The performance effects of head tracking and viewpoint correction in a task-based environment

Daniel Davison
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente, Enschede,
The Netherlands
davisondp@gmail.com

ABSTRACT
In this paper a first attempt is made to couple new, cheap head tracking techniques with an existing virtual human framework, resulting in a realistic viewpoint correction mechanism. The research focuses on an environment in which the agent gives users tasks, by pointing at specific objects and locations laid out on a tabletop. The idea is that when a user is able to “look around” the pointing arm of the agent, it is less difficult to identify the correct location from which to pick up or put down an object. During an experiment, using an overhead photo camera the exact locations of selected and placed items are recorded. The results are compared between two groups of participants: one group with, and one without head tracking. Based on a selection of the measured variables there exists a significant accuracy advantage for the head tracking group, in particular the ability to select a specific, correct item from a group of identical items. In a different category of measurements the results are inconclusive, requiring further research on the topic of accurately placing objects once they have been picked up.

Keywords
Head tracking, WiiMote, viewpoint correction, task-based environment, virtual human, Eckerlyc, pointing gestures, selection and placement of objects

1. INTRODUCTION
Over the past decade, there have been many breakthrough researches concerning realistic virtual humans, or agents. Over the last years more and more home users have access to high performance PC’s capable of rendering highly realistic 3-dimensional environments. The ever-increasing computing power could also be utilized to explore other aspects of immersive virtual reality, which before were only possible in expensive laboratory setups: a cheap head tracking solution is available in the form of the Nintendo WiiMote, giving programmers a tool to implement viewpoint correction while working with a 3D model. This paper aims to research the effects on human performance in task-based scenarios, when combining these technologies with the representation of a traditional agent. In particular, the focus of this research lies on the ability of a user to accurately identify the correct location where an agent is pointing. Using viewpoint correction in conjunction with a head tracking device can create the illusion of 3D, and more interestingly enables the user to “look around” any objects on the screen. In the case of a pointing virtual agent, the user will be able to shift his body position to judge the pointing location more accurately. Ultimately the cheap head tracking solution could be deployed into environments where a human is working side-by-side with a virtual agent, while performing accuracy-dependent tasks.

1.1 Problem statement
At the time of writing, two important 3D related technologies have matured to a point that they can be deployed into common households. These technologies are summed up as follows:

- Graphical Processing Unit. GPU’s are becoming more and more specialized in dealing with general parallel computations (in addition to generating graphics) [2], resulting in an offload of work from the CPU, which can now be fully utilized for other specialty virtual reality related calculations, like head tracking.
- Tracking techniques. Since the introduction of the Nintendo WiiMote, researchers have discovered its built-in ability to perform as a low-cost, high-precision spatial head tracking device. J.C. Lee is among the pioneer researchers to publish a paper describing the WiiMote tracking technique. [6]

As the research on virtual humans is ever advancing, there is still little to no scientific experience when it comes to combining the above head tracking techniques with a communication experiment between a real human and his virtual counterpart. In particular the field of task-oriented communications, in which the agent instructs the human to perform certain tasks, could benefit from this combined technology. The performance of completing these tasks can be measured in terms of accuracy. Since there are no directly similar researches available, a first glance at the (dis)advantages for a user has been investigated in this paper.

1.2 Research question
The problem described above leads to the following research question:

When communicating with an agent in a task-based environment, does a human benefit from the addition of viewpoint correction?

The following sub questions will be taken into account when researching this topic:

1. How is the correct selection of items influenced by the addition of the head tracking and viewpoint correction?
2. How is the precision of placing items influenced by the addition of head tracking and viewpoint correction?

1.3 Research method
An experiment was used to compare the differences between a setup with, and without viewpoint correction. During the experiment a virtual human would ask the test subject to pick up and place various items on the table in front of him. At each command the agent would point at the specific item, signaling the human which one is meant to be picked up. Using an overhead camera the selection and placement of the items was recorded for later analysis.

Multiple scenarios were setup in a similar fashion to explore the effects in both an easy and a difficult environment. Both scenarios are setup following the same protocol: Three groups of items were placed in front of the test subject: items A, B and C. The agent then instructed the human to perform the following tasks:

1. Item of group A is picked up and moved to a new location
2. Item of group B is picked up and moved to a new location
3. Item of group C is picked up and is applied on the item of group A

The test subjects are split in two groups: one control group and one group using head tracking. All test subjects will perform all scenarios.

