This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2019.2922134, IEEE Transactions on Learning Technologies

Automatic Deictic Gestures for Animated Pedagogical Agents

Sri Rama Kartheek Kappagantula, Nicoletta Adamo-Villani, Meng-Lin Wu, Voicu Popescu

***Abstract*—We present a system that automatically generates deictic gestures for Animated Pedagogical Agents (APAs). The system takes audio and text as input, which define what the APA has to say, and generates animated gestures based on a set of rules. The automatically generated gestures point to the exact locations of elements on a whiteboard nearby the APA, which are calculated by searching for keywords mentioned in the speech. We conducted a study with 100 subjects, in which we compared lecture videos containing gestures automatically-scripted by the system to videos of the same lecture containing manually-scripted gestures. The study results show that the manually-scripted and automatically-scripted lectures had comparable number of gestures, and that the gestures were timed equally well.**

***Index Terms*—Animation, Intelligent agents, Automatic Programming, Computer-assisted instruction, Educational simulations, Instructor interfaces, Learning management systems, Speech analysis, Gesture.**

I. Introduction

ANIMATED Pedagogical Agents (APAs) are on-screen instructor avatars embedded into e-learning environments to facilitate instruction [1], [2]. Compared to an e-learning activity that relies only on text and images, an activity presented by an APA can lead to more learning, especially for younger learners [3], for learners not yet proficient in English [4], and for learners with special needs [5]. Like a real-life instructor, an APA can capture, guide, and maintain student attention. Compared to a video-based e-learning activity, an activity presented by an APA promises two important advantages. First, the APA's appearance and teaching style can be customized to match learner's affinities, which can advance research in learning personalization. Second, an APA-delivered activity can be interactive, leveraging research in intelligent tutoring.

Furthermore, APA-delivered learning activities can have a dramatically lower production cost than video-based activities. Consider the case of a learning activity on mathematics. Once an initial activity is generated with mathematical examples, creating additional activities involving different mathematical examples should be substantially easier. Recent work has demonstrated such e-learning content creation scalability with the problem instance [6].

Many aspects of e-learning content creation scalability remain open problems. First, content creation should scale in terms of number of lectures generated for a problem instance. Even creating the first instance of an APA delivered e-learning activity poses substantial challenges. The content creator, who is typically an educator with little programming expertise, first needs to create a lesson script, and then needs to animate the APA according to the script. This is typically done by annotating the script with animation instructions, using an animation scripting language, which can be difficult to do for someone without programming background. Even with the needed animation scripting language proficiency, achieving a good timing of the APA animation with the APA speech remains a time consuming task. Second, content creation should scale across problem types i.e. from mathematics equivalence to linear equations, to polynomial multiplication, and so on, and even across domains, i.e. from math to physics, to engineering, and beyond.

Researchers point out that one factor that will drive future demand for pedagogical agents is the need for scalable online learning that engages and retains students [7]. Scalable e-content delivered by multimodal, engaging APA will help motivate and retain students in online courses. In this paper we present a step towards achieving e-learning content creation scalability by automating the generation of APA deictic gestures. Given as input the script of a lesson to be delivered by an APA standing in front of a whiteboard, and a human voice recording of what the APA needs to say, our system automatically augments the lesson script with animation instructions such that the APA delivers the lesson by pointing to the elements displayed on the whiteboard, at the appropriate time, as they are mentioned in speech. The generality of our solution hinges on the fact that, just like in conventional classroom instruction, an APA in front of a whiteboard can teach many topics in math and beyond. Furthermore, research in instructor gesture has shown that a large percentage of instructor gestures are deictic gestures [8], and that instructor deictic gestures are essential for learning [9]. Hence, supporting this type of gesture goes a long way towards preserving instructor gesture benefits as e-learning content creation is scaled up.

Our system identifies the keywords in the lesson script and checks if the corresponding targets to the keywords exist on the whiteboard. The 3D location of the target is computed from the known mechanism for displaying elements on the whiteboard, and from the known location of the whiteboard. The time at which the gesture should be made is derived from a mapping of text-to-audio of the APA speech. The actual gesture animation is computed using inverse kinematics (IK) animation algorithms, which control the APA movements in

S. Kappagantula and N. Adamo-Villani are with the Department of Computer Graphics Technology, Purdue University, West Lafayette, IN, 47907, USA.
E-mail: kartheek.kappagantula@gmail.com, nadamovi@purdue.edu
M. Wu and V. Popescu are with the Department of Computer Science, Purdue University, West Lafayette, IN, 47907, USA.
E-mail: {wu223, popescu}@purdue.edu

1939-1382 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2019.2922134, IEEE Transactions on Learning Technologies

front of the whiteboard. The algorithms ensure that the APA is within pointing reach of the target, the APA leans to reach the target without occluding the content on the board, the APA rotates the body towards the board to enable pointing, and the APA looks at the target while pointing to it.

We demonstrate our system's ability to automatically generate APA deictic gestures in the context of linear equations. In Figure 3, the top row shows the frames from an e-learning activity where the APA animation was generated automatically by our system. The bottom row shows frames from the same e-learning activity generated by manually scripting the APA animation. It can be seen from the figure that the deictic gestures of the APA are almost indistinguishable across the two rows. In other words, when both the manually-scripted and the automatically-scripted animation ask the APA to point to a certain location on the board, the results are very similar. However, it is not our goal to replicate precisely the manually-scripted animation, nor is such 100% replication possible. Figure 4 shows cases when the APA in manually-scripted animation makes a deictic gesture and the APA in automatically-scripted animation does not, and vice versa.

This paper does not discuss whether automatically-scripted lectures have a direct impact on student learning, or how realistic the APA animation looks when compared to human motion. These topics are out of the scope of the work presented in this paper. Rather, the goal of our work is to generate automatically-scripted animation that augments lesson delivery at a quality similar to manually-scripted animation, while enjoying the advantage of a near zero production cost.

We have conducted a user study with 100 participants and a small expert evaluation with 4 Psychology researchers. In both studies, subjects were asked to compare the timing of gestures and number of gestures of automatically-scripted animation to manually-scripted animation. The results of both the user study and expert evaluation show (1) that the automatically-scripted animation and manually-scripted animation are equivalent in terms of number of gestures and timing of gestures, (2) that subjects' major field of study and years of experience had no significant effect on their evaluations and (3) that more than two-thirds of the subjects did not have a preference between the two types of animation, when asked about including them in online lectures. We also refer the reader to the video accompanying our paper[1].

The paper is organized as follows. In section II, we present a review of prior work relevant to automation of APA gestures. In section III, we present an overview of our system. In section IV, we describe the user study we conducted and discuss the results, and in section V we report the expert evaluation. Limitations of our system and potential future work are discussed in section VI.

II. Relevant Literature

A. Animated Pedagogical Agents

Animated Pedagogical Agents (APAs) are animated characters embedded within a computer based learning environment to facilitate student learning. Early examples of APAs are Cosmo [10], Herman [11], STEVE [12], PETA [13] and the "Thinking Head" [14]. In addition to APAs, animated signing agents have also been used to teach mathematics and science to young deaf children using sign language, e.g. Mathsigner and SMILE [5].

Many studies confirm the positive learning effects of systems using these agents [15], [16], [17], [18]. One of the first researchers who studied the use of animated agents in learning and communication, developed the Embodied Conversational Agent, an interactive virtual agent that can speak and exhibit nonverbal behaviors [19], [20]. It was argued in these studies that well-designed embodied pedagogical agents could enrich one's learning experience and foster motivation. Studies also suggest that APAs could be employed in e-learning environments to enhance users' attitude towards online courses [21].

Over the years, APA studies have become more focused, and researchers have begun to examine which specific APA characteristics promote learning, in which contexts, and for what types of learners. A few studies suggest that APAs have a positive impact on students with low prior knowledge of the subject, and have no impact and sometimes, a negative impact on students with high prior knowledge [22], [23]. Agents interacting using multiple modalities appear to lead to greater learning than agents that interact only in a single channel [24], [25].

A few studies have investigated the effect of different APAs' features on student's learning, engagement, and perception of self-efficacy. Some researchers examined whether the degree of embodiment of an APA had an effect on students learning of science concepts [26]. Findings showed that students learned better from a fully embodied human-voiced agent that exhibited human-like behaviors than from an agent who did not communicate using these human-like actions. In addition, students reported stronger social reactions to the fully-embodied agent. A recent study revealed that the visual style of an animated signing agent had an effect on student engagement. The stylized agent was perceived more engaging and "fun" than the realistic one, but the degree of stylization did not affect the students' ability to recognize and learn American Sign Language signs.

