Digital Learning Activities Delivered by Eloquent Instructor Avatars: Scaling With Problem Instance

Saikiran Anasingaraju,* Meng-Lin Wu,† Nicoletta Adamo-Villani,‡ Voicu Popescu§
Purdue University

Susan Wagner Cook¶
The University Of Iowa

Mitchell Nathan,‖ Martha Alibali**
University of Wisconsin-Madison

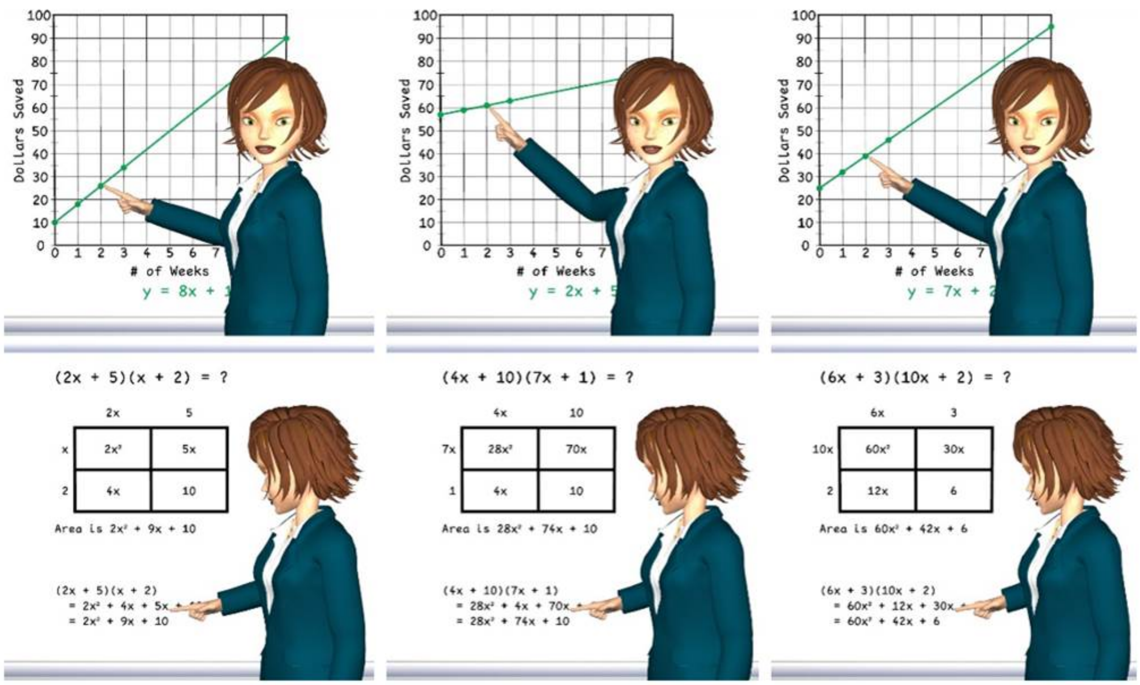

Figure 1: *Linear equation (top) and polynomial multiplication (bottom) learning activities delivered by our instructor avatar: original problem instances (left), and two new problem instances generated automatically (middle and right).*

Abstract

We present an approach for achieving scalable authoring of digital learning activities, without sacrificing delivery eloquence. A computer animation character serves as an instructor avatar that not only speaks but also makes deictic, iconic and charisma gestures. The avatar is controlled via a text script, without the prerequisites of computer programming or animation expertise. Given a script for a problem, the system automatically generates scripts for additional instances of the problem, by adapting the targets of the deictic gestures, the speech, and the synchronization between speech and gestures. Starting from initial learning activities of a few minutes, the system can automatically generate hours of quality on-line learning activities. An evaluation by computer graphics, computer animation, and education research experts reveals that the automatically generated learning activities have animation quality that is comparable to that of the original activities.

Keywords: instructor avatar, instructor gesture, digital learning activity authoring, scalability.

Concepts: •**Applied computing** → **E-learning;**

*e-mail:sanasing@purdue.edu
†e-mail:wu223@purdue.edu
‡e-mail:nadamovi@purdue.edu
§e-mail:popescu@purdue.edu
¶e-mail:susan-cook@uiowa.edu
‖e-mail:mnathan@wisc.edu
**e-mail:mwalibali@wisc.edu

SA '16 Symposium on Education - Papers ,, December 05-08, 2016, ,

1 Introduction

On-line learning activities have great potential for improving learning at and after school. An important bottleneck on the path of realizing this potential is the scarcity of quality digital learning materials. Whereas a modern tablet or even a smart phone can now easily run an interactive learning activity with high quality graphics, authoring learning activities remains a challenge.

