Animation Killed the Video Star

Voicu Popescu
Purdue University
popescu@purdue.edu

Nicoletta Adamo-Villani
Purdue University
nadamovi@purdue.edu

Meng-Lin Wu
Purdue University
wu223@purdue.edu

Suren D. Rajasekaran
Purdue University
srajase@purdue.edu

Martha W. Alibali
University of Wisconsin
mwalibali@wisc.edu

Mitchell Nathan
University of Wisconsin
mnathan@wisc.edu

Susan Wagner Cook
University of Iowa
susan-cook@uiowa.edu

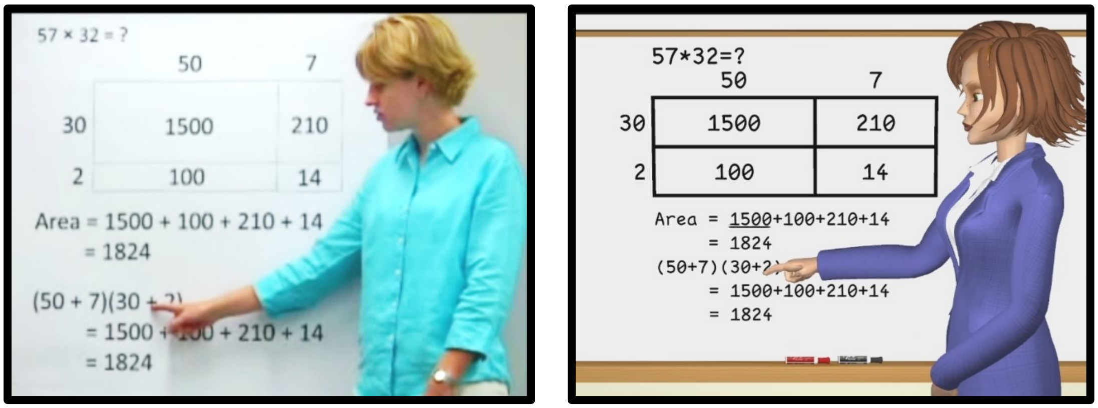

Figure 1: Visual stimulus used in instructor gesture research: video (left), animation (right).

ABSTRACT

In this position paper, we describe a novel approach for creating visual stimuli for research on gesture in instruction. The approach is based on a system of computer animation instructor avatars whose gesture is controlled with a script. Compared to video recording instructor actors, the approach has the advantage of *efficiency*—once the script is written, it is executed automatically by the avatar, without the delay of script memorization and of multiple takes, and the advantage of *precision*—gesture is controlled with high fidelity as required for each of many conditions, while all other experiment parameters (e.g. voice tone, secondary motion) are kept constant over all conditions, avoiding confounds. We have begun implementing the approach, and we will test it in the context of connecting mathematical ideas in introductory algebra.

Author Keywords

Instructor gesture, instructor avatar, computer animation, connecting mathematical ideas, introductory algebra.

ACM Classification Keywords

H.5.1 Multimedia Information Systems—Animations
H.5.m. Information interfaces and presentation (e.g., HCI)—Miscellaneous.

INTRODUCTION

Computer animation has the potential to become a powerful platform for research on gesture in instruction, overcoming many of the challenges associated with the traditional approach of creating lesson stimuli by video recording instructors. We have begun implementing a system of instructor avatars that we plan to test in introductory algebra (Figure 1).

Importance of gesture in education

Communication is an integral element of nearly all forms of instruction, including tutoring, peer collaboration, and classroom instruction. Instructors use a range of modalities to communicate, including language, drawing, writing, and non-verbal behaviors such as gestures.

Many recent, naturalistic studies have highlighted the importance of teachers' gestures, particularly in the domains of mathematics and science. For example, researchers have argued that teachers use gestures as a "semiotic resource" for developing and refining ideas [e.g., 2, 7], as a means to link ideas and representations [1], and as a means of fostering joint attention or shared understanding [8], particularly in cases where students are having difficulty with the material.

However, empirical data that address whether teachers' gestures are actually beneficial for students' learning are relatively scarce. A handful of studies have compared lessons with gestures to lesson without gestures, and shown that lessons with gestures lead to greater student learning [e.g., 3, 4, 9] Although these studies represent a valuable first step, they do not address the range of variation in teachers' gestures, or whether some types of gestures are more effective than others. What is needed are empirically validated recommendations about what type of gestures are most effective. However, one challenge in research on gesture in instruction is creating appropriate lesson stimuli.