The preparation and execution of the experiment was done in collaboration with B.F. Postema [7], however his research focused primarily on the user’s perception of the virtual agent, comparing various subjective factors via a validated questionnaire.

In chapter three of this paper a detailed description is given about the various tools, frameworks and implementations used in the preparation and execution of the experiment. Various choices for specific equipment and frameworks are explained, giving the reader an insight to advantages over other available tools. In the fourth chapter the setup is described in which the experiments and measurements are conducted. Also the various methods used for calibration are explained in detail.

In the fifth chapter gives an overview of the results of this experiment, focusing on the relevant data necessary for answering the research question. These results are further explained and discussed in the sixth chapter, where the reader will also find a description about difficulties and shortcomings encountered in this research. A possible explanation is outlined for any unexpected results, resulting in a suggestion for enhancements in future research.

Finally a conclusion is given, where the research question is answered based on the provided results. The limiting constraints of this research are also outlined, as well as possible directions for future research.

2. RELATED WORK
The work of R.J. Rienks et al. [9] describes an experiment where virtual agents are combined with a head tracking system, in order to provide interactive gaze adjustment. In this research the subtle differences of head orientation of listening and speaking agents are used to investigate the ability for a human observer to distinguish between the two.

In a different area of research, the work of C. Youngblut et al. [11] describes the relationship between perceived presence and measured performance in a task-based environment. In an experiment test subjects were asked to complete certain training missions in a virtual world. Each testing group would complete the training missions aided by a varied degree of presence, after which a similar real-life mission was to be completed. The performance in this real-life mission was a measure for the effectiveness of the respective level of presence used in the training missions.

In a related research by M. Gandy et al. [4] the effects of presence on task performance are studied in an Augmented Reality environment. The participants in the experiment were asked to drop virtual balls on virtual targets, while maneuvering over a three stories deep pit, which was generated using augmented reality. Because of the realism of the frightening pit, a high sense of presence was achieved. Using five different rendering frame rates the level of presence was varied within subjects. The task performance for each frame rate was measured by calculating the accuracy of which the balls were dropped on the targets. Also multiple questionnaires were used to determine the subjective level of presence.

3. MEASUREMENT TOOLS
The execution of the experiment was dependent on a number of tools and implementations, all of which were necessarily working together to produce the desired results.

3.1 Head tracking
First of all an accurate head tracking tool was needed, which should fill the following requirements:
- High resolution and accuracy
- High refresh rate

Since this experiment provided real-time viewpoint correction based on the head tracking data, it was important to have a sufficient refresh rate and resolution accuracy to reduce lag and jittering in the image. In general, a frame rate of at least 15 fps should provide a smooth enough video stream to enable a human to work with. [3]

Based on the above requirements, the Nintendo WiiMote was the head tracking tool of choice. It boasts a refresh rate of 100HZ and provides a captured resolution of 1024*768 pixels in a 20 degree angle of view. [6] Also the WiiMote provides a number of added advantages over conventional head tracking systems:

- Small, lightweight and mobile unit which can be deployed on location within minutes
- Communicates with workstation via wireless Bluetooth
- Open source framework support for many programming languages including Java, C++ and C#
- With a retail price of 40 euro this device is less expensive than most other head tracking solutions

A comparable alternative to the WiiMote is the Microsoft Kinect motion capture device. It supports a lower frame rate and a resolution of only 640x480 pixels, but has the advantage for the user of not needing infrared LED’s for tracking. At the time of writing however, this technology has only just been released and only a few drivers have been developed. [5]

The WiiMote uses an infrared (IR) camera that can capture up to 4 distinct infrared sources, transmitting these as (x, y) points back to the workstation. A special pair of glasses has been constructed, with an IR LED on each side of the head. The WiiMote will capture and transmit locations of both LEDs.

