

HEAD TRACKING BASED AVATAR CONTROL FOR VIRTUAL ENVIRONMENT TEAMWORK TRAINING

Stefan Marks

Department of Computer Science, The University of Auckland, Auckland, New Zealand

John A. Windsor

Department of Surgery, Faculty of Medicine and Health Sciences, The University of Auckland, Auckland, New Zealand

Burkhard Wünsche

Department of Computer Science, The University of Auckland, Auckland, New Zealand

Keywords: Virtual Environments, Head Tracking, Nonverbal Communication, Teamwork Training, Surgery.

Abstract: Virtual environments (VE) are gaining in popularity and are increasingly used for teamwork training purposes, e.g., for medical teams. One shortcoming of modern VEs is that nonverbal communication channels, essential for teamwork, are not supported well. We address this issue by using an inexpensive webcam to track the user's head. This tracking information is used to control the head movement of the user's avatar, thereby conveying head gestures and adding a nonverbal communication channel. We conducted a user study investigating the influence of head tracking based avatar control on the perceived realism of the VE and on the performance of a surgical teamwork training scenario. Our results show that head tracking positively influences the perceived realism of the VE and the communication, but has no major influence on the training outcome.

1 INTRODUCTION

In recent years, virtual environments (VEs) have become increasingly popular due to technological advances in graphics and user interfaces (Messinger et al., 2009). One of the many valuable uses of VEs is teamwork training. The members of a team can be located wherever it is most convenient for them (e.g., at home) and solve a simulated task in the VE collaboratively, without physically having to travel to a common simulation facility. Medical schools have realised this advantage and, for example, created numerous medical simulations within Second Life or similar VEs (Danforth et al., 2009).

When looking at teamwork, communication is a vital aspect. An ideal VE would therefore facilitate all communication channels that exist in reality – verbal as well as non-verbal. Due to technical limitations, this is not possible, and therefore, existing communication in VEs is currently mostly limited to voice. Other channels like text chat, avatar body gestures, facial expressions have to be controlled manually and thus do not reflect the real-time communicative behaviour of the user.

Analysis of communication in medical teamwork has shown that nonverbal communication cues like gesture, touch, body position, and gaze are equally important to verbal communication in the analysis of the team interactions (Cartmill et al., 2007). VEs that do not consider those nonverbal channels are likely to render the communication among the team members less efficient than it would be in reality.

We propose an inexpensive extension of a VE by camera-based head tracking to increase the “communication bandwidth”.

Head tracking measures the position and the orientation of the user's head relative to the camera and the screen. The *rotational* tracking information can be used to control the head rotation of the user's avatar. That way, other users in the VE can see rotational head movement identical to the movement actually performed physically by the user, like nodding, shaking, or rolling of the head.

The *translational* tracking information can be used to control the view “into” the VE. This so called (HCP) enables intuitive control, like peeking around corners by moving sideways, or zooming in by simply moving closer to the monitor. The use of head track-

ing information has therefore the potential to simplify the usage of a VE by replacing non-intuitive manual view control by intuitive motion-based view control. Especially in medical applications, this has the potential to free the hands of the user, enabling the use of other simulated instruments or tools, e.g., an endoscope.

This paper presents the results of an experimental study designed to measure any influence that the introduction of head tracking has on teamwork communication, teamwork efficiency, and perceived realism of the VE.

2 RELATED WORK

The reproduction of nonverbal communication in virtual environments is an ongoing research topic that benefits from technological advancement. In the early stages, data was captured by using magnetic tracking sensors, facial markers, data gloves, etc. (see, e.g., (Ohya et al., 1993)). With the development of better hardware and advanced tracking algorithms, it became possible to use only one camera instead of multiple systems.

(Cordea et al., 2002) track the head of the user with a camera to extract the rotational information. This can easily and efficiently be transmitted during a video call to simulate the head movement of the user's avatar on the receiver's side. However, the focus of this paper is more on information reduction than on virtual environments.

(Wang et al., 2006) use only the 2-dimensional position of the face within the camera image to control the 2D-movement of a game character.

This idea is extended into the 3rd dimension by (Yim et al., 2008), where a head-mounted LED line is used to track the position and rotation of the user's head. This information is used again to control a game instead of an avatar.

Using only a single camera, (Sko and Gardner, 2009) present a range of interaction techniques based on 3-dimensional translation and rotation tracking data. A predefined set of head gestures is recognized and associated with certain actions in a game. Slightly tilting the head sideways is used for peering around a corner. Leaning forwards is interpreted as zooming. Head rotation is used for a slight change in the view direction, whereas head translation is used for HCP. These techniques focus on a single user, but the authors have not extended their research on the possibilities for multi-user scenarios.

In a recent study, (Marks et al., 2010) found that users can clearly perceive head motions, facial ex-

pressions, and mouth movement of avatars, and that camera-based HCP is more natural and intuitive to use than manual view control.

3 QUESTIONS AND HYPOTHESES

In this paper, we are going to validate the following hypotheses:

- H1 When head tracking is enabled, the participants demonstrate more overall head movement than when head tracking is disabled.
- H2 When head tracking is enabled, the amount of head movement is greater when participants talk to each other compared to when head tracking is disabled.
- H3 When head tracking is enabled, the participants perceive the other avatars as more natural than when head tracking is disabled.
- H4 When head tracking is enabled, the participants perceive the communication with the other participants as more natural than when head tracking is disabled.
- H5 When head tracking is enabled, the participants look more at each other's avatars than when head tracking is disabled.
- H6 When head tracking is enabled, the participants solve the teamwork task better than when head tracking is disabled.

4 DESIGN

The following factors were involved in the design of the study:

- The number of participants was limited to three due to the amount of available game engine licenses.
- It should be impossible or at least infeasible for only one or two participants to solve the scenario efficiently and instead require the effective teamwork of all three users involved.
- To facilitate comparison of the results, the scenario should be of equal length when repeated.
- To avoid too much of a learning curve effect, the scenario should vary at least in parts.