Some researchers explored whether the instructional role of the agent had an effect on students' learning and motivation [27]. Findings showed that the motivational agents (Motivator and Mentor) led to increased learner self-efficacy. However, the affective support was not sufficient for learning. The agents with expertise (Expert and Mentor) increased learning outcomes and were perceived as more effective. A study revealed that female students preferred as a learning companion an agent that developed social relationship during the learning activities rather than an agent that was strictly task-oriented [28]. Additional empirical studies showed that peer-like agents helped enhance positive affect and motivation for females who learned STEM topics [18], [29]. An experiment showed that teachable agents in educational games could help achieve deeper levels of mathematics learning for elementary and middle school children [30]. Other studies suggest that agent's features such as voice and appearance [31], [32], visual pres-

[1]https://youtu.be/XGiyaNt9c1Q

1939-1382 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2019.2922134, IEEE Transactions on Learning Technologies

ence [33], non-verbal communication [34] and communication style [35] could impact learning and motivation. Researchers also studied the extent to which agent persona affects students learning using path analysis. Findings showed that perceptions measured by the Agent Persona Instrument (API) had no significant effect on learning [36]. However, agent persona did affect perceived information usefulness [36].

A few researchers have investigated whether APAs are more effective for certain learner populations as compared to others. A study revealed that middle grade females and ethnic minorities improved their self-efficacy in learning algebraic concepts after working with the APA, and improved learning significantly compared to white males [29]. High school students preferred to work with an agent with the same ethnicity more than with a different one [37], [38], [39]. College students of color felt more comfortable interacting with a similar agent than with a dissimilar one [37].

A 2013 meta-analytic review of 43 papers showed that APAs enhance learning in comparison with learning environments that do not feature agents [1]. A more recent meta-analysis of 20 experiments revealed that gesturing pedagogical agents in multimedia environments had a small-to-medium impact on near transfer of knowledge and retention of learning [40]. A 2015 review that examined findings from studies on the efficacy of affective APAs in computer-based learning environments shows that the use of affect in APAs had a significant and moderate impact on students' motivation, knowledge retention and knowledge transfer [41].

Although there appears to be a growing consensus on the positive effects of APAs on learning outcomes, a few studies have failed to find significant improvements with using APAs in learning environments. In a study conducted on education and psychology college students, researchers concluded that the inclusion of APA had no effect on motivation or learning [31].Results of an experiment conducted with 5th and 6th graders showed no significant differences in retention and transfer test scores when gesturing APAs and non-gesturing APAs were compared [42].

In summary, research has shown that APAs can be effective in promoting learning, but not equally for all learner populations, learning subjects, and contexts. Despite growing evidence in support of the positive value of pedagogical agents, many questions still remain unanswered and additional studies need to be conducted. Easy-to-use APA systems, like the one presented in this paper are necessary in order to conduct future studies that can advance our understanding of the effects of APA on learning.

B. APA gestures

A beneficial effect of gesture on learning has been demonstrated in multiple domains, including mathematics, science, and foreign language education [3], [43]. For example, research on mathematical learning suggests that certain gestures such as point, sweep and balance promote conceptual understanding [44]. Gestures also make the accompanying speech more memorable, supporting learning of new content. In a wide variety of studies, speech accompanied by gesture has been shown to be more likely to be subsequently remembered than speech that is not accompanied by gesture [45], [46].

A substantial body of research has investigated the effects of APA gestures on learning, motivation, and social perception. A 2017 study investigated the extent to which APA gestures affect elementary school students learning of foreign language grammar and agent social acceptance [42]. The experiment tested three agent conditions, full gesture condition (deictic, iconic, metaphoric, and beat gestures), deictic gesture condition, and no gesture condition. Results showed that there were no significant differences in learning across the three conditions. However the full gesture agent was perceived more engaging and human-like. A further analysis of findings revealed that social acceptance features were also negative predictors of learning outcomes.

Some researchers argue that the gestures of an APA provide deictic believability [47] to the agent, making it more lifelike. A few studies suggest that gestures, in addition with facial expressions and body movements define the persona of an agent [48]. A recent experiment [40] concluded that gestures influence the agent persona with a small-to-medium effect size. The enhancement that gestures bring to the agent persona validate the signaling effect, which suggests that APAs can support student learning by signaling to the instructional material in a learning environment. In an experiment [49] conducted with 159 middle school students, it was found that students who learned with the help of an animated arrow and students who learned with the help of deictic movements of an APA outperformed students who learned without any visual attention-guiding method (control group), thus validating the signaling effect. In the experiment, the APA student group produced a significant difference than the control group but the animated arrow student group didnt, suggesting that the persona of the APA might have contributed to increased motivation in students. Gestures of an APA are also a key part of the embodiment principle [50], which states that people learn more deeply when an APA exhibits human-like characteristics such as facial expressions, eye gaze and gestures. Thus, it is important to understand how gestures aid the learning process by conducting more studies. APAs have been used to this effect in some studies to study the effects of gestures in mathematics education [51], [52], [3].

Research on generating gestures automatically requires producing two sets of lectures as stimuli, one set having gestures and one set not having gestures [53], [54], [55], [56]. Using humans as teachers in live experiment settings [53], [54] or using video recordings of humans as lectures [55], [56] can add confounding variables. When a researcher wants gestures to be the only difference between the two sets of lectures, it is difficult to maintain several other parameters like tone, gaze, facial expressions and stride length constant. Thus, it becomes difficult to attribute the results of a study to gestures alone. APAs have been used to overcome this challenge as all the features of the APA can be controlled using a computer. A recent study [3] used APAs as instructors and concluded that when all the other channels except gestures were maintained constant, children who observed the gesturing APA learned more. Our system can be used in such research studies.

1939-1382 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2019.2922134, IEEE Transactions on Learning Technologies

While some studies on APAs have resulted in new models to conduct APA research [57] and guidelines to design APAs [27], not many studies focused on developing actual systems to make the process of conducting APA research easier across multiple disciplines. In this paper we present one such system that facilitates researchers with no animation or programming background to conduct APA studies easily.

C. Existing APA systems

APA systems usually include a scripting language [51], [58]. Users write scripts in a given language in order to animate the agents. Although controlling a virtual character through scripts is now possible, users from non-programming backgrounds might find it difficult to learn the scripting language. Systems like the one presented in this paper can alleviate the problem by automatically generating the scripts that define the APA's behavior.

A majority of prior APA systems used talking heads rather than fully embodied agents [27], [59], [60], [61], [62], [29]. A few systems provided full-bodied human-like APAs that used gaze and gestures to direct user's attention [63], [64], [65], [26]. However, the agents were confined to a specific area on the screen and, other than minor position shifts, could not move across the learning environment. Our system allows for creating human-like animated agents that not only communicate with speech, gaze and gestures but also move in the virtual environment in order to point at different objects.

A few early systems created APAs that could navigate through the virtual environment and generate deictic gestures [10], [12], [66]. However, these systems had several limitations: (1) the APAs could only point to a limited set of domain related objects, (2) some of them required the user to enter goals and steps for each task, (3) some systems used simplified, non human-like characters, and (4) some did not provide authoring capabilities to take inputs and generate appropriate animations for APAs. In contrast, our system uses a full-bodied human-like character placed in a classroom environment, and does not require the user to write any execution plans. The user however, needs to provide content on the white board as input (details will be discussed in the next section).

D. Prior research on automatic generation of animation in virtual agents

Some prior research on generating agent animations from speech has focused on lip sync, facial expressions and head and eye movements [67], [68], [69], [70], [71]. Different approaches have been used to generate whole body animations from speech. One approach uses variations in speech prosody, e.g. variations in volume and pitch to generate the agents head and body motions. A few systems based on this approach were able to generate full body gestures in concurrence with speech [72]. One system synthesized body language in real time by doing a prosodic analysis of speech and used motion capture data for training purposes [73]. Our system avoids the overhead for motion capture data by using predefined animations and inverse kinematic algorithms. The main problem with prosody-based approaches is that they do not capture semantics and therefore the generated gestures do not augment the meaning being conveyed by the speech with information not present in the spoken message [74].

Another approach generates gestures (locations and type) based on the text of the speech [75]. One of the first systems that used this approach was the BEAT toolkit [76], which generated nonverbal behavior and synthesized speech, taking text as input. One other system synthesized deictic gestures (describing directions) from speech [77]. These systems produced animations for cartoon characters and simple 3D models that had only face and arms. In our system, we generate whole body animations from speech for a full-bodied human-like APA. One advantage of the text-based approaches is that the generated gestures augment the semantics of the utterances. One drawback is that some amount of manual work is usually required to create the rules.

