Shared virtual training in an affect-sensitive environment

Lori Malatesta
Image, video and multimedia systems lab
National Technical University of Athens
lori@image.ntua.gr

Amaryllis Raouzaiou
Image, video and multimedia systems lab
National Technical University of Athens
araouz@image.ntua.gr

Kostas Karpouzis
Image, video and multimedia systems lab
National Technical University of Athens
kkarpou@image.ntua.gr

Stefanos Kollias
Image, video and multimedia systems lab
National Technical University of Athens
stefanos@cs.ntua.gr

CHI 2008, April 5 – April 10, 2008, Florence, Italy
This paper is part of the 03ED853 research project, implemented within the framework of the "Reinforcement Programme of Human Research Manpower" (PENED) and co-financed by National and Community Funds (25% from the Greek Ministry of Development-General Secretariat of Research and Technology and 75% from E.U.-European Social Fund).

Abstract
In this paper we present our work on an affect-sensitive training interface that can be used by multiple users in different locations simultaneously. It allows for two or more users to share a virtual jogging experience while each one is in front of his/her personal computer. The system is sensitive to specific affective parameters of the user's movements and adapts environment variables accordingly.

Keywords
Affective computing, exertion interfaces, motion capture, computer vision.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): H.5.2 User Interfaces -- Input devices and strategies. I.4 Image processing and computer vision

Introduction
The application of affective computing principles in human computer interfaces provides endless options. In the current work we focus on integrating affective state recognition techniques with a virtual sports-training interface. Motivation for this approach arose from the idea to enhance exercise experience both by tailoring it to the user's current affective state and by

providing the means to share the experience with someone in a remote location. Motion capture techniques as well as computer vision algorithms allow for detailed monitoring of human movement with either limited or no sensor intrusion. Nevertheless in some cases detailed and sophisticated tracking of movement is not necessary whereas qualitative information about what the movement *expresses* makes more sense. Such is the case of our shared virtual training environment. The novelty of the idea resides in the fact that the users' movements/ gestures will be tracked in a qualitative way. The results of this tracking process will determine both the manipulation of the user's avatar as well as the conditions of the virtual environment.

Expressivity Parameters

Expressivity of behavior is an integral part of the communication process as it can provide information on the current emotional state, mood, and personality of a person [6]. Many researchers have investigated human motion characteristics and encoded them into dual categories such as slow/fast, small/expansive, weak/energetic, small/large, unpleasant/pleasant. To model expressivity, in our work currently in press [1], we use the six dimensions of behavior described in [2], as a more accomplished way to describe the expressivity, since this approach tackles all the parameters of expression of emotion. Five parameters modeling behavior expressivity have been defined at the analysis level: (1) Overall activation, (2) Spatial extent, (3) Temporal, (4) Fluidity, (5) Power.

Overall activation is considered as the quantity of movement during a conversational turn. In our case it is computed as the sum of the motion vectors' norm. Spatial extent is modeled by expanding or condensing the entire space in front of the user that is used for gesturing and is calculated as the maximum Euclidean distance of the position of the two hands. The temporal expressivity parameter of the gesture signifies the duration of the movement while the speed expressivity parameter refers to the arm movement during the gesture's stroke phase (e.g., quick versus sustained actions). Fluidity differentiates smooth/graceful movements from sudden/jerky ones. Power actually is identical with the first derivative of the motion vectors calculated in the first steps.

Virtual Training Environment

The virtual training environment is comprised of a virtual world developed using a game engine platform [5]. The user can navigate the environment with the restriction of being 'forced' to follow a specific path from which there is no option to deviate. Thus, the only control the user has on his/her movement is his/her speed. This parameter is controlled by tracking the user's movement and gestures while the user is running 'in place' in front of his/her PC.

At the same time as a result of the tracking process the application continuously calculates values for all user expressivity parameters over time and makes adjustments in environment variables in order to be consistent with the recognized affective state. Background music is changed to a more mellow version when the user's movements are found more fluid and power values measured are lower or it becomes more up-beat if the opposite requirements are met. Changes also take place in the sky and weather conditions by dynamically relighting the scene as necessary in order to make it 'match' the user's current mood. This mapping of the participant's expressivity onto the environment is easily defined programmatically. The rules chosen for the mapping process (i.e. high power + overall activation leads to a cloudy sky) are fairly intuitive/ arbitrary at the moment and deem a formal evaluation. The application allows for multiple users to access the same training environment from remote locations. Each user can see other users moving along the path with their corresponding speed.

User Tracking Methods

The training environment provides two tracking options for potential users. One is fully based on computer vision techniques and relies on a simple webcam for data collection while the second tracking method is based on motion capture using the Wiimote [4], a state of the art wireless controller with a motion sensing capability through the use of accelerometer and optical sensor technology. More information on the image processing algorithms used in the first method can be found in [1][3]. Both methods produce the required data output for the calculation of the specific expressivity parameters presented in the following section. The collected data is used to calculate expressivity parameters' values. These values in turn determine environment conditions in our virtual training environment.

Evaluation

The work presented here is yet to be formally evaluated. The computer vision tracking interface has already been displayed in the 2nd International Conference on Affective Computing and Intelligent Interaction 2007 as way to control characters in a game environment. We plan to conduct an experiment in order to collect feedback from participants using this virtual training environment and in order to back currently arbitrary choices. Choices to be questioned are the ones regarding way the user's affective state can influence the virtual environment (i.e. darker sky when the power of user's movements is high) as well as how users perceive other users and activity sharing. What would make sense/ be perceived as rendering the experience more pleasant/ engaging?

Future Work

At the moment, users participate in the virtual training environment by running in place. It is in our immediate future plans to integrate either treadmill or stepper equipment in order to render the training experience more realistic as well more effective exercise-wise. Such a solution would provide more precise information on the actual speed of the participant and thus would allow for a more just competition between users.

References

[1] Caridakis, G., Raouzaiou, A., Bevacqua E., Mancini, M., Karpouzis, K., Malatesta, L., Pelachaud, C. (in press) Virtual agent multimodal mimicry of humans. Journal on Language Resources and Evaluation.

[2] Hartmann, B., Mancini, M., and Pelachaud, C. (2005a). Implementing expressive gesture synthesis for embodied conversational agents. In Gesture Workshop, Vannes.

[3] Malatesta, L., Caridakis, G., Raouzaiou, A., Karpouzis, K. Gesture-Based Character Control. Demo in ACII07 proceedings.

[4] Nintendo Wii http://wii.nintendo.com/controller.jsp

[5] Torque Game Engine http://www.garagegames.com

[6] Wallbott, H. G. and Scherer, K. R. (1986). Cues and channels in emotion recognition. Journal of Personality and Social Psychology, 51(4)