4.1 Overview

We chose a simplified surgical procedure for the simulation scenario. It involves three participants: a surgeon, an anaesthesiologist, and a nurse. Each of these three roles has very clear and defined tasks:

- The surgeon performs the necessary steps of the surgical operation on the patient.
- The anaesthesiologist monitors the patient's vital signs and stabilises the patient when critical events occur.
- The nurse is responsible for supplying the surgeon and the anaesthesiologist with the correct instrument respectively medication.

This clear definition of the roles and tasks together with the necessary interactions is depicted in Figure 1. Every role has to communicate with every other role throughout the scenario. Ineffective communication would inevitably lead to an ineffective teamwork.

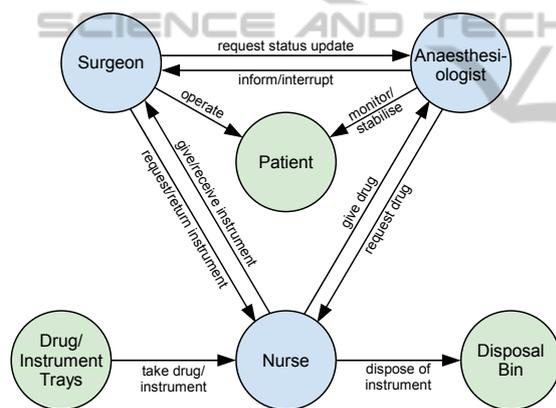


Figure 1: Diagram of the relations and interactions between the three roles and the patient.

For measuring the effectiveness of the teamwork of a group of participants, we used the following two metrics:

- The total time from the beginning of the first step of the surgical procedure until the completion of the last step.
- The relative amount of time that the patient's vital signs are critical (e.g., blood pressure too low, heart rate too high).

For each group of three participants, we conducted six experiments, rotating the roles and disabling/enabling the head tracking so that no participant would have the same role in two successive experiments, and every participant would experience each role with head tracking disabled and enabled.

4.2 Surgical Procedure

The selection of the nature of the simulated surgical procedure was mainly based on the abilities of the game engine we used for creating the VE, a modified version of the Source Engine (Valve Corporation, 2004).

In general, larger objects can be handled more easily in this VE than small objects. Therefore, we chose the relatively large torso of the patient as the operating field, as opposed to the leg or other extremities. This allowed us to design relatively large instruments that would cause less problems for the participants to move and position.

For the same reason, we chose to design a rather large incision along the middle of the upper body, revealing organs of the digestive system like stomach, liver, pancreas, small intestines, and colon.

Located in the centre of this incision are the stomach and the pancreas. After consultation with staff of the surgical department, we decided on the removal of dead and infected tissue from the pancreas of the patient as the surgical procedure to simulate. In medical terms, this procedure is called *pancreatic necrosectomy*.

4.3 Role Description

The positions of the three roles involved in the surgical procedure and the layout of the room can be seen in Figure 2.

4.3.1 Surgeon

The surgeon operates on the right side of the patient. The task of the surgeon is to apply the right instruments in the right order to the patient to complete the procedure step by step.

The steps of the simulated surgical procedure are visualised in Figure 3. To get from one step to the next, the indicated instrument has to be used. The surgeon uses the instrument on the patient by touching the operating field with it.

When the instrument is the correct one for the task, it starts to blink blue and a progress bar gets visible in the lower right corner of the screen. Each step takes a certain amount of time, also listed in Figure 3. When the instrument has been applied for that amount of time, the progress bar disappears, the instrument blinks green once and can now be disposed of by the nurse.

When the surgeon applies a wrong instrument, it blinks red and has no effect.

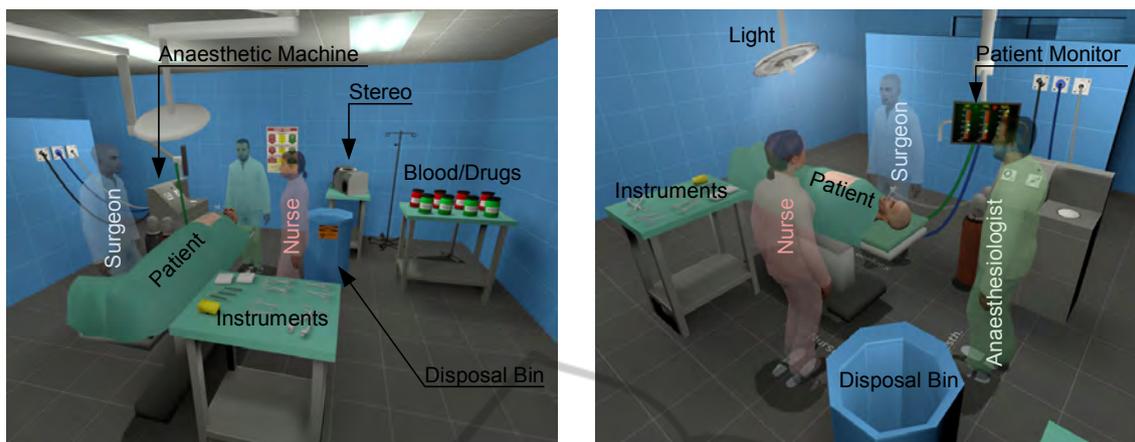


Figure 2: Two views of the operating theatre implemented in our simulation.

4.3.2 Anaesthesiologist

The anaesthesiologist is located at the head of the patient, close to the airways. The role of the anaesthesiologist is to observe the patient's vital signs on the patient monitor and to counteract four different types of events that will occur at random times during the operation:

1. The oxygen level of the patient's blood falls lower than 80%. This will cause a *hypoxaemia* alarm. In this case, the anaesthesiologist has to press the "O₂" button of the anaesthetic machine to supply additional oxygen and cause the blood oxygen level to rise back to normal again.
2. The blood pressure drops due to bleeding during the procedure, resulting in *hypotension*. In this case the anaesthesiologist has to start a blood infusion to stabilise the blood pressure again. The blood pressure will continue to drop as long as the bleeding is not stopped by the surgeon who has to apply pressure by using pads.
3. An increasing heartbeat rate is an indicator of *tachycardia* and has to be treated by administering β -Blockers. The procedure does not have to be stopped during this condition, but as it is the nurse who has to hand the drug to the anaesthesiologist, the surgeon might have to wait for the next instrument.
4. A decreasing heartbeat rate is a sign of *bradycardia* and has to be treated by administering Atropine to the patient. Similar to tachycardia, the procedure does not have to be stopped, but might be delayed by the nurse having to pass the drug to the anaesthesiologist.