Given a math problem in textual form, it is straightforward to generate a large set of instances of the original problem by changing the numbers involved in the problem. For example, a learning activity involving a linear equation $y = 2x + 48$ can be easily changed to any linear equation by changing the slope (i.e., 2) and the y-intercept (i.e., 48). However, putting the learning activity in textual form misses the richness of real-world instructors who, in addition to writing on a whiteboard, also use speech and gesture. Speech and gesture are particularly important as a source of information for young learners, who might not be proficient readers, and for learners who are not fluent in English. One approach is to video-record a talented instructor who eloquently delivers the learning activity in a supportive context. However, the approach does not scale with the problem instance: changing the video without re-recording the instructor for every problem instance is challenging if not impossible.

We are a team of computer science and education researchers and we have developed a system of computer animation characters that serve as instructor avatars. The avatar stands in front of a whiteboard as it explains a math problem. The avatar speaks by lip-synching to a pre-recorded audio file. The avatar also gestures: it points at the various elements of the problem on the whiteboard, it makes iconic gestures, such as a balance gesture to indicate the equilibrium between the left and the right side of an equation, and it makes beat gestures to convey a charismatic and engaging personality, using synchronous-parallel-outward hand gesture. The avatar is controlled with a text script that specifies what and when the avatar speaks and gestures. The script is executed to create an animation that shows the instructor avatar delivering the learning activity. The system of instructor avatars is used both as a research testbed, to generate stimuli for studies that identify which instructor gestures are more effective and for which learners, and as an authoring tool for on-line learning activities.

In this paper we present our results on achieving on-line learning activity authoring scalability with problem instance: given an avatar script for an original instance of a problem, additional scripts for other problems of the same type are generated automatically, without sacrificing the eloquence of the delivery by the instructor avatar. In fact, the animations for the new instances of the problem are visually indistinguishable from that of the original problem.

We demonstrate scalability with problem instance in the context of two mathematical concepts: linear equation, and polynomial multiplication. Figure 1 shows the avatar delivering the original instances of the problems, for which the scripts were written (left), and two new instances for each problem type (middle and right), created automatically by adjusting the original scripts to the new problem instances. The content of the whiteboard (e.g. the expressions, the graph, the polynomial), the targets of the deictic gestures (e.g. the points on the graph or the expression terms where the avatar points), the speech, and the synchronization between gestures and speech are automatically adapted to the new instances of the problems. We also refer the reader to the accompanying video.

Macao
ISBN: 978-1-4503-4544-6/16/12
DOI: http://dx.doi.org/10.1145/2993352.2993355

We conducted a survey to compare the quality of automatically generated learning activities to that of original, hand scripted learning activities. Generated polynomial multiplication and linear equation activities were shown to 41 participants with technical (i.e. computer science, computer graphics, animation) and with education research backgrounds. The participants were asked to rate the quality of the animation and of the audio on a scale from 1 (lowest) to 5 (highest). The average animation and audio scores were 4.4 and 3.83, respectively. When the original and automatically generated learning activities were shown side by side, without audio, 82.5% of the participants found no difference between the animation quality of the two. We conclude that the animation generated by automatic scripting is of high quality and it is comparable to the original animation, and that the automatically generated audio is of acceptable but lower quality than the original audio.

Given the script for an original problem instance, a large number of learning activities can be created automatically based on new problem instances. For example, the system can generate hundreds of linear equation learning activities with different slopes and y-intercepts, and hundreds of polynomial multiplication learning activities with different polynomial coefficients. The original five minute scripts turn into hours of online activities, which can be leveraged for adaptive instruction, and personalized for individual learners.