Other work uses statistical methods to predict the gestures that the character will preform. One recent approach [78] uses a deep neural network for 3D gesture motion generation and a Bi-Directional Recurrent Neural Networks (Bi-Directional RNN) with Bi-Directional Long Short-Term Memory (Bi-Directional LSTM) for natural language processing. A study with thirty participants showed that the predicted gestures generated by the system were ranked higher than the original gestures for naturalness, but lower for time and semantic consistency.

Other automatic gesture generation methods use a combination of prosody and text to produce the agent nonverbal behavior [79], [80]. One recently proposed approach generates the agents metaphoric gestures from spoken text using Image Schemas [81], [82]. The system extracts Image Schemas from spoken texts, aligns them with text using a set of rules based on lexical and prosodic models, translates the image schemas into gesture invariants, combines the invariants into full gestures, and then syncs the gestures with speech. Another fairly recent system [83] uses deep learning in order to map gestures from text and prosody.

In summary, existing systems use different methods for generating full-bodied human-like agents that perform some types of gestures. In some systems the agents have the ability to navigate through the virtual environment, some systems generate deictic gestures, and in some systems the agents animations are generated directly from speech. An APA in a classroom setting needs to have all of the above features and prior systems that combined all these features and used them in APA studies are scarce. In this paper, we present one such system that not only has all the above features but also produces accurate deictic gestures that point to the exact locations of objects in the environment. Furthermore, our system allows users without animation expertise to control the deictic gestures of the APA by defining a set of keywords that specify the possible targets of the deictic gestures, set that can be modified to cover additional learning topics and domains.

III. System Design

The flowchart in Figure 1 gives an overview of the animation system. The *input module* provides the text and

1939-1382 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2019.2922134, IEEE Transactions on Learning Technologies

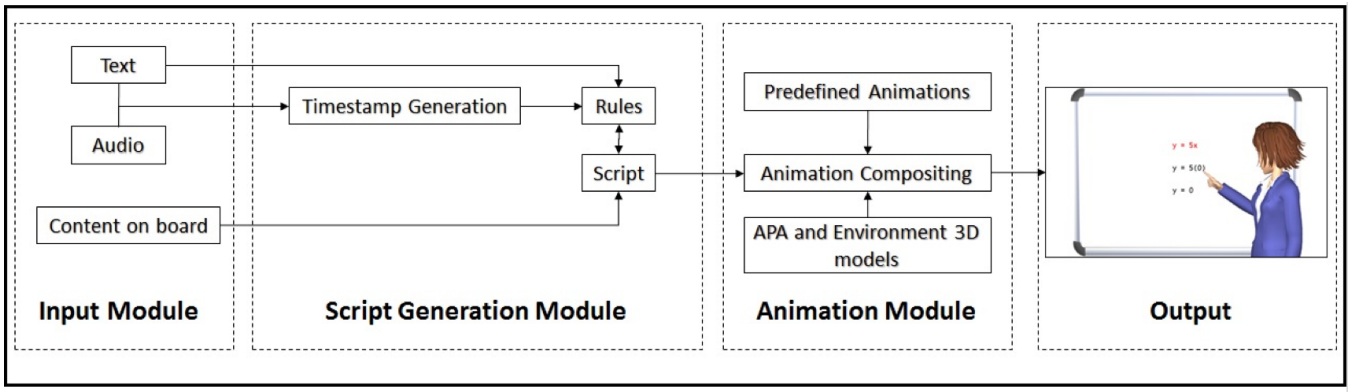


Fig. 1. System Design

the audio recording of what the APA has to say to deliver the lesson, without any animation instructions. The *script generation module* applies a set of rules to generate a script that augments the input text script with APA animation. The *animation module* takes the script and executes it to animate the APA using predefined animations and inverse kinematics algorithms.

A. Input Module

The inputs to the system are the text and audio of the speech of the APA as it delivers the lesson. In addition to these, the input text script also defines the content on the whiteboard. The whiteboard evolves during the lesson with elements being added and erased as needed [51], [52].

B. Script generation module

1) Timestamp generation: The *timestamp generation unit* gets the timestamps of utterances of each word in the input audio using an external software [84]. These timestamps are then used by the *rules* units to generate a script file that contains APA animation commands synchronized to speech. Users who want to change or fine-tune the animation timing can do so easily by editing the script with a text editor.

2) Rules: The *rules unit* parses through the text of the APA speech to generate animation based on rules. We distinguish between primary rules, which generate the deictic gestures of the APA for it to point to the elements represented on the whiteboard as it mentions them in speech, and secondary rules, which improve the overall animation of APA, as it executes the deictic gestures.

Primary rules: Research on instructor gesture has shown that a large number of important instructor gestures are deictic gestures [85], [86], which connect the graphical representation of lesson elements, appearing on a whiteboard, on paper, or in a textbook, to their utterance within the instructor speech. Therefore, the primary animation generation rule of our system is to make the APA point to the whiteboard location of the elements it mentions in speech. In addition to its importance and frequent appearance in the instructors non-verbal communication vocabulary, the rule is also low-level, with general applicability that depends little on content or context. This makes the automatic implementation of the rule tractable, without the prerequisite of a high-level understanding of the lesson material. Our system implements automatic deictic gesture animation by finding in the script provided as input instances when a lesson element is both mentioned in the APA speech and drawn on the whiteboard. The set of possible lesson elements that can constitute the target of deictic gestures is specified by the lesson author as a list of keywords. Consider an example where the input script contains the lines:

```
@ 0 GraphInsertText b 3.8 3 y = 8x
@ 19.5 GraphDeleteText b 3.8 3
```

The first line displays the text “y = 8x” at time 0s, with black ink, at whiteboard location (3.8, 3). The prerecorded audio, which is provided as an input to the system contains a sentence “James saves eight dollars per week”. Since, in our case, all numbers from 0 to 1000 are part of the list of keywords, script parsing connects the “eight” in the sentence to the 8 written on the whiteboard, and generates a pointing gesture. The script with the automatically generated gesture animation instructions looks like:

```
@ 0 GraphInsertText b 3.8 3 y = 8x
@ 5.703 RightPoint GraphCoord 5.792 2.7
@ 19.5 GraphDeleteText b 3.8 3
```

The automatically generated script line makes the APA point with its right hand at whiteboard location (5.792, 2.7), which corresponds to the location of the “8” in “y = 8x”. The location is computed from the parameters of the drawing command GraphInsertText. The time when the APAs right index touches the whiteboard underneath the “8” is 5.703s, which is derived from the timestamp of the word “eight”.

For most of the keywords, the APA points with its right hand, as described in the example above. Additional rules generate more complex deictic gestures for specific keywords. For example, the keywords “x axis” and “y axis” make the avatar automatically trace the respective axes. Another example is a rule that triggers pointing with both hands to indicate the coordinates of a point that was previously drawn on the graph. For example, the keywords “two on x axis and one on y axis” trace perpendiculars to a point with coordinates (2, 1) from each of the two axes, with each of two hands.

1939-1382 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2019.2922134, IEEE Transactions on Learning Technologies

Secondary rules: In order to deliver the payload of deictic gestures in a way that is natural and convincing, the primary rules are enhanced with a set of secondary animation generation rules.

One secondary rule makes the APA move to get sufficiently close to the whiteboard location of a target to which it has to point. This way the APA will always be able to indicate the deictic gesture target by making contact with it. Otherwise, asking the student to extrapolate the pointing direction and to estimate the intersection between the pointing direction and the whiteboard could make the target ambiguous, reducing the benefit of the gesture.

Another secondary rule makes the APA look at the target to which it is pointing. Whereas the APA has perfect knowledge of the whiteboard and could very well point precisely at the target without looking at it, while facing the audience, such a "stunt" would be distracting. Another secondary rule avoids repetitive pointing to the same target;once that target was pointed to, it is flagged to avoid pointing again to it in the near future, even if the speech mentions it repeatedly. This is done to avoid the APA look robotic.

Another group of secondary rules resolve any animation conflict by analyzing the generated script in a second and final pass. A conflict arises when the APA has to point to several targets in rapid succession. Having the APA completely return to the neutral position every time just to immediately engage in the next gesture is unnatural. Conflict resolution makes the avatar hold its pointing gesture at the current target for the short time until it needs to point to a different target. This results in a more fluid, ergonomic motion, with the avatar switching from the current target directly to the next target.