In order to provide true 3D head tracking one must be able to measure the following coordinates, relative to the screen: x, y and z. Here x and y represent the horizontal and vertical position.
of the head, and $z$ represents the distance from the screen. The following equations were executed to calculate the coordinates:

$$ \text{headX} = \sin (\text{radiansPerPixel} \times (\text{moteX} - 512)) \times \text{headZ} $$

$$ \text{headY} = -0.5 + \sin (\text{radiansPerPixel} \times (\text{moteY} - 384) + \text{moteVerticalAngle}) \times \text{headZ} $$

$$ \text{headZ} = (\text{IRDotDistance} / 2) / \tan (\text{radiansPerPixel} \times (\text{detectedDotDistance} / 2) / \text{screenHeight}) $$

The variables presented in the above formulas are explained as follows:

- **radiansPerPixel**: A constant represented by the formula: $\pi / 4 / 1024$
- **moteX, moteY**: The x and y locations of the head, in the 2d captured image from the WiiMote
- **moteVerticalAngle**: The angle in which the WiiMote is positioned
- **IRDotDistance**: The constant distance between both IR dots on the head tracking glasses
- **detectedDotDistance**: The WiiMote measured distance (in pixels) between both IR dots
- **screenHeight**: The height of the screen in millimeters

### 3.2 Virtual human
The experiment required an implementation of a 3D virtual human, or avatar, that can be instructed to perform simple animations accompanied by a text-to-speech mechanism. Furthermore, the environment of the avatar should be easy to manipulate by control software, for example, changing the camera viewpoint and orientation.

The Elckerlyc [8] framework, designed and built by the HMI department of the University of Twente, offers all the above features in a single package. It is programmed in Java and can be extended to support more specific features. Also, its Behavior Markup Language (BML) realizer implements all necessary behavioral commands for conducting the experiment.

The camera position could be adjusted via a Java subclass, directly onto the camera object. After synchronizing these camera position updates with the global refresh timer of the rendering engine, the viewpoint within Elckerlyc will be adjusted in near real-time to reflect the position of the subject’s head.

### 3.3 Connecting it all together
The Bluetooth connection between the workstation and the WiiMote was handled by the WIDCOMM stack driver, which in its turn was controlled by the Java BlueCove stack implementation. Using this BlueCove library it is possible to execute and read Bluetooth commands from within a Java program.

The open source MoteJ library was used for communicating with the WiiMote. This framework can be instructed to perform the basic commands like “connect” and “disconnect”. Also it provides a readout mechanism for the available sensors.

IR point coordinates provided by MoteJ were converted into a 3D Elckerlyc camera position using the equations presented in chapter 3.1. A slight camera rotation was applied, to compensate for the horizontal and vertical shift of the camera, while still allowing the correct viewpoint correction.

Because of the relative heavy load of the Elckerlyc framework, the Bluetooth connection and raw head tracking data collecting were conducted on a physically different workstation. Both workstations were then connected together using a Client-Server model, where the Client in fact controlled the Server by sending it head tracking coordinates.

In Figure 1 an overview of the entire setup is captured, showing the relative locations of the screen running Elckerlyc, the WiiMote and the person wearing the IR LED glasses.

![Figure 1. Overview of head tracking setup](image)

### 3.4 Viewpoint calibration
For accurate and immersive viewpoint correction one must calibrate the relationship between the calculated head position and the camera position within the virtual world. Since the focus lies on a precise pointing action performed by an agent, first rough calibrations were done by aligning the avatar’s pointing arm with a physical object. Different parameters were tested making sure the pointing direction was consistent when switching viewpoints.

The initial rough calibration was further fine-tuned by comparing rendered onscreen objects with physical objects placed behind the screen for reference. Once the perspective of both virtual and physical objects lined up from multiple viewing angles, calibration was complete.

It must be noted however, that this calibration method relies on the same subjective principles as are present during the experiment, namely the perceived viewpoint from the eyes of the user. Although in the case of the calibration, a higher accuracy was achieved by comparing physical objects with their rendered counterparts, and by testing multiple pointing locations from various viewpoints. It is therefore likely that the calibration results for the viewpoint correction offered sufficient realism for usage during the experiment.

### 4. MEASUREMENT SETUP
In order to answer the research questions a number of measurements were performed during the testing sessions, namely the selection of the (correct or incorrect) items and the re-placement of the items.

#### 4.1 Tabletop grid
The experiment was performed in a static, controlled environment where the use of a one-time calibrated grid was justified. In order to persuade the test subjects to place the items freely as opposed to placing them on a cross point the gridlines were not drawn fully, but were marked on the outer edge of the tabletop. See Figure 1 from [7].