2 Prior Work

Several examples of teaching avatars can be found in the literature and a substantial body of research supports their pedagogical efficacy. Early examples include Cosmo [Lester et al. 1997b], a cosmonaut who explains how the internet works, Herman [Lester et al. 1999b; Lester et al. 1999a], a bug-like creature that teaches children about biology, and STEVE [Johnson and Rickel 1998; Johnson et al. 1998; Rickel and Johnson 1999], who trains users in operating complex machinery using speech, pointing gestures, and gaze behavior. In Virtual Human project [Reithinger et al. 2006] a teacher gives astronomy lessons by following different pedagogical paradigms and using a variety of nonverbal behaviors. PETA, a 3-D animated human head that speaks by synthesizing sounds while conveying a variety of facial articulations [PET 2008], teaches children how to acquire a new language in a spontaneous, unconscious manner. A similar example is the Thinking Head [Davis et al. 2007], a virtual anthropomorphic agent that speaks and displays emotion through complex facial expressions, vocal prosody, and gestures. Gesturing avatars have also been used to teach sign language, mathematics, and science to young deaf children using sign language [Adamo-Villani and Wilbur 2008].

Many studies confirm the intended positive influences on education by systems using these agents [Holmes 2007; Moreno and Mayer 2007; Lusk and Atkinson 2007]. Students appear to attend primarily to agents' faces much like they attend to the faces of live conversational partners [Louwerse et al. 2009; Theonas et al. 2008]. Studies also suggest that teaching avatars could be employed in e-learning environments to enhance users' attitudes towards online courses [Annetta and Holmes 2006], as animated agents help to create a positive learning environment [Lester et al. 1997a]. Animated characters that use multiple modalities (e.g. speech and gestures) appear to lead to greater learning than agents that interact only in a single channel [Lester et al. 1997a; Lusk and Atkinson 2007]. Including gesture in instruction with human teachers increases learning of a wide variety of mathematical concepts, such as conservation [Church et al. 2004; Ping and Goldin-Meadow 2008], symmetry [Valenzeno et al. 2003], and mathematical equivalence [Church et al. 2004; Singer and Goldin-Meadow 2005]. In addition to promoting conceptual understanding, 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 [Church et al. 2007; Cohen and Otterbein 1992; Cutica and Bucciarelli 2008; Feyereisen 2006; Galati and Samuel 2010]. Gestures may also influence how learners perceive their instructors [Williams and Ware 1977].

A few e-learning content authoring tools exist that enable users to incorporate animated avatars into digital education materials with text, images, and videos [Pappas a; Pappas b]. However, the avatar models are very simple and the animations quite limited. One avatar authoring system lets the users design customized avatars for web deployment (with limited animation) [Lab], and one avatar system allows K-12 children to improve their writing skills by designing storytelling avatars and placing them in virtual worlds [LLC].

Creating believable and expressive 3D avatars that speak and gesture is a challenging task that has been approached from various directions: (1) the avatar can be manually animated by a skilled artist, (2) the motion and facial deformations can be captured directly from an actor using motion capture equipment and applied to the character, or (3) the animation can be scripted, e.g the character's speech and gestures are automatically generated from input text. Methods 1 and 2 are time consuming, demand artistic talent and require expensive equipment, hence they are not easily scalable. Method 3 promises high scalability, as it provides a fast and accessible method for creating a large number of e-lessons delivered by instructor avatars. Method 3 also allows scaling of both speech and gesture, which is desirable given the benefits of gesture on learning.

To date, we have developed a system that provides a computer animation instructor avatar that is animated quickly and effectively based on a text script [Cui et al. 2014]. The script, created by the e-Learning content creator with a conventional text editor, specifies what and when the avatar does and says, and is executed automatically to obtain the desired animation. The animation is produced quickly, and without the requirements of artistic talent, of familiarity with complex animation software, and of programing expertise. To our knowledge, no other system currently exists that allows for scalable creation of e-lessons delivered by expressive 3D avatars that speak and gesture.

3 Scalability With New Problem Instance

The new problem instance is generated according to the piepline shown in Figure 2. The input is the script text file (*Script*), which specifies what and when the avatar says and gestures, the script audio file (*Audio*), which is a recording of what the avatar has to say to deliver the learning activity, and an audio file (*Numbers audio*), which contains the recording of all possible values of the script parameters. For example, for a linear expression $y = ax + b$, with $0 < a < 10$, $0 \leq b \leq 50$, and $0 \leq x \leq 10$ the *Numbers audio* file has to cover the 0 to 140 range.

The script is preprocessed to identify the parameters of the problem. For example, the original script might contain the linear expression $y = 2x + 48$ and the text "After 2 weeks James would have saved 52 dollars", which are changed to $y = ax+b$ and to "After 2 weeks James would have saved $a \times 2 + b$ dollars". The audio files are preprocessed using a forced aligner [of Pennsylvania Linguistics Lab] to identify the starting and ending timestamps of all the words. This preprocessing is done once per problem, after which any number of problem instances can be generated automatically.