C. Animation Module

The *animation module* takes the script file generated in the *script generation module* as input and generates APA animations. The animation commands in the script file are converted into APA animations using inverse kinematics algorithms and predefined animations [52]. Predefined animations were used to define the high level motion attributes like stride length and pace whereas inverse kinematics were used to define low level motion attributes like how the joints should work in unison to achieve the required motion.

The basic inverse kinematics algorithm makes the APA place the tip of the index finger of a given hand at a given location in 3D space. Pointing gestures are implemented based on the whiteboard location of the pointing target. When the APA cannot reach the target to make contact with it, the APA points with the extended arm towards the location of the target. However, this does not happen in practice as the APA is instructed to move first in a position from where the target is reachable. More complex deictic gestures, such as tracing gestures, are implemented by calling the basic pointing algorithm repeatedly, with a target that follows the lesson element to be traced.

Input: Text and audio of the APA speech
Output: Script file that generates deictic gestures for APA

```
for each word in speech text do
    if word is a keyword or part of a key-phrase then
        if target to corresponding keyword/key-phrase is present on board then
            if target is not pointed to yet then
                Write a command to perform a deictic gesture to the script file
            end if
        end if
    end if
end for
return the script file consisting of deictic gesture commands
```

Fig. 2. Algorithm for single hand pointing gesture generation (Script Generation Module)

IV. User Study

We have conducted a user study with 100 participants, comparing automatically-scripted animations to manually-scripted animations. The study used a within-subjects design; the independent variable was the method used to generate the lectures (manually-scripted versus automatically-scripted); the dependent variables were perceived correctness of the timing of gestures, perceived appropriateness of the number of gestures, lecture preference for inclusion in online lectures(manually-scripted versus automatically-generated). The manually-scripted animations were developed in collaboration with educational psychology researchers who study the role of instructors gestures in education, and they are used as a golden standard in our study. We hypothesized that our system can generate animated lectures that are comparable in quality to these manually-scripted animated lectures. The system was used to generate 6 lectures on the topic of linear equations, which are discussed in Section IV.A. A description of the sample population is presented in section IV.B, and the study procedure is described in section IV.C. The statistical analysis conducted on the data obtained from user study is discussed in Section IV.D.

A. Materials

The rules that were written for the system presented in this paper were used to create 3 animated lectures, in which the APA explained the concept of linear equations with the help of graphs. Four Clips from these lectures were used as stimuli for the user study. These four clips were compared to 4 other clips taken from manually-scripted "golden standard" lectures. The golden standard material was generated through an iterative process where the gestures were manually fine-tuned in timing and frequency until the educational psychology researchers found them to convey the learning material effectively, based on the experts experience in instructor gesture, which is derived from direct measurements of learning with children.

Each lecture explained a linear equation by substituting example numbers into algebraic equations. The linear equations explained in the three lectures were "y = 5x", "y = 8x" and "y = 8x+10". All the three lectures substituted example values of 0,1,2 and 10 for x and calculated the values of y. The

1939-1382 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2019.2922134, IEEE Transactions on Learning Technologies

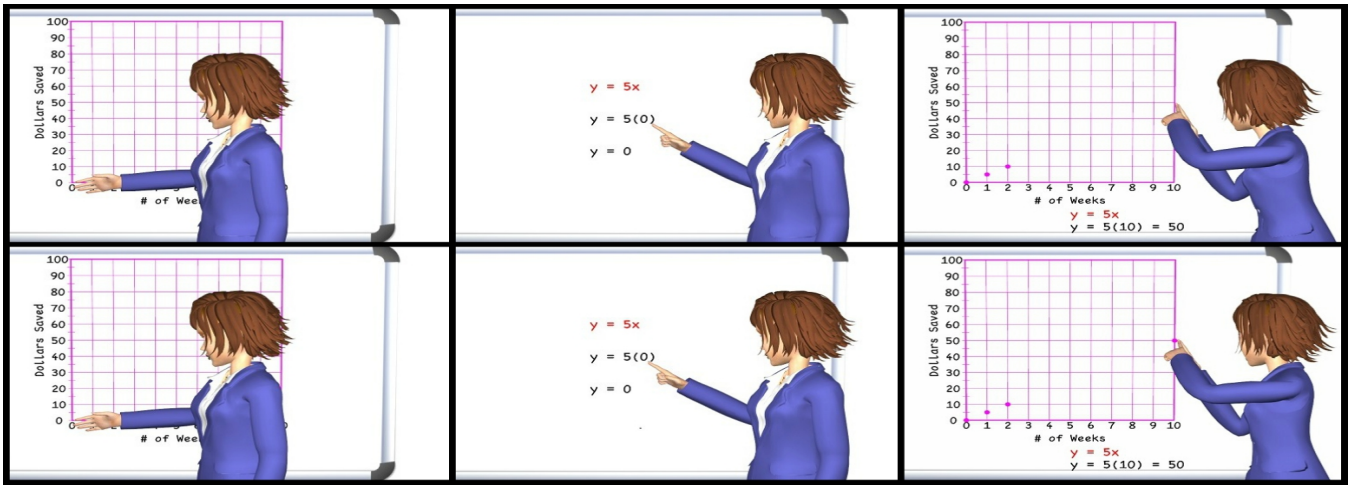


Fig. 3. Frames when automatically-scripted animation (top frames) and manually-scripted animation (bottom frames) produce almost identical gestures

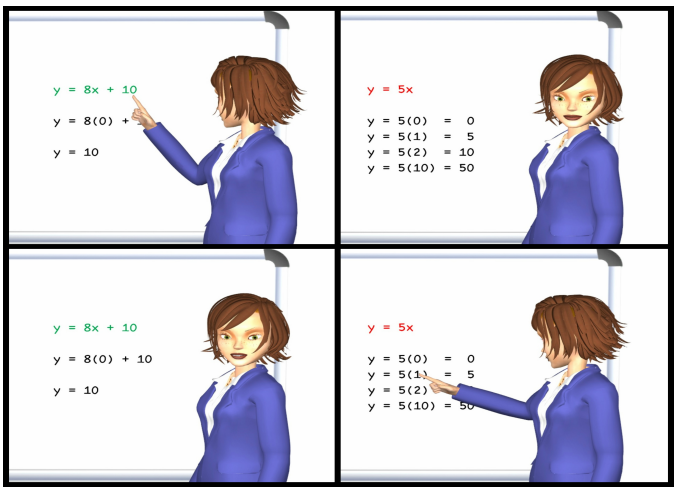


Fig. 4. Frames when automatically-scripted (top frames) and manually-scripted (bottom frames) lectures differ in gestures.

lectures were 117, 70 and 75 seconds of duration and the four clips taken from those lectures were 12, 13, 13 and 15 seconds of duration. The deictic gestures made by the APA included single handed pointing, two handed pointing (right most frames in Figure 3) and line tracing(left most frames in Figure 3).

We have made sure that the chosen clips for the study varied in both timing of gestures and number of gestures between the manual and automated conditions. The four manually-scripted clips contained 3, 1, 8 and 2 gestures, whereas the automatically-scripted clips contained 5, 1, 7 and 1 gestures. We have ensured that the overall number of gestures(14) stayed the same between both conditions, so that we do not add a bias to our study. When it comes to timing of gestures, the differences are much subtler. The execution time of a gesture, i.e., the time from the start of the gesture to the completion of gesture stayed the same for all gestures. The start time of the gestures differed slightly between the two conditions (500 milliseconds on average). Of the 12 gestures that were common between the two conditions, 6 gestures started earlier in the manual condition and 6 gestures started earlier in the automated condition.

In addition to the graph lectures, three equation lectures were also generated using our system, where the APA explained the concept of linear equations using algebraic equations alone. All the deictic gestures made by the APA were single hand pointing gestures like the ones shown in the middle frames of Figure 3. However, we did not include clips from the equation lectures in the online survey to avoid the confounding affect of the type of lecture (graph or equation) on the study.

B. Participants

A total of 100 subjects (N = 100) participated in the study. The subjects included undergraduate students, graduate students and faculty of our university. 56 of the subjects had a Computer Graphics Technology major, 40 had a Computer Science major, and 4 had other majors. 19 participants had less than one year of experience in their major field of study, 30 had 1-2 years of experience, 46 had 2-5 years and 5 had more than 5 years of experience.

C. Procedure

The subjects were sent an email with a brief description of the research and a link to an online survey. In the online survey, subjects first answered two questions about their major field of study and their years of experience. Then they were presented with four webpages in randomized order. Each webpage included two videos, one from an automatically-scripted lecture and one from a manually-scripted lecture. The audio, lip sync and the content on the whiteboard were maintained the same in both the manual and automated conditions, so that they do not have a confounding effect on the results of the study. We do not display any significant facial expressions for the APA other than eye blinks, which also happen at the same times in both conditions.

