Enhancing Virtual-Environment-Based Teamwork Training with Non-Verbal Communication

Stefan Marks
Graphics Group, Department of Computer Science, Faculty of Science
The University of Auckland, New Zealand
smar189@aucklanduni.ac.nz

John Windsor
Faculty of Medical and Health Sciences, Department of Surgery, Advanced Clinical Skills Center,
The University of Auckland, New Zealand
j.windsor@auckland.ac.nz

Burkhard Wünsche
Graphics Group, Department of Computer Science, Faculty of Science
The University of Auckland, New Zealand
burkhard@cs.auckland.ac.nz

Abstract: Virtual reality simulations for individual training of surgical skills are increasingly used in medical education and have been shown to improve patient outcome. Since recent research suggests that a large percentage of mistakes in clinical settings are due to problems with non-technical skills like communication, teamwork training simulators are developed and used to address this problem.

Virtual-environment-based teamwork training simulators are very cost-efficient and allow for non-co-located settings, but have their limitations in communication among the participants. We present an inexpensive camera-based system for capturing aspects of non-verbal communication of participating users and projecting these onto the avatars in the simulation. This additional information has the potential of increasing the realism of the simulation and the effectiveness of team communication, resulting in a better training outcome – for all kinds of simulation that involves human communication.

Introduction

In 1999, the Committee on Quality of Health Care in America released a report estimating that 98,000 people die each year because of medical errors occurring in hospitals (Kohn et al., 1999). This report caused massive changes to the medical education system, including the increased use of medical simulation for training and assessment. As a result, the market now offers a variety of simulators, ranging from inexpensive bench models made of rubber for training suturing and other basic technical skills to expensive mannequins that can be used for the training of emergency response teams.

With realism improving through increasing computational and graphics power, computer-based tools have also secured their position among medical simulators (Scott et al., 2008). Numerous research projects have contributed to mathematical models for the realistic simulation and manipulation of soft tissue. Based on these models and increasingly realistic rendering capabilities of modern graphics cards, commercially available products offer training opportunities in areas like endoscopic procedures (Undre and Darzi, 2007), obstetrics, cardiology, or other related fields.

However, regardless of the technical features, these computer-based simulators only serve the purpose of training technical skills of a single person at a time. Surgery and other medical procedures are usually performed as a team and thus also require non-technical skills like communication, teamwork, leadership, and decision-making (Yule et al., 2006). Research indicates that failure in communication occurs in 30% of team exchanges and that one third to 70% of these failures result in situations putting the patient's life at risk (Guise and Segel, 2008, Lingard et al., 2004). Therefore, to increase patient safety, training, and improvement of communication among team members has to be an important aspect of clinical simulation.

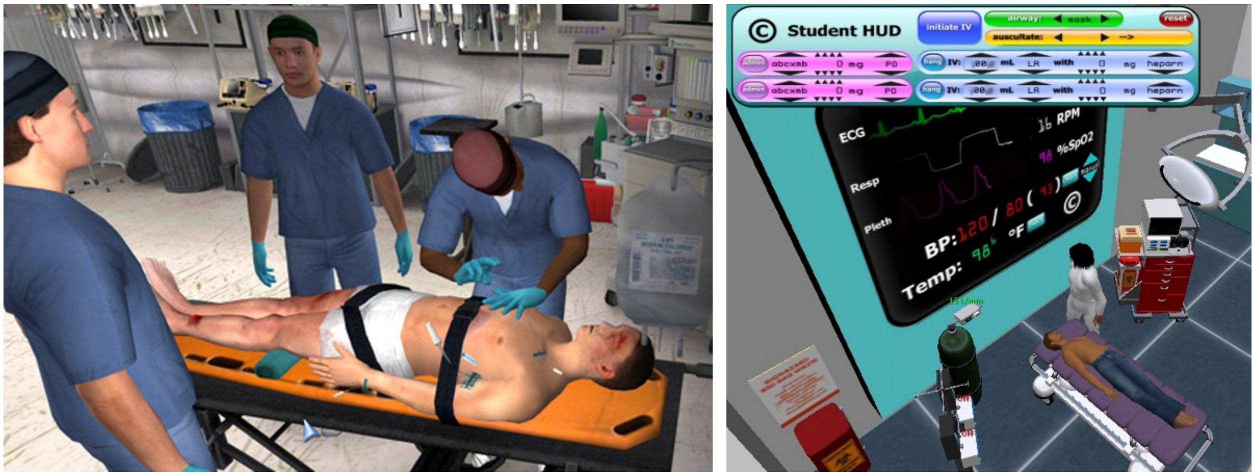

Figure 1: Virtual Environments used for training of medical procedures. *Left side:* 3DiTeams (Source: Virtual Heroes, 2008), *Right side:* Emergency response training room in Second Life.

