stimuli leveraging the user input and the 3-D model and animation resources. The avatar speaks, moves, and gestures according to script, which is interpreted as described in Section V. The avatar is lip-synced automatically to the audio. Complex animation (e.g. complex gestures or transitions between poses) are retrieved from the resources database, whereas simple animation is computed on the fly. Secondary motion is added according to pseudo-random patterns. The avatar, the environment, and the audio are rendered together to create the stimuli.

IV. Resources

A. 3-D models

Our long term goal is for our system to provide multiple instructor avatars, of different age, gender, ethnicity, and personality, covering multiple topics and participants. The work reported 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 they are [35]. Consequently we have developed an instructor avatar, called Jason, which corresponds to a fifteen year old boy (Figure 4, left). The environment is a classroom with a white-board that can be scripted to display mathematical equivalence exercises

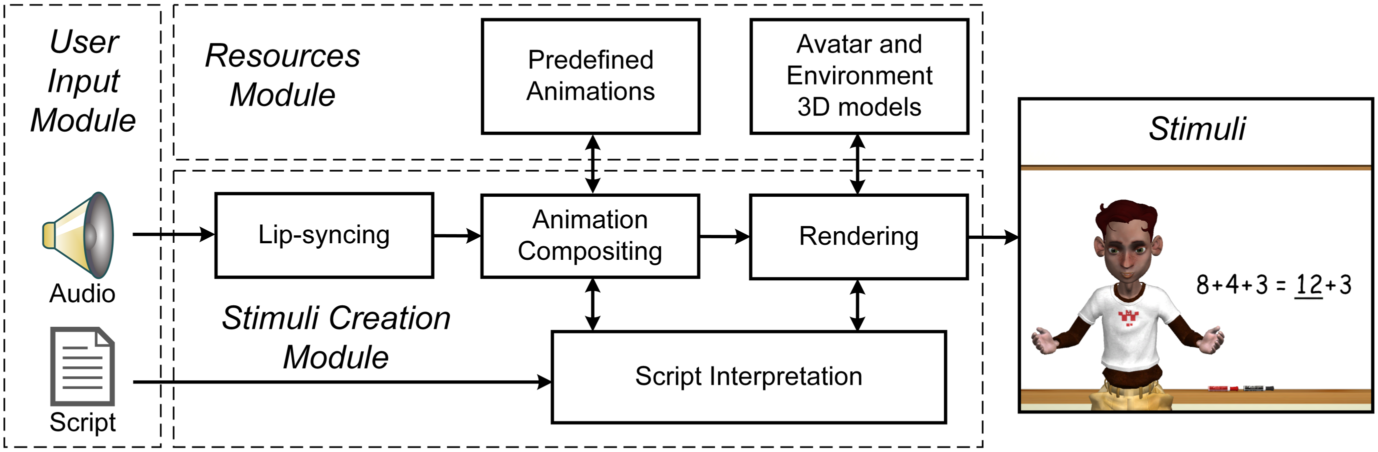

Fig. 3: System architecture overview

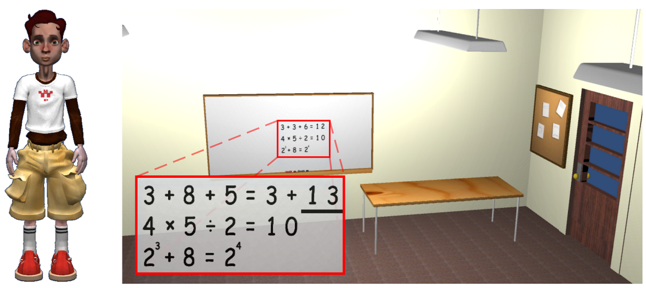

Fig. 4: Instructor avatar Jason (left) and classroom environment with scripted equations on the board (right).

(Figure 4, right). The avatar is a partially segmented character with 12 polygon meshes, and a total polygon count of 107K. 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 number of influences of 5.

