Evaluating the Impact of Interface Agents in an Intelligent Tutoring Systems Authoring Tool

Maria Moundridou

DEPARTMENT OF INFORMATICS
UNIVERSITY OF PIRAEUS
80, Karaoli & Dimitriou Street
Piraeus, Greece
mariam@unipi.gr

Maria Virvou

DEPARTMENT OF INFORMATICS
UNIVERSITY OF PIRAEUS
80, Karaoli & Dimitriou Street
Piraeus, Greece
mvirvou@unipi.gr

SUMMARY

This paper describes the evaluation of a speech-driven cartoon agent that has been embodied in the interface of a Web-based Intelligent Tutoring Systems (ITSs) authoring tool. The agent is mainly used in the student interface of the ITSs that are produced by the authoring tool and represents either a student or an instructor. The agent was evaluated during an experiment in terms of the effect that it could have on students' interaction with the generated ITSs. The participants in the experiment were divided into three groups: one third of them worked with a version of the ITS in which the agent was supposed to be an instructor, another third worked with a version of the ITS in which the agent was acting as a co-student and the rest worked with an agent-less version. The results from this study seem to confirm the hypothesis that a pedagogical agent incorporated in an ITS can improve students' experience. On the other hand, the hypothesis that the presence of the agent significantly enhances learning was rejected based on our data. Comparing the two agent versions of the co-student and instructor, the most important finding was that students seem to feel more comfortable working with a virtual co-student than with a virtual instructor.

KEYWORDS : animated pedagogical agents, evaluation, intelligent tutoring system, authoring tool, student interface.

INTRODUCTION

Over the last years a new paradigm has emerged in the field of computer-based learning concerning the incorporation of animated pedagogical agents in learning environments (e.g. [4], [8]). These agents are characters embodied in learning environments and exhibiting life-like behaviour through speech, facial expressions, gestures and body movements. Empirical investigations that have been conducted to measure the effect of animated agents on students' learning and motivation, although in some aspects controversial, have shown encouraging results towards using animated pedagogical agents.

Here we will report on WEAR, a Web-based authoring tool for Intelligent Tutoring Systems, which embodies an animated talking cartoon in its student interface. This talking cartoon of WEAR represents either a virtual instructor or a virtual co-student. In an initial evaluation study that we conducted concerning WEAR's interface with the students, three groups of student interacted with different versions of the system: the first group worked with an agent-less version, the second with a version in which the talking cartoon represented the instructor and the third group worked with a version in which the talking cartoon was supposed to be a co-student. In this paper we will report on the results of this evaluation study.

In the next section we will briefly discuss animated pedagogical agents. We will then report on our system and the agent incorporated in it. Finally, we will describe the evaluation study we conducted, discuss its results and draw some conclusions.

ANIMATED INTERFACE AGENTS IN LEARNING ENVIRONMENTS

Animated pedagogical agents [11] are a special case of animated interface agents aiming at improving students' interaction with a learning environment by exhibiting life-like behaviour, e.g. speaking, expressing emotions, making head or body movements, etc. They are presented as three-dimensional graphics, real video or two-dimensional cartoon-style drawings. Their responsibilities range from acknowledging a student's action to providing help in a problem-solving situation.

Several educational systems exist which embody animated pedagogical agents. For example, Adele which stands for Agent for Distance Learning: Light Edition [10], operates in a Web-based distributed simulation environment, where she guides and assesses students as they work through clinical cases. Adele consists of two components: the animated persona which is a Java applet that runs in a separate window, and the reasoning engine which monitors the student's actions and decides how Adele should respond to each of these actions. Another animated pedagogical agent is Herman the Bug [3].

Herman the Bug inhabits Design-A-Plant, a learning environment for the domain of botanical anatomy and physiology. Herman observes students' actions as they build plants that can thrive in a given set of environmental conditions and provides explanations and hints. In the process of explaining and hinting, Herman performs various actions, such as walking, flying, swimming, teleporting, etc. PPP Persona [9] is an animated interface agent that presents multimedia material to the user. While the user views the presentation, the agent can comment on particular parts and highlight them through pointing gestures. The repertoire of the persona's presentation gestures includes gestures expressing approval or disapproval, warning or recommendation, etc.