4.3.3 Nurse

The nurse is located opposite, on the other side of the patient. The role of the nurse is to

- switch on the lighting of the operating field,
- hand the requested instrument to the surgeon,
- take back the used instrument from the surgeon and dispose of it, and
- hand the requested drugs to the anaesthesiologist.

4.3.4 Role Enforcement

Physical and logical barriers prevent any of the users from taking over the role of any other user.

- The anaesthesiologist cannot reach the table with the drug and blood bottles, because the stereo (see Section 4.7 for details), a small tray, an IV stand, and the disposal bin are in the way.
- The surgeon and the anaesthesiologist cannot easily reach the position of the other, because they are blocked by the gas lines to the anaesthetic machine.
- The surgeon can see the instruments and guide the nurse by looking at them, but it is impossible to reach them from the other side of the patient.
- When roles other than the surgeon try to apply an instrument on the patient, it will not have any effect.
- Running and jumping is disabled to prevent participants with a higher experience in first person perspective games to jump over the patient or objects to reach other users' locations.

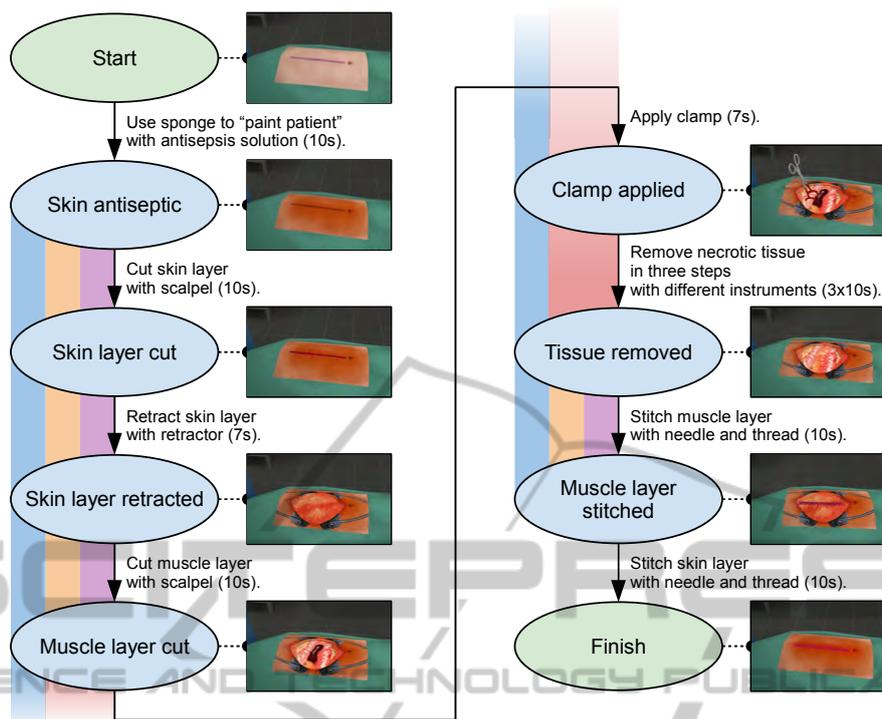


Figure 3: Overview of the steps of the simulated surgical procedure and the instruments necessary to reach the next step. The coloured areas symbolise the sections of the procedure during which the events in Figure 4 can occur.

4.4 Events

During certain phases of the procedure, the events described in Section 4.3.2 will occur. In Figure 3 and Figure 4, the times at which the events can happen are visualised by different colours.

During a certain phase of the operation, especially while the clamp is applied or the necrotic tissue is being cut, unexpected and severe bleeding might occur. This has to be stopped by applying pressure with a pad. The operation can only be continued when the bleeding has stopped. Bleeding is visualized by a red particle system simulating a medium intense blood flow inside the operating field. At the same time, the anaesthesiologist will notice a loss of blood pressure. Therefore, bleeding is an event that is easy to recognise by all three roles.

At random times before and after this part, tachycardia and bradycardia will occur. Those events are most apparent to the anaesthesiologist who will see the pointers on the patient monitor rise or fall. The surgeon and the nurse will only have the audio cue of the beeping heartbeat monitor becoming faster or slower.

Finally, at another random point in time, the oxygen saturation will drop significantly. The anaesthesiologist has to keep an eye on this value to prevent

hypoxaemia. This event is apparent only to the anaesthesiologist.

All four events are spaced temporally so that they do not occur at the same time. Only when the participants do not react to an event and instead continue with the operation, the effects of multiple events will eventually build up.

4.5 Face Validity

The simulated procedure is a very simplified version of a real surgery. We had to find a compromise between face validity, the medical term for realism, and simplicity for several reasons:

- The game engine has limitations in terms of physical simulation and animation.
- The interaction of the user with the VE has to be kept simple.
- Some of the participants were non-medical students who did not have a solid background in medicine.

For both participant groups, medical and non-medical, an explanation was given beforehand that this scenario is a very simplified simulation and does not always realistically reflect procedures, devices, and effects.

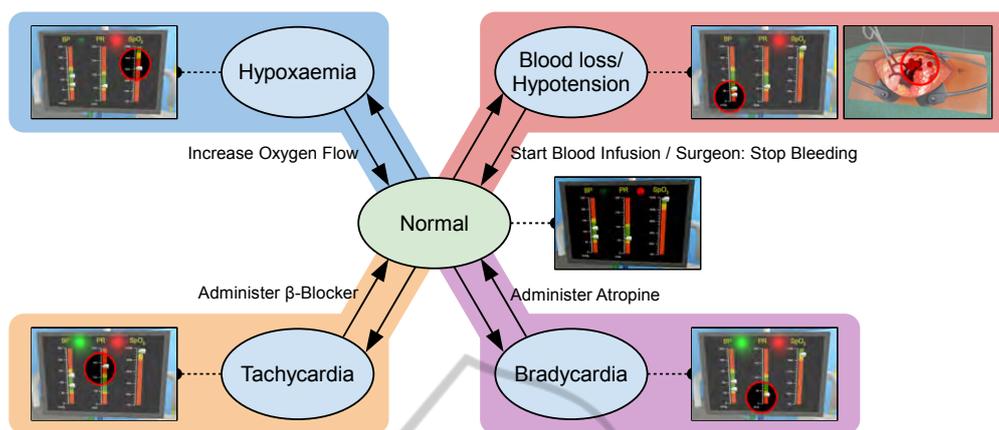


Figure 4: Overview of the four events that the team has to react to during the simulated surgical procedure. The events with a coloured background occur at a random point of time during the steps of the procedure in Figure 3 with the same coloured background. To return to the “Normal” state, the action next to the arrow has to be taken.