B. Predefined animation

The first studies supported by our system investigate the benefits of instructor gestures in the contexts of making the instructor more engaging and of improving learning of mathematical equivalence. Animation too complex for being computed on the fly was created by animators in advance, and includes:

Charisma gestures. Psychology research indicates that public speakers who gesture with their hands and arms during discourse are perceived as more charismatic and more engaging [9]. As such, our avatar can make vertical, inward or outward gestures, with the left hand, with the right hand, with both hands in parallel, or with both hands unsynchronized, and with full or half amplitude. This results in $3 \times 4 \times 2 = 24$ charisma gestures (Figure 5).

Embodied cognition gestures. Embodied cognition gestures are one of a kind and they depend on the learning context. Research on mathematical equivalence learning suggests that the balance gesture is important. In addition to the balance gesture (Figure 1, left) our database also contains a beat gesture (Figure 6, left) that can be invoked for the avatar to segment and emphasize verbal explanations.

Poses and transitions. The avatar has to be able to face the students and it has to be able to point anywhere on the white-board. These requirements are satisfied with three poses: one frontal pose (pose A), one profile pose for pointing nearby (pose B), and one profile extended pose for pointing at the far end of the board (pose C), see Figure 6.

Secondary motion. Secondary motion includes blinking, gaze shifts, head turns, and minor torso motions. Secondary motion can be switched on or off through the script. When on, secondary motion is added in pseudo-random 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.

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

V. Script Language and Script Interpretation

Education researchers are used to creating video stimuli by writing textual scripts. In order to change this process as little as possible, we have developed a script-based interface to our avatar system. The scripting language allows controlling what is written on the white-board, what the avatar says, what gestures the avatar makes, and how speech and gestures are synchronized.

White-board commands control what is displayed on the white-board and when. They also allow defining interaction targets which are used by reference gestures. The script snippet below defines the white-board used in Figure 6, middle and right. 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 number 13, which are named automatically T1 and T2. The number 13, referred to as T2, is underlined.

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

Fig. 5: Vertical (left), inward (middle), and outward (right) two-hand parallel charisma gestures with full amplitude. Two extremal states are shown for each gesture.

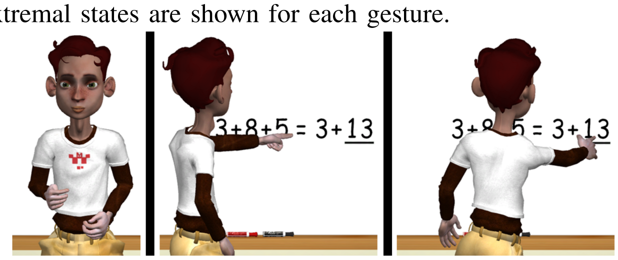

Fig. 6: Frontal (left), profile (middle), and profile extended (right) poses of the avatar.

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, transparently to the user. The avatar can speak all or only part of an audio recording, see script snippet below, where the second command makes the avatar speak the three seconds starting at second 22. It is up to the user whether an audio file corresponds to an individual sentence, to a paragraph, or to 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.

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

Gesture synchronization commands allow attaching 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 it 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 commands *point, underline*, and *circle* allow referring to a target defined by a write command with a pointing, underlining, and circling gesture. The command *gesture* invokes an animation database gesture. The command *move* transitions the avatar from the current to the new pose specified. The command *pause* suspends speech for the duration specified.

```
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
```

The script snippet above has the avatar speak the words in the ”singleSentenceAudio” resource. After 0.5s from the beginning of the speech sequence the avatar moves from the initial pose A to pose B. At 1.2s into the sequence, the avatar points with its left hand to target T1, attached to the equal sign in the script snippet above. 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 0.1s before actually reaching pose A, and the gesture is performed accelerated by a factor of 1.8×.