Related Work

Computer-based *teamwork* training simulators have been used by the military and the aviation industry for years (Hussain et al., 2007). Only recently, this kind of simulation has found its way into the domains of healthcare and medicine.

3DiTeams (Taekman et al., 2007, Fig. 1, left side) simulates a military operating theatre, including all necessary devices, medical instruments, and the patient. In the virtual environment (VE), team members are represented by avatars and can interact with each other, the devices, and the patient. Each participant uses an individual client computer that is connected to a central server by a network. This server receives all actions and movements of the users from the clients, runs the simulation of physical objects and the patient's physiology, and synchronises the clients with updated information.

The simulation uses a game engine as underlying framework. Such an engine provides a well-tested and stable base capable of realistic graphics, animation, sound, physics, and networking support (Marks et al., 2007). The principle of exchanging the content of a computer game by content with educational or training purpose is referred to as "Serious Gaming" (Serious Games Initiative, 2007).

With the focus on teamwork and communication aspects, it is important to examine the amount of communication channels that a simulation provides for all participating users (Tab. 1). Modern game engines support verbal communication via microphones and headsets or speakers or predefined textual commands that are accessible by pressing certain keys.

Despite the advances in technology, the support for other communication channels is still limited. Therefore, communication and interaction within VEs still appears unnatural and inefficient (Garau et al., 2001, Vertegaal et al., 2001). Without gaze direction and head orientation, for example, communication suffers from a 50% decrease in deictic references to persons, like "you" or "him/her" (Vertegaal et al., 2000). The lack of non-verbal communication channels has to be compensated by other channels, for example, by replacing deictic references to objects and persons by explicitly saying their names and positions (Fig 2, left side).

By introducing head orientation and gaze direction, users can simply look at objects or other users instead of referring to them by voice (Fig. 2, right side). Advanced animation capabilities of modern game engines or VEs allow for detailed facial expressions like joy, anger, or fear, or gestures like pointing or shrugging shoulders. A description of how those facial expressions and/or gestures could be controlled by using explicit user interfaces can be found in (Slater et al., 2000).

We propose the use of an inexpensive camera (for example, a webcam) to acquire the necessary data in real-time. This has several advantages. Manual control of the aspects of non-verbal communication is only possible if the user is aware of them, which is not necessarily the case. The camera requires no control from the user and can capture conscious and unconscious elements. In addition, the user can completely concentrate on the simulation content and is not distracted by the additional controls. A second advantage is the temporal immediacy of the captured data. Momentarily raised eyebrows during an emphasis in a spoken sentence can be perceived by all others users at the

same time. If this optical clue would follow the verbal clue with a delay, for example, when using manual control, it would degrade or even counteract the purpose of the gesture.
We present a method for the enhancement of VE based teamwork simulations with aspects of non-verbal communication that can be captured simply by using an inexpensive camera.

Figure 2: *Left side:* Without eye gaze, it is difficult to refer to other users or objects in the simulation. Verbal communication has to become more complicated in structure and content to compensate for the lack of non-verbal communication channels. *Right side:* When eye gaze is included, the use of deictic references (for example, “you” or “that”) simplifies the communication.

Component	*Reality*	*Virtual Environment*
Occulesics (gaze direction, duration, focus)	+	+
Language		
– Text based chat	-	+
– Sound effects	-	+
– Speech, paralanguage (tone, pitch, etc.)	+	+
Facial expression		
– Facial Action Coding System	+	+
– Emotion (anger, joy, etc.)	+	+
Spatial Behaviour (orientation, proximity, etc.)	+	+
Kinesics (head movement, gestures, etc.)	+	+
Physical appearance (skin, hair, clothes, etc.)	+	+
Physical contact/Haptics	+	-
Olfactics (scent, odour)	+	-
Environment (using/exchanging artefacts, etc.)	+	+

Table 1: Aspects of communication and their appearance in reality (“+”: used naturally, “-”: could be used, but unnatural) and technical feasibility in virtual environments (“+”: available using standard equipment,”-”: requires special hardware and/or software).