Challenges of video

A common approach is to create video recordings of lessons that vary in whether and how the teacher uses gestures; lessons are scripted so that, ideally, only gesture varies between lessons. This allows for strong inferences to be drawn about whether the gestures contribute to students' learning. In some cases, the same audio track is used across conditions in a study, and the teacher-actor "lip-syncs" the speech while producing the scripted gestures. This approach has the advantage of realism—video depicts the instructors exactly as students see them in classrooms.

However, this methodological approach also has many challenges. It is often difficult for teacher-actors to memorize all the needed scripts for each condition in a study. Even more problematic, teacher-actors have to keep all aspects of behavior other than gesture the same between conditions, including eye gaze, posture, and facial expression. Because it is effortful to simultaneously follow the script and manage other aspects of behavior, the resulting videos often look unnatural, and they often require many "takes".

The challenges do not end there. It is also often challenging for teacher-actors to manage auxiliary visuals, such as writing on the board, while also producing scripted gestures. For this reason, many experiments give up on developing visuals incrementally, in real time, therefore losing the benefits of focusing student attention and of controlling complexity by conveying the solution progressively.

ANIMATION: POTENTIAL, CHALLENGES, SOLUTIONS

Potential of animation

Advances in technology have enabled developing and deploying computer animation applications on consumer level computing technology such as personal computers, laptops, tablets, and smart phones. A computer animation character is well suited to serve as an instructor proxy in studies of gesture: a humanoid character can make any gesture a human instructor can make, and an animation character has perfect memory and infinite energy.

Once the script is complete, the instructor avatar can render the stimulus in one "take", with no errors. Slight modifications of the script produce the stimuli for any number of additional conditions, with perfect control over secondary parameters. The avatar does not have to worry about remembering what to say and do, and when, and the resulting lessons can be, paradoxically, less contrived and more natural than in the case of human teacher-actors.

Instructor avatars are malleable, so they simplify generating stimuli for studies that take into account instructor characteristics such as gender, personality, and age. Finally, if desired, animation can be bigger than life. The whiteboard can magically animate concepts, the avatar can have perfect drawing skills, and the avatar can have cartoon-character-like qualities, such as a stylized appearance and an extroverted personality reflected through speech and action (e.g., backflips to celebrate correct answer).

Challenges of animation

The first question is whether what is learned about gesture in the context of instructor avatars does transfer to human instructors. Preliminary analyses of data collected in our first studies indicate that, yes, benefits of gesture previously measured with human instructors are replicated when instructor avatars are used. Moreover, effective instructor avatars are valuable in and of themselves, enabling the creation of effective digital learning materials.

However, before instructor avatars can become an effective platform for gesture research, several challenges have to be addressed. First, the animation needs to be of sufficient quality. We define animation quality in this context both at a low level, which includes rendering nuanced gestures with precision, clarity, and life-like quality, and at a high level, which includes conforming to avatar behavior and speech rules for effective social and teaching interactions [12, 13]. Second, for the approach to scale, one has to be able to create stimuli without the prerequisites of artistic and programming talent. Gesture researchers need to be able to create their own experiment stimuli without assistance from digital artists and programmers.

Solutions to animation challenges

A solution to these challenges is a system of computer animation instructor avatars that are controllable through a script. The script is a text file that contains directions for the avatar, much the same way scripts are used by researchers to direct human teacher-actors when generating video stimuli. No programming expertise is needed; instead, the approach relies on scripting, an interface already familiar to researchers. Compared to the natural language and loose rules used when scripting a human actor, the scripting of the avatar has to be done in a language that can be interpreted automatically, i.e., an English-like language that is complete, concise, and unambiguous.

The need for artistic talent is removed by using a database that already contains all the digital art assets needed for generating the stimuli. These assets include the computer animation characters for the avatars, pre-recorded joint

