Evaluation of the Effectiveness of Head Tracking for View and Avatar Control in Virtual Environments

S. Marks[1], J. A. Windsor[2], B. Wünsche[1]

[1]Department of Computer Science, The University of Auckland, New Zealand.
[2]Department of Surgery, The University of Auckland, New Zealand.
Email: smar189@aucklanduni.ac.nz

Abstract

Virtual environments (VE) are gaining in popularity and are increasingly used for teamwork training purposes, e.g., for medical teams. We have identified two shortcomings of modern VEs: First, nonverbal communication channels are essential for teamwork but are not supported well. Second, view control in VEs is usually done manually, requiring the user to learn the controls before being able to effectively use them. We address those two shortcomings by using an inexpensive webcam to track the user's head. The rotational movement is used to control the head movement of the user's avatar, thereby conveying head gestures and adding a nonverbal communication channel. The translational movement is used to control the view of the VE in an intuitive way. Our paper presents the results of a user study designed to investigate how well users were able to use our system's advantages.

Keywords: Virtual Environments, Nonverbal Communication, Face Tracking

1 Introduction

In recent years, VEs have become increasingly popular due to technological advances in graphics and user interfaces [1]. One valuable use 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 potential and, for example, created numerous medical simulations within Second Life or similar VEs [2].

Communication is a vital aspect of teamwork, so an ideal VE would 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 [3]. Communication in a VE that does not consider those nonverbal channels is likely to render the communication among the team members less efficient than it would be in reality.

We propose an inexpensive extension of the communication within a VE by camera-based head tracking.

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 Head Coupled Perspective (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 tracking 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.

This paper presents a subset of the results of a user study designed to analyse how well participants actually were able to control their own avatar and observe other avatar's behaviours.

978-1-4244-9631-0/10/$26.00 ©2010 IEEE

2 Related Work

There is ongoing research about how to control the view into a VE or a game by head tracking.

The authors of [4] 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.

The researchers in [5] 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 in [6], 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, the authors of [7] present a range of interaction techniques based on 3-dimensional translation and rotation tracking data. A predefined set of head gestures is recognised 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.

To the best of our knowledge, we have not yet found any research combining the rotational and translational data from the head tracking into one VE with a focus on support for nonverbal communication.

3 Questions and Hypotheses

The goals of our experiments were to find out

- how well nonverbal clues like head movement and head direction can be perceived, and
- how intuitive and affective HCP is to the user.

We want to analyse the following hypotheses in detail.

3.1 Head Movement

Rotational data from the head tracking can be used to convey head gestures like nodding, shaking, or rolling. We would like to find out whether these gestures are perceived correctly, and how the perception is influenced by jitter introduced by imprecisions of the head tracking.

H1 Head movement can be perceived and correctly identified/categorised by the user.

H2 The head direction can be perceived and the target identified by the user.

H3 Jittery head movement

H3a reduces the chances of correct identification of movement/direction compared to jitter-free head movement.

H3b is perceived as being more unnatural than jitter-free head movement.

3.2 Head Coupled Perspective

By using the translational data of the head tracking, the user can control the view by simply moving around physically, e.g., zooming in by moving closer to the screen, peeking around a corner by moving the head sideways.

We would like to find out if this type of control, called Head Coupled Perspective (HCP) is intuitive and improves the experience of the VE.

H4 Users can control their view faster using HCP compared to manual control.

H5 The accuracy of 3D perception is better when using HCP compared to manual view control.

H6 The speed of 3D perception is improved when using HCP compared to manual view control.

H7 HCP is easier/more intuitive to use than manual view control.

H8 The user feels more immersed in the VE when using HCP compared to manual view control.

4 Experiment Design

In order to verify our hypotheses, we have designed three experiments. Each experiment is conducted in a separate virtual room. Table 1 illustrates, which room is used to verify which hypothesis. Some rooms are specifically designed for just one hypothesis, some rooms contribute data for several hypotheses at once.

For the evaluation, we are automatically logging keyboard and mouse movement, tracking data, and events. Additionally, we hand out a questionnaire which the participant have to fill out before, throughout, and after the experiments.

