Effective Pedagogical Agent Behaviour

Xiaojie Zha

Beijing University of Posts and Telecommunication
Xitucheng Road, Haidian District, Beijing, China 100876
charleszha@bupt.edu.cn

Marie-Luce Bourguet

Queen Mary University of London
Mile End Road, London, UK E1 4NS
marie-luce.bourguet@qmul.ac.uk

This paper describes a small experimental study where pedagogical agents (avatars) are added to voice-over-slide learning materials. Four different agent behaviours are tested: Avatar A performs all the upper-body gestures of the lecturer; Avatar B is animated using few random gestures in order to create a natural presence, but one that is unrelated to the speech; Avatar C only performs the lecturer's pointing gestures; finally, Avatar D performs "lecturer-like" gestures, but these are desynchronised with the speech. Preliminary results indicate that Avatar C is the most effective at facilitating learning. An agent that displays a more active behaviour, even though the behaviour is an exact representation of the lecturer's behaviour (such as Avatar A), is distractive and does not support learning as much.

Learning Materials, Video Lecture, Avatar, Gestures, 3D Pointing.

1. INTRODUCTION

An alternative to full lecture recording that limits the size of the data is voice recording over slides (i.e. the lecturer's speech is played together with the accompanying lecture's slides). Voice-over-slide lecture material preserves many of the advantages of a recorded lecture, except that one important component is missing: the lecturer. Indeed, it has been shown that lecturers' behaviour, in particular their arm and hand gestures, can effectively reflect their pedagogical intentions and participate in the efficient delivery of the lecture's message [3][5]. Pointing gestures, in particular, have been linked to pedagogical significance. Therefore, the goal of the work presented in this paper is to augment voice-over-slide presentations with 3D avatars that can effectively replace the lecturers' presence. Such pedagogical agents have been shown to be effective at supporting learning [2][4], however, our contribution here is to elicit a better understanding of the type of avatar's behaviour (upper-body movements) that best fosters students' satisfaction and understanding of a lecture.

The approach we have taken is to use a low-cost depth camera to capture the lecturer's behaviour during a lecture and to detect their pointing gestures. We then use the data to generate various versions of an avatar, which represents the lecturer and is added to voice-over-slide lecture material. Finally, an experiment is conducted to find out which avatar's behaviour is best received by the audience and contribute most to a smooth and effective lecture delivery.

2. BUILDING THE PEDAGOGICAL AGENTS

Four lectures from three different lecturers were recorded using both a digital video camera and a depth camera (a Kinect sensor). The Kinect avatar controller from the Kinect SDK Unity3D package [1] was then used to generate avatars that move according to the joint positions of the lecturers. Images generated from PowerPoint lecture slides were also added to the 3D scene so they can be played synchronously with the lecturer's recorded speech (see Figure 1).

Several versions of the avatar, displaying different behaviours, were generated in order to experiment with the effect of their behaviour on the students. One avatar ("Avatar A") was generated that reproduces exactly the lecturer's behaviour, i.e. Avatar A's upper-body movements are completely controlled from the Kinect data and are an exact representation of the lecturer's upper-body movements. They are also synchronised with the lecturer's speech.

Avatar B displays gestures and movements that are generated from a Unity3D animation library. It was designed to use low amplitude movements, giving an impression of natural and relatively quiet presence. No Kinect data is used to generate Avatar B. Avatar B has no relation with the lecturer's behaviour nor the speech.

Avatar C performs only the pointing gestures that are detected from the recorded lecturer's Kinect data, and these are synchronised with the speech. It remains static (natural standing posture) the rest

of the time (see Figure 1). Avatar C is meant to display only the gestures that are indicating pedagogical significance in the accompanying lecture speech.

Figure 1: Avatar C standing idle (left image) and then pointing towards the slide (right image)

Finally, Avatar D displays gestures that have been generated from Kinect data, but these are played at random times, and thus are desynchronised with the speech (we used the Kinect recordings to create our own animation library).

3. EXPERIMENTAL STUDY

The purpose of the experimental study is to test the four different avatar behaviours on a student audience. Twenty Engineering students from the same cohort of the Multimedia programme at the Beijing University of Posts and Telecommunications participated in the experiment (4 women and 16 men). All participants shared very similar background. Four one-minute video clips were extracted from two recorded lectures. Each of the four clips was then used to generate voice-over-slide videos, which were then augmented with an avatar, making a total of 16 different avatar / lecture clip combination conditions.

