What are you looking at?

Effects of a movable tablet on gaze direction detection

Author:

Vincent Slieker

Supervisors:

Dr. Pim Haselager

Dr. Jason Farquhar

August 22, 2012

Abstract

In 2007 Schrammel et al. did research on how well people could estimate the gaze location of an embodied agent. They found that participants misinterpreted the agents gaze location with errors up to thirty five centimeter. The agent they used consisted of an avatar displayed on a fixed monitor. We suspected due to previous literature that not being able to physically turn towards a point of interest might be one of the causes for these errors. Bailenson et al. showed in 2002 that participants derived information about a person's gaze behavior from head movements. In our research, we looked at whether physical head like movements of a tablet could improve the gaze direction detection of that agent. We conducted an experiment similar to the one of Schrammel et al. in 2007, using a table on which participants had to locate the gaze location of an agent in front of them. We compared three scenario's labeled static neck, dynamic neck and person. Static neck was a scenario in which participants had to locate the gaze direction of a face shown on a fixed tablet screen, like in the experiment of Schrammel et at. in 2002. Dynamic neck, in which a face was shown on a table's screen which physically turned with the help of a robotic arm toward the gaze location. As last, the person scenario in which we seated a person in front of the participants so that they where face to face. We had 9 participants who all conducted 8 trails in each scenario. When looking at the distance between the targeted and the estimated gaze location, we found some significant differences between the scenarios. The error in distance was significantly smaller in the person scenario than in the static and dynamic neck scenario's, showing that participants could estimate a gaze location of another person better when that person was sitting in front of them rather than being displayed on a screen. We could not find a significant difference between the two tablet conditions, though this might be due to the small number of participants. However, the effect was not significant participants estimated the gaze location in the dynamic neck scenario more consistently than in the static neck scenario, which suggests that physical moments of a tablet do improve a tablet's ability to transfer a gaze location. Making a tablet movable seems to have potential in terms of improving the accuracy in which a gaze direction can be detected, though follow up research in this field is required to provide more conclusive results.

Keywords: Gaze detection, videoconferencing, head-movements, embodiment.

Table of contents

Acknowledgements

Introduction

Video conferencing is increasingly used to communicate over a distance; with the program Skype alone reaching over 663 million registered users in 2011 (Skype grows FY revenues 20%, reaches 663 mln users, 2011). Video conferencing can be used to support long distance business meetings and is being used by governments, companies, and individuals. In 2009, the U.S. Social Security Administration conducted 86,320 videoconference hearings, a 55 percent increase over 2008. The use of video conferencing can be explained by several of the benefits it has. It enables communication over large distance, which reduced the need to travel and provides more visual information that a telephone conversation or mail. Although video conferencing has its benefits, non-verbal communication like your body language and posture do not convey as well through a video camera as they do in person (D.T.Moore, 2008).

Being able to tell where someone is looking can have an effect on how we behave in a conversation. Bailenson et al. (2002) showed that when participants where in a group meeting the length and duration a participant got to speak was depend on the gaze behavior of that participant. These effects produced by gaze behaviors could be amplified by supporting head movements of the participants; this was presumably because participants derived extra information about another participants gaze location by looking at their head orientation.

In 2007 Schrammel et al. did research on how well participants could estimate the gaze location of an avatar displayed in a video conferencing setting. The test results showed that it was difficult for participants to estimate a gaze location projected on a grid in front of them, producing offsets up to thirty-five centimeters. In their research they didn't look for means to improve the estimations of their participants nor did they compare it to a face-to-face situation in which participants had to estimate gaze locations. There are differences between talking to someone face to face or through a video conferencing device. One of the limitations is that video conferencing devices such as laptops are unable to physically move in support of the conversation in the same way a head does. In 2008 Nakanishi et al. published a paper showing that even small movements of a video conferencing device had an effect on how it was perceived by participants. Effects were also shown for the operators of the device; they had a greater feeling of being present in the same room as the observer when being allowed to move their device. Though physical movement has an effect on video conferencing it hasn't been researched whether it could be used to improve the detection of an agents gaze direction.

The aim of our research is to investigate whether the physical movement of a video conferencing screen has an effect on the accuracy with which people can determine their conversations partners gaze location. To do this we built a robotic neck which can move a tablet in head like movements and placed it in a setting comparable to the research setup of Schrammel et al. (2007). To look at whether physical movements have an effect on the detection of gaze direction three different scenarios where compared: video conferencing with a static tablet, a video conferencing setting with dynamic neck making head like movements and a face-to-face scenario as reference. If movement of a tablet has an effect on the gaze direction detection it could be an improvement to video conferencing, helping to establish better non verbal communication in regard to gaze location.