The grid itself was divided into 10 horizontal and 7 vertical coordinates, resulting in 70 possible locations on which to place the initial items. This method was extremely flexible and has proven useful during pilot tests to adjust the scenarios on-the-fly, without having to recalibrate after every change.
4.2 Grid calibration

In order to calibrate all locations on the grid an initial calibration was made of the four corner points. The Elckerlyc agent would point roughly at the location of a corner while noting the virtual coordinates for later reference. This process was repeated for all four corners.

All intermittent grid coordinates could now be calculated by interpolating the results from 3 corner points. As a means of checking the accuracy, the location of the fourth corner point was calculated from the other three. Comparing this calculated result with the calibrated coordinates for that point gave an indication of the accuracy of the calibration. The overall calibration was repeated until this fault was negligible.

An independent volunteer was asked to verify the resulting calibration by comparing the location of actual grid points with the location of where the agent was pointing.

As noted with the viewpoint calibration, this grid calibration method also relies on the same perceptive principles as those used during the experiment. In the discussion chapter an alternative grid calibration method is suggested. For this pilot experiment however, the calculated check of the alignment fault combined with the verification of the volunteer, should provide an accurate enough grid for the test subjects. Also, since the initial grid calibration was kept constant during the execution of the experiment, all participants would be subject to the same misalignment (if any).

4.3 Photographic results

In order to determine the selection and placement of the items, numerous photos were made during the experiment. A digital camera was suspended directly above the tabletop, aiming downwards. Using a remote trigger the images were recorded without interfering with the experiment. Because of the static nature of the setup, the gridlines are equivalent on all images and can be compared individually.

After uniformly straightening and resizing the images of the grid in a post-processing step, the resulting placement of each individual item was noted for each test subject.

4.4 Subject information

So as not to influence the performance of the subjects, caution was taken while giving instructions about the test. The test subjects were not informed that the accuracy of selection was a measuring factor, nor were they told what the goal of the research was.

At the beginning of the test the virtual agent would first introduce herself and give a short explanation about the experiment. During this explanation the test subjects were able to get used to the voice and appearance of the agent, as well as the viewpoint correction. The agent used a text-to-speech converter and was able to speak entire sentences in natural language. Animation of facial muscles during speech was automatically provided by the Elckerlyc framework.

After this initial instruction the subjects received further explanation from the researchers.

4.5 Test scenarios

As described in chapter 1.3 the experiment consisted of two scenarios: one easy and one difficult.

In the easy scenario (see Figure 2) all items were arranged parallel to the screen and the test subject. This setup was chosen to eliminate the depth factor from the pointing gestures during the item selection instructions.

Figure 2. Bird’s eye view of item groups in easy scenario

The difficult scenario (see Figure 3) placed the items perpendicular to the screen, thus making the pointing gestures for the selection of the objects fully reliant on the depth perception of the test subject.

Figure 3. Bird’s eye view of item groups in difficult scenario

Each scenario consisted of three groups of items: A, B and C. All items in the same group were identical. Correct items must be selected, picked up and moved to new locations, following commands given by the agent.

A description of the various different tasks for each item is given in the introduction of this paper. However the inner workings of each scenario are of less interest when comparing the performance results, thus these details are not repeated here.

The research focuses on two different performance factors:

- **Selection of correct item.** The agent pointed at one particular item in each of the item groups, asking the human to pick up that item. Whether or not the correct item is selected by the test subject is an indication for his or her performance.

- **Accurate placement of item.** After selecting and picking up an item, the agent would ask the test subject to place it back on the table, pointing at an empty location in front of the screen. The exact location of the placement is recorded and can be analyzed to reveal the accuracy as a performance measure. In both scenarios the target locations for the placement of the items were identical. This was done to make sure the results could be compared between the scenarios.
5. RESULTS
In total 31 test subjects participated in the experiment, about 80 percent was male and all were aged between 18 and 26 years. All subjects had a technical background and were students at the University of Twente at the time of testing.

Of the 31 test subjects 16 formed the head tracking group, while the rest formed the control group. As previously mentioned, all tests were identical for each participant. This enables us to effectively compare all data between the groups as well as comparing specific results.

All results and conclusions of the various T-tests are based on an alpha value of 0.05

5.1 Selection of items
After each experimentation session the recorded images of all subjects were studied and the results were gathered. For each correctly picked up item ‘1’ was noted, while a ‘0’ was noted for each incorrect item. Counting all ‘1’ s yields the number of correctly selected items, as is illustrated in Table 1.