The problem instance generation begins with instantiating the script parameters. Values for the input parameters, i.e., a and b in our example, are generated randomly within an allowable range. The internal parameters of the script are computed by propagating the input parameter values. If the randomly generated values for a and b are 2 and 3, then "$a \times 2 + b$ dollars" becomes "7 dollars". The script with known parameter values is used to update the whiteboard, e.g. to display the correct linear expression and the correct points on the graph, to update the target of deictic gestures, e.g. to point to the correct point on the graph, and to assemble the audio file of the problem instance. The numbers for the parameter values are spliced into the original audio file leveraging the timestamps found through forced alignment. The lip sync animation is updated according to the new numbers. The new numbers have different time lengths in the audio file, so the gesture timeline is shifted to allow gestures to remain synchronized to speech. Once the script is finalized, the animation of the new problem instance is rendered.

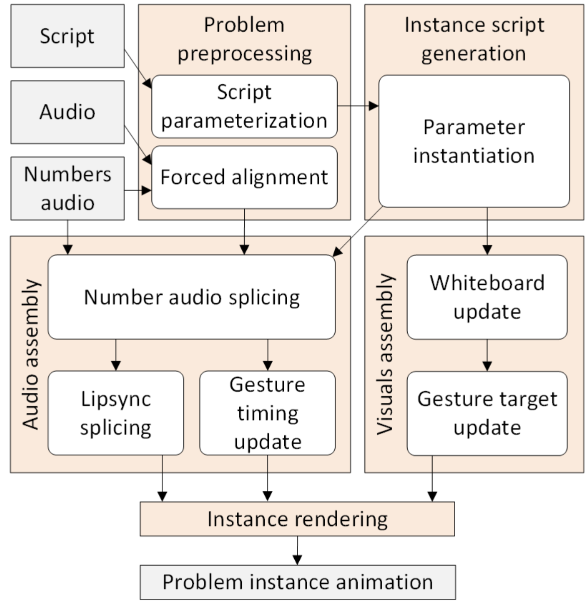

Figure 2: *Generation of new problem instance.*

Our system currently has two instructor avatars, one corresponding to a professionally dressed young adult (Figure 1), and one corresponding to a casually dressed teenager. We have developed lessons and practice problems covering mathematical equivalence, polynomial multiplication, and linear equation concepts. Starting from a learning activity of a few minutes, our approach opens the door to creating hours of eloquently delivered learning activities. The new problem instances are indistinguishable from the original problem except for small audio imperfections caused by the splicing, additional audio processing can likely decrease these imperfections. Our approach has great potential in the context of learning personalization that accounts for individual learner differences, where it can be used to practice problem instances matching the current knowledge level of the learner.

4 Results and Discussion

We have implemented and evaluated our approach for scalability of on-line learning activity authoring in the context of two mathematics topics (Section 4.1). We have conducted an expert evaluation with 41 participants with computer graphics and education research backgrounds to compare the quality of automatically generated learning activities to that of the original learning activities (Section 4.2). A statistical analysis of the results of the expert evaluation is given in Section 4.3.

4.1 Scalable authoring of math learning activities

We have implemented and evaluated our approach in the context of two topics in mathematics education: linear equation and polynomial multiplication. For both topics, the lesson uses two alternative representations that appear simultaneously on the virtual whiteboard and that are linked by the instructor through gesture. Education research has shown that when the instructor uses multiple representations of a concept, and when the instructor explicitly connects the multiple representations through gestures, students learn more. Consequently it is essential that the new, automatically generated on-line learning activities are delivered by an instructor avatar that gestures accordingly.

Linear equation

The original linear equation learning activity takes about 10 minutes. The linear equation is shown both in the algebraic expression and in the graph representation. The instructor avatar gestures and their synchronization with speech are precisely defined through an original text script created with the help of the education researchers on our team. The learning activity contains a linear equation of the form $y = ax + b$. Parameter a can take any value from 1 to 9, and parameter b can take any value from 0 to 50. The system can generate a learning activity for any one of the $9 \times 51 = 459$ possible linear dependencies, turning the original 10 minutes into over 75 hours.