Background

The Importance of human gaze

People interact with each other and their environment in many ways. When people interact, they can communicate verbally or in a non-verbal way, where non-verbal communication is often visually transmitted. Visual non-verbal communication has an important impact on human to human interaction. Take for example a group meeting where people can't talk all at the same time and non-verbal communication between the participants helps in keeping the discussion structured. A participant can influence the turn taking process by a non verbal-communication channel like gaze. The direction and length of a person's gaze influence the duration of a turn and the amount of times he or she gets a turn, as showed in research of Bailenson et al. (2002) The length and direction of a person's gaze indicate on what that person's attention is focused. Being able to tell what other people are focused on is important for humans and is learned at a very early age. Newborn babies already look longer at facial photographs that show direct gaze than the ones showing an averted gaze (Farroni, 2002). Baby's actually engage in a dyadic interaction: a mutual gaze between the newborn and the mother. Later in life, infants interact with people in a triadic way which may be used to establish joint attention, the common look at a point of interest and the following and directing of gaze. Gaze following can be seen as looking where some else is looking, this in order to identify a point of interest for example (Butterword, 1991). Some of the main functions of gaze is to regulate face to face social interaction and to draw attention, which are attributes that are also important in a meeting using a video interaction. However, research done by Gemmell et al. (2000) showed that the effects of a person's gaze on a conversation while interacting through a video feed alone is diminished compared to face to face interaction. In a video conversation, people are less effective in establishing a non-verbal communication (D.Hogan, 2008).

Gaze awareness in teleconferencing

Many devices and software enable people to interact with each other through some forms of video feed. Mobile phones and computers can enable video conversation thanks to software like Skype (Skype grows FY revenues 20%, reaches 663 mln users, 2011). One of the limitations of video conferencing is that you cannot use the properties of non verbal communication as much as in a face to face interaction. Body language and posture aren't going to convey as well through the video camera as they do in person. Also turning your head has a different impact because the screen in which it is broadcasted is physically unable to move. Head movements are an important way for people to direct their attention and even when they are only displayed on a screen, they give a good indication about where the given person is looking at. In 2007 Schrammel et al published the paper "Look! – Using the Gaze Direction of Embodied Agents" about the effect of gaze on embodied agents. An embodied agent refers to an agent that can interact with an environment through a physical body within that environment, in Schrammels research this was represented by an avatar. In one of their experiments, they looked at how accurately a test subject could follow the gaze of an avatar on a static screen. They concluded that it was easier for subjects to follow the gaze direction of the avatar when it was looking toward an object placed right in front of it, though subjects still made a significant error in guessing the spatial location of the gaze. When the avatar was looking to its right or left side, the accuracy in which the subjects were able to follow the gaze decreased drastically. The results of this experiment are shown on figure 1, taken from their paper. To conclude, a gaze can give subjects a sense of direction, but it is difficult for subjects to pinpoint it to an exact location (J.Schrammel, 2007). If people can pinpoint the gaze direction of another distant person through a telepresence device in a better way, this might provide interaction some advantages in joint attention tasks. The capacity to apprehend and follow the gaze in direct human to human interaction is regarded as a critical component in joint attention skills (G. Butterworth, 1991).
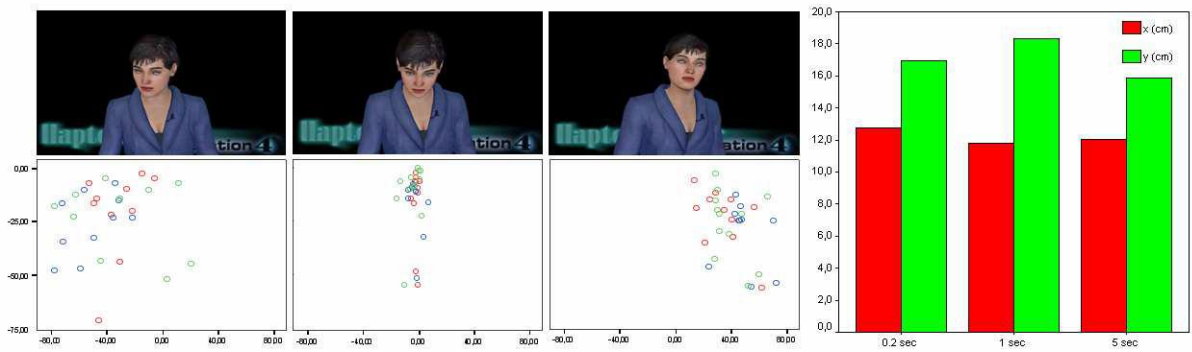
(Figure 1: Scatter plots showing the identified locations of three examples pose (J.Schrammel, 2007))