Methodology

Figure 3 depicts the schematic design of our system. The user participating in the simulation is captured by a webcam, mounted close to the monitor. A user monitor application detects the user's face in the video and calculates parameters like head orientation, gaze direction, and facial expression. Via an extension, called plug-in, the application delivers this data to the simulation client, in our case a game engine. The client sends this information to the simulation server together with other data, for example, mouse movement, or pressed keys. The server receives the data from all clients, applies it to the simulation, and in turn synchronises all clients with the updated information, including the aspects of non-verbal communication. The clients receive these updates and display it in form of changed positions of objects, a changed patient state, changed gaze directions and facial expressions of the avatars.

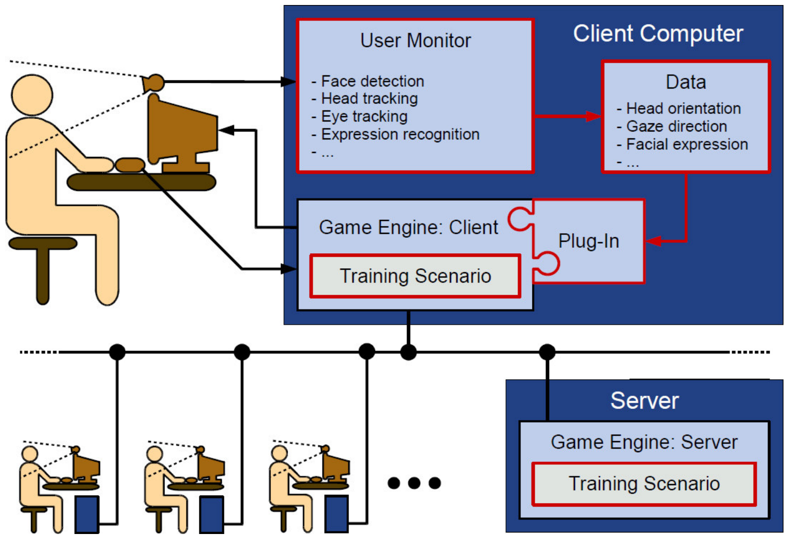

Figure 3: Functional blocks of our framework. Red borders indicate the parts that we have developed.

Based on the Rich Interaction Model (Manninen, 2004), we have derived a list of aspects of communication in real and virtual environments and constructed a data model which defines the necessary parameters to represent these aspects. Table 1 lists the components of this model, their occurrence in reality, and the technical feasibility in VEs, given the current state of computer technology. Not all aspects of the real world can be implemented in the virtual world, for example, olfactics. In addition, not all aspects are equally important and we have to weigh the benefit of adding an aspect against the technological challenge in acquiring the data.

Depending on the choice of the simulation platform, we are connecting our user monitor application to, several aspects of the data model are already implemented. A modern game engine usually enables verbal communication by support of headsets and microphones, spatial behaviour by the movement and orientation of avatars, physical appearance by a selection of avatars that can be adjusted to the user's needs, and environment by physically simulated objects that can be grabbed or moved. Data input for these aspects comes from the keyboard or mouse movement, and does not have to be provided by our user monitor application.

Our framework adds occulesics, kinesics, and facial expression by defining an interface between the application and the simulation that exchanges data about the gaze direction, the head orientation, and the facial expression of the user.

By increasing the viewing angle of the camera or placing it further away from the user, the user monitor application could capture not only the face but also the upper body movement and hand gestures. This could add support for gestures to the simulation like shrugging of shoulders or pointing and increase the communication bandwidth even further.

Results

Figure 4 gives an impression of our work. The head rotation and facial expression of the avatar in the left figure is controlled by the user monitor application shown next to it. An inexpensive standard webcam captures a video of the user's head. Shaking the head, nodding, or other head movements in reality are immediately transferred into the VE and can be observed by other users. The user monitor application allows for a large range of head rotation and translation (X-Axis, "nodding": $-30° < X < +60°$; Y-Axis, "shaking head": $-90° < Y < +90°$, Z-Axis, head tilt: $-90° < Z < +90°$). With the help of controllers, the user can change the facial expression into one of the six basic emotions happiness, sadness, anger, fear, surprise, disgust with varying strength. This will happen automatically in a later stage of our work.