The speech commands are interpreted one at the time. All gestures synchronization commands associated with a given speech command are sorted based on their absolute starting time with respect to the start of the speech and are executed in order. Inverse kinematics algorithms are used to compute the animation for pointing, underlining, and circling gestures. It is the script's responsibility to position the avatar such that the intended target is within the reach of the avatar. This way the script can ask the avatar to point in the direction of a target without actually touching it. The stimuli can be generated on the fly by executing the script during the study or it can be saved to a video file.

VI. Results

We have implemented the system using a third party game engine, i.e. Unity 3D [36]. Unity 3D scripting was done in C#. The 3-D model and animation resources were developed in a commercial animation system, i.e. Maya [37], 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. Lip-syncing was done using the Maya plug-in Voice-O-Matic [38]. The system was developed at the computer science department of one university, with formative feedback from psychology and education psychology researchers from two other universities. The system has then been used to generate stimuli for two different studies on instructor gestures in the context of mathematical equivalence. The studies were conducted at the psychology and education psychology departments of the two universities that provided feedback and not at the university where the system was built. Sample stimuli and the gesture vocabulary are also shown in the video accompanying this paper.

A. Study 1: which gestures make the instructor more appealing?

Teachers who gesture can increase student engagement. However, 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 motion that confounds the studies [11], [13].

Developing the gesture stimuli for this study was an iterative process. First, real-world examples of relevant gestures were located in on-line video clips of charismatic speakers. These examples were then replicated for the instructor avatar by a digital artist. The avatar gestures were then refined several times with feedback from psychology researchers. Once the gestures were deemed acceptable, animation stimuli were created through 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 level of parallelism (single arm, two arms unsynchronized, two arms in parallel), direction of movement (vertical, inward, outward), and movement amplitude (full, partial). Once one script is created, the other scripts are created quickly through minor modifications of the first script.

The stimuli were then shown to undergraduate students ($n = 56$). Each student saw all 18 stimuli in 1 of 14 preselected counterbalanced orders, in which the amplitude alternated and no more than two levels of parallelism or movement appeared next to each other. After each stimulus, the participant made a number of ratings of their impressions of the avatar. Participants also answered questions about their own charisma, personality, mood, and demographics. A factor analysis of the result data isolated two dimensions of perception rating – charisma and attractiveness. Then analyses of variance showed that parallel outward focused gestures are particularly charismatic. This is the first time gestural charisma has been isolated in a true experimental research design.

Compared to previous studies using human instructor actors, the avatar system enables testing the effects of individual channels of communication while holding other channels constant in an efficient and convenient way, 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. Along this line of research, we are currently examining whether individual differences are associated with sensitivity to the avatars' gestures, and whether viewers spontaneously imitate the gestures that the avatar made when asked to explain the math problem.

B. Study 2: do instructor gestures promote learning?

It has been difficult to empirically probe the effect of gesture on learners because it is difficult for human actors to produce controlled yet naturalistic 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 also vary with gesture. Accordingly, researchers have moved towards using prerecorded video stimuli to afford experimental control. Our study used the gesturing avatar to address the question of whether or not adding gesture to instruction increases learning. Using the avatar, we were able to control for prosody, eye gaze speech rate and other nonverbal behaviors across our condition while ensuring that the gesture produced were naturalistic.

Third and fourth grade children ($n = 51$) participated in a tutorial on mathematical equivalence delivered by an avatar that either gestured or did not gesture. We included children who got only 0 or 1 problem (out of 6) correct on a pretest in our analyses ($n = 30$). 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 children who saw the gesturing avatar performed much better than those who saw the non-gesturing avatar (Posttest: 76% versus 61%, Transfer: 82% versus 63% and Conceptual: 94% versus 86%), which was reflected in a significant effect of condition ($\beta = -2.05$, $p = 0.024$) and no interaction with test.

These findings are expected given similar findings using live and prerecorded stimuli [9], [10], [6]. However, these findings go beyond prior work in that they provide more convincing evidence that the effect observed is due to hand gesture 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 quite 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.