In summary, the following aspects of the simulation are not realistic:

- The avatars don't wear face masks which, in reality, are mandatory for infection control (Lipp and Edwards, 2002).
- A medical team is usually comprised of more than only three members. In our case, the roles of, e.g., the running nurse (responsible for fetching instruments and gear from outside of the operating theatre) and anaesthesiology nurse (mainly responsible for assistive tasks required by the anaesthesiologist) are not covered.
- The anaesthetic machine and the patient monitor have a very simplified functionality and a different look compared to real machines.
- In case of blood loss between 15% and 30% of the total blood volume, defined as *Class II hemorrhage* by the American College of Surgeons, saline fluid is usually sufficient to stabilise the blood pressure (O'Leary et al., 2008, p. 564). A blood transfusion is only necessary when the patient has lost larger amounts of blood. In addition, due to its viscosity, blood would not flow fast enough into the circulatory system of the patient.
- In case of blood loss, the blood pressure does not immediately decrease. Instead, the heart starts to beat faster to maintain the blood pressure despite the loss of blood volume (O'Leary et al., 2008, p. 564). During this phase, only the diastolic blood pressure starts to drop, caused by the tachycardia and contracting blood vessels, while the systolic pressure stays more or less the same.
- The incision is in general too large for an abdominal procedure. In addition, removal of necrotic

pancreas is usually not done by an open operation. A multi-centre study (van Santvoort et al., 2010) has shown that endoscopic procedures result in less complications and a lower mortality rate among patients compared to open procedures.

- In the simulation, the pancreas is immediately visible. In reality, a larger part of the pancreas is hidden behind the stomach.
- The treatment for the critical events in the simulation is extremely simplified.
- Instruments that have been dropped on the floor can be used again on the patient without the risk of infection (see further commentary to this issue in 7.3).
- A team coming together for a real world medical simulation already has a reasonable background of medical knowledge and experience. During the simulation, this knowledge and understanding of their roles enables them to completely focus on the teamwork and the task. In contrast, several of the participants in our VE scenario had no medical knowledge and had to learn their roles *as well as* having to work together as a team. For some of them, having to learn the use of the VE was another additional task.

4.6 Room Design

The overall design of the room was based on photos of real operating theatres. 3D modelling of the instruments and other objects in the operating theatre was done in the 3D open-source software Blender (Blender Foundation, 2007). As for the room, we used photos and schematic diagrams for textures and

measurements, but enlarged the majority of the instruments by 20% to 50% to enable easier handling in the VE.

4.7 Sound

To disrupt voice communication, we added two sound sources to the simulation. Random announcements can be heard in intervals of 20s to 40s through a speaker mounted in the ceiling. Those announcements are meant to disrupt the communication between the participants on a regular, but short basis.

A second sound source is a CD player with speakers in the operating theatre, which is common practice in operating theatres (Ullmann et al., 2008). It randomly plays one of six music tracks that are all about 6min to 8min long and differ in style: Rock, Funk, Folk, Jazz, Trip Hop, and Ambient. The anaesthesiologist can change the track, but cannot switch it off or alter the volume.

The music is intended to be a continuous source of noise, making it slightly harder for the participants to fully rely on pure voice communication. Therefore, we purposely violate a general recommendation that, during a surgical procedure, the music should be turned to a low volume in general and completely off in cases of emergency (Hodge and Thompson, 1990).

5 SETUP

For the VE, we use the Source Engine, running a heavily modified version of a multi-player death-match game. The head tracking is done by an external program that uses the commercial tracking engine faceAPI (Seeing Machines, 2009). This program runs in the background and communicates with the VE, continuously transmitting updates of the user's head position and rotation.

We used three DELL Optiplex 745 desktop computers¹ as clients and the development machine² as the server.

The simulation was run with a 60Hz framerate on the server as well as on the clients.

The client desktops were equipped with a Logitech QuickCam Pro 9000 webcam. This camera provided a stable framerate of 30fps with only the avail-

¹Intel Core™2 Duo CPU, 2.13GHz, 3GB memory, Windows Vista Enterprise 32bit, Radeon X1300 graphics card with 256MB graphics memory, DELL 2007FP 20.1inch flat panel monitor with 1600x1200pixel resolution.

²Intel Core™2 Quad CPU, 2.4GHz, 4GB memory, Windows XP Professional, NVIDIA Quadro FX 570 graphics card with 256MB graphics memory.

able room lighting. Autofocus ensured that even with movement away and towards to the monitor, the heads of the participants always stayed in focus.

It was not possible to organise three separate rooms for the participants, so we had to use other methods to ensure that they could not hear or see each other directly. One solution was the use of Logitech G35 headsets which covered the ears of the participants completely and provided enough isolation from outside sounds. That way, the participants had to rely solely on communication channels provided by the game engine.

To prevent the participants from seeing each other or each other's screens, we set up office walls that occluded direct line of sight between the three client workstations and in addition provided further acoustic separation of the participants. A photo of the complete setup is shown in Figure 5.

6 METHODOLOGY

We recruited participants by printed advertisements in the buildings of the departments of Computer Science and the Medical School, and by emails sent to classes and postgraduate students of the department of Computer Science.

On average, one user study took around 90min to complete. Due to this extended duration, we reimbursed the participants by handing out grocery vouchers valued 10 NZD. We assumed that this amount would encourage participants sufficiently without introducing the risk that they would "just do it for the money".