For humans, it is important to be able to move their head while following a gaze direction. In addition, research of Biguer et al. (1984) showed that when subjects were unable to move their head, they would make a larger error in their accuracy while trying to point to a specific place with their hands. Head direction is important for humans for the orientation towards other people's gaze direction. In the human brain, there are some regions of the temporal cortex that respond strongly to head orientations (DI.Perrett, 1992). In the research of Schrammel et al, a static display was used which was unable to imitate physical head movements (J.Schrammel, 2007). Adding physical movements to a display in order to enable that agent to mimic head movements might influence how well the subjects are able to follow this gaze. One way to look at the effects of enabling head movements is by comparing an agent in two different conditions. One condition in which the agents embodiment allows physical head moments, and one condition in which the agent's embodiment doesn't allow for physical head movement.

Embodiment
An agent is influenced in many ways by its embodiment. There are the basic physical abilities that are or aren't allowed by a certain type of embodiment, but there are also some more complex effects on the agent or its environment. There are many findings that connect social cognition to embodiment. Barsalou et al mentioned four categories where social cognition is influenced by embodiment in their paper about social embodiment (L.W. Barsalou, 2003). First, the perception of social stimuli produces cognitive as well as bodily states. Second, perceived bodily states in others produce bodily mimicry. Third, affective states are produced by bodily states in the self. Fourth, the compatibility of cognitive states and bodily states modulate the performance effectiveness. These four statements about social cognition refer to the embodiment as states of the body such as facial expressions, arm movements and postures that can also apply to telepresence devices. In "Minimum Movement Matters" Nakanishi et al conducted experiments with teleprocess devices and the impact movement had on the operator and the observers of those devices (H.Nakanishi, 2008). In one of the experiments, a telepresence device was used in four different conditions. The first condition was to create the conditions where the telepresence device was static, meaning that no movement was allowed. The second condition allowed the device to rotate his head. The third condition allowed movements through the room but no rotations of the head. The fourth and last condition allowed movements through the room and rotation of the head. In these conditions, the operator of the device was located in a different room than the observer. Comparison revealed a strong effect of movement on the social presence experienced by both the operator and the observer of the device. Both operator and observer had the feeling of being more in the same room. The effect these forms of embodiment have on social presence can be compared to the earlier mentioned effects on cognition by Barsalou et al. Having an agent embodied in such a

way that he or she can move has an influence on the interaction between that agent and another person.

For our research we used an agent that was embodied in a robotic neck with a tablets screen, this embodiment allowed for both physical movements of the screen by moving the robot neck and digital movements of de agents displayed on the screen. With digital movements we refer to the head movements of the agents showed in the screens display. Having an agent embodied more physically is different from displaying an agent's body digitally as shown in research of Wainer et al. In their work "Embodiment and Human-Robot Interaction", they conducted experiments in order to evaluate whether physical embodiment has a measurable effect on performance of participants on a game and whether it influenced their impression of social interaction (J. Wainer, 2008). In order to test these effects, they used a robot capable of giving instructions on the game Tower of Hanoi. Participants of the experiment were asked to play the game while being coached by the robot. In one of the conditions, the robot would be physically present in front of them giving automated feedback on the game. In another condition, the robot was placed somewhere else and was shown through a live video feed though which it conducted the same behavior as in the first condition.

(a) The participant interacts directly with a physical robot.

(b) The participant interacts with a physical robot over a real-time video-conferencing link.

(Figure 2: Experimental conditions in the research of Wainer et al (J. Wainer, 2008))

A time-line was recorded in order to register the interaction done by the robot and the moves made by the participant. The total time spent on the task was used in order to measure the effectiveness of the different conditions. The participants were also given a questionnaire that was used for the evaluation of social aspects. The research concluded that the participant's impressions of a robot's watchfulness, helpfulness, and enjoy ability where significantly affected by different embodiments.