We realized that the differences between the manually-scripted and automatically-scripted videos were too subtle to distinguish on the first attempt. So, the videos were presented

1939-1382 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2019.2922134, IEEE Transactions on Learning Technologies

side by side to make it easy to play the videos any number of times the user wanted, instead of going back and forth on two pages on the web survey, a problem we encountered in some of our past studies. The subjects were not told which video was automated and which video was manual. The order in which the two videos appeared in each webpage was also randomized. Subjects were allowed to watch the videos any number of times they wanted and in any order they wanted. The side-by-side visualization of the two videos, which differ only in terms of gesture, and not in terms of audio, topic, or domain, is indeed the most revealing and therefore rigorous comparison possible, with any difference in gesture being immediately obvious.

Three questions were asked in each webpage following the videos: (1) Do you agree that the timing of gestures in the video was correct (one response each for the two videos)? (2) Do you agree that the number of gestures in the video is appropriate to support the speech (one response each for the two videos)? (3) Which of the two videos can be included in an online lecture? The subjects were not provided any prior training or examples on what a correct timing of gesture is, or what an appropriate number of gestures is. Those bars were left to the subjects to set and the underlying assumption was that the subjects would use their day-do-day human interactions as references to set those bars.

Subjects answered the first two questions using a 5-point Likert scale ranging from strongly agree to strongly disagree. Subjects were asked to answer the first two questions for both videos separately. The third question was a multiple choice question with four options: (a) only left video, (b) only right video, (c) either of the two videos and (d) neither of the two videos. While the responses to the third question alone would have been sufficient to show that manually-scripted and automatically-scripted videos were comparable in quality, we wanted to address how similar the quality was. We needed some objective metrics for this purpose so that they could be collected individually for both conditions and compared. We chose timing of gestures and number of gestures as the metrics. It must be emphasized that we did not select these two metrics because they correlate to the realism of the animations. As stated in the introduction, producing realistic animation was not the goal of our system. We selected the metrics because they are important and can be easily quantified for an objective comparison. Further, realistic animation cannot just be defined with timing and number of gestures alone, as it depends on many other intrinsic details.

D. Statistical Analysis

We conducted a series of equivalence tests on the responses collected. After that, we conducted a series of ANOVA tests to check for any bias due to the subjects' major field of study and years of experience.

The five-point Likert scale was converted to a numerical scale. Strongly agree corresponds to -2, somewhat agree corresponds to -1, neither agree nor disagree corresponds to 0, somewhat disagree corresponds to +1, and strongly disagree corresponds to +2. For each of the first two questions

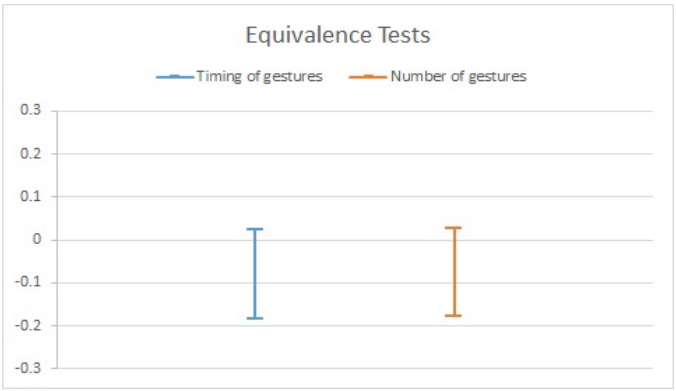


Fig. 5. Confidence intervals of equivalence tests show that the manually-scripted videos and automatically-scripted videos are equivalent in terms of number of gestures and timing of gestures

(about timing of gestures and number of gestures), the mean scores for the four manual videos and four automated videos were calculated. The difference between the two mean values was calculated for each participant. A delta value of 0.2 was considered to build the null and alternate hypotheses for a TOST (Two One-Sided Test) test as given below:

H_0: The absolute difference between the mean scores of automatically-scripted and manually-scripted videos is greater than or equal to 0.2
H_a: The absolute difference between the mean scores of automatically-scripted and manually-scripted videos is less than 0.2

The delta value was chosen as 0.2, as the delta interval (-0.2,0.2) would be just 10% of the total interval (-2,2), which would mean that the videos were 90% similar. Since our goal was not to replicate the manually-scripted lectures but to create automatically-scripted lectures that would be comparable to the manual ones, we have decided to go with 0.2 and set it as the delta value before the start of the experiment.

1) Timing of gestures: The difference between the mean responses of the manually-scripted and automatically-scripted videos was calculated for each participant in regard to timing of gestures. The mean value of the differences for 100 participants was -0.08 (m = -0.08). The standard deviation of the differences for 100 participants was 0.6194 (sd = 0.6194). For the chosen significance value of $\alpha = 0.05$ and 99 degrees of freedom, $t_{0.95,99} = 1.6604$. A 90% confidence interval was calculated using these values, which was (-0.1828, 0.0228) as shown in Figure 5. Since the confidence interval lies strictly within the delta interval of (-0.2,0.2), we rejected the null hypothesis and concluded that the timing of gestures in automatically-scripted videos was equivalent to the timing of gestures in manually-scripted videos.

One-way ANOVA tests were conducted to see if the major field of study and the years of experience had any significant effect on the subjects' responses about the timing of gestures. For a significance level of $\alpha = 0.05$, the p-values calculated were 0.122 and 0.971 for major field of study and years of experience respectively. Since the p-values were greater than

1939-1382 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2019.2922134, IEEE Transactions on Learning Technologies

the chosen significance level, we concluded that there is no bias induced on the responses about timing of gestures because of subject's major field of study and years of experience.

2) Number of gestures: The difference between the mean responses of the manually-scripted and automatically-scripted videos was calculated for each participant in regard to number of gestures. The mean value of the differences for 100 participants was -0.075 (m = -0.075). The standard deviation of the differences for 100 participants was 0.6139 (sd = 0.6139). For the chosen significance value of $\alpha = 0.05$ and 99 degrees of freedom, $t_{0.95,99} = 1.6604$. A 90% confidence interval was calculated using these values, which was (-0.1769, 0.0269) as shown in Figure 5. Since the confidence interval lies strictly within the delta interval of (-0.2,0.2), we rejected the null hypothesis and concluded that the number of gestures in automatically-scripted videos was equivalent to the number of gestures in manually-scripted videos.

One-way ANOVA tests were conducted to see if the major field of study and the years of experience had any significant effect on the subjects' responses about the number of gestures. For a significance level of $\alpha = 0.05$, the p-values calculated were 0.216 and 0.523 for major field of study and years of experience respectively. Since the p-values were greater than the chosen significance level, we concluded that there was no bias induced on the responses about number of gestures because of subject's major field of study and years of experience.

3) Inclusion in online lectures: We presented four sets of videos in the online survey, with each set showing one manually-scripted video and one automatically-scripted video. For each set, we asked the subjects which of the two videos could be included in online lectures. 67% of the participants did not have a preference for one video over the other. 18.75% of the participants preferred only automated video to be included and 14.25% of the participants preferred only manual video to be included in online lectures. These results show that the majority of subjects did not have a preference between the manually-scripted and automatically-scripted lectures. For those who did express a preference, the difference between the two videos was very small (4.5%). This suggests that while there were some periods of times in the lectures where one type of lecture (manual or automated) might be perceived more effective than the other type in regard to timing or number of gestures, the two lectures overall were equivalent. In other words, the responses to the third question show that the equivalence proved by TOST tests is primarily because of the videos being identical and not because of equal number of subjects preferring one type of lectures over the other.

In summary, the equivalence tests conducted showed that the timing of gestures and number of gestures in automated videos were equivalent to those in manual videos. ANOVA tests conducted showed that the major field of study and years of experience had no significant effect on the subjects' evaluation of the videos. More than two-thirds of the 100 participants had no preference over the manual and automated videos to be included in online lectures. These results suggest that the system presented in this paper is capable of generating animated lectures that are comparable in quality to lectures animated through manual scripting.