Hypothesis	Room(s) 1	Room(s) 2	Room(s) 3	Post-Test
H1	D			
H2			D	
H3a	D/Q			
H3b	Q			
H4		D		
H5			D	
H6			D	
H7		(D)[1]		Q
H8				Q

Table 1: This table indicates which hypothesis is covered by which experiment. 'D' indicates that the validity of the hypothesis is checked by analysing numerical data, e.g., logfiles. 'Q' indicates that the validity of the hypothesis is checked by evaluating questionnaires.

Before the actual set of experiments starts, we let the participants complete a simple task in an introduction room. This task is designed to familiarise the participants with the use of the mouse and the keyboard for moving the avatar and interacting with the environment.

4.1 Room 1

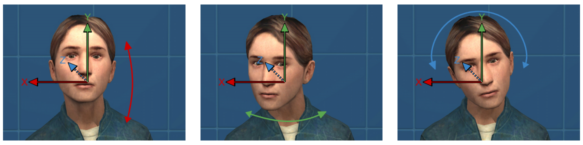
Room 1 is designed to verify whether avatar head movements are perceived by the user, and how much jitter influences the naturalness and recognition rate of head movement.

The objective of the participant is to observe and classify three possible head movements of the opposite avatar in different relative positions to the user. As depicted in Figure 1, the three possible head movements correspond with a rotation around one principal coordinate axis of the head:

X-Axis: Moving up and down (nodding).

Y-Axis: Moving left and right (shaking).

Z-Axis: Rolling left and right (rolling).



(a) Nodding (b) Shaking (c) Rolling

Figure 1: Screenshots showing the three different head movements performed by the avatar in room 1.

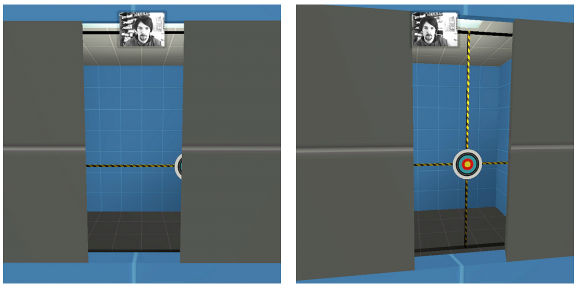
The avatar does each of the possible head movements four times in random order with four varying ratios of amplitude and jitter. The jitter simulates noise in the rotation data that might be introduced by head tracking. By varying the ratio of jitter and movement amplitude, we want to determine the amount of noise that is still tolerable for safe determination of head movements. In addition, each set of movements is repeated with the avatar facing left, right, and away from the user's avatar.

The participant determines the observed head movement by pressing a button with a movement specific icon on it within the VE.

4.2 Room 2

Room 2 is designed to evaluate the influence of HCP on the user (see Figure 2).

The room is split into two halves by a wall with a small viewing slot. This slot is changed in its size during the experiment. Through the slot, the user can see a moveable target disc in the other half of the room. When the user 'looks at the disc' by pointing the viewpoint indicator exactly at the centre of it, the target disc moves to another random location and the user has to follow that movement as fast as possible.



(a) The target disc has moved out of view. (b) By moving the head sideways, the user is now able to see the target disc again.

Figure 2: Screenshots of room 2, showing the necessary head movement for the user to see the target disc in extreme left and right positions.

After every disc movement, the sliders close the viewing slot a little bit more, so that during the course of the experiment, the sliders are more and more likely to obstruct the direct view to the target disc. Then the user has to manually move the avatar slightly left or right to be able to see the disc again.

In another pass of the experiment, the video camera is activated and HCP is enabled. Now, the translational physical movement of the user's head controls the viewport. The participant can now physically 'peek' around the corner of the slot walls.

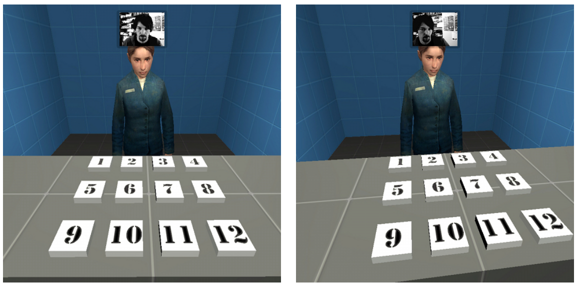
[1] Observation of the participants in room 2 also contributed to the support for the hypothesis. See Section 5.3 for details.