We explained the overall goal of the study to the participants, mentioning the head tracking and the possibility to control the view and the head movement of the avatar. Afterwards, we gave a brief overview of the procedure and the roles and their responsibilities. The participants were encouraged to interrupt and ask questions at any stage.

Before the actual study took place, We introduced the participants to the VE by running an introduction simulation without recording any data. The participants were able to get accustomed to the environment and to experiment with objects, instruments, and devices. Each participant had a set of three information sheets in front of them that described their role, the necessary instruments or devices, and the mouse and keyboard controls.

After the introduction and any further questions from the participants, we enabled data logging of the simulation and started the actual series of six experiments. The participants were *not* informed whether



Figure 5: Photo of the experimental setup with three participants. The participants are isolated visually by office walls and acoustically by headsets that completely cover the ears.

head tracking was disabled or enabled.

After each experiment, the participants filled in a short questionnaire about their opinion on teamwork and communication aspects of the simulation. During that time, we changed the information sheets according to the rotating roles of the participants.

At the end of the six experiments, the participants filled out an additional questionnaire page with general questions about the VE.

7 RESULTS

7.1 Participants

More than 30 people answered to the invitation emails and advertisements. However, due to financial and time constraints, we had to limit the amount of participants to 27. The age of the participants, 6 of them being female, covered a range from 20 to around 44 years (see Figure 6 for more details).

All participants have used a computer often or more. 5 of them have not played computer games at all within the three months before the study. 22 of the participants have had experience in using a webcam, e.g., by using it for video conferencing. 11 of them have not used a VE before.

7.2 Hypotheses

Hypothesis H1: *When head tracking is enabled, the participants demonstrate more overall head movement than when head tracking is disabled.*

For the verification of this hypothesis, we calculated the total amount of translational and rotational head movement in 1s steps and compared the values

from experiments with head tracking disabled to the values from experiments with head tracking being enabled.

The resulting data is not normally distributed, but instead strongly skewed towards lower values and tailing off towards higher values. For that reason, it is not possible to apply the t-test to this data. Instead, we chose the Wilcoxon rank sum test (McCrum-Gardner, 2008).

The results show a statistically significant but only minute difference in translational movement (median values 28.2mm / 26.7mm, Wilcoxon rank sum test with continuity correction, $p < 0.001$, 95% CI 0.71mm to 1.51mm) and a minute and statistically insignificant difference in rotational movement (median values 11.2° / 11.1°, Wilcoxon rank sum test with continuity correction, $p = 0.500$, 95% CI -0.07° to 0.16°). Based on this outcome, Hypothesis H1 has to be falsified. When head tracking is active, the participants do *not* increase the amount of their overall head movement. On the contrary, the data suggests that the amount of movement is *the same* irrespective of head tracking being disabled or enabled.

Hypothesis H2: *When head tracking is enabled, the amount of head movement is greater while participants talk to each other compared to when head tracking is disabled.*

For the analysis of this hypothesis, we prepared the data in the same way as described for the hypothesis before, but selected only the subset of the measurements when the participants were talking. Again, we had to apply the Wilcoxon ranked sum test due to the skewed distribution of the data.

Like for the previous hypothesis, the results of the test for head movement of the participants when they were talking do either indicate statistically sig-

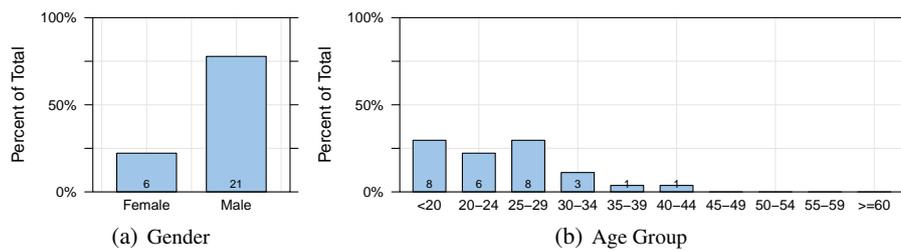


Figure 6: Demographic data of the participants of the multi user study.

nificant but minor differences, or statistically insignificant differences. Translational head movement differs only by -1.0mm (median values $35.9\text{mm} / 34.9\text{mm}$, Wilcoxon rank sum test with continuity correction, $p = 0.054$, 95% CI -0.00 to 1.53), and rotational movement by -0.2 (median values $14.0 / 13.8$, Wilcoxon rank sum test with continuity correction, $p = 0.183$, 95% CI -0.06 to 0.42).

Similar to the previous one, Hypothesis H2 has to be falsified. While talking, the participants do *not* increase their head movement significantly when head tracking is enabled. Instead, the results suggest that the amount of movement is *the same* for head tracking disabled respectively enabled.

Hypothesis H3: *When head tracking is enabled, the participants perceive the other avatars as more natural than when head tracking is disabled.*

For the verification of this hypothesis, we evaluated the questionnaires that the participants had to answer after each of the six experiments (see Figure 7).

We converted the answers to the 7-point Likert scales on the questionnaires to a numerical value using a translation scale of -3 for “Strongly Disagree” to $+3$ for “Strongly agree”. We then compared the answers of experiments 1, 3, and 5 (head tracking disabled) to the answers of experiments 2, 4, and 6 (head tracking enabled) using a t-test.

The comparison reveals a statistically non-significant difference of 0.2 (mean values $1.4 / 1.6$, Welch Two Sample t-test, $p = 0.282$, 95% CI -0.45 to 0.13), thereby weakly supporting the validity of the hypothesis. The participants perceive the other avatars slightly more natural when head tracking is enabled compared to when head tracking is disabled.

Hypothesis H4: *When head tracking is enabled, the participants perceive the communication with the other participants as more natural than when head tracking is disabled.*

We evaluated the answers to the question about how natural the communication is perceived (see Figure 8) in a similar way to the one described for the

previous hypothesis. Again, we compared the results of experiments 1, 3, and 5 (head tracking disabled) to the answers of experiments 2, 4, and 6 (head tracking enabled) using a t-Test.