Polynomial multiplication

The original polynomial multiplication learning activity also takes about 10 minutes. The polynomial multiplication is shown both in the algebraic expression form, where each term of the first polynomial is distributed to each term of the second polynomial, and in rectangle area form, where the first polynomial is associated with the width of a rectangle and the second polynomial is associated with the height of the rectangle (see Fig. 1, bottom). The learning activity uses the multiplication of two binomials of the form $(ax + b)(cx + d)$. Parameters a and c can have any value from 1 to 9 and b and d can have any value from 0 to 9, so the system can generate any one of the 8,100 possible combinations. This totals 1,350 hours of on-line learning activities, from which to select activities with appropriate difficulty for each evolving individual learner.

4.2 Expert evaluation

We evaluated and compared the original and the automatically generated learning activities with regard to quality of the gestures, quality of the audio, and quality of the animation through an online survey. The survey used one minute original and generated animation sequences from the polynomial multiplication and from the linear equation learning activities. We provide the four animation sequences as additional material accompanying our submission.

Participants

There were 41 participants in the survey, including the seven authors of this paper. 30 participants had primary expertise in computer science/computer graphics/computer animation, and they were recruited from the computer graphics and animation laboratories of the computer science and the computer graphics technology departments of University A. 11 participants had primary expertise in education research, and they were recruited from the educational psychology laboratories of universities B and C.

Evaluation Instrument

The web survey consisted of 11 screens and was structured in the following way: (1) participants were first asked to answer a question regarding their background; (2) the linear equation and the polynomial multiplication generated animation sequences were presented in randomized order, and, for each sequence, participants were asked to rate the quality of the audio of the verbal explanation, and the quality of the gestures made by the instructor; (3) the original and generated animation sequences were shown side by side and muted, for each of the two math topics, in randomized order, and, for each side by side video, the participants were asked to identify the sequence with better animation quality; the possible answers were left, right, no difference; if the participant preferred a sequence the next screen prompted for the reason behind the participant's choice.

Procedure

Participants were sent an email invitation to participate in the survey, and the link to the web survey. Participants completed the survey using their own computers. A participant completed the survey in about 10 minutes.

Findings

The results of the survey are given in Table 1. The first row gives the primary expertise of the 41 participants. The top and bottom halves of the table give the survey data for the two learning activities.

For each learning activity, the quality of the animation (i.e. video) and of the audio of the generated lesson were scored separately. The participants used a scale from 1 (lowest quality) to 5 (highest quality). The video quality received high overall scores, of 4.19 and 4.6 for polynomial multiplication and for linear equation, respectively. There was agreement between the participants with a technical background (i.e. CS, graphics, and animation) and the participants with an education research background. The audio received lower scores than the video, with overall scores of 3.66 and 4.0. The participants with an education research background were more critical of the audio quality than the participants with a technical background.

When asked to compare the generated lesson to the original lesson in a side-by-side muted video, 80% and 85% of the participants thought that the two animations were of the same quality, for the polynomial multiplication and the linear equation activities, respectively.

For the polynomial multiplication learning activity, 33 participants thought that the side-by-side lessons had the same animation quality, 5 participants preferred the original lesson and 3 participants preferred the generated lesson. Interestingly, out of the 3 education research participants who preferred a lesson, 2 preferred the generated lesson and only 1 preferred the original lesson. When a participant preferred one lesson over the other, they were asked to explain their preference. The explanations reveal that the preferences are only "slight" preferences. One participant who preferred the original lesson states that "movements seem more natural", a participant who preferred the generated lesson states that the lesson "seemed more fluid, though it's hard to tell". Since the overwhelming majority of participants found the two lessons of similar quality (i.e. 80%), since those who preferred a lesson were split between the two and had only a slight preference, we conclude that the animation in the generated polynomial multiplication lesson is of similar quality to the animation in the original lesson.

For the linear equation learning activity, 35 participants thought that the side-by-side lessons had the same animation quality, 6 participants preferred the original lesson, and no participant preferred the generated lesson. Out of these 6 participants, 5 have technical expertise, and 1 has education research expertise. Examining the par-

Table 1: *Results of survey comparing original to automatically generated instance of learning activity, for two math learning topics.*