V. Expert Evaluation

In addition to the user study, we also conducted a small expert evaluation. We presented the same online survey used in the user study to 4 Psychology professors, who have more than 20 years of experience in their fields. Two of the experts are Cognitive Psychologists, one is an Educational Psychologist and one is a Psychologist who specializes in nonverbal behavior. The four experts were selected because of their expertise and experience, and because they could be potential users of the system. The survey included 4 manually-scripted videos and 4 automatically-scripted videos, which translates to 16 total responses about timing of gestures and number of gestures each for the expert analysis. The experts gave the same ratings in regard to the number of gestures for both manual and automated conditions in all the 16 responses. For timing of gestures, the experts gave same ratings for 15 out of 16 responses for both manual and automated conditions. Even in the one comparison that differed, the automatically-scripted video was given a better rating (Strongly Agree) compared to the manually scripted one (Somewhat agree). When asked which video could be included in online lectures, all 16 responses said either video could be used. This analysis, intended as an additional validation, proves that our system is capable of producing automatically-scripted videos that are comparable in quality to manually-scripted videos.

VI. Limitations and Future work

The system presented in this paper is unique because it tries to use the cues in a speech input to generate APA gestures that point to the locations of objects in the APA environment. There are several challenges in this process that make this a complex problem to solve. In this section, we discuss some of these challenges and how they were overcome by our system, as well as the limitations of our system and directions for future work.

A. Ambiguity in selecting pointing-targets

Our system finds a target corresponding to a keyword or key-phrase on the whiteboard and makes the APA point to it. However, multiple targets that can correspond to a keyword or key-phrase result in ambiguous conditions. For example, if there are two equations "y = 5x" and "y = 5(0)" on the whiteboard and the APA says "five dollars", the system needs to know which of the two fives to point to. Currently, we follow a top-down approach where the APA points to the first target if it has not already been pointed to. In the above example, the APA would point to the five in the equation "y = 5x", if the five had not already been pointed to. This approach however doesn't always work. If nothing had been pointed to in the same example above, when the APA says "five times zero", it should point to the five in "y = 5(0)", but it will point to the five in "y = 5x". The selection of the right target to point to depends on the pattern of the speech and more examples of lectures are needed to come up with rules to resolve this ambiguity.

1939-1382 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2019.2922134, IEEE Transactions on Learning Technologies

B. Generation of instructional material

The system presented in this paper generates deictic gestures for an APA that points to objects in the virtual world. But it does not automatically generate the instructional material that needs to be pointed to. For example, when the APA says "y equals five times x", a rule will ask the APA to point at a target "5". But the system does not generate the equation "y = 5x" and write it on the board. In the current version of the system, the user has to specify manually the content that should be present on the board at different times using a script [51], [52]. Automatically generating instructional content, such as equations on the board directly from speech is a complex problem in itself to solve. Sometimes, it might be straightforward to generate simple material like the "y = 5x" equation directly from speech. But in other situations, the material might be a complex sequence of equations that is difficult to deduce from speech, as the speech might not contain sufficient cues. We intend to address this limitation in our future work.

The work reported in the paper did not address gesture style, as embraced by different teachers. One interesting direction of future research could look into providing high-level control of gesture type style, e.g. through parameters that modulate gesture frequency and amplitude, or through the loading of gesture profiles.

C. Scalability across topics and domains

Our system picked low-hanging fruit when it comes to automatic animation of APAs non-verbal communication for eloquent lesson delivery. The system automates deictic gestures, which are frequent and important beyond linear functions and beyond mathematics. Furthermore, the deictic gestures are triggered by simple rules, which only need to establish that the same lesson entity is both represented on the whiteboard and mentioned in speech. As such, the present work is a good start towards scalability across domains.

The list of keywords that define the lesson elements that should serve as target for the deictic gestures is provided by the lesson author and can be easily expanded/modified to cover other mathematics lessons and lessons in other domains. For example, a mechanics lesson on objects moving on an incline under the gravity and friction forces will reuse the ability to point at the parameters and unknowns of an equation developed for our example and will need to expand the keyword list with vector arrows and angle arcs. Since the APA can point at anything on the whiteboard, it only needs to be told what lesson elements are the possible pointing/underlining/circle targets.

Our user study compared snippets from the linear algebra lessons. Using the same topic is required for a valid comparison between the manually and automatically animated sequences. We have demonstrated that deictic gestures can be automatically added to other mathematics topics. As mentioned before, the system does not work as is for topics with a different set of deictic gesture targets—these have to be specified through an updated list of keywords.

D. High-level gesture rules

As mentioned, our system takes advantage of low-level gesture rulespointing at the lesson elements present on the whiteboard that are also mentioned in speech is effective and tractable, i.e. it benefits students and it is tractable from an implementation standpoint, as it bypasses challenging content and context analysis. Of course, the holy grail of automatic APA animation is to devise and implement high-level rules for gesture production, which generate specific gestures based on content, context, and even student characteristics.

Almost a century ago, Edward Sapir noted that we "respond to gestures with an extreme alertness" according to "an elaborate and secret code that is written nowhere, known to none, and understood by all". Whereas gesture research has made significant advances, more work is needed to synthesize a set of high-level gesture production rules to accompany the delivery of educational content. We are part of a multidisciplinary team where our system is used to generate stimuli for the experimental research on gesture. An alternative approach is a computational, machine learning, approach where videos of skilled instructors are analyzed to extract gesture production rules.

In addition to finding these high-level gesture production rules, implementing them also requires advances in natural language processing, to be able to analyze the lesson script to determine when a complex set of preconditions are met for the production of a gesture.

VII. Conclusion

In this paper, we presented a system that can automatically generate deictic gestures for an animated pedagogical agent. The system takes the audio and text of the speech as inputs and produces deictic gestures that point to the exact locations of objects in the environment surrounding the APA. The APA moves in the environment in order to reach the targets to point to. The system was used to generate 6 lectures in which the APA explained the concept of linear equations with the help of algebraic equations and graphs. We conducted a user study with 100 participants and a small expert evaluation with 4 Psychology researchers to compare the quality of videos automatically generated by our system versus videos that were produced manually. The results of both the user study and the expert evaluation show that the timing of gestures and the number of gestures in the automated and manual videos were equivalent. The major field of study and the years of experience had no effect on the evaluations of the participants. More than two-third of the participants had no preference over manual and automated videos for inclusion in online lectures, and those who had a preference were split almost evenly among the two types of animation.

The research work reported in this paper provides a solution for e-learning content creation scalability by automating the generation of APA deictic gestures. Our system connected the three important pieces in an e-learning activity: instructor speech, instructor gestures and instructor environment, thus forming a pipeline that takes speech as input and delivers an e-learning activity as the output. There are many other

1939-1382 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2019.2922134, IEEE Transactions on Learning Technologies

problems yet to be solved to make e-learning content creation completely scalable and more research studies are needed to accomplish that. We hope that our system will be used in such studies and will inspire more research in this direction.

ACKNOWLEDGMENT

The authors would like to thank all the subjects who participated in the user study. The research presented in this paper is 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 presented in this paper are those of the authors and do not represent views of the Institute of Education Sciences, the U.S. Department of Education, or the National Science Foundation.

REFERENCES

[1] N. L. Schroeder, O. O. Adesope, and R. B. Gilbert, "How effective are pedagogical agents for learning? a meta-analytic review," *Journal of Educational Computing Research*, vol. 49, no. 1, pp. 1–39, 2013.

[2] W. L. Johnson, J. W. Rickel, J. C. Lester *et al.*, "Animated pedagogical agents: Face-to-face interaction in interactive learning environments," *International Journal of Artificial intelligence in education*, vol. 11, no. 1, pp. 47–78, 2000.

[3] S. W. Cook, H. S. Friedman, K. A. Duggan, J. Cui, and V. Popescu, "Hand gesture and mathematics learning: lessons from an avatar," *Cognitive science*, vol. 41, no. 2, pp. 518–535, 2017.

[4] H.-C. Yang and D. Zapata-Rivera, "Interlanguage pragmatics with a pedagogical agent: the request game," *Computer Assisted Language Learning*, vol. 23, no. 5, pp. 395–412, 2010.

[5] N. Adamo-Villani and R. Wilbur, "Two novel technologies for accessible math and science education," *IEEE MultiMedia*, vol. 15, no. 4, 2008.

[6] S. Anasingaraju, M.-L. Wu, N. Adamo-Villani, V. Popescu, S. W. Cook, M. Nathan, and M. Alibali, "Digital learning activities delivered by eloquent instructor avatars: scaling with problem instance," in *SIGGRAPH ASIA 2016 Symposium on Education*. ACM, 2016, p. 5.

[7] W. L. Johnson and J. C. Lester, "Pedagogical agents: Back to the future." *AI Magazine*, vol. 39, no. 2, 2018.