Several studies have investigated the effect of animated interface agents on human-computer interaction. However, these studies have resulted in conclusions that provide a diversity of opinions concerning the effect of animated interface agents. Most of them seem to agree that animated agents render the educational applications more pleasant and thus more attractive. However, they do not agree as to whether these agents improve the educational benefits of the applications or not. Indeed, Dehn & van Mulken in [1] provide a comprehensive and systematic overview of the empirical studies conducted so far on this subject and notice that empirical investigations on the effect of animated agents are still small in number and differ with regard to the measured effects. The authors also argue that the possible effects of animated interface agents on users can and should be classified in: (i) effects on the user's subjective experience of the system, (ii) effects on the user's behaviour while interacting with the system and (iii) effects on the outcome of the interaction as indicated by performance data.

Examples of empirical studies that evaluate the presence of animated interface agents include studies by Walker et al. [17], Lester et al. [2] and van Mulken et al. [13]. Walker et al. investigated subjects' responses to a synthesised talking head displayed on a computer screen in the context of a questionnaire study. Their findings showed that compared to subjects who answered questions presented via text display on a screen, subjects who answered the same questions spoken by a talking head spent more time, made fewer mistakes, and wrote more comments. The study of Lester et al. with different versions of Herman the Bug revealed the persona effect, which is that "the presence of a lifelike character in an interactive learning environment – even one that is not expressive – can have a strong positive effect on student's perception of the learning experience". Van Mulken et al. performed an empirical study to examine the effect of PPP Persona both on subjective and objective measures. The results of their study indicate that the presence of the agent has neither a positive nor a negative effect on comprehension and recall performance. However, even the mere presence of the Persona causes presentations and tests to be experienced by users as more entertaining and less difficult.

BRIEF DESCRIPTION OF THE SYSTEM

WEAR (WEb-based authoring tool for Algebra Related domains) is a Web-based authoring tool for the construction of Intelligent Tutoring Systems ([14],[15]). WEAR incorporates knowledge about the construction of exercises and a mechanism for student error diagnosis that is applicable to many domains that can be "described" by algebraic equations. Such domains could be chemistry, economics, physics etc.

In particular, the tool takes input from a human instructor about a specific equation-related domain (e.g. physics). This input consists of knowledge about variables, units of measure, formulae and their relation. When the human instructor wishes to create exercises s/he is guided by the system through a step by step procedure. At each step of this procedure the instructor specifies values for some parameters needed to construct an exercise. Such parameters could be for example what is given and what is asked in the exercise to be constructed. After the completion of this procedure the tool constructs the full problem text and provides consistency checks that help the instructor verify its completeness and correctness.

WEAR also allows the authoring of electronic textbooks by instructors and delivers them over the WWW to learners [6]. These textbooks offer navigation support to students, adapted to their individual needs and knowledge. WEAR associates with each student a level of knowledge according to his/her past performance in solving problems with the tool. The tool suggests each student to try the problems and study the topics corresponding to his/her level of knowledge. When a student attempts to solve an exercise the system provides an environment where the student gives the solution step by step (Figure 1). The system compares the student's solution to its own. The system's solution is generated by the domain knowledge about algebraic equations and about the specific domain in which the exercise belongs (e.g. economics). While the student is in the process of solving the exercise the system monitors his/her actions. If the student makes a mistake, the diagnostic component of the system will attempt to diagnose the cause of it.