The differences in embodiment can have a significant effect on the interaction with that agent. In our research we looked at the effects of a movable tablet on gaze direction detection. We wanted to know if physical head-like-movements of a tablet would have an influence on the precision in which participants could pinpoint the gaze direction of an agent shown on that tablets screen. Therefore, we choose to use an embodiment only consisting in a head and a neck. In one setup, the embodiment enables the neck to move and in another setup, the neck is not allowed to move. I will refer to setup as dynamic when physical movement is allowed and as static when it is not allowed to move. A static screen which projects a video feed of facial expressions is not as effective in directing gaze as it is the case in a face to face conversation (J.Schrammel, 2007).The differences between real life face to face interactions and teleconferencing are partly created because of the limitation in the way a telepresence device embodies a person. One of the limitations you see in video conversations is that in case of

embodiment on a static screen, there is no way to turn physically towards or away from something. The assumption is therefore that the dynamic setup which allows for physical movements will embody a person better.

In order to test whether there is an effect of (robotic) head-movements on the detection of the gaze direction in video based telepresence robots, we were fortunate to have a robot neck built for us by the *Technical Support Group* (TSG), capable of carrying a tablet thereby allowing head-like movements. The robot can be preprogrammed with some series of movements that can be replayed afterwards. The robotic neck holds a tablet that can display a video stream with or without neck-movements and which has 3 degree's of freedom.
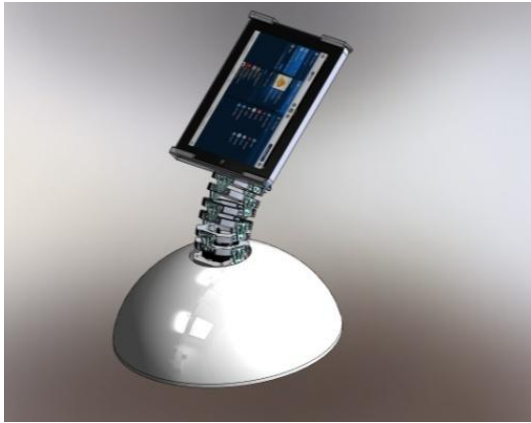
(Figure 3: Virtual representation of our robotic neck)

The research setup

The setup

We looked at whether the physical embodiment of head movements can enhance the precision in which the gaze direction of an agent can be detected. Based on the literature discussed earlier we had the following hypothesis. First, a person can detect the gaze direction of another person more accurately when that person is physically present instead of represented by a video on a tablet. Second, a person can detect the gaze direction of agent represented on a movable tablet better that of on represented on a static tablet. These hypotheses are based on our assumption that neck movements have a significant effect on how precisely a gaze direction can be detected. In order to test our expectations we created 3 different scenarios.

The three scenarios:
- Scenario A: The static neck. The robot's tablet will display a video of a face that looks towards a specific point while the tablet is kept in a default position by the robot neck.
- Scenario B: The dynamic neck. The robot's tablet will display a video of a face that looks straight ahead. The physical movements of the robot neck will be the only indication of its gaze direction.
- Scenario C: A person. A person will direct its gaze towards a specific point on the grid using his neck while his eyes look in the same direction as his head is pointing.

In each scenario a participant had to sit in front of a grid with a sender in front of him or her. The sender was represented by an agent or our preprogrammed robot-neck showing a prerecorded video of that agent on its tablet. In each scenario the sender looked at specific point on the grid for 3 seconds after which he returned to a default position. At the beginning and the end of the 3 seconds a sound indicated the start and end time for a participant to make a decision. Within the given time a participant had to determine as accurately as possible what the

senders gaze location was by putting a colored block on that location on the grid. Participants received no feedback on the accuracy of their assessments during the experiment. To evaluate how accurate the guesses of the test subject where, we used the coordinates of the locations on the grid where the blocks where placed. All actions of the participants on the grid where recorded so they could be evaluated with sufficient time afterwards.
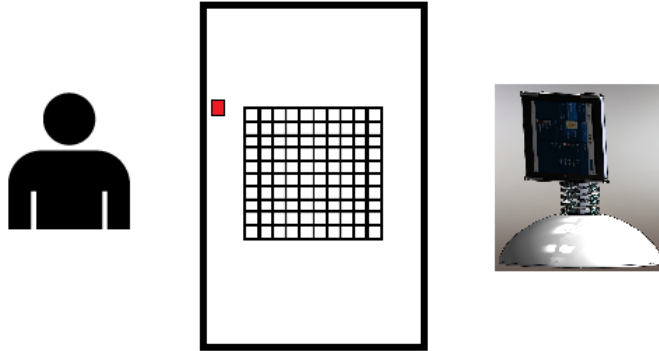