The participants were divided into four groups, each made of 4 male and 1 female students. Each group was given to watch 4 of the 16 short videos and we made sure that each group saw the 4 avatars as well as the 4 lecture clips. All 16 conditions were used. The order in which the different avatars were presented varied from one group to another. After watching a video, the participants were required to rate their understanding of the lecture and comment on the avatar behaviour. Towards the end of the experiment, they were asked to rank the four videos in their order of preference and provide explanations.

The experimental results show that Avatars A and C are generally preferred over Avatars B and D. 16 of the 20 participants ranked Avatar A as their first or second choice when ranking the avatars by order of preference. 14 participants ranked Avatar C as their most or second most preferred avatar.

The participants commented that Avatar A displays the most realistic behaviour, which was the main reason for their preference. However, because it is constantly moving (apparently lecturers never

stand idle!), it could also be distracting. For Avatar C, some participants qualified it as boring (it is idle a lot of the time), but overall, the feeling was that it was useful. Conversely, Avatar B (with random but relatively quiet behaviour) was deemed unnecessary. The students preferred no avatar to the presence of Avatar B. Finally, Avatar D was found very confusing and preventing them from focusing on the content of the lecture.

Table 1 shows the “degree of understanding” the students acquired of the lecture’s content, depending on the avatar. The participants were asked to judge their understanding of the short lecture by attributing a mark ranging from 0 to 5.

	A	B	C	D
Means	3.3	3.5	3.65	3.35
Standard Deviation	1.2689	1.2042	1.2757	0.9631

Table 1: Degree of understanding on a scale up to 5, for videos displaying Avatar A, B, C or D

Avatar C seems to facilitate understanding compared to the other avatars. A bit more surprisingly, avatar A, who is an exact representation of the lecturer’s behaviour is the least effective at supporting learning. Students mentioned that it was “moving too much”. More experimentation is needed to confirm these results as the number of participants is not large enough to enable us to judge the significance of these results. However, they seem to indicate that an avatar that displays either the exact behaviour of the lecturer (A), or a behaviour that is “lecturer-like” (D), is not as effective as an avatar that only displays the lecturer’s pointing gestures (B). They also seem to indicate that an avatar with “quiet” behaviour (B and C) is more supportive than an avatar with a more active but distracting behaviour (A and D).

4. CONCLUSION

Preliminary experimental results indicate that an avatar, which behaves quietly and only performs the lecturer’s gestures that indicate pedagogical significance is the most effective at facilitating learning. Conversely, an avatar that displays a more active behaviour, even though the behaviour is an exact representation of the lecturer’s behaviour, is distractive and does not support learning as much. In the future, larger experiments will be conducted to confirm these results. Future work also include improving the behaviour of Avatar B by making the pointing gesture directed to a precise location on the slide. Finally, in order to become independent from the availability of the Kinect data, the lecturer’s speech will have to be analysed to find the significant parts of a lecture

and then generate the avatar's pointing gestures accordingly, i.e. in synchrony with the speech.

5. REFERENCES

- [1] Kinney, P. 2011. Kinect Wrapper Package for Unity3D [online]. Available: http://wiki.etc.cmu.edu/unity3d/index.php/Microsoft_Kinect_-_Microsoft_SDK#Integrating_with_Unity [Accessed 2014-4-14].
- [2] Rickel, J. & Johnson, W. L. 1997. Integrating pedagogical capabilities in a virtual environment agent. In *Proceedings of the first international conference on Autonomous agents*. ACM. DOI= <http://doi.acm.org/10.1145/267658.267664>
- [3] Tian, Y. & Bourguet, M.L. 2016. Lecturers' Hand Gestures as Clues to Detect Pedagogical Significance in Video Lectures. In *Proceedings of European Conference on Cognitive Ergonomics*, ECCE 2016, Nottingham, UK
- [4] Towns, S. G., Callaway, C. B., Voerman, J. L. & Lester, J.C. 1998. Coherent Gestures, Locomotion, and Speech in Life-Like Pedagogical Agents. In *Proceedings of the 3rd international conference on Intelligent user interfaces*. NY, 13-20. DOI= <http://doi.acm.org/10.1145/268389.268393>
- [5] Zhang, J. R. 2012. Upper body gestures in lecture videos: indexing and correlating to pedagogical Significance. In *Proceedings of the 20th ACM international conference on Multimedia*. 1389-1392. ACM. DOI= <http://doi.acm.org/10.1145/2393347.2396499>