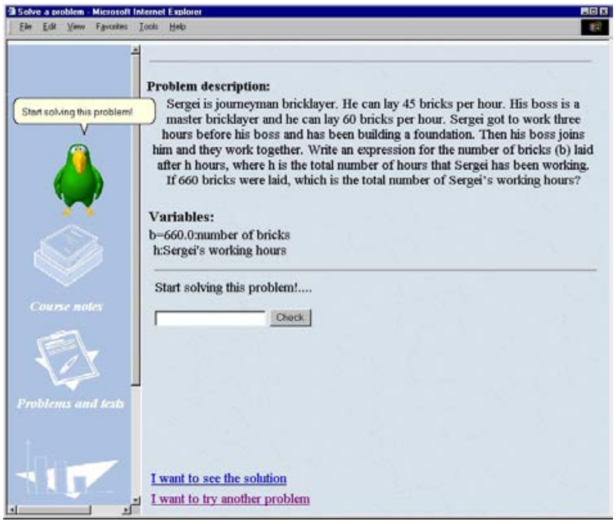

Figure 1: Solving a problem while in student's mode.

THE ANIMATED INTERFACE AGENT

WEAR used to use a talking head in its interface with the students [16]. This talking head component used to use speech synthesis to automatically produce speech output from text using MBROLA, a freely available speech synthesiser. However, in a subsequent version of WEAR after we conducted an empirical study concerning the type of animated agent that we should use, we replaced the anthropomorphic agent with a cartoon character. The empirical study that we conducted was among students of various ages and backgrounds. Among other questions we also asked them what type of animated agent they would prefer in an interface. The majority of students seemed to have been in favour of a cartoon character rather than an anthropomorphic one. The main reasons for this was that they considered cartoon characters added more sense of humour to the application, they were more pleasant and cute and less stressful. As a consequence the new speech-driven animated agent was Microsoft's Peedy the parrot [5].

This speech-driven agent is responsible for guiding the student to the environment and providing feedback messages to her/him while s/he is solving problems or studying a section of the electronic textbook. The animated agent renders the interface quite attractive to students through the sound of speech. Moreover, since WEAR is an authoring tool for ITSs there are a lot messages that are dynamically formed during the execution of the application. Therefore, the authoring tool could not use pre-stored material for the speech feature but rather a speech synthesiser.

When the student begins a new session of interaction with the system, s/he is requested to choose whether s/he wishes to work with a simulated "co-student" or with the "instructor". These two different choices are restricted to the level of the user interface. This means that they use the same underlying reasoning abilities from the diagnostic component of WEAR.

If the student selects the mode of the "co-student" then the talking cartoon provides very friendly messages as a peer to the student. This simulated student is responsible for providing positive feedback when the student's actions are correct and for pointing out the student's underlying misconception in case of an erroneous action. The information concerning which actions are considered correct or not and also the messages that the simulated student should say are provided by the diagnostic component of WEAR. The animated agent as a simulated student is aimed at increasing the student's attention and possibly collaboration attitude (although in a limited form). Indeed as VanLehn et al. have pointed out in [12], peer learning even in a setting where the other peer is a simulated student may increase students' collaboration skills.

If the student selects the mode of the "instructor" then the talking cartoon provides messages similar to the ones provided by the "co-student" but they are more formal and the diagnosis of misconceptions goes one step further to resolve ambiguities using the long-term student model. Indeed, there may be cases where a student's erroneous action may be attributed to more than one misconception. In such cases, the diagnostic component first consults the long-term student model and then through the animated agent asks the student a question to determine his/her underlying misconception and resolve the ambiguity. The benefits of directly asking the student are twofold: firstly, the system may find out the real reason for the erroneous action and provide appropriate feedback and secondly, the student through explaining why s/he acted in that way gains more knowledge and understanding. It is a common finding by many researchers that explaining things either to oneself or to another student helps one's understanding ([18], [7]).

EVALUATION STUDY

The aims of the experiment we conducted were twofold: To examine if the presence of the speech-driven cartoon agent would affect students' experience, behaviour and performance and to examine if there were differences on these aspects when the agent represented an instructor and when it represented a co-student.