Table 1. Total number of correctly selected items per scenario

<table>
<thead>
<tr>
<th>Scenario 1 no HT</th>
<th>Item A</th>
<th>Item B</th>
<th>Item C</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Scenario 2 no HT</td>
<td>10</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Scenario 1 with HT</td>
<td>4</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Scenario 2 with HT</td>
<td>11</td>
<td>13</td>
<td>3</td>
</tr>
</tbody>
</table>

Since these results are not fully conclusive, a further concatenation of the results is made to show only the total number of correctly selected items for each scenario, this is illustrated in Figure 4.

Figure 4. Total number of correctly selected items for each test group

An independent-samples T-test shows that it is significantly more likely that a person with head tracking and viewpoint correction will select and pick up the correct item, when compared to the control case: T(29)=2.327, p=0.027. This confirms the expectation that, overall, the addition of head

tracking and viewpoint correction will result in a more accurate selection of items by the test subject. The means and standard deviation for both groups are given in Table 2.

Table 2. Mean and standard deviation of the total selection of correct items

<table>
<thead>
<tr>
<th>selection_sum</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>15</td>
<td>3,0000</td>
<td>84515</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>3,8125</td>
<td>104682</td>
</tr>
</tbody>
</table>

Further analysis of the selection data shows that the second scenario was indeed more difficult than the first, resulting in less correctly selected items. Using a paired-samples T-test the following values are concluded: T(30)=2.327, p=0.027. This confirms the expectation that it is more difficult to see the depth of where the agent is pointing, compared to the pointing breadth.

However, comparing the statistics for each of the scenarios from the head tracking group with the non-head tracking group yields no significant results: T(29)=1.529, p=0.137 for the first scenario and T(29)=1.437, p=0.161 for the second scenario.

5.2 Placement of items
The results extracted from the captured images are unusable in their raw form. Therefore each coordinate was first subtracted from the target location’s coordinate, giving the x and y offset. Then, using the Pythagoras formula, the absolute distance from the target is calculated. The resolution of the coordinates is based on the resolution of the captured image: 4272 * 2848.

For each test subject the total offset is calculated by summing up distances of the four individual placements of the items. This resulting total offset is not significantly different when comparing both groups with and without head tracking. However the means and standard deviation are given in Table 3 for clarification purposes.

Table 3. Mean and standard deviation of the total placement offset for both testing groups

<table>
<thead>
<tr>
<th>Total offset</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>Mean</td>
<td>3143.59</td>
<td>2881.82</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>1455.660</td>
<td>1259.869</td>
</tr>
<tr>
<td>Std. Error Mean</td>
<td>375.850</td>
<td>314.967</td>
</tr>
</tbody>
</table>

Instead of looking at the total displacement for each participant, more focus will lie on each placement offset individually. For each test subject four separate placement measurements are taken, resulting in a total result set of 124 (60 without, and 64 with head tracking).

An independent-samples T-test, comparing the two test groups, reveals no significant results when run over both scenarios: T(122)=0.909, p=0.365.

However when the T-test is executed on only the second scenario, a trend does seem to emerge: T(60)=1.781, p=0.08.
Although this cannot be concluded as a significant result, it is an indication that further research could prove to be interesting. In the discussion chapter a possible explanation is given for the obvious difference between the first and second scenario.

The selection results of chapter 5.1 suggest that it is easier to recognize the x coordinate of the pointing location, than to recognize the y coordinate. This same outcome is also true for the placement test, where average total displacement per test subject on the x-axis is 625 pixels while on the y-axis this is 2868 (Std. Dev. 277 and 1383 respectively). A paired-samples T-test reveals this difference is significant: T(30)=-9.385, p<0.000.

When the displacement on only the x-axis between the two groups (with and without viewpoint correction) is compared using an independent-samples T-test, one finds that items are placed significantly more accurately when viewpoint correction is used: T(29)=2.093, p<0.045.

This is however not the case when the results on the y-axis are compared. No significant difference has been measured in the depth perception.

5.3 Placement graphs

For illustration purposes the locations of all placed items have been graphed in Figure 5. In this graph all coordinates of both items have been plotted, giving a good overview of the spread. The intended locations for the first and second items were (2687, 660) and (1546, 660) respectively. One might notice that these target coordinates are not in the center of the spread; this shows that most participants were quite a bit off on the y-axis, because of the depth judgment difficulties.