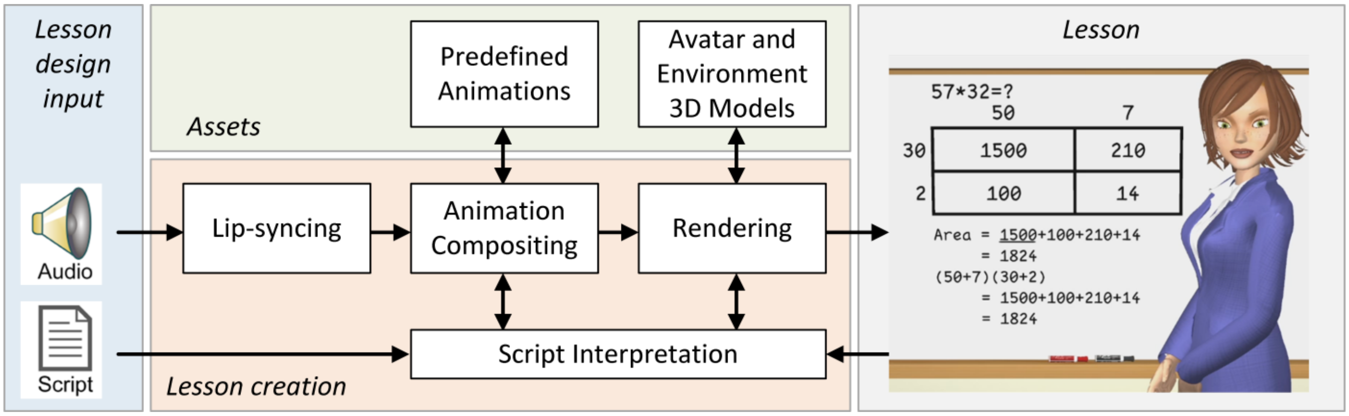

Figure 2: Overview of system of computer animation instructor avatars for generating lesson stimuli for gesture research.

angle values for complex animations, and auxiliary visuals (e.g., a whiteboard capable of displaying any polynomial multiplication). Simple animations, such as pointing at a specific location on the board, are computed automatically on-the-fly, as needed, using animation algorithms.

For animated avatars to be effective and believable instructors, their animation should be of life-like quality. To achieve this, the animation must adhere to fundamental principles. The 12 principles of animation [5] are a set of procedures taught as fundamental rules of the "language of movement" at the Walt Disney Studio in the late 1930s. Five of these principles are essential in our context.

Anticipation. Anticipation is preparation for action. Actions are preceded by smaller actions in the opposite direction (e.g., to jump with joy, an avatar has to bend its knees first). Without anticipation, gestures appear abrupt, stiff, and unnatural.

Follow-through and overlapping action. Follow-through is the termination of an action. "Actions very rarely come to a sudden and abrupt stop, they are generally carried past their termination point" [6]. For example, when the avatar lands after jumping, it has to bend its knees again before assuming the standing pose. Moreover, movements of different parts of a living creature do not occur at the same time; motions start, move, and stop at different points and at different rates. Lack of adherence to this principle yields mechanical motion with an undesirable "stop-start" quality.

Secondary Action. Any movement of a living creature is composed of a primary action and multiple secondary actions. For example, when the avatar makes a beat gesture with its right hand, the avatar might sketch a similar gesture of reduced amplitude with its left hand. Representing different types of interconnected motions is fundamental to the creation of "a believable whole within any movement" [6]. Failure to adhere to this principle results in disjointed, puppet-on-a-string-like characters; the motion does not appear to come from within but rather to be applied from an external source.

Slow-out and slow-in. Body parts do not usually move at constant speed; they show a certain degree of acceleration and deceleration. Actions should progressively accelerate out of a key pose and decelerate into a resting pose. Failure to adhere to this principle confuses the viewer with sheer defiance of the laws of physics.

Arcs. Living creatures cannot perform linear movements. The same way the laws of physics do not tolerate abrupt changes in velocity magnitude, they do not tolerate abrupt changes of velocity direction. Failure to adhere to this principle results in mechanical, robotic motion.

A SYSTEM OF INSTRUCTOR AVATARS

We are developing a system of animation avatars with the overall architecture given in Figure 2.

The lesson designer provides two types of input: instructor audio and lesson script. The designer records the instructor audio using voice talent that matches the instructor avatar. Synthesizing the audio directly from text would result in a robotic voice. The audio file is used by a lip-syncing module to derive the facial animation of the instructor avatar needed to utter the recorded words.