Agent versus Non-Agent ITS

The participants in the first part of the experiment were 19 college students from the University of Piraeus. Twelve of them were studying Informatics and the rest Economics. The students were randomly divided into two groups: the Agent group (group SA) consisting of 10 subjects and the Non-Agent group (group NA) with 9 subjects. Both groups participated in the experiment in the same way. The difference between them was the version of the ITS they interacted with: the version used by group SA embodied the animated cartoon (as a "co-student"), whereas group NA worked with an agent-less version. To be able to explore the agent's effect on students' learning, the information that both versions of the ITS passed to students (feedback messages, guidance to the environment, etc.) was completely identical.

To generate the ITS that we used as testbed, we fed WEAR with several math problems, similar to the ones contained in a GMAT test. An example of such a problem is the following: "Sergei is a journeyman bricklayer. He can lay 45 bricks per hour. His boss is a master bricklayer and he can lay 60 bricks per hour. Sergei got to work three hours before his boss and has been building a foundation. Then his boss joins him and they work together. Write an expression for the number of bricks (b) laid after h hours, where h is the total number of hours that Sergei has been working. If 660 bricks were laid, which is the total number of Sergei's working hours?" (This problem is based on a problem we found at http://www.algebratutor.org/big_list.html).

The experiment started with a pre-test, consisting of five problems. After completing the pre-test, the students worked with the system to solve similar (as of their difficulty) problems. While working with the system, the students' actions were logged. The post-test that followed consisted again of five similar problems. Students were finally asked to complete a questionnaire concerning their experience from interacting with the system.

Students' responses to the questionnaire were used to measure their attitude towards the system. The corresponding questions were the following:

1. Did you enjoy working with the system?

2. Was the system easy to use?

3. Were the problems that you were asked to solve with the system difficult?

4. Did the system help you to improve your problem-solving skills?

Students' answers to these questions were scored on a five point Likert scale.

In all questions the ratings that group SA gave are numerically higher than the ratings of group NA. However, the significant difference between the two group means concerns only the enjoyment rating (t=4.98; p=.000) and the perceived problem difficulty (t=4.05; p=.001).

As Dehn & van Mulken mention in [1], an aspect that is usually investigated in order to examine the effect of animated agents on user's behaviour is the amount of attention that the user directs towards the system or the task s/he should perform. In our experiment, to test the user's attentiveness we investigated the time students spent with the system and the number of problems that they attempted to solve. As for the time spent, group SA needed on average more time to solve each problem than group NA. However, the difference between the means of the two groups is far from being statistically significant. This is also the case with the number of attempted problems.

To test whether the presence of the agent improves or not the learning outcomes we compared data collected from the pre- and post-tests and from the interaction of students with the system. When comparing the pre-and post-tests results we came up with the following findings:

- Both groups (SA and NA) scored similarly in the pre-test and needed about the same time to complete it. This is an indication of both groups being equally capable of dealing with the tasks they should perform.

- Both groups (SA and NA) improved their scores in the post-test, as well as the time they needed to complete it. Based on this, we can assume that the tutoring system achieved its goal to improve the students' performance.

However, the pointed improvement in time and grade differed between the two groups: In average group SA spent 30% less time to complete the post-test and achieved 10% higher grade than the pre-test. In contrast to group SA, the improvement of group NA's average time and grade was 18% and 5% respectively. Although numerically higher, the improvement both in time and grade is not significantly different between groups (time improvement: t=-0.95; p=0.181, grade improvement: t=0.59; p=0.280).

As for the students' performance when solving problems with the system, this was almost identical for both groups.

Table 1 illustrates a summary of the results of this part of the experiment. The results in bold letters indicate that these particular differences were statistically significant.

Variable	**Results**
Experience	**Enjoyement(SA) > Enjoyement(NA)** Ease-of-use(SA) > Ease-of-use(NA) **Perceived-problem-difficulty (SA) < Perceived-problem-difficulty (NA)** Usefulness(SA) > Usefulness(NA)

Behaviour	Time-spent(SA) > Time-spent(NA) Attempted-problems(SA) > Attempted-problems(NA)
Learning	Time-improvement(SA) > Time-improvement(NA) Grade-improvement(SA) > Grade-improvement(NA)

Table 1: Agent as "co-student" (SA) versus Non-Agent (NA) version.