(Figure 4: Setup of the table for the gaze tracing experiment)

After the experiment, the participants had to fill in a form with some question about their background and experience in this area, in order to explain possible anomalies in the results. Our form included questions about the more social aspect of the experiment to examine how well the subjects believed they were able to follow a gaze and how present they perceived the sender to be within a certain setup. The questions about social aspects where included in order to draw more specific conclusions regarding whether physical mobility with tablets would be an improvement in joint attention tasks. The questionnaire can be found in appendix A.
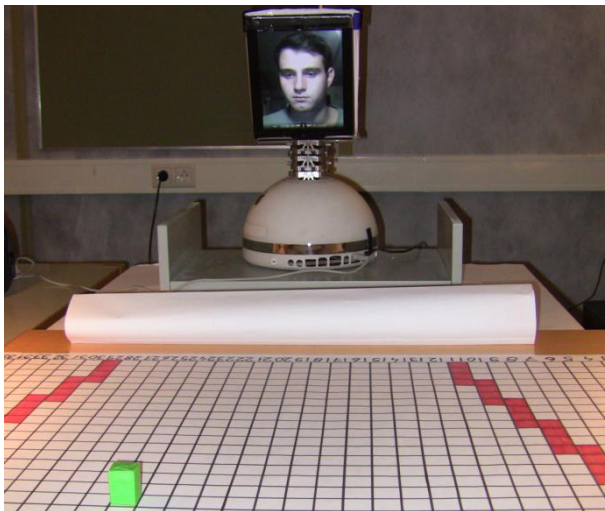
The grid

Schrammel et al. (2007) used a squared grid when trying to measure how well participants could determine the gaze location of a virtual agent. Difference between the guessed and targeted gaze location where measured on an x and y- axis. Participants determent the gaze locations of the agent with errors up to a distance of nearly thirty-five centimeter between the targeted and the gaze location. Figure 6 shows the grid in one of the experimental setups used in their research. Our experimental setup resembles that of Schrammel et al. (2007) with having a participant sitting in front of an agent with a grid between them. We used these results and a pilot experiment to determine a suitable field size of 90 by 65 centimeter with a grid made up of squire with a diameter of 2, 5 centimeter. Though a smaller squire size could increase the resolution of our results, it would also make it more difficult for our human sender to accurately gaze at the target point. The sides of the field where also marked with horizontal value's between 1-35 and vertical value's between A-X in order to reduce the chances that our human sender would misjudge a targeted gaze location. Though an error of judgment by our sender would be detectable in the video recording of the experiment later on we choose to minimize that risk.
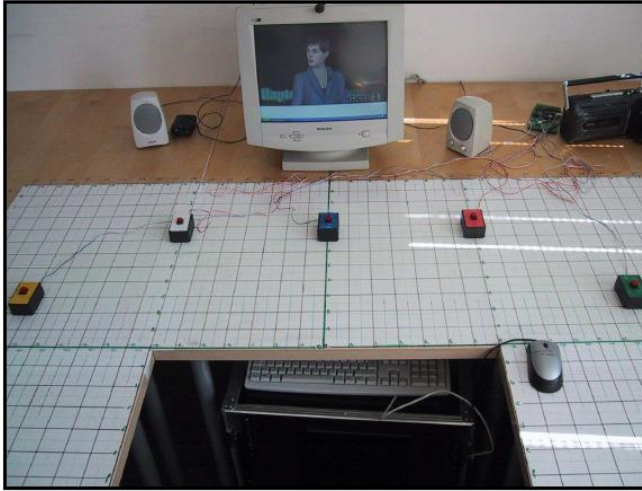
(Figure 5: Experimental setup used by Schrammel at al. (2007))

During pilots we concluded that the points on the grid that we could have our senders target differed between scenario's, with the static robot-neck as most limited. We choose to mark an area on the grid that was a viable option of possible target points for all three scenarios, resulting in a V-shaped outline. Three sets of 8 points were selected on the grid for the sender to target. In order to ensure that all three sets contained points that were well separated over the grid and of a comparable difficulty we divided the grid into 8 sub-areas. Each set of points was then selected by randomly picking a point out of each of the 8 sub-areas of the field. The same three set of points where used with every participant, though the order in which the sets where presented was different between participants.