The order of the two passes for manual and tracking-based control is determined randomly to avoid influence of the learning curve on the results.

4.3 Room 3

The direction that an avatar's head is pointing at can be used to indicate objects or people that can then be referred to in a deictic manner ('Can you give me *that*?'). The experiment in room 3, shown in Figure 3, is designed to evaluate how well users can actually perceive this head direction.

The objective of the user is to identify which of the 12 buttons in the middle of the room the opposite avatar is looking at. To get a better idea of the head direction of the opposite avatar, the participant can look at the scene from different perspectives. During one pass of the experiment, the user has to manually move the own avatar sideways. During another pass, similar to the experiment in room 2, HCP is activated and the user can physically control the view. Again, the order of those two passes is determined randomly, to avoid influence of the learning curve.



(a) Centre view. (b) Head has been moved to the left.

Figure 3: Screenshots of room 3, showing sideways head movement for the user to judge the target of the opposite avatar's gaze.

5 Results

5.1 Participants

Participants were recruited 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. No previous knowledge of VEs was required. No financial incentive was given.

We received feedback from 31 people, 10 ($\approx$ 33%) of them being female. The age of the participants covered a range from 20 to more than 60 years, with 15 participants being in the age band of 25-29 years. All participants have used a computer often or more. 11 of them ($\approx$ 35%) indicated that they don't play computer games at all or not any more, 15 ($\approx$ 49%) play between 1 and 5 hours per week, and the other 5 ($\approx$ 16%) play computer games for more than 5 hours per week. 20 of the participants ($\approx$ 64%) have not used a VE before.

5.2 Room 1

The experimental results of the experiment in room 1 are displayed in Figure 4. For each head movement type, the bars represent the amount of correct answers by the participants. The results are grouped by the jitter/movement ratio.

For all ratios < 0.4, the head movement is correctly identified in more than 95% of the cases. Rolling head movement starts to become a problem when the ratio increases to 0.444. Finally, for jitter/movement ratios > 0.5, the rate of correct identification drops down to around 50%.

Interestingly, rolling head movement is more often categorised wrong than shaking or nodding. The reasons for this might be of cultural origin, as this kind of movement is used less often than, for example, shaking as a gesture of negation, or nodding as a gesture of approval.

Analysis of the questionnaires indicated that the participants themselves found it easy to identify the head movement without jitter, but less easy with jitter. The same tendency applies for the realism. Jittery head movement made the avatar appear less realistic than jitter-less movement. A common comment of the participants was that they thought the other avatar had Parkinson's disease.

Hypothesis H1: *Head movement can be perceived and correctly identified/categorised by the user.*

Overall, 84.3% of the avatar head movements were correctly identified by the participants (One Sample t-test, $p < 0.001$, 95% CI 79.93% to 88.61%). This is significantly higher than the value for pure guessing (3 choices = 33.3%) and strongly supports the validity of Hypothesis H1.

We also analysed the results with respect to the direction that the opposite avatar was looking at. The results of a Tukey's HSD test did not show any significant differences for the four directions. The participants were able to perceive the head movements equally well from any direction.

Hypothesis H3a: *Jittery head movement reduces the chances of correct identification of movement/direction compared to jitter-free head movement.*

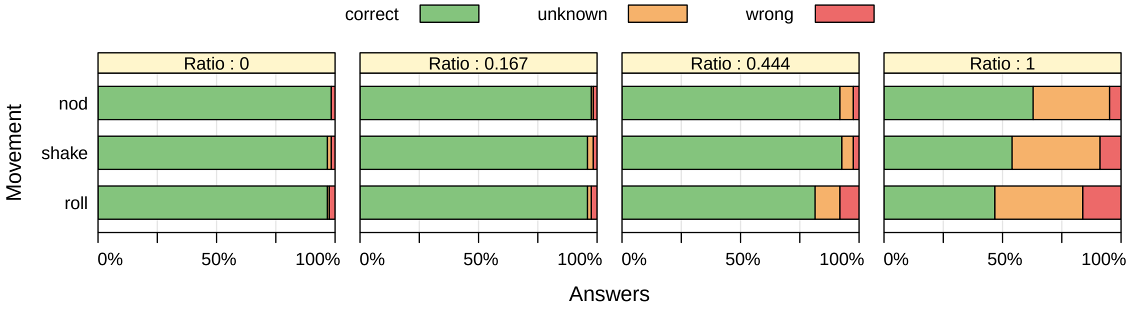


Figure 4: Distribution of correct, 'unknown', and wrong answers to the head movement of the avatar. The results are grouped by the ratio of jitter to the amplitude of the head movement.