The VE is based on the Source Engine, a game engine that provides a Software Development Kit (SDK) for modification of the content (Marks et al., 2008). We added the plug-in necessary for the communication with the user monitor application and modified parts of the engine for the above listed aspects of non-verbal communication and interaction between avatars, objects, and a simulated patient.

The right figure shows the avatar in the simulation, gazing at objects on a shelf. The gaze direction and the target of the gaze can be visualised. In addition, information about what objects or avatars a user has looked at and how long is logged into a file that can automatically be analysed by software suitable for observational studies, e.g. Observer XT (Noldus Information Technology, 2009).

Our data model is universal and independent of the VE it is connected to. If support for detailed facial expression is provided, for example, if the VE implements the Facial Action Coding System or the MPEG-4 standard, the corresponding low-level parameters of the data model (e.g., raise left mouth corner, lower right eyebrow) can be used with only minor adaptations. This is the case for the majority of modern game engines, for example, the Source Engine (Valve Corporation, 2004) or the Unreal Engine 3 (Epic Games, 2006).

If the support is not as fine, the data model provides high-level parameters (e.g., emotion-type: angry, emotion-level: 50%) that can be used for an alternative way of displaying the emotional state of the user, for example by changing the texture of the avatar's face. This flexibility of our data model enables the connection to a wide range of VEs and game engines. An example for such limited support of facial expressions is Second Life (Linden Research, Inc, 2008). This VE has gained much popularity in the last years and is also increasingly used for teamwork training (Ullberg et al., 2007, Fig. 1, right side). The control provided by Second Life over the avatar's facial animation is not as fine grained as for modern game engines and is limited to displaying pre-defined animations. In this case, the interpreted parameters of our data model can be used to display basic emotions in form of short animations.

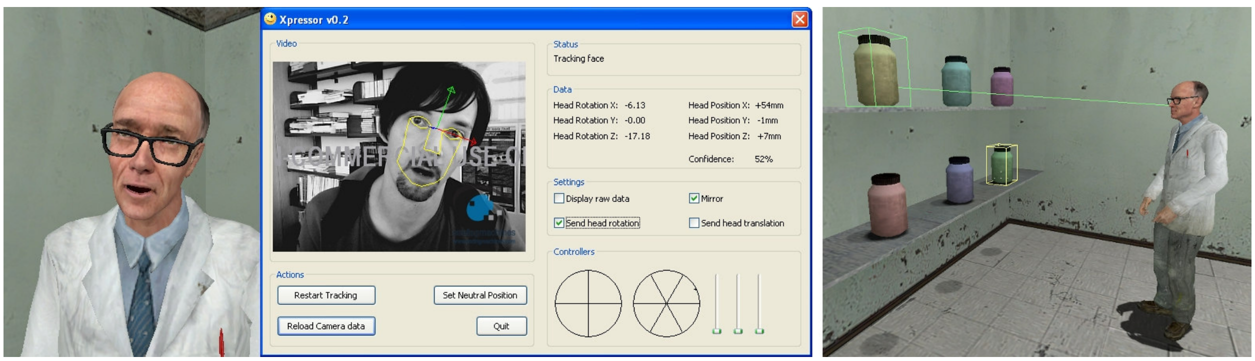

Figure 4: The left picture is a screenshot of the user monitor program capturing the head rotation and the user's avatar reflecting this rotation. It is also possible to change the facial expressions of the avatar. The right picture is a screenshot of an avatar gazing at objects with visual aids showing the “gaze ray” and the object the user is looking at.

Conclusion and Future Work

We have presented a framework that allows the enhancement of VE based teamwork simulations by aspects of non-verbal communication. The necessary data is captured in real-time by an inexpensive webcam. Our framework is flexible, extendible, independent of the used simulation engine. It is applicable not only to medical simulation, but to all kinds of teamwork simulations that involve human communication, e.g., emergency response, military teams, or power plant operators.

We have received positive feedback from medical professionals and developers of teamwork simulations in Second Life (Fig. 1, right side) about the use and the potential of our application. Currently, we are performing more detailed user studies to measure the performance of the user monitoring application, increase of communication bandwidth, to verify the ease of use of the camera based data acquisition, and to improve and extend our data model.