[8] T. Koumoutsakis, R. B. Church, M. W. Alibali, M. Singer, and S. Ayman-Nolley, "Gesture in instruction: evidence from live and video lessons," *Journal of Nonverbal Behavior*, vol. 40, no. 4, pp. 301–315, 2016.

[9] A. B. Hostetter, "When do gestures communicate? a meta-analysis." *Psychological bulletin*, vol. 137, no. 2, p. 297, 2011.

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

[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, no. 1, pp. 1–44, 1999.

[12] W. L. Johnson, J. Rickel, R. Stiles, and A. Munro, "Integrating pedagogical agents into virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 6, pp. 523–546, 1998.

[13] D. Powers, R. Leibbrandt, M. Luerssen, T. Lewis, and M. Lawson, "Peta: a pedagogical embodied teaching agent," in *Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2008, p. 60.

[14] C. Davis, J. Kim, T. Kuratate, J. Chen, D. K. Burnham *et al.*, "Making a thinking-talking head," in *AVSP 2007: International Conference on Auditory-Visual Speech Processing 2007*, 2007.

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

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

[17] M. M. Lusk and R. K. Atkinson, "Animated pedagogical agents: Does their degree of embodiment impact learning from static or animated worked examples?" *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, vol. 21, no. 6, pp. 747–764, 2007.

[18] Y. Kim and A. L. Baylor, "Pedagogical agents as social models to influence learner attitudes," *Educational Technology*, pp. 23–28, 2007.

[19] J. Cassell, *Embodied conversational agents*. MIT press, 2000.

[20] ——, "Embodied conversational agents: representation and intelligence in user interfaces," *AI magazine*, vol. 22, no. 4, pp. 67–67, 2001.

[21] 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, no. 8, pp. 27–43, 2006.

[22] A. M. Johnson, G. Ozogul, R. Moreno, and M. Reisslein, "Pedagogical agent signaling of multiple visual engineering representations: The case of the young female agent," *Journal of Engineering Education*, vol. 102, no. 2, pp. 319–337, 2013.

[23] A. M. Johnson, G. Ozogul, and M. Reisslein, "Supporting multimedia learning with visual signalling and animated pedagogical agent: moderating effects of prior knowledge," *Journal of Computer Assisted Learning*, vol. 31, no. 2, pp. 97–115, 2015.

[24] 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.

[25] M. Alseid and D. Rigas, "Three different modes of avatars as virtual lecturers in e-learning interfaces: a comparative usability study," *The Open Virtual Reality Journal*, vol. 2, no. 1, pp. 8–17, 2010.

[26] R. E. Mayer and C. S. DaPra, "An embodiment effect in computer-based learning with animated pedagogical agents." *Journal of Experimental Psychology: Applied*, vol. 18, no. 3, p. 239, 2012.

[27] A. L. Baylor and Y. Kim, "Simulating instructional roles through pedagogical agents," *International Journal of Artificial Intelligence in Education*, vol. 15, no. 2, pp. 95–115, 2005.

[28] A. Gulz and M. Haake, "Social and visual style in virtual pedagogical agents," in *Workshop on Adapting the Interaction Style to Affective Factors associated with the 10th International Conference on User Modeling*, 2005.

[29] Y. Kim and J. H. Lim, "Gendered socialization with an embodied agent: Creating a social and affable mathematics learning environment for middle-grade females." *Journal of Educational Psychology*, vol. 105, no. 4, p. 1164, 2013.

[30] L. Pareto, "A teachable agent game engaging primary school children to learn arithmetic concepts and reasoning," *International Journal of Artificial Intelligence in Education*, vol. 24, no. 3, pp. 251–283, 2014.

[31] S. Domagk, "Do pedagogical agents facilitate learner motivation and learning outcomes?" *Journal of media Psychology*, 2010.

[32] R. E. Mayer, "Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles." 2014.

[33] R. B. Rosenberg-Kima, A. L. Baylor, E. A. Plant, and C. E. Doerr, "Interface agents as social models for female students: The effects of agent visual presence and appearance on female students attitudes and beliefs," *Computers in Human Behavior*, vol. 24, no. 6, pp. 2741–2756, 2008.

[34] A. L. Baylor and S. Kim, "Designing nonverbal communication for pedagogical agents: When less is more," *Computers in Human Behavior*, vol. 25, no. 2, pp. 450–457, 2009.

[35] N. Wang, W. L. Johnson, R. E. Mayer, P. Rizzo, E. Shaw, and H. Collins, "The politeness effect: Pedagogical agents and learning outcomes," *International journal of human-computer studies*, vol. 66, no. 2, pp. 98–112, 2008.

[36] N. L. Schroeder, W. L. Romine, and S. D. Craig, "Measuring pedagogical agent persona and the influence of agent persona on learning," *Computers & Education*, vol. 109, pp. 176–186, 2017.

[37] R. Moreno and T. Flowerday, "Students choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity," *Contemporary educational psychology*, vol. 31, no. 2, pp. 186–207, 2006.

[38] E. A. Plant, A. L. Baylor, C. E. Doerr, and R. B. Rosenberg-Kima, "Changing middle-school students attitudes and performance regarding engineering with computer-based social models," *Computers & Education*, vol. 53, no. 2, pp. 209–215, 2009.

[39] Y. Kim and Q. Wei, "The impact of learner attributes and learner choice in an agent-based environment," *Computers & Education*, vol. 56, no. 2, pp. 505–514, 2011.

[40] R. O. Davis, "The impact of pedagogical agent gesturing in multimedia learning environments: A meta-analysis," *Educational Research Review*, 2018.

[41] Y. R. Guo and D. H.-L. Goh, "Affect in embodied pedagogical agents: Meta-analytic review," *Journal of Educational Computing Research*, vol. 53, no. 1, pp. 124–149, 2015.

1939-1382 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2019.2922134, IEEE Transactions on Learning Technologies

[42] R. Davis and P. Antonenko, "Effects of pedagogical agent gestures on social acceptance and learning: Virtual real relationships in an elementary foreign language classroom," *Journal of Interactive Learning Research*, vol. 28, no. 4, pp. 459–480, 2017.

[43] E. L. Congdon, M. A. Novack, N. Brooks, N. Hemani-Lopez, L. O'Keefe, and S. Goldin-Meadow, "Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention," *Learning and instruction*, vol. 50, pp. 65–74, 2017.

[44] M. W. Alibali, M. J. Nathan, M. S. Wolfgram, R. B. Church, S. A. Jacobs, C. Johnson Martinez, and E. J. Knuth, "How teachers link ideas in mathematics instruction using speech and gesture: A corpus analysis," *Cognition and Instruction*, vol. 32, no. 1, pp. 65–100, 2014.

[45] I. Cutica and M. Bucciarelli, "The deep versus the shallow: Effects of co-speech gestures in learning from discourse," *Cognitive Science*, vol. 32, no. 5, pp. 921–935, 2008.

[46] A. Galati and A. G. Samuel, "The role of speech-gesture congruency and delay in remembering action events," *Language and Cognitive Processes*, vol. 26, no. 3, pp. 406–436, 2011.

[47] J. C. Lester, J. L. Voerman, S. G. Towns, and C. B. Callaway, "Deictic believability: Coordinated gesture, locomotion, and speech in lifelike pedagogical agents," *Applied Artificial Intelligence*, vol. 13, no. 4-5, pp. 383–414, 1999.

[48] H. L. Woo, "Designing multimedia learning environments using animated pedagogical agents: factors and issues," *Journal of Computer Assisted Learning*, vol. 25, no. 3, pp. 203–218, 2009.

[49] R. Moreno, M. Reislein, and G. Ozogul, "Using virtual peers to guide visual attention during learning: A test of the persona hypothesis." *Journal of Media Psychology: Theories, Methods, and Applications*, vol. 22, no. 2, p. 52, 2010.

[50] R. E. Mayer, "based principles for designing multimedia instruction," *Acknowledgments and Dedication*, p. 59, 2014.

[51] N. Adamo-Villani, J. Cui, and V. Popescu, "Scripted animation towards scalable content creation for elearninga quality analysis," in *International Conference on E-Learning, E-Education, and Online Training*. Springer, 2014, pp. 1–9.

[52] J. Cui, V. Popescu, N. Adamo-Villani, S. W. Cook, K. A. Duggan, and H. S. Friedman, "Animation stimuli system for research on instructor gestures in education," *IEEE Computer Graphics and Applications*, vol. 37, no. 4, pp. 72–83, 2017.

[53] 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.

[54] S. W. Cook and S. Goldin-Meadow, "The role of gesture in learning: Do children use their hands to change their minds?" *Journal of cognition and development*, vol. 7, no. 2, pp. 211–232, 2006.