			Participant expertise					
			CS graphics animation		**Education research**		**Overall**	
		Participants	30	73%	11	27%	41	100%
Polynomial multiplication	**Generated lesson quality**	**Video**	4.2/5.0		4.18/5.0		4.19/5.0	
		Audio	3.97/5.0		2.82/5.0		3.66/5.0	
	Side-by-side comparison	**Same**	25	83%	8	73%	33	80%
		Original	4	13%	1	9%	5	12%
		Generated	1	3%	2	18%	3	7%
Linear equation	**Generated lesson quality**	**Video**	4.7/5.0		4.3/5.0		4.6/5.0	
		Audio	4.3/5.0		3.3/5.0		4.0/5.0	
	Side-by-side comparison	**Same**	25	83%	10	91%	35	85%
		Original	5	17%	1	9%	6	15%
		Generated	0	0%	0	0%	0	0%

ticipant provided explanations for the preferences, it appears that the instance of the linear equation used in the original lesson is judged to be more suitable for an eloquent illustration, compared to the instance used in the generated instance. Participants noted that the original lesson had a "more noticeable location difference for points on the line", that "gestures seemed more natural with the steeper line", or that "the slope of that function is larger, so more illustrative". In other words, it appears that the few participants who preferred the original lesson did not do so due to a lower quality of the animation of the generated lesson, but rather because the values of the parameters of the linear equation used in the original lesson resulted in a graph with greater illustrative value. Although this conclusion absolves the automatic animation from lowering the quality of the generated lesson, it remains that not all possible combinations of parameter values are equally useful in learning activities. An interesting direction of future work is to derive rules that define the subset of the parameter space that results in eloquent illustrations.

4.3 Statistical analysis

A series of t-tests and ANOVA tests were conducted for both problems, linear equation and polynomial multiplication. These tests were conducted to determine whether the differences in the perceived quality of gestures and the perceived quality of audio between the original sequences and the automatically generated ones were statistically significant.

With regard to the perceived quality of the 'gestures', the hypotheses were the following for both problems:

H_0: The mean perceived quality of the gestures of the automatically generated sequence $< 75\%$ of the mean perceived quality of the gestures of the original sequence. H_1: The mean perceived quality of the gestures of the automatically generated sequence $\geq 75\%$ of the mean perceived quality of the gestures of the original sequence.

With α = 0.05, degrees of freedom = 40 and $t_{0.95}$ = 1.684, results of the t-test show that t = 1.854 for linear equation and t = 2.4749 for polynomial multiplication. Since t $>$ 1.684 for both problems, we reject the null hypotheses. Thus, for both problems, the perceived quality of the gestures for the automatically generated sequence is not significantly less than 75% of the perceived quality of the gestures for the original sequence.

With regard to the perceived quality of the 'audio', the hypotheses were the following:

H_0: The mean rating of the perceived audio quality of the automatically generated sequence $<$ 3.2 (out of 5) H_1: The mean rating of the perceived audio quality of the automatically generated sequence $\geq$ 3.2

With α = 0.05, degrees of freedom = 40 and $t_{0.95}$ = 1.684, results of the t-test show that t = 5.3102 for linear equation and t = 2.1409 for polynomial multiplication. Since t $>$ 1.684 for both problems, we reject the null hypotheses. Thus, for both problems, the mean rating of the perceived audio quality of the automatically generated sequence is not significantly less than 3.2/5.

ANOVA tests were conducted for both problems to determine the effect of participants' backgrounds on the perceived quality of the 'gestures'. At an alpha level of 0.05, probability values of 0.832324 and 0.0552218 were calculated for the problems linear equation and polynomial multiplication respectively. Since $P > 0.05$ for both problems, there is no significant difference in perceived quality of the gestures among the different groups of participants (namely psychology, computer science, computer graphics/animation, and education experts).

Similarly, ANOVA tests were conducted for both problems to determine the effect of participants' backgrounds on perceived quality of the 'audio'. At an alpha level of 0.05, probability values of 0.012839 and 0.058315 were calculated for the problems linear equation and polynomial multiplication respectively. For polynomial multiplication, since $P > 0.05$, there is no significant difference in perceived audio quality among the different groups of participants.

Since $P < 0.05$ for linear equation, there is a significant difference in the perceived audio quality among different groups of participants. Comparisons between individual groups were made using the Turkey HSD test. Since there are four groups, we had six one-to-one comparisons. The groups psychology and computer science had a significant difference between their audio quality as their HSD test' s p-value is less than 0.05. Similarly, the groups psychology and computer graphics/animation had a significant difference between their audio quality as their HSD test's P-value is also less than 0.05. All the other four comparisons between the groups turned out to be nonsignificant. Future work can focus on understanding the influence of participants' background on the perceived quality of the audio.