6. DISCUSSION

A number of questions can be answered when looking at the results. For example it becomes clear that in fact the addition of head tracking and viewpoint correction will (in general) significantly improve the ability for a human to distinguish and select one of several objects an avatar is pointing at.

Furthermore, the data obtained from the selection tests shows that the second scenario (where the objects are placed perpendicular to the screen) is significantly more difficult to complete than the first scenario. This result is further backed-up by the data obtained from the placement tests, where participants performed significantly more difficult to judge the depth of placement (on the y-axis), than judging the breadth of placement (on the x-axis). It is however not conclusive whether or not the addition of viewpoint correction gives an advantage in either of these specific scenarios, when looking at the selection of items. But based on the significant results obtained over the entire dataset (enveloping the results from both scenarios) it is in the line of expectation that, also for both scenarios individually, the addition of viewpoint correction is beneficial. Further research, with a larger resulting dataset, is however necessary in order to prove this expectation.

When looking at the placement tests, in particular the displacement on the x-axis, the results conclude that test subjects were placing the objects significantly more accurately when aided by head tracking. The displacement on the y-axis was however unaffected by the addition of head tracking and viewpoint correction. Perhaps future research could focus on this factor.

In the previous chapter it is suggested (with a p of 0.08) that in the second scenario, objects were placed more accurately (on both the x- and y-axis combined, i.e. absolute displacement) by the head tracking group when compared to the control group. In the first scenario there is no significant difference between both groups. This phenomenon can perhaps be explained by the learning effect that has taken place while wearing the head tracking device. In the second scenario test subjects had already become used to the head tracking and were starting to use it to their advantage to place the items. Further analysis of the data proves that the group wearing head tracking performed significantly better in the second scenario than in the first: T(15)=2.364, p=0.032. The control group, however, performs equally well in both scenarios.

The fact that there is no significant difference in placement over the y-axis differs from the expected result. Perhaps this can be

Figure 5. Placement graph illustrating the location spread for each item, with or without head tracking.
explained by the setup of the experiment, in which the participants were seated in a relative stationary position from where mostly only sideways movement was possible. The expectation is that in an environment where the subjects were able to move more freely (i.e. height and distance variation from screen), the difference in accuracy on the y-axis between both groups would become more apparent.

6.1 Obstacles encountered during setup
Because this experiment was the first to combine several techniques and technologies, there was little knowledge and experience as to how the various frameworks could be connected. A total of three existing technologies/techniques were coupled together in order to execute the experiments: the WiiMote provided head tracking, from this the viewpoint correction was calculated, which was then applied to the Eickerlyc framework.

The limited time-frame of eight weeks in which the research was conducted meant that not all bugs in the implementation could be fixed; instead a quick workaround would be set in place. Also several possible enhancements to the experiment setup could not be implemented.

For example the Bluetooth communication channel between the WiiMote and the PC proved to be somewhat temperamental, sometimes refusing to give a connection at all. Because of the experimental nature of the MoteJ framework (used to communicate with the WiiMote), very specific Bluetooth drivers were necessary, requiring a number of extra configuration settings in order to make the connection. Also, the frame rate of the viewpoint correction in Eickerlyc was lower than expected, usually giving only 15 fps. Participants remarked that this frame rate was less than desirable, and attention should be given to raise this to an acceptable rate. However, it should be noted that research shows that even frame rates as low as 10 fps could still provide enough performance. [3]

In chapter 4.2 we remarked that the executed grid calibration method was perhaps too heavily reliant on subjective perception. In a follow up experiment, perhaps a more objective calibration method could be used. For example relying on a secondary WiiMote sensor mounted above the grid. Using a single IR LED, the locations of the corner grids could be saved. Then the formulas used for head tracking could be slightly rewritten and applied to these obtained locations, in order to objectively map the real world coordinates onto the Eickerlyc coordinates. A different calibration method is proposed by R.Y.D. Xu [10]: A pan-tilt-zoom camera scans the surface for a grid-pattern and can calculate the calibration point accordingly. Using a transformation matrix, the obtained grid locations could be mapped onto the virtual coordinates.

Some participants also remarked that the IR LED glasses were bulky and did not fit very well, resulting in a distraction from the actual experiment. Perhaps, in future research, a more advanced pair of IR glasses could be used.