The lesson script is a text file that controls the avatar and the auxiliary visuals. The scripting language supports a small number of commands, each with subcommands and arguments. For example the command SAY *audioA* 1.2 10.5 makes the avatar say the instructor audio recording stored in file *audioA*, from seconds 1.2 to 10.5. The command GESTURE DB *idkGesture* makes the avatar gesture a pre-defined "I don't know gesture" by raising its shoulders, bending its arms at 90º, and turning its hands such that the palms point upwards. The command GESTURE POINT LEFT INDEX *targetA* makes the avatar point with its left index to *targetA,* which was previously defined to correspond, for example, to a point on the graph. The scripting language also allows precisely synchronizing gesture and speech by defining time points relative to the beginning of an audio file, which can then be used as starting times for gestures.

The script is interpreted automatically. The necessary animations are either retrieved from the assets database (for GESTURE DB commands) or from the lip-syncing module, or they are computed on the fly through inverse kinematics animation algorithms (e.g., for the GESTURE POINT command). The various animation elements are composited and applied to the avatar computer animation character.

We have developed an avatar named Julie (Figures 1 and 2). Julie is a partially segmented 3D animation character comprised of eight polygonal meshes with a total polygon count of 171,907. Julie's body is rigged with a skeletal deformation system that includes 69 joints. Her face is rigged with a combination of joint deformers and blendshape deformers. 29 joints are used for opening and closing the mouth and for eye/eyebrow deformations, 20 blendshape targets are used for facial deformations, and eight targets provide the visemes required for animation of dialogue (four for the consonant sounds and four for the vowel sounds). The polygon meshes are skinned to the skeleton with a dual quaternion smooth bind with a maximum number of influences of five.

The avatar is rendered in the environment to obtain the lesson stimuli. Initially, we will focus on non-interactive lessons. Interactive lessons will be supported by collecting input from learners to which the avatar will react, according to the script. Once a script is written for a first condition, the scripts for the other conditions are generated by modifying the first script. This procedure is efficient and keeps all parameters constant except for the gesture conditions that are the target of study.

Whenever possible, a gesture is implemented algorithmically through inverse kinematics. This includes all deictic gestures which abound in instructors' non-verbal communication repertoires. The advantage of algorithmic animation is efficiency—one inverse kinematic algorithm animating the arm allows pointing anywhere on the whiteboard by simply changing the target parameter. The challenges of algorithmic animation are lower quality, as rigorous conformance to animation principles cannot be achieved in target independent fashion, and lack of support for complex gestures, which have to be animated manually by digital artists.

The reliance on the database of predefined animations brings scripting language simplicity. The language does not have to support the definition of new avatar poses and the linking of poses to form new gestures. The tradeoff is that users cannot define new gestures, which can be a potential bottleneck preventing scalability to other concepts, subject matters, and student age groups. Whenever a new gesture is needed, that gesture has to be defined by a digital artist using an animation software system (e.g. Maya) and added to the database.

One option is to extend the scripting language to support a precise and inherently complex description of the behavior of the avatar, as was done for example in the SmartBody system which relies on SAIBA framework's Behavior Markup Language standard [14].

However, we believe that the important advantages of scripting language simplicity and similarity to the scripts already used by gesture researchers when creating video stimuli do not have to be sacrificed. We anticipate that the database of gestures shared online will quickly grow to cover all gestures needed. For example we have already defined a complex vocabulary of charisma gestures covering 18 parameter combinations: inward, vertical, or outward; one hand, two hands unsynchronized, or two hands parallel; small or large amplitude. Moreover, gestures can be retargeted to any animation character with the same skeletal structure.

As prior research indicates, the precise timing between gesture and utterances is essential [10, 11]. The scripting language does allow defining time steps with fine granularity (i.e., milliseconds). However, finding the precise time step presently requires a trial and error approach. Proceeding with a binary search requires ten tries to find a time stamp with millisecond accuracy within a one second time interval. We will investigate direct synchronization using the textual representation of speech. We will leverage the availability of the text, which simplifies the audio to text mapping.

INSTRUCTOR GESTURE RESEARCH

We will use experimental lessons created with the instructor avatars to address a several key questions about how teachers' gestures contribute to student learning. Our initial focus will be on instruction in algebra, which typically involves instructors drawing links across multiple related representations. For example, in a lesson about polynomial multiplication, an instructor might seek to connect a procedure for multiplying polynomials that is expressed in symbolic form with a procedure that is depicted using an area model (see Figure 1). We will eventually expand our focus to other types of mathematical and statistical content (e.g., geometry, confidence intervals). Where possible, we will compare our findings to those obtained in research with human teachers.