If jitter is distracting the user and preventing correct identification of head movement, we expect to see a negative correlation between the amount of jitter and the amount of correct answers. The results confirm this negative correlation ($\rho = -0.63$, Pearson's product-moment correlation, $p < 0.001$, 95% CI -0.73 to -0.51) and therefore strongly support the validity of Hypothesis H3a.

This result is also backed by the questionnaire answers of the participants. For the evaluation, we converted the five steps of the Likert scale to numeric values, ranging from -2 for 'Strongly Disagree' over 0 for 'Neutral' to $+2$ for 'Strongly agree'. Using this conversion, the mean value of the answers to the question 'I could easily identify the head movement without jitter' is 1.7 – the participants agreed very strongly. The question 'I could easily identify the head movement with jitter' is only answered with a mean value of -0.3 – the participants tended slightly towards disagreement (Welch Two Sample t-test, $p < 0.001$, 95% CI 1.59 to 2.41).

Hypothesis H3b: *Jittery head movement is perceived more unnatural than jitter-free head movement.*

We also asked the participants if they perceived the opposite avatar as realistic. They agreed for the avatar without head jitter, and slightly disagreed for the avatar with head jitter (mean values 1.0 / -0.4, Welch Two Sample t-test, $p < 0.001$, 95% CI 0.94 to 1.77). This validates Hypothesis H3b.

5.3 Room 2

The results of the experiment in room 2 are shown in Figure 5. The box-and-whisker plots visualise the average reaction times of all participants throughout the experiment.

Hypothesis H4: *Users can control their view faster using HCP compared to manual control.*

To validate this hypothesis, we measure the correlation between reaction time and the sequence number of the disc movement. For both rooms, the reaction time increases during the experiment, as the slot becomes more and more narrow. However, for HCP, the increase is less pronounced ($\rho = 0.11$, Pearson's product-moment correlation, $p < 0.001$, 95% CI 0.05 to 0.17) than for manual view control ($\rho = 0.15$, Pearson's product-moment correlation, $p < 0.001$, 95% CI 0.09 to 0.21).

We interpret these correlation factors as an indicator for the validity of Hypothesis H4, especially when when detailed control over the view is required.

We also observed that a lot of the participants already moved their heads sideways during the experiment, even when HCP was not yet enabled. This is a strong indicator for how intuitive this method of perspective control is and adds further support for Hypothesis H7.

When participants first encountered the new way of controlling their view with the camera, excitement and positive surprise was a common reaction. We also received some comments about whether it would make sense to also use the head rotation to control the viewing direction. We did some experiments during early stages of development, but this type of control proved problematic because of the jitter in the rotational data. It was not possible to look precisely at smaller objects, even with filtering of the tracking data.

5.4 Room 3

Because of the arrangement of the buttons relative to the avatar and the user, we expected that the participants would have more difficulties identifying the correct button row than the correct button column. The results, shown in Figure 6, proved this assumption correct.