In cooperation with researchers and educators from the medical school, we are going to design surgical training scenarios that are to be used with our application. Furthermore, our user monitor application and the data model is suitable for teamwork training applications beyond the field of medicine and surgery. For this reason, we are also in cooperation with developers for emergency response teamwork training in Second Life, giving us the opportunity to collect valuable information with their simulation scenarios.

References

Epic Games (2006). Unreal Engine 3. URL: http://www.unrealtechnology.com/html/technology/ue30.shtml

Garau, M., Slater, M., Bee, S., and Sasse, M. A. (2001). The impact of eye gaze on communication using humanoid avatars. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, 309–316, New York, NY, USA. ACM.

Guise, J.-M. and Segel, S. (2008). Teamwork in obstetric critical care. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 22(5):937–951.

Hussain, T. S., Weil, S. A., Brunyé, T. T., Sidman, J., Alexander, A. L., and Ferguson, W. (2007). *Eliciting and Evaluating Teamwork within a Multi-Player Game-Based Training Environment*, chapter 5, 77–103. Elsevier, first edition.

Kohn, L. T., Corrigan, J. M., and Donaldson, M. S., editors (1999). *To Err is Human: Building a Safer Health System*. National Academy Press, Washington, DC, USA.

Linden Research, Inc (2008). Second Life. URL: http://secondlife.com/

Lingard, L., Espin, S., Whyte, S., Regehr, G., Baker, G. R., Reznick, R., Bohnen, J., Orser, B., Doran, D., and Grober, E. (2004). Communication failures in the operating room: An observational classification of recurrent types and effects. *Quality & Safety in Health Care*, 13(5):330–334.

Manninen, T. (2004). *Rich Interaction Model for Game and Virtual Environment Design*. PhD thesis, University of Oulu, Finland.

Marks, S., Windsor, J., and Wünsche, B. (2007). Evaluation of Game Engines for Simulated Surgical Training. In *GRAPHITE '07: Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia*, pages 273–280, New York, NY, USA. ACM.

Marks, S., Windsor, J., and Wünsche, B. (2008). Evaluation of Game Engines for Simulated Clinical Training. In *New Zealand Computer Science Research Student Conference 2008*, 92–99.

Noldus Information Technology (2009). Observer XT. URL: http://www.noldus.com/human-behavior-research/products/the-observer-xt

Scott, D. J., Cendan, J. C., Pugh, C. M., Minter, R. M., Dunnington, G. L., and Kozar, R. A. (2008). The Changing Face of Surgical Education: Simulation as the New Paradigm. *Journal of Surgical Research*, 147(2):189–193.

Serious Games Initiative (2007). Serious Games Initiative. URL: http://www.seriousgames.org

Slater, M., Howell, J., Steed, A., Pertaub, D.-P., and Gaurau, M. (2000). Acting in Virtual Reality. In *Proceedings of the Third International Conference on Collaborative Virtual Environments*, 103–110.

Taekman, J., Segall, N., Hobbs, E., and Wright, M. (2007). 3DiTeams – Healthcare Team Training in a Virtual Environment. *Anesthesiology*, 107(A2145):A2145.

Ullberg, L., Monahan, C., and Harvey, K. (2007). The New Face of Emergency Preparedness Training: Using Second Life to Save First Lives. In Livingstone, D. and Kemp, J., editors, *Second Life Education Workshop 2007*, 96–99.

Undre, S. and Darzi, A. (2007). Laparoscopy Simulators. *Journal of Endourology*, 21(3):274–279.

Valve Corporation (2004). Valve Source Engine Features. URL: http://www.valvesoftware.com/sourcelicense/enginefeatures.htm

Vertegaal, R., Slagter, R., van der Veer, G., and Nijholt, A. (2001). Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 301–308, New York, NY, USA. ACM.

Vertegaal, R., van der Veer, G., and Vons, H. (2000). Effects of Gaze on Multiparty Mediated Communication. In *Graphics Interface*, 95–102.

Virtual Heroes (2008). Virtual Heroes Inc – Serious Games and Andvanced Learning Systems. URL: http://www.virtualheroes.com

Yule, S., Flin, R., Paterson-Brown, S., and Maran, N. (2006). Non-technical skills for surgeons in the operating room: A review of the literature. *Surgery*, 139(2):140–149.