As a starting point, we will investigate whether students learn more about links between ideas when teachers gesture to corresponding ideas sequentially or when they do so simultaneously (i.e., pointing to two corresponding ideas with the two hands). We will also investigate whether students learn more about links between ideas when teachers use gesture to make element-by-element mappings or more global mappings. We will also ask whether deictic or representational gestures are more communicatively effective. The avatar system will allow us to design focused tests of these and many other research questions.

LONG TERM GOALS

Our long term goal is to provide a powerful system for creating stimuli for research on gesture and beyond. Although video *did* kill the radio star (as claimed by the Buggles in their hit song in 1979), we do not foresee that animation will ever completely replace video recordings of human instructors. Video stimuli are by definition photorealistic, and the talent of a gifted instructor will continue to elude lossless translation into scripted animation for the foreseeable future.

The instructor avatar approach opens the door for research on learning personalization, where the avatar offers the fine-grained instructor control needed to adapt to a variety of learner characteristics. Our initial focus is on introductory algebra with middle and high school learners, but the approach can be extended to other gestures, topics, disciplines, and learner age groups. The system of instructor avatars can also become a powerful platform for creating interactive digital learning activities, and for creating materials for pre- and in-service teacher professional development. Beyond standard education, the system of instructor avatars can also support multimodal instruction for special education. Finally, our system can be repurposed for developing non-verbal communication skills for other forms of public speaking, beyond instruction.

ACKNOWLEDGMENTS

We thank Jian Cui, Howard Friedman, and Katherine Duggan who have helped develop the initial version of the computer animation instructor avatar system on which the present system is based. We are particularly grateful to Jian Cui for his help as we extended the initial system. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A130016, and by the National Science Foundation, through Grant 1217215. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences, of the U.S. Department of Education, or of the National Science Foundation.

REFERENCES

1. Alibali, M. W., Nathan, M. J., Wolfgram, M. S., Church, R. B., Johnson, C. V., Jacobs, S. A., & Knuth, E. J. (2014). How teachers link ideas in mathematics instruction using speech and gesture: A corpus analysis. *Cognition & Instruction, 32*(1), 65-100.

2. Arzarello, F., Paola, D., Robutti, O., & Sabena, C. (2009). Gestures as semiotic resources in the mathematics classroom. *Educational Studies in Math., 70*(2), 97-109.

3. Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning? *International J. of Bilingual Education & Bilingualism*, 7, 303–319.

4. Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Dev., 84*(6), 1863-1871.

5. Johnston, O., & Thomas, F. The illusion of life: Disney animation. *Disney Editions;* Rev Sub edition, Oct 5, 1995.

6. Lasseter, J. Principles of traditional animation applied to 3D computer animation. *Computer Graphics* (1985), 21(4): 35-44.

7. Rasmussen, C., Stephan, M., & Allen, K. (2004). Classroom mathematical practices and gesturing. *J. of Mathematical Behavior, 23*, 301-324.

8. Reynolds, F. J., & Reeve, R. A. (2001). Gesture in collaborative mathematics problem solving. *J. of Mathematical Behavior, 20*(4), 447-460.

9. Valenzeno, L., Alibali, M. W., & Klatzky, R. L. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemp. Educ. Psych., 28*, 187–204.

10. Striegnitz, K., Tepper, P., Lovett, A. & Cassell, J. (2008). Knowledge representation for generating locating gestures in route directions. In K.R. Coventry, T. Tenbrink & J. Bateman (Eds.), *Spatial Language and Dialogue (Explorations in Language and Space)*. Oxford: Oxford University Press.

11. Nass, C., Isbister, K. & Lee, E.-J. (2000). Truth is beauty: Researching embodied conversational agents. In *Embodied conversational agents* (pp.,374-402). Cambridge, MA: MIT Press.

12. Finkelstein, S., Ogan, A., Walker, E., Muller, R., & Cassell, J. (2012). Rudeness and rapport: Insults and learning gains in peer tutoring. In *Proceedings of Intelligent Tutoring Systems*,.Berlin: Springer..

13. Ogan, A., Finkelstein, S., Mayfield, E., Matsuda, N., & Cassell, J. (2012). Oh dear Stacy! Social interaction, elaboration, and learning with teachable agents. In *Proceedings of CHI 2012.* Austin, TX.

14. Thiebaux, M., et al., (2008). SmartBody: behavior realization for embodied conversational agents. In *Proceedings of 7th international conference on autonomous agents and multiagent systems* (pp. 151-158). Estoril, Portugal.