The comparison suggests that the participants perceive the communication slightly more natural when head tracking is enabled than when it is disabled (mean values $1.7 / 1.9$, Welch Two Sample t-test, $p = 0.059$, 95% CI -0.55 to 0.01), thereby supporting the hypothesis.

Hypothesis H5: *When head tracking is enabled, the participants look more at each other’s avatars than when head tracking is disabled.*

The logged data from the simulation also recorded which user has been looking at which object or avatar for how long. We used this information to calculate the relative amount of time each participant had been looking at one of the team members. The results, grouped by role, are visualized in Figure 9.

Though there are differences in the comparisons, those are again only minor and statistically insignificant. When head tracking is enabled, the surgeon spends 0.4% more time looking at the colleagues (mean values $25.0\% / 25.4\%$, Welch Two Sample t-test, $p = 0.849$, 95% CI -5.15% to 4.25%), the anaesthesiologist 1.6% (mean values $11.7\% / 13.3\%$, Welch Two Sample t-test, $p = 0.430$, 95% CI -5.71% to 2.47%), and the nurse 1.4% (mean values $23.9\% / 25.3\%$, Welch Two Sample t-test, $p = 0.576$, 95% CI -6.05% to 3.40%) compared to when head tracking is disabled.

Overall, these results add only very weak support for the hypothesis. Nevertheless, it is interesting to see how the values reflect the different observation patterns of the three roles. It is apparent that the anaesthesiologist has a different pattern than surgeon and nurse. This difference is most probably caused by the primary task of the role of the anaesthesiologist: to watch the patient monitor for changes in the vital signs.

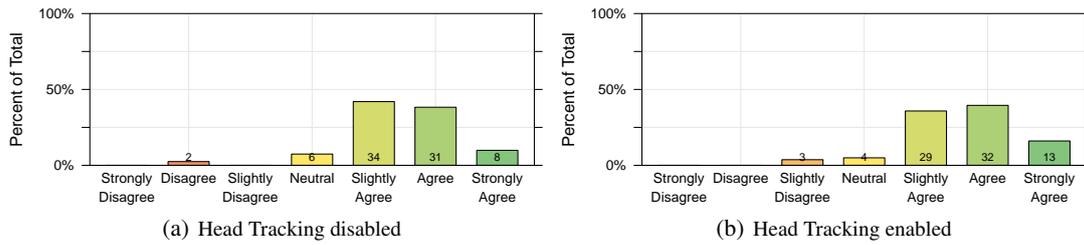


Figure 7: Results of the questionnaires about how natural the participants perceive the teammate's avatars.

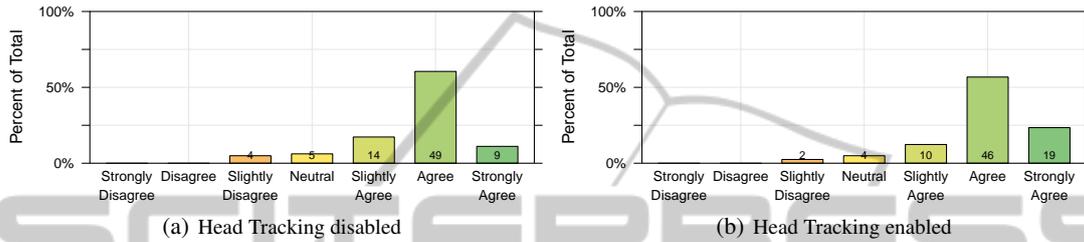


Figure 8: Results of the questionnaires about how natural the participants perceive the communication with their team members.

Hypothesis H6: *When head tracking is enabled, the participants solve the teamwork task better than when head tracking is disabled.*

As described in Section 3, we calculated the efficiency of the teamwork based on two values that are extracted from the data:

- The total time for completion of each experiment.
- The relative amount of time that the patient is in a critical state.

For all two values, we expected the influence of the head tracking to reflect on the results like shown in Figure 10. The participants were repeating a similar task six times, so we anticipated a learning curve that would demonstrate a tendency towards faster completion with each repetition, with a gradient that would ease out towards the end. The positive influence of head tracking would be visible by average values that are *below* the fitting curve for experiments 2, 4, and 6, whereas experiments 1, 3, and 5 without head tracking would produce results *above* the fitting line.

However, the measured results look very different. The learning curve is visible in experiments 2 to 6, but experiment 1 has a significantly lower completion time than experiments 2 to 3. The reason for this outlier is most probably the practice run that the participants had to complete before the first experiment was conducted.

For the practice run, they performed the teamwork task without their results being recorded and with help and minor intervention from the simulation administrator. After the practice run, the first experiment

started with the participants having the same roles than in the practice run. They had gotten used to their roles and were able to finish the teamwork task relatively fast.

For experiment 2 and 3, the roles were rotated so that every participants had to learn the requirements and tasks of the new roles. The influence of this new learning process reflects in the similar results for experiments 2 and 3.

Finally, for experiments 4 to 6, the participants were repeating former role combinations. Because by now, they had experienced every role, they were better able to focus on the task.

Due to the unusual values of the boxplots results, it is not possible to model an average curve for the comparison of the results with and without head tracking. Instead, we grouped the values by head tracking “off” (experiments 1, 3, and 5) and “on” (experiments 2, 4, and 6) and compared the mean values using a t-test. The values do not differ significantly (mean values 285.2s /285.9s, Welch Two Sample t-test, $p = 0.974$, 95% CI -45.21s to 43.74s). Instead, the high p-value suggests that head tracking makes *no difference at all* on the completion times of the teamwork task.

Likewise, the comparison of the relative amount of time that the master alarm was on results in a statistically insignificant difference (mean values 16.7% / 17.3%, Welch Two Sample t-test, $p = 0.821$, 95% CI -6.68% to 5.33%), suggesting additional support for the hypothesis that head tracking makes *no difference at all*.