Given more time, the overall presentation of the virtual agent could perhaps have been tweaked to provide a more natural experience. Parameters such as the voice, the appearance and the animation of the agent could be changed in order to give a less distracting computer generated feeling. Participants regularly commented on the obvious computer generated speech and relative stiff movement patterns. The paper of B.F. Postema [7] also mentions these factors as being of possible influence for the perceived presence of the virtual human.

6.2 Unclear tasks
The participants were purposely not informed about the hypothesis of the experiment, nor were they explicitly informed that their performance, in terms of selection and placement accuracy, would be measured. For some test subjects it was therefore unclear that an object, once picked up, should be placed in a different location than from where it originated. These subjects would place the selected item back in the original location, without giving further notice to the pointing direction of the agent.

When these participants were afterwards confronted with these observations, they confirmed that the commands given by the agent were unclear to them. In a future research care must be taken to ensure a clear set of tasks.

These results were not discarded in the analysis, however, because both groups were affected equally by these unclear commands and could still be compared.

6.3 Improvements to the experiment
Summing up the obstacles encountered and the suggestions given by participants leads to the following list of possible improvements for future experimentation:

- Bluetooth connection stability
- Viewpoint correction frame rate
- Objective calibration process
- Comfortable IR LED glasses
- Natural representation of agent
- Clear unambiguous tasks

7. CONCLUSIONS

This research was carried out to give an initial analysis of the possible performance advantages gained from adding head tracking and viewpoint correction in a task based environment. These advantages were researched via a pointing agent, where performance was measured by the accuracy with which the tasks were completed by the human. The results show a significant performance gain in certain aspects of the measured tasks, while other aspects require further testing.

The correct recognition and selection of a single object from a group of identical objects is significantly less difficult when aided by head tracking and viewpoint correction. These findings in fact answer the first sub question: How is the correct selection of items influenced by the addition of the head tracking and viewpoint correction?

The second sub question, which focuses on the placement of the objects, cannot conclusively be answered with the obtained results from this experiment; the data suggests there is an advantage in locating the correct location on the x-axis, but further testing is necessary in order to achieve significant results. In the relatively static setup of this experiment there were no significant advantages with respect to the placement on the y-axis.

Overall, the main research question can be answered positively: there are in fact performance advantages to be gained when adding head tracking and viewpoint correction to an environment in which a virtual agent gives uses pointing gestures to give tasks to humans.

7.1 Limitations of these results

Because of the homogenous composition of the testing group the results only apply to a small group of the population. In fact, because of the technical nature and required spatial awareness inherent to head tracking and viewpoint correction, the results could prove very different when these experiments are conducted on a broader selection of the population with a less technical background.

In this research the focus lies on the ability to correctly perform tasks given by a virtual agent. As a performance factor the
selection and placement of objects was measured and compared between the different groups. This is only a subset of the possible task-based interactions between a human and a virtual agent, so future research could also focus on different aspects.

This research only focuses on a one-way communication in which the agent gives commands to the human. There is no mechanism for the human to give feedback to the agent, which may be experienced as unnatural for the human when compared with real life applications. In a more elaborate setup these factors could be further explored using a two-way communication mechanism.

7.2 Future work
In a follow up experiment, focus could lie on a more free experiment setup, where test subjects are less constrained to sitting in front of a tabletop and screen. Instead they could perhaps walk freely in front of a life-sized projection of the virtual agent. This setup would allow the y-axis of the placement of objects to be researched, as well as the x-axis and overall displacement. In combination with a clear set of tasks and more participants, this could give more significant results.

Another direction for future research would be to investigate the addition of viewpoint correction in a similar setup, where the human and agent communicate with each other via a two way channel, as opposed to the one-sided communication in this experiment.

In all future research the improvements presented in chapter 6.3 should give a more stable and pleasant environment for the human participant to work in.

Also the results presented in this paper could be compared to a similar setup using (for example) a fully immersive Augmented Reality virtual agent.

Ultimately the future research could focus on real-world applications where the addition of viewpoint correction could for example speed up assembly lines or repair procedures.

8. ACKNOWLEDGEMENTS
Our thanks to all participants for their time and patience during the experiment, to Dennis Reidsma and Betsy van Dijk for their (technical) support and to Björn F. Postema for the good cooperation during the preparation and execution of the experiment.

9. REFERENCES