[55] L. Valenzeno, M. W. Alibali, and R. Klatzky, "Teachers gestures facilitate students learning: A lesson in symmetry," *Contemporary Educational Psychology*, vol. 28, no. 2, pp. 187–204, 2003.

[56] L. Rueckert, R. B. Church, A. Avila, and T. Trejo, "Gesture enhances learning of a complex statistical concept," *Cognitive Research: Principles and Implications*, vol. 2, no. 1, p. 2, 2017.

[57] S. Heidig and G. Clarebout, "Do pedagogical agents make a difference to student motivation and learning?" *Educational Research Review*, vol. 6, no. 1, pp. 27–54, 2011.

[58] T. Noma and N. Badler, "A virtual human presenter."

[59] A. C. Graesser, P. Chipman, B. C. Haynes, and A. Olney, "Autotutor: An intelligent tutoring system with mixed-initiative dialogue," *IEEE Transactions on Education*, vol. 48, no. 4, pp. 612–618, 2005.

[60] Y. Kim, "Desirable characteristics of learning companions," *International Journal of Artificial Intelligence in Education*, vol. 17, no. 4, pp. 371–388, 2007.

[61] M. Haake and A. Gulz, "A look at the roles of look & roles in embodied pedagogical agents–a user preference perspective," *International Journal of Artificial Intelligence in Education*, vol. 19, no. 1, pp. 39–71, 2009.

[62] S. K. DMello and A. Graesser, "Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features," *User Modeling and User-Adapted Interaction*, vol. 20, no. 2, pp. 147–187, 2010.

[63] E. Shaw, R. Ganeshan, W. L. Johnson, and D. Millar, "Building a case for agent-assisted learning as a catalyst for curriculum reform in medical education," in *Proceedings of the International Conference on Artificial Intelligence in Education*, 1999, pp. 509–516.

[64] A. L. B. E. A. PLANT, "Pedagogical agents as social models for engineering: The influence of agent appearance on female choice," *Artificial intelligence in education: Supporting learning through intelligent and socially informed technology*, vol. 125, p. 65, 2005.

[65] A. L. Baylor, S. Kim, C. Son, and M. Lee, "Designing effective nonverbal communication for pedagogical agents," in *Proceedings of the 2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology*. IOS Press, 2005, pp. 744–746.

[66] J. C. Lester, L. S. Zettlemoyer, J. P. Grégoire, and W. H. Bares, "Explanatory lifelike avatars: performing user-centered tasks in 3d learning environments," in *Proceedings of the third Annual Conference on Autonomous Agents*. ACM, 1999, pp. 24–31.

[67] C. Pelachaud, N. I. Badler, and M. Steedman, "Generating facial expressions for speech," *Cognitive science*, vol. 20, no. 1, pp. 1–46, 1996.

[68] M. Brand, "Voice puppetry," in *Proceedings of the 26th Annual Conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999, pp. 21–28.

[69] I. Albrecht, J. Haber, and H. Seidel, "Automatic generation of non-verbal facial expressions from speech," in *Proc. Computer Graphics International*, 2002, pp. 283–293.

[70] C. Busso, Z. Deng, M. Grimm, U. Neumann, and S. Narayanan, "Rigid head motion in expressive speech animation: Analysis and synthesis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1075–1086, 2007.

[71] J. Jia, Z. Wu, S. Zhang, H. M. Meng, and L. Cai, "Head and facial gestures synthesis using pad model for an expressive talking avatar," *Multimedia Tools and Applications*, vol. 73, no. 1, pp. 439–461, 2014.

[72] S. Levine, P. Krähenbühl, S. Thrun, and V. Koltun, "Gesture controllers," in *ACM Transactions on Graphics (TOG)*, vol. 29, no. 4. ACM, 2010, p. 124.

[73] S. Levine, C. Theobalt, and V. Koltun, "Real-time prosody-driven synthesis of body language," in *ACM Transactions on Graphics (TOG)*, vol. 28, no. 5. ACM, 2009, p. 172.

[74] M. Neff, "Hand gesture synthesis for conversational characters," *Handbook of Human Motion*, pp. 1–12, 2016.

[75] M. Lhommet and S. C. Marsella, "Gesture with meaning," in *International Workshop on Intelligent Virtual Agents*. Springer, 2013, pp. 303–312.

[76] J. Cassell, H. H. Vilhjálmsson, and T. Bickmore, "Beat: the behavior expression animation toolkit," in *Proceedings of the 28th Annual Conference on Computer graphics and interactive techniques*. ACM, 2001, pp. 477–486.

[77] M. E. Sargin, O. Aran, A. Karpov, F. Ofli, Y. Yasinnik, S. Wilson, E. Erzin, Y. Yemez, and A. M. Tekalp, "Combined gesture-speech analysis and speech driven gesture synthesis," in *Multimedia and Expo, 2006 IEEE International Conference on*. IEEE, 2006, pp. 893–896.

[78] D. Hasegawa, N. Kaneko, S. Shirakawa, H. Sakuta, and K. Sumi, "Evaluation of speech-to-gesture generation using bi-directional lstm network," in *Proceedings of the 18th International Conference on Intelligent Virtual Agents. ACM*, 2018, pp. 79–86.

[79] J. Lee and S. Marsella, "Nonverbal behavior generator for embodied conversational agents," in *International Workshop on Intelligent Virtual Agents*. Springer, 2006, pp. 243–255.

[80] S. Marsella, Y. Xu, M. Lhommet, A. Feng, S. Scherer, and A. Shapiro, "Virtual character performance from speech," in *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. ACM, 2013, pp. 25–35.

[81] B. Ravenet, C. Clavel, and C. Pelachaud, "Automatic nonverbal behavior generation from image schemas," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2018, pp. 1667–1674.

[82] B. Ravenet, C. Pelachaud, C. Clavel, and S. Marsella, "Automating the production of communicative gestures in embodied characters," *Frontiers in psychology*, vol. 9, 2018.

[83] C.-C. Chiu, L.-P. Morency, and S. Marsella, "Predicting co-verbal gestures: a deep and temporal modeling approach," in *International Conference on Intelligent Virtual Agents*. Springer, 2015, pp. 152–166.

[84] I. Rosenfelder, J. Fruehwald, K. Evanini, and J. Yuan, "Fave (forced alignment and vowel extraction) program suite," *URL http://fave. ling. upenn. edu*, 2011.

[85] W.-M. Roth, "Gestures: Their role in teaching and learning," *Review of educational research*, vol. 71, no. 3, pp. 365–392, 2001.

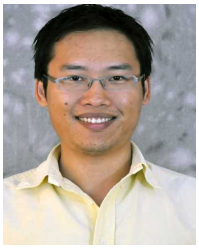
[86] M. W. Alibali and M. J. Nathan, "Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures," *Journal of the learning sciences*, vol. 21, no. 2, pp. 247–286, 2012.

1939-1382 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TLT.2019.2922134, IEEE Transactions on Learning Technologies

Sri Rama Kartheek Kappagantula received the Master's degree in Computer Graphics Technology from Purdue University and Bachelor's degree in Computer Science from BITS-Pilani, India. His research focus is on procedural animation, animated pedagogical agents and building scalable solutions for e-learning content creation.

Nicoletta Adamo-Villani is a Professor of Computer Graphics Technology and a Faculty Scholar at Purdue University. She is an award-winning animator, graphic designer and creator of several 2D and 3D animations that aired on national television. Her area of expertise is character animation and character design and her research interests focus on the application of 3D animation technology to education, Human Computer Communication, and visualization. She is a co-founder and director of the IDEA Laboratory at Purdue University



Meng-Lin Wu is a Ph.D. student in the Computer Science Department of Purdue University. He received B.S. and M.S. degrees in physics from National Taiwan University, Taiwan in 2005 and 2007 while participating in experimental high energy physics at KEK, Japan. Prior to joining Purdue, he programmed game physics at International Games System. His research is concerned with occlusion and visibility management in visualization and augmented reality. His current projects include camera model design, 3D scene acquisition and synthesis, and educational gesturing avatars.

Voicu Popescu is an associate professor of Computer Science at Purdue University. His research interests span the areas of computer graphics, visualization, and computer vision. His current research projects develop novel camera models for efficient and effective rendering of complex visual effects, a system for rapid photorealistic 3D modeling of large-scale real-world environments, a system that aims to make distance education an integral but unobtrusive part of on-campus education, and a method for high-fidelity general-purpose-visualization of large-scale computer simulations.

1939-1382 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.