Summary

In summary, we conclude that the quality of the generated animation is high and comparable to that of the original animation, and the quality of the generated audio is good, but clearly below that of the original audio.

5 Conclusions and Future Work

We have presented an approach for authoring on-line learning activities that scales with the problem instance, without sacrificing delivery eloquence. A computer animation character serves as an instructor avatar that not only speaks but also gestures as it explains math problems. The instructor avatar is controlled with a text script. Given the script for an initial instance of a problem, the system automatically generates new instances of the problem by changing the numbers of the problem and by adapting the instructor avatar speech and gesture to the new problem instance. We have conducted an expert evaluation that reveals that the animation of the automatically generated learning activities is indistinguishable from that of the initial hand-scripted activities. We have demonstrated our approach in the context of the mathematics concepts of linear equation and of polynomial multiplication. Starting from initial learning activities of a few minutes, the system can automatically generate hours of quality on-line learning activities.

Our group has focused so far on the visual rendering of the instructor avatar, and we have achieved good scalability of graphics and animation with problem instance. One direction of future work is to improve the scalability of the audio, which is now the weakest link of the system. The goal is to minimize the amount of audio recording needed, while maximizing the continuity of the spliced audio. One option is to record the numbers that are spliced in with several tonalities and in combination with succeeding words (e.g. 'x', 'dollar'), while another option is to attempt to process one generic recording automatically to achieve continuity with the surrounding audio.

The learning activities generated with our system are ready to be posted on-line to benefit learners. To ensure consistent quality of generated lessons, Pedagogical Content Knowledge [Shulman 1986], or PCK, can be incorporated as part of an evaluation filter after the combinatorial generation of all possible problem items. PCK includes knowledge of selecting instructive examples. In the future, PCK can be involved for the need of variations in whole lessons composed of these generated items to keep learners engaged.

Another direction of future work is to use our system to cover additional learning topics in mathematics and beyond, and to cater to learners in all age groups. We do not foresee any difficulties in extending our system to other quantitative domains, such as physics, where, for example, the illustration of a mechanics problem depends on numerical parameters. However, an obvious limitation of our approach is the lack of applicability to domains such as history or literature, where the lesson content is defined by more complex, subjective parameters. A third direction of future work is to couple our system with an intelligent tutor that models the learner's general background and specific knowledge, to dynamically personalize the learning activity sequence to maximize learning for each individual learner.

Acknowledgements

We thank the respondents of the survey for their expert evaluation of our system, and our reviewers for their help with improving our paper. 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

ADAMO-VILLANI, N., AND WILBUR, R. 2008. Two novel technologies for accessible math and science education. *IEEE Multimedia Special Issue on Accessibility*, 38–46.

ANNETTA, L. A., AND HOLMES, S. 2006. Creating presence and community in a synchronous virtual learning environment using avatars. *Intern. J. Instruct.Technol. Dist. Learn. 3*, 27–43.

CHURCH, R. B., AND AYMAN-NOLLEY, S. 2008. The role of gesture when instruction is given by adults versus same-age peers. In *Proceedings of the Jean Piaget Conference*.

CHURCH, R., AYMAN-NOLLEY, S., AND MAHOOTIAN, S. 2004. The role of gesture in bilingual education: Does gesture enhance learning? *International Journal of Bilingual Education and 7(4)*, 303–319.

CHURCH, R., GARBER, P., AND ROGALSKI, K. 2007. The role of gesture in social communication and memory. *Gesture 7*, 137–157.

COHEN, R., AND OTTERBEIN, N. 1992. The mnemonic effect of speech gestures: Pantomimic and non-pantomimic gestures compared. *European Journal of Cognitive Psychology 4*, 113–139.

CUI, J., ADAMO-VILLANI, N., AND POPESCU, V. 2014. Charismatic and eloquent instructor avatars with scriptable gesture. In *Proceedings of Siggraph 2014 Talks*.

CUTICA, I., AND BUCCIARELLI, M. 2008. The deep versus the shallow: Effects of co-speech gesture in learning from discourse. *Cognitive Science 32*, 921–935.

DAVIS, C., KIM, J., KURATATE, T., AND BURNHAM, D. 2007. Making a thinking-talking head. In *In: Proc. Of the International Conference on Auditory-Visual Speech Processing (AVSP 2007)*.

FEYEREISEN, P. 2006. Further investigation on the mnemonic effect of gestures: Their meaning matters. *European Journal of Cognitive Psychology 18*, 185–205.