VII. Conclusions and Future Work

We have described a system for creating animation stimuli for the study of gestures in education. Compared to the traditional method for creating stimuli by video-recording and instructor actor performing a script, the system allows creating stimuli faster, as it removes the problem of having to memorize long scripts, and with higher precision, as all parameters can be precisely controlled for each condition. The system has been used to create stimuli for two studies. The system is becoming a powerful tool for research on gesture in education. We are currently conducting three more studies with stimuli created with the system.

One direction of future work is to reduce the scripting learning curve. One possibility is to further simplify the scripting language to make it more user friendly. For example, we consider introducing time targets attached to spoken words in order to relieve the script author from trial-and-error fine tuning of the synchronization time offsets. Another possibility is to abandon creating the script in a text editor but rather to use a graphical editor that allows the script author to populate each field of each line of the script using drop-down menus, which will eliminate the possibility of syntax errors.

Another important direction of future work is to enhance the scripting language and the system to produce interactive stimuli that collect answers from students and alter the subsequent execution of the script accordingly. Whereas conditional and loop statements are well known in computer science, the challenge consists in having these formalisms adopted by education researchers.

Our system was developed with extensibility in mind. The system will become even more useful as it is extended to additional characters, environments, and gestures in research studies and in real-world educational contexts.

VIII. Acknowledgement

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." 1927.

[2] L. M. Flevares and M. Perry, "How many do you see? the use of nonspoken representations in first-grade mathematics lessons." *Journal of Educational Psychology*, vol. 93, no. 2, p. 330, 2001.

[3] S. Goldin-Meadow, S. Kim, and M. Singer, "What the teacher's hands tell the student's mind about math." *Journal of Educational Psychology*, vol. 91, no. 4, p. 720, 1999.

[4] R. Núñez, "Embodied cognition and the nature of mathematics: Language, gesture, and abstraction." Lawrence Erlbaum Associates, 2004, pp. 36 – 37.

[5] X.-l. Wang, R. Bernas, and P. Eberhard, "Engaging adhd students in tasks with hand gestures: a pedagogical possibility for teachers," *Educational studies*, vol. 30, no. 3, pp. 217–229, 2004.

[6] M. A. Singer and S. Goldin-Meadow, "Children learn when their teacher's gestures and speech differ," *Psychological Science*, vol. 16, no. 2, pp. 85–89, 2005.

[7] S. Cook, "Abstract thinking in space and time: Using gesture to learn math," *Cognitie, Creier, Comportament*, vol. 15, no. 4, 2011.

[8] S. Goldin-Meadow and M. Alibali, "Gesture's role in speaking, learning, and creating language," *Annual Review of Psychology*, vol. 64, pp. 257–283, 2013.

[9] R. B. Church, S. Ayman-Nolley, and S. Mahootian, "The role of gesture in bilingual education: Does gesture enhance learning?" *International Journal of Bilingual Education and Bilingualism*, vol. 7, no. 4, pp. 303–319, 2004.

[10] S. W. Cook, R. G. Duffy, and K. M. Fenn, "Consolidation and transfer of learning after observing hand gesture," *Child development*, 2013.

[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, pp. 203–211, 1988.

[12] P. C. Abrami, L. Leventhal, and R. P. Perry, "Educational seduction," *Review of educational research*, vol. 52, no. 3, pp. 446–464, 1982.

[13] 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." *Journal of Educational Psychology*, vol. 74, no. 1, p. 126, 1982.

[14] R. G. Williams and J. E. Ware, "An extended visit with dr. fox: Validity of student satisfaction with instruction ratings after repeated exposures to a lecturer," *American Educational Research Journal*, vol. 14, no. 4, pp. 449–457, 1977.

[15] T. Noma, L. Zhao, and N. I. Badler, "Design of a virtual human presenter," *Computer Graphics and Applications, IEEE*, vol. 20, no. 4, pp. 79–85, 2000.

[16] E. André, J. Müller, and T. Rist, "The ppp persona: a multipurpose animated presentation agent," in *Proceedings of the workshop on Advanced visual interfaces*. ACM, 1996, pp. 245–247.