Co-student Agent versus Instructor Agent

In the second part of the experiment 9 more students worked with the agent version of the system, representing now a virtual instructor (group IA). The rest of the experiment was completely identical with the first part of it: students of group IA sat a pre-test, worked with the system, sat a post-test and completed the questionnaire.

Comparing the responses to the questionnaire of group IA and group SA we did not come up with any significant differences. However, it is worth to notice some tendencies: group IA rated the problems as more difficult than group SA but gave higher rating to the system concerning its ability to help them improve their problem-solving skills.

An interesting finding is that group IA attempted to solve the same number of problems with group SA but when facing difficulties group IA gave up easier than group SA and either requested to see the solution of the problem or went on to a new problem. This reminds real educational settings where students often hesitate to answer to their instructor feeling that their possibly erroneous response will make a bad impression to the instructor or even lead to a lower grade.

Comparing the pre- and post-test results, no significant difference was noticed between group SA and group IA concerning their improvement in time and grade. As for the students' performance when solving problems with the system, group SA did better than group IA. This was to be expected since group IA often quitted after giving a few erroneous answers.

In Table 2 a summary of the results of this part of the experiment (Co-student Agent versus Instructor Agent) is illustrated.

Variable	**Results**
Experience	Enjoyement(IA) < Enjoyement(SA) Ease-of-use(IA) = Ease-of-use(SA) Perceived-problem-difficulty(IA) > Perceived-problem-difficulty(SA) Usefulness(IA) > Usefulness(SA)
Behaviour	Time-spent(IA) < Time-spent(SA) Attempted-problems(IA) = Attempted-problems(SA)
Learning	Time-improvement(IA) = Time-improvement(SA) Grade-improvement(IA) < Grade-improvement(SA)

Table 2: Agent as "co-student" (SA) versus Agent as "instructor" (IA) version.

CONCLUSIONS

Some of the findings of our case study were positive for the existence of animated agents in the interface of ITSs. In particular, students feel more enjoyment when interacting with a system that embodies an animated interface agent that speaks the system's response messages. Moreover, the group of students working with the agent version of the system found the problems that they should solve less difficult than the other group did. This finding is also met in the study presented in [13]. However, there was also an important but not positive result. This was that the presence of the interface agent did not manage to improve significantly the learning outcomes. Furthermore, the students' attentiveness to the system was not promoted by the animated talking agent, as was indicated by not significant differences in time spent and number of attempted problems between the two groups.

Comparing two different agent versions of the system (agent being an instructor and agent being a co-student) the most interesting finding was that students interacting with a virtual instructor do not seem to feel as comfortable as these interacting with a virtual peer do. This was indicated by the students' tendency to drop out from a problem.

However, as Dehn & van Mulken point out in [1], the type of agent used, the type of information it provides and the domain in which the system is set are factors that moderate the way in which an animated interface agent influences the users' attitude or performance. Therefore, since there is a high possibility that these results might be different in different settings, the generalisation of them should only be made with caution.

BIBLIOGRAPHY

1. Dehn, D. & van Mulken, S. *The impact of animated interface agents: a review of empirical research*. International Journal of Human-Computer Studies, Vol. 52, 2000, pp. 1-22.
2. Lester, J., Converse, S., Kahler, S., Barlow, S., Stone B. & Bhogal R. The persona effect: affective impact of animated pedagogical agents. In S. Pemberton (ed.) *Proceedings of ACM Conference on Human Factors in Computing Systems CHI'97*, ACM Press, New York, 1997, pp. 359-366.

3. Lester, J., Stone, B. & Stelling, G. *Lifelike Pedagogical Agents for Mixed-Initiative Problem Solving in Constructivist Learning Environments*. User Modeling and User-Adapted Interaction, Vol. 9, No. 1-2, 1999, pp. 1-44.