GALATI, A., AND SAMUEL, A. G. 2010. The role of speech-gesture congruency and delay in remembering action events. *Language and Cognitive Processes*.

HOLMES, J. 2007. Designing agents to support learning by explaining. *In: Comput. Educ 48*, 523–547.

JOHNSON, W., AND RICKEL, J. 1998. Steve: An animated pedagogical agent for procedural training in virtual environments. *SIGART Bulletin 8*, 16–21.

JOHNSON, W., RICKEL, J., STILES, R., AND MUNRO, A. 1998. Integrating pedagogical agents into virtual environments. *Presence: Teleoperators and Virtual Environments 7*, 523–546.

LAB, V. C. Avatar authoring tool. http://vcl.iti.gr/avatar-authoring-tool/.

LESTER, J., CONVERSE, S., STONE, B., KAHLER, S., AND BARLOW, T. 1997. Animated pedagogical agents and problem-solving effectiveness: A large-scale empirical evaluation. In *Proceedings of the eighth world conference on artificial intelligence in education*, 23–30.

LESTER, J., VOERMAN, J., TOWNS, S., AND CALLAWAY, C. 1997. Cosmo: A life-like animated pedagogical agent with deictic believability. In *In Notes of the IJCAI '97 Workshop on Animated Interface Agents: Making Them Intelligent*, 61–70.

LESTER, J., STONE, B., AND STELLING, G. 1999. Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *In: User Modeling and User-Adapted Interaction 9(1-2)*, 1–44.

LESTER, J., TOWNS, S., AND FITZGERALD, P. 1999. Achieving affective impact: Visual emotive communication in lifelike pedagogical agents. *In: The International Journal of Artificial Intelligence in Education 10(3-4)*, 278–291.

LLC, W. Avatar storytellers. http://www.avatarstorytellers.com/default.asp?iId=HILHG.

LOUWERSE, M., GRAESSER, A., MCNAMARA, D., AND LU, S. 2009. Embodied conversational agents as conversational partners. *Applied Cognitive Psychology 23*, 1244–1255.

LUSK, M. M., AND ATKINSON, R. K. 2007. Varying a pedagogical agents degree of embodiment under two visual search conditions. *Applied Cognitive Psychology 21*, 747–764.

MORENO, R., AND MAYER, R. 2007. Interactive multimodal learning environments. *In: Education. Psychol. Rev. 19*, 309–326.

OF PENNSYLVANIA LINGUISTICS LAB, U. Fave-align. http://fave.ling.upenn.edu/usingFAAValign.html.

PAPPAS, C. Free authoring tools for elearning. http://elearningindustry.com/free-authoring-tools-for-elearning.

PAPPAS, C. Html5 elearning authoring tools. http://elearningindustry.com/the-ultimate-list-of-html5-elearning-authoring-tools.

2008. Peta a pedagogical embodied teaching agent. In *In: Proc. of PETRA08: 1st international conference on PErvasive Technologies Related to Assistive Environments*.

PING, R., AND GOLDIN-MEADOW, S. 2008. Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology 44*, 1277–1287.

REITHINGER, N., GEBHARD, P., LCKELT, M., NDIAYE, A., PFLEGER, N., AND KLESEN, M. 2006. Virtualhuman - dialogic and emotional interaction with virtual characters. In *In: Proc. of the Eighth International Conference on Multimodal Interfaces (ICMI 2006)*.

RICKEL, J., AND JOHNSON, W. 1999. Animated agents for procedural training in virtual reality: perception, cognition, and motor control. *Applied Artificial Intelligence 13*, 343–382.

SHULMAN, L. 1986. Pedagogical content knowledge (PCK).

SINGER, M., AND GOLDIN-MEADOW, S. 2005. Children learn when their teacher's gestures and speech differ. *Psychological Science 16(2)*, 85–89.

THEONAS, G., HOBBS, D., AND RIGAS, D. 2008. Employing virtual lecturers facial expressions in virtual educational environments. *In: Intern. J. Virtual Real. 7*, 31–44.

VALENZENO, L., ALIBALI, M., AND KLATZKY, R. 2003. Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology 28(2)*, 187–204.

WILLIAMS, R. G., AND WARE, J. E. 1977. An extended visit with dr. fox: Validity of student satisfaction with instruction ratings after repeated exposures to a lecturer. *American Educational Research Journal 14*, 449–457.