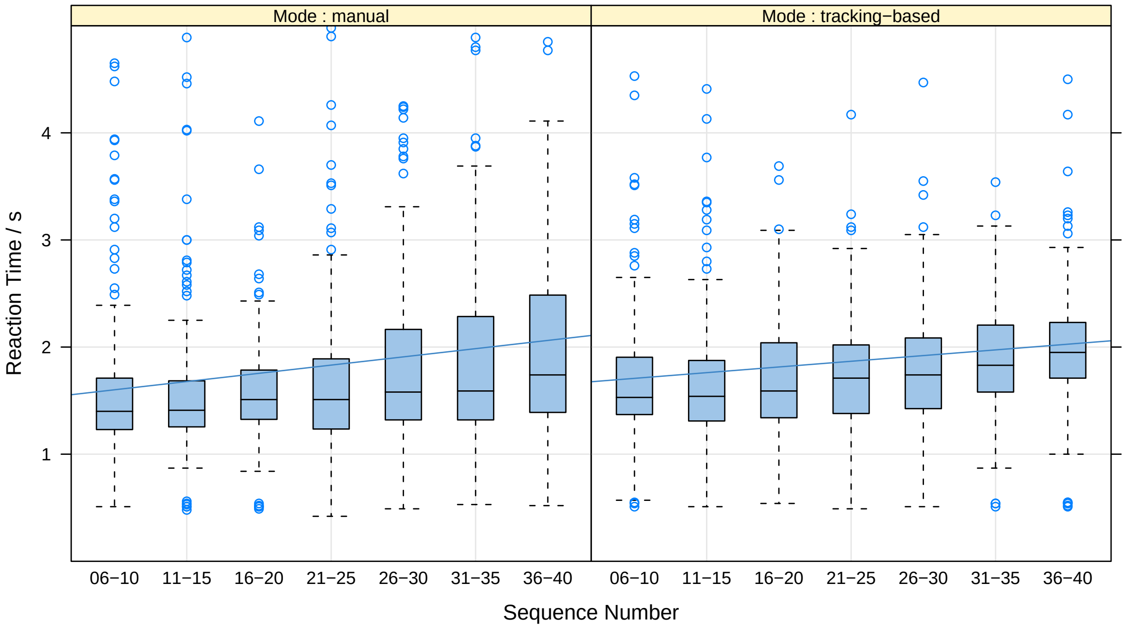


Figure 5: Plot of the reaction time of the participants for both rooms. The initial five results are removed to account for the fact that some participants required some further explanation during these first passes.

Hypothesis H2: *The head direction can be perceived and the target identified by the user.*

Overall, the participants managed to identify the correct button in 45.3% of all cases (One Sample t-test, $p < 0.001$, 95% CI 41.56% to 48.99%). This value is significantly above random guessing (12 choices = 8.3%) and strongly supports the validity of Hypothesis H2.

The participants had significantly less problems in identifying the correct column (mean value 92.2%, One Sample t-test, $p < 0.001$, 95% CI 89.33% to 95.11%). In contrast, the results for identifying the correct row (mean value 47.7%, One Sample t-test, $p < 0.001$, 95% CI 44.32% to 51.10%) are close to pure guessing (3 choices = 33.3%).

Hypothesis H5: *The accuracy of 3D perception is better when using HCP compared to manual view control.*

The mean values of the frequency of correct button choices do not differ significantly for manual control and HCP (mean values 46.0% / 44.6%, Welch Two Sample t-test, $p = 0.670$, 95% CI −5.11% to 7.89%). The null-hypothesis cannot be rejected, therefore, we cannot confirm Hypothesis H5 for the data collected in this experiment.

Hypothesis H6: *The speed of 3D perception is improved when using HCP compared to manual view control.*

For the verification of this hypothesis, we compare the reaction times for manual control and HCP. The average reaction time for manual view control is 0.3 s faster than for tracking-based control (mean values 4.5 s / 4.8 s, Welch Two Sample t-test, $p = 0.138$, 95% CI −0.66 s to 0.09 s). Based on these results, we have to reject Hypothesis H6 for the results of this experiment. When judging the target of an avatar's head direction, tracking-based view control is slightly *slower* than manual.