4. Lester, J., Voerman, J., Towns, S. & Callaway, C. *Deictic believability: Coordinating gesture, locomotion, and speech in lifelike pedagogical agents*. Applied Artificial Intelligence, Vol. 13, 1999, pp. 383-414.

5. Microsoft Corporation , MS Agent. Accessible at http://msdn.microsoft.com/msagent/

6. Moundridou, M. & Virvou, M. Authoring and Delivering Adaptive Web-Based Textbooks Using WEAR. In *Proceedings of ICALT-2001, IEEE International Conference on Advanced Learning Technologies*, 2001, to appear.

7. Pressley, M., Wood, E., Woloshyn, V.E., Martin, V., King, A. & Menke, D. *Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning*. Educational Psychologist, Vol. 27, 1992, pp. 91-109.

8. Rickel, J. & Johnson, W. *Animated agents for procedural training in virtual reality: Perception, cognition, and motor control.* Applied Artificial Intelligence, Vol. 13, 1999, pp. 343-382.

9. Rist, T., André, E. & Müller, J. Adding Animated Presentation Agents to the Interface. In J. Moore, E. Edmonds & A. Puerta (eds.) *Proceedings of the 1997 International Conference on Intelligent User Interfaces*, ACM Press, New York, 1997, pp. 79-86.

10. Shaw, E., Ganeshan, R., Johnson, W. L. & Millar, D. Building a Case for Agent-Assisted Learning as a Catalyst for Curriculum Reform in Medical Education. In S. Lajoie & M. Vivet (eds.) *Proceedings of the 9th World Conference on Artificial Intelligence in Education - AIED'99*, Frontiers in Artificial Intelligence and Applications, IOS Press, Amsterdam, Vol. 50, 1999, pp. 509-516.

11. Stone, B. & Lester, J. Dynamically sequencing an animated pedagogical agent. In *Proceedings of the 13th National Conference on Artificial Intelligence*, 1996, pp. 424-431.

12. VanLehn, K., Ohlsson, S., & Nason, R. *Applications of simulated students: An exploration*. Journal of Artificial Intelligence and Education, Vol. 5, No. 2, 1994, pp. 135-175.

13. van Mulken, S., André, E. & Müller, J. The Persona Effect: How Substantial Is It? In H. Johnson, L. Nigay & C. Roast (eds.) *People and Computers XIII: Proceedings of HCI'98*, Springer, Berlin, 1998, pp. 53-66.

14. Virvou, M. & Moundridou, M. An authoring tool for Algebra-related domains. In H.-J. Bullinger, J. Ziegler (eds.) *Human-Computer Interaction: Communication, Cooperation, and Application Design, Proceedings of the 8th International Conference on Human-Computer Interaction - HCI International '99*, Lawrence Erlbaum Associates, Mahwah NJ, Vol. 2, 1999, pp. 647-651.

15. Virvou, M. & Moundridou, M. *A Web-Based Authoring Tool for Algebra-Related Intelligent Tutoring Systems*. Educational Technology & Society, Vol. 3, No. 2, 2000, pp. 61-70.

16. Virvou, M., Sgouros, N., Moundridou, M. & Manargias, D. Using a speech-driven, anthropomorphic agent in the interface of a WWW educational application. In J. Bourdeau & R. Heller (eds.) *Proceedings of ED-MEDIA 2000, World Conference on Educational Multimedia, Hypermedia & Telecommunications*, AACE, Charlottesville VA, 2000, pp. 1724-1726.

17. Walker, J.H., Sproull, L. & Subramani, R. Using a Human Face in an Interface. In B. Adelson, S. Dumais & J. Olson (eds.) *Proceedings of ACM Conference on Human Factors in Computing Systems CHI'94*, ACM Press, New York, 1994, pp. 85-91.

18. Webb, N. *Peer interaction and learning in small groups*. International Journal of Educational Research, Vol. 13, 1989, pp. 21-40.