[17] J. C. Lester, J. L. Voerman, S. G. Towns, and C. B. Callaway, "Cosmo: A life-like animated pedagogical agent with deictic believability," 1997.

[18] J. C. Lester, S. G. Towns, P. J. Fitzgerald *et al.*, "Achieving affective impact: Visual emotive communication in lifelike pedagogical agents," *International Journal of Artificial Intelligence in Education (IJAIED)*, vol. 10, pp. 278–291, 1998.

[19] 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, no. 1-2, pp. 1–44, 1999.

[20] W. L. Johnson and J. Rickel, "Steve: An animated pedagogical agent for procedural training in virtual environments," *ACM SIGART Bulletin*, vol. 8, no. 1-4, pp. 16–21, 1997.

[21] J. Rickel and W. L. Johnson, "Animated agents for procedural training in virtual reality: Perception, cognition, and motor control," *Applied artificial intelligence*, vol. 13, no. 4-5, pp. 343–382, 1999.

[22] N. Reithinger, P. Gebhard, M. Löckelt, A. Ndiaye, N. Pfleger, and M. Klesen, "Virtualhuman: dialogic and affective interaction with virtual characters," in *Proceedings of the 8th international conference on Multimodal interfaces*. ACM, 2006, pp. 51–58.

[23] J. C. Lester, S. A. Converse, S. E. Kahler, S. T. Barlow, B. A. Stone, and R. S. Bhogal, "The persona effect: affective impact of animated pedagogical agents," in *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. ACM, 1997, pp. 359–366.

[24] A. L. Baylor *et al.*, "The impact of three pedagogical agent roles," in *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*. ACM, 2003, pp. 928–929.

[25] J. Holmes, "Designing agents to support learning by explaining," *Computers & Education*, vol. 48, no. 4, pp. 523–547, 2007.

[26] R. Moreno and R. Mayer, "Interactive multimodal learning environments," *Educational Psychology Review*, vol. 19, no. 3, pp. 309–326, 2007.

[27] R. K. Atkinson, "Optimizing learning from examples using animated pedagogical agents." *Journal of Educational Psychology*, vol. 94, no. 2, p. 416, 2002.

[28] M. Lusk and R. Atkinson, "Varying a pedagogical agents degree of embodiment under two visual search conditions," *Applied Cognitive Psychology*, vol. 21, pp. 747–764, 2007.

[29] L. A. Annetta and S. Holmes, "Creating presence and community in a synchronous virtual learning environment using avatars," *International journal of instructional technology and distance learning*, vol. 3, pp. 27–43, 2006.

[30] S. Barlow, S. Converse, S. Kahler, J. Lester, and B. Stone, "Animated pedagogical agents and problem solving effectiveness: A large scale empirical evaluation," in *Artificial Intelligence in Education, 1997: Knowledge and Media in Learning Systems: Proceedings of AI-ED 97, World Conference on Artificial Intelligence in Education, Kobe, Japan*, vol. 39. IOS Press, 1997, p. 23.

[31] (2011) Character builder. [Online]. Available: http://www.mediasemantics.com/

[32] (2011) Noah animated avatar technology. [Online]. Available: http://noahx.com/index.asp

[33] (2010) Codebaby. [Online]. Available: http://www.codebaby.com/products/elearning-solutions

[34] (2010) Gesture builder. [Online]. Available: http://www.vcom3d.com/index.php?id=gesturebuilder

[35] S. Ellis and B. Rogoff, "The strategies and efficacy of child versus adult teachers," *Child development*, pp. 730–735, 1982.

[36] (2013) Unity 3d. [Online]. Available: http://unity3d.com/

[37] (2013) Maya. [Online]. Available: http://www.autodesk.com/products/autodesk-maya/overview

[38] (2013) Voice-o-matic. [Online]. Available: http://www.di-o-matic.com/products/plugins/VoiceOMatic/