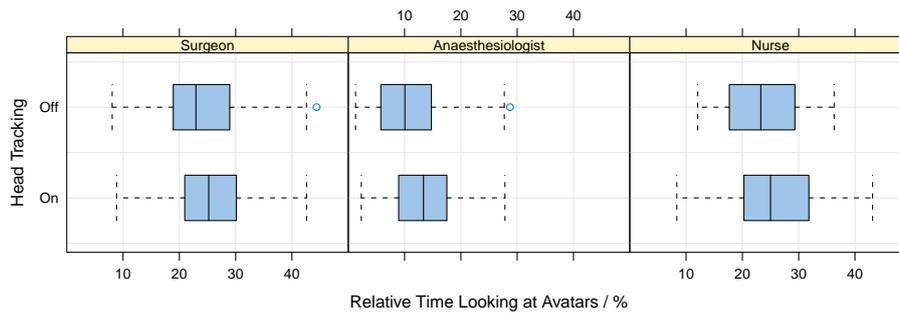


Figure 9: Plots of the relative amount of time the participants have been looking at the avatars of their team members.

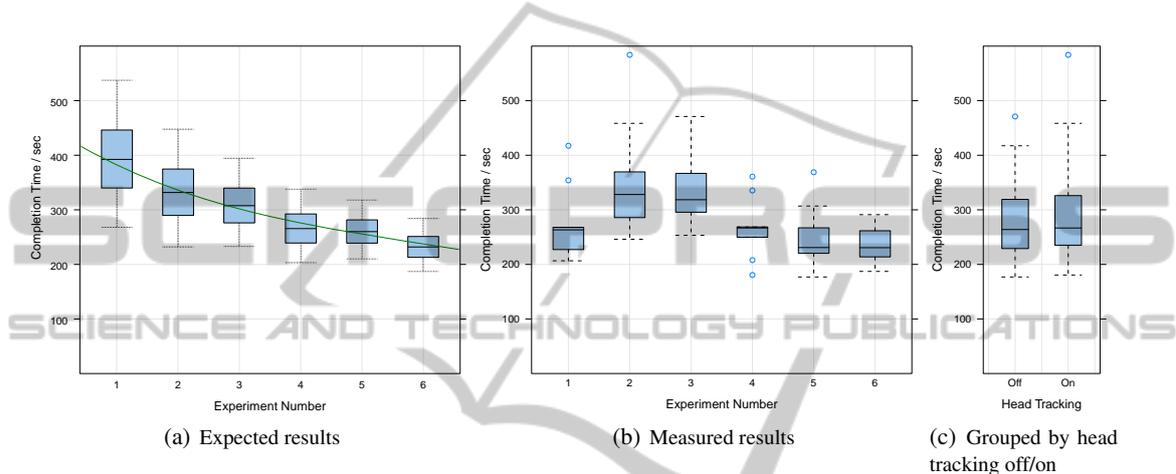


Figure 10: Box plots of the times the participants needed to complete the task.

7.3 Overall Impression

Overall, the participants enjoyed to work in teams within the simulation and had no major problems with the head tracking. In fact, the tracking was so unobtrusive, that several participants later commented that they were not at all aware of the tracking and the head movements.

When being asked about the stressfulness of the roles, the majority of the participants agreed that the nurse was the most stressful role, in contrast to the anaesthesiologist, who was considered the most relaxed role.

7.3.1 Head Tracking

There were slight problems resulting from a combination of head tracking jitter and handling of small objects like the scalpel. In these cases, some participants had problems passing on the scalpel from one to another. The tracking jitter of one simulation client would move the object in a slightly random pattern, making it more difficult for the opposite user to grab it. In their comments, several participants agreed on the fact that small objects were most difficult to han-

dle, especially the scalpel.

On the other hand, head tracking was mentioned as being very useful in precise placement of instruments or for zooming in on the patient monitor.

In general, the rotation of the avatar’s heads by head tracking was perceived, although most participants stated that they were too preoccupied by the task to pay attention to the head movements. Occasions where the tracking would stand out were when the participants would physically look sideways to their role description sheets and their avatar would also perform that sideways look. Other occasions were, for example, head bobbing to the rhythm of the music.

7.3.2 Involvement

The participants demonstrated a high level on involvement, characterized by utterances like “Oh No!”, when the master alarm would go off, or “Don’t die, please!” when the bleeding could not be stopped immediately. Sometimes, participants would even engage in role play and joke around or tease their team colleagues.

On some occasions, when teamwork was not optimal, and the group was especially competitive, the

tone turned slightly aggressive. In one group, for example, the nurse ignored the request of the surgeon for the pad and instead decided to give a bottle of blood to the anaesthesiologist first. In terms of efficiency, this step was not optimal, because as long as the bleeding was not stopped, the patient continued to lose blood. So the first priority of the team should have been to stop the bleeding first (i.e., give the surgeon a pad) and then to stabilise the blood pressure. The surgeon, being aware of this ineffectiveness, commented in a slightly annoyed tone: “We need another nurse!”. In return, the nurse defiantly countered: “I think we need another surgeon!”.

7.3.3 Unexpected Behaviours

A disadvantage of the rotation of roles was that on several occasions, participants would take over responsibilities of other roles. A nurse who was the anaesthesiologist in the previous run would keep an eye on the patient monitor and grab the necessary drug before the current anaesthesiologist would notice and require it. However, in a training environment where roles are predetermined and not rotated after every simulation, this problem would occur only to a lesser extent or not at all.

Another unexpected behaviour, observed especially with participants who indicated a high amount of computer game playing time, was the exploration of similarities and differences between the simulation and reality. When participants had, for example, dropped the scalpel on the floor, there was an initial hesitancy to use that scalpel again on the patient, caused by basic knowledge of the risk of infection in reality. However, when the participants found out that, in the simulation, this re-use did not have any negative consequences, they continued to use dropped instruments on the patient in favour to demanding a new and “clean” instrument from the nurse and thereby losing time.

8 CONCLUSIONS

Table 1 summarizes the hypotheses and the results of our findings. In summary, the evaluation of experimentally obtained data is showing

- that avatars with tracking-based head movement are perceived as more natural than avatars without tracking-based head movement, and
- that avatars with tracking-based head movement facilitate a communication that is perceived as more natural than with avatars without tracking-based head movement,

However, these results are either only weakly supported by the data or the difference in the participants’ perception is only marginal.

Table 1: This table summarises the results of the multi user study. “+” indicates that the results support the hypothesis. “-” indicates that the results falsify the hypothesis.