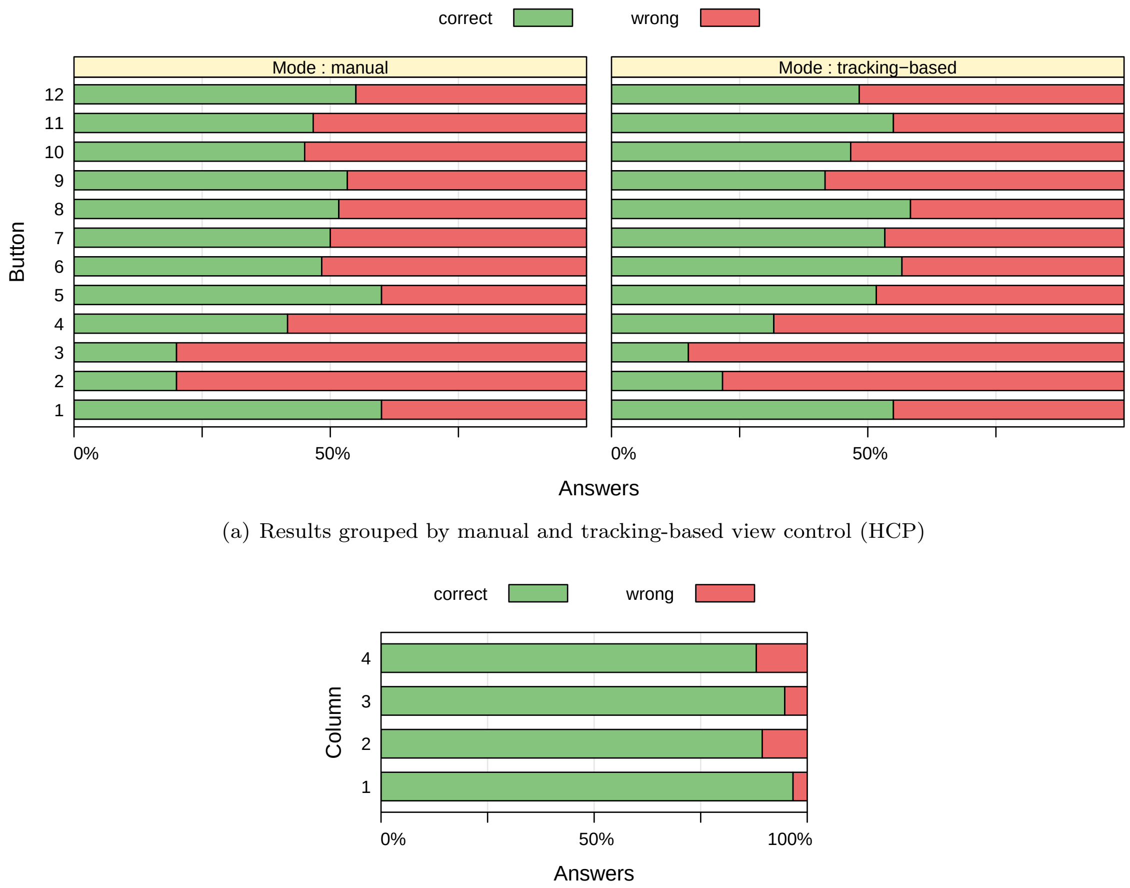
5.5 Post Questionnaire

Hypothesis H7: *HCP is easier/more intuitive to use than manual view control.*

The post-test questionnaire addressed questions especially related to the head tracking. The answers to four of the six questions are evaluated to support this hypothesis.

The participants were asked how much they agree that they had no problems with manual control respectively HCP. The results do not show any significant difference in the answers (mean values 1.1 / 1.2, Welch Two Sample t-test, $p = 0.879$, 95 % CI −0.47 to 0.41). Instead, the high p-value suggests that the participants found manual control and HCP rather equal in ease of use.

When asked for how natural the view control felt, the participants answered with a significant difference in favour of HCP (mean values 0.5 / 1.3, Welch Two Sample t-test, $p < 0.001$, 95 % CI −1.29 to −0.45).



(a) Results grouped by manual and tracking-based view control (HCP)

(b) Results for choosing the correct column

Figure 6: Distribution of correct and wrong choices of the button looked at by the opposite avatar in room 3.

In total, we find support for Hypothesis H7 in the results. HCP does not appear to be easier or more difficult than manual control, but the participants agreed that it is definitely more intuitive.

Hypothesis H8: *The user feels more immersed in the virtual environment when using HCP compared to manual view control.*

We want to acknowledge that two simple questions in the post-test questionnaire cannot cover the amount of nuances and facets of the term ‘presence’. A very thorough questionnaire purely designed for measuring the user’s experience with a simulation can be found in [8].