Hypothesis	Experimental Result
H1	-
H2	-
H3	+
H4	+
H5	+
H6	-

Unfortunately, there is only very weak support for the hypothesis that participants look at each other’s avatars more regularly when head tracking is enabled.

Finally, contrary to the original hypotheses, the analysis of the data also shows

- that the participants do *not* change the amount of head movement when head tracking is enabled compared to when head tracking is disabled,
- that, while talking, the participants do *not* change the amount of head movement when head tracking is enabled compared to when head tracking is disabled, and
- that the teamwork performance does *not* improve when head tracking is enabled compared to when head tracking is disabled.

While especially those last three statements have the potential to discouragement, they nevertheless indicate an important fact: Although a new technology has been introduced with several possibilities for failure and measurement errors, the participants were largely unaware of that new technology. They perceived an improvement in the naturalness of the VE and an improvement of the quality of communication, but did not experience any major negative limitations imposed upon them by this new technology.

We assume that the main reason for the lack of influence of head tracking was the fact that the simulated scenario required more focus on the task and on handling the instruments than on the communication.

In addition, not all possibilities of this new technology have been explored yet. For our experiment, the head rotation was simply mapped onto the avatars without any further semantic analysis. One future extension will be to calculate the actual target of the user’s head direction on the screen and to use that information to transform the tracking data so that the user’s avatar looks at the same target in the VE. An

additional level of realism would be introduced by eye tracking and a similar transformation of the avatar's gaze.

Automatic recognition of facial expressions is another future extension. This feature would not increase the complexity of the physical setup and would also not require any major changes in our simulation engine. However, we would expect that the inclusion of this communication channel can further increase at least the perceived naturalness and ease of use – if not more.

REFERENCES

- Blender Foundation (2007). Blender. <http://www.blender.org>.
- Cartmill, J., Moore, A., Butt, D., and Squire, L. (2007). Surgical Teamwork: Systemic Functional Linguistics and the Analysis of Verbal and Nonverbal Meaning in Surgery. *ANZ Journal of Surgery*, 77(Suppl 1):A79–A79.
- Cordea, M. D., Petriu, D. C., Petriu, E. M., Georganas, N. D., and Whalen, T. E. (2002). 3-D Head Pose Recovery for Interactive Virtual Reality Avatars. *IEEE Transactions on Instrumentation and Measurement*, 51(4):640–644.
- Danforth, D., Procter, M., Heller, R., Chen, R., and Johnson, M. (2009). Development of Virtual Patient Simulations for Medical Education. *Journal of Virtual Worlds Research*, 2(2):3–11.
- Hodge, B. and Thompson, J. F. (1990). Noise pollution in the operating theatre. *The Lancet*, 335(8694):891–894.
- Lipp, A. and Edwards, P. (2002). Disposable surgical face masks for preventing surgical wound infection in clean surgery. *Cochrane Database of Systematic Reviews* 2002, 1(CD002929).
- Marks, S., Windsor, J., and Wnsche, B. (2010). Evaluation of the Effectiveness of Head Tracking for View and Avatar Control in Virtual Environments. *25th International Conference Image and Vision Computing New Zealand (IVCNZ) 2010*.
- McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery*, 46(1):38–41.
- Messinger, P. R., Stroulia, E., Lyons, K., Bone, M., Niu, R. H., Smirnov, K., and Perelgut, S. (2009). Virtual Worlds – Past, Present, and Future: New Directions in Social Computing. *Decision Support Systems*, 47(3):204–228.
- Ohya, J., Kitamura, Y., Takemura, H., Kishino, F., and Terashima, N. (1993). Real-time reproduction of 3D human images in virtual space teleconferencing. In *Virtual Reality Annual International Symposium, 1993*, pages 408–414.
- O'Leary, J. P., Tabuenca, A., and Capote, L. R., editors (2008). *The physiologic basis of surgery*. Wolters Kluwer Health/Lippincott Williams & Wilkins, 4 edition.
- Seeing Machines (2009). faceAPI. <http://www.seeingmachines.com/product/faceapi>.
- Sko, T. and Gardner, H. J. (2009). Human-Computer Interaction - INTERACT 2009. In Gross, T., Gulliksen, J., Kotz, P., Oestreicher, L., Palanque, P., Prates, R. O., and Winckler, M., editors, *Lecture Notes in Computer Science*, volume 5726/2009 of *Lecture Notes in Computer Science*, chapter Head Tracking in First-Person Games: Interaction Using a Web-Camera, pages 342–355. Springer Berlin / Heidelberg.
- Ullmann, Y., Fodor, L., Schwarzberg, I., Carmi, N., Ullmann, A., and Ramon, Y. (2008). The sounds of music in the operating room. *Injury*, 39(5):592–597.
- Valve Corporation (2004). Valve Source Engine Features. <http://www.valvesoftware.com/sourcelicense/engine-features.htm>.
- van Santvoort, H., Besselink, M., Bakker, O., Hofker, H., Boermeester, M., Dejong, C., van Goor, H., Schaaferder, A., van Eijck, C., Bollen, T., van Ramshorst, B., Nieuwenhuijs, V., Timmer, R., Lamris, J., Kruijt, P., Manusama, E., van der Harst, E., van der Schelling, G., Karsten, T., Hesselink, E., van Laarhoven, C., Rosman, C., Bosscha, K., de Wit, R., Houdijk, A., van Leeuwen, M., Buskens, E., Gooszen, H., and Dutch Pancreatitis Study Group (2010). A Step-up Approach or Open Necrosectomy for Necrotizing Pancreatitis. *The New England Journal of Medicine*, 362(16):1491–1502.
- Wang, S., Xiong, X., Xu, Y., Wang, C., Zhang, W., Dai, X., and Zhang, D. (2006). Face-tracking as an augmented input in video games: enhancing presence, role-playing and control. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1097–1106, New York, NY, USA. ACM.
- Yim, J., Qiu, E., and Graham, T. C. N. (2008). Experience in the design and development of a game based on head-tracking input. In *Future Play '08: Proceedings of the 2008 Conference on Future Play*, pages 236–239, New York, NY, USA. ACM.