We asked the participants if they thought that they were more immersed when using tracking-based view control instead of manual control. The participants agreed rather uniformly to this statement (mean value 1.3, One Sample t-test, $p < 0.001$, 95% CI 1.07 to 1.58).

We also were interested if the users subjectively thought that tracking-based view control improved their 3D perception. Again, the answer was thoroughly positive (mean value 1.2, One Sample t-test, $p < 0.001$, 95% CI 0.93 to 1.40).

Based on the answers to these two simple questions, we support the validity of Hypothesis H8.

6 Conclusion

Table 2 summarises the hypotheses and the results of our findings.

In this paper, we have experimentally proven

- that an avatar’s head rotation is perceived and communicates meaningful information,
- that the avatar’s head direction can be used to point in a certain direction that other users can perceive,
- that HCP is neither less nor more accurate than manual view control,
- that HCP is slightly slower than manual view control,
- that HCP is more intuitive for the user than manual control,

- and that HCP has a positive effect on immersion.

With the inclusion of the head rotation as a nonverbal communication channel, we expect an improvement of the communication within a teamwork training scenario, which is likely to result in a better training outcome. We are conducting further studies to support this hypothesis.

Furthermore, HCP simplifies the use of a VE by introducing an intuitive way to control the view. Specifically users who have no experience with using VEs will find it easier to navigate within the environment and to interact with other users and objects. Another advantage is the ability to be able to operate manual devices (e.g. Wii Remote) with both hands and still be able to control the view with the head.

Based on these results, we will conduct further studies to verify a positive influence of camera-based head tracking an a VE-based teamwork training scenario.

Hypothesis	Experimental Result
Head Movement	
- H1	++
- H2	++
- H3a	++
- H3b	++
Head Tracking	
- H4	+
- H5	0
- H6	-
- H7	+
- H8	++

Table 2: This table indicates which hypothesis has been supported by our experimental results. '++' indicates that our results strongly verify the hypothesis. '+' indicates that our results support the hypothesis. '0' indicates that our results can not verify nor falsify the hypothesis. '-' indicates that our results suggest that the hypothesis is not valid. '--' indicates that our results strongly falsify the hypothesis.

References

[1] P. R. Messinger, E. Stroulia, K. Lyons, M. Bone, R. H. Niu, K. Smirnov, and S. Perelgut, "Virtual Worlds – Past, Present, and Future: New Directions in Social Computing," Decision Support Systems, vol. 47, no. 3, pp. 204–228, Jun. 2009.

[2] D. Danforth, M. Procter, R. Heller, R. Chen, and M. Johnson, "Development of Virtual Patient Simulations for Medical Education," Journal of Virtual Worlds Research, vol. 2, no. 2, pp. 3–11, Aug. 2009. [Online]. Available: https://journals.tdl.org/jvwr/issue/view/72

[3] J. Cartmill, A. Moore, D. Butt, and L. Squire, "Surgical Teamwork: Systemic Functional Linguistics and the Analysis of Verbal and Nonverbal Meaning in Surgery," ANZ Journal of Surgery, vol. 77, no. Suppl 1, pp. A79–A79, May 2007.

[4] M. D. Cordea, D. C. Petriu, E. M. Petriu, N. D. Georganas, and T. E. Whalen, "3-D Head Pose Recovery for Interactive Virtual Reality Avatars," IEEE Transactions on Instrumentation and Measurement, vol. 51, no. 4, pp. 640–644, Aug. 2002.

[5] S. Wang, X. Xiong, Y. Xu, C. Wang, W. Zhang, X. Dai, and D. Zhang, "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. New York, NY, USA: ACM, 2006, pp. 1097–1106.

[6] J. Yim, E. Qiu, and T. C. N. Graham, "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. New York, NY, USA: ACM, 2008, pp. 236–239.

[7] T. Sko and H. J. Gardner, Human-Computer Interaction – INTERACT 2009, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, Aug. 2009, vol. 5726/2009, ch. Head Tracking in First-Person Games: Interaction Using a Web-Camera, pp. 342–355.

[8] B. G. Witmer and M. J. Singer, "Measuring Presence in Virtual Environments: A Presence Questionnaire," Presence, vol. 7, no. 3, pp. 225–240, Jun. 1998.