(Figure 6: Our Experimental setup showing the grid and our robotic neck)

Gaze directing through a tablet
There can be several problems when video-teleconferencing with regard to the precision of gaze. One of the problems we encountered was a gaze offset introduced by the location of our camera. We used a basic laptop and tablet that both had a camera installed above the centre of the screen. The problem introduced by these cameras was that even though you looked in a video conversation to the middle of your screen you would appear to be looking down. This problem we encountered was also described by Yang et al in their paper Eye Gaze Correction with Stereovision for Video-Teleconferencing, where they argue for a software-based solution for this problem (R.Yang, 2001).
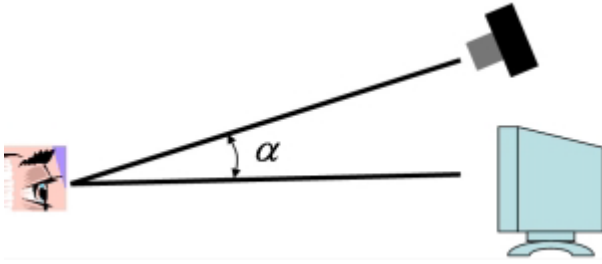
(Figure 7 : Camera-Screen displacement causes the loss of eye-contact. From the paper of Yang et al (2001))

The offset in gaze due to the camera's angle was so large that it made it impossible for participants to make any significant distinctions between gaze locations during pilot tests because all targeted locations seemed to be at the bottom of the grid. We did not have the hardware available to avoid this problem so we adjusted our gaze direction. The agent who's gaze was being filmed was asked to look a few centimeters higher that the point that was actually being targeted on the middle of the screen. This new gaze location was marked by a grid that was placed on the screen after several pilots that provided the correct offset data. New pilots showed that with this adjustment all points on the grid could be target by the agents gaze in a way that clearly distinguished between lower and higher located points. In addition it was shown that a trained accomplice was able to locate more that 90% of the points with an error of less than 7 centimeter showing that our setting provided sufficient feedback for participants to work with. In order to keep the experiment consistent we used the same counter offset in all tablet scenarios. The tablet itself was oriented by the robotic neck so that the center of the agents eyes where in was a straight line with a 90 degree angle from the tablet to an assigned grid location. For scenario A (the static neck) the assigned grid location was the center of the field, where for scenario B (the dynamic neck) the assigned location was the same as the senders target location.

(Figure 8: Robot neck in a static setting left and a dynamic setting right)

In figure 8 you can see both tablet scenarios. On the left you can see scenario A where the sender is targeting a point to the left by turning his head while keeping the tablet in the default position pointing towards the center of the field. On the right you can see scenario B where the agents is looking straight ahead while using the robotic neck to turn the tablet towards a target point.

Data acquisition

Participants

Nine students participated in our experiments, aging between 19 and 24 (6 male). When asked about their handedness six participants reported to be right handed. None of our participants had uncorrected problems with his or her eye sight. Participants that had sight problems but wore glasses or had lenses where considered to have corrected sight. None of our participants had concentration issues, which we checked for in order to minimize the risk of getting a faulty trail. Eight of our participants had little experience with tablets and one had no experience with the use of tablets. When questioned about video conferencing three participants reported to have little to none experience with video conferencing. In summary we can describe our group of participants as inexperienced with tracing gaze locations from a tablet.

Instructions

Participants where only allowed in the experiment room when it was their turn. Once a participant entered the room he or she was requested to fill in a form about their characteristics and continue reading the instruction if no eye or concentration issue's where reported. If a participant had a question about the form or was finished reading it he or she was instructed to signal the experimenter.

The instruction:

"You will be seated in front of a grid where an agent will be sitting in front of you. There will be three rounds in which the agent will gaze at specific points on the grid with a small break between them. Once the agents is gazing toward a location on the grid you will hear a sound, after the first sound you will 3 seconds to determine the location of the agents gaze as accurate as possible. Once the 3 seconds are over you will hear on other sound indicating that the time is up, is you haven't made a decision by then just guess as accurate as you can when the targeted gaze location was. If you have selected a location place the green block on that spot on the grid within a square. Keep the block in place for one second so we have time to recorded your choose and don't to take the block from the grid after each exercise. You will be asked to guess the location of 8 targeted gaze locations after which there will be a small break before the next round. In total there will be 3 round consisting of a total of 24 points. After the final round please don't forget to fill in the back of the evaluation form. "

After the experiment participants were asked to answers the following questions for each scenario's on the following scale [Strongly agree, agree, uncertain, disagree, strongly disagree].

1. I found it difficult to locate the target points
2. I believe my guesses where accurate
3. I had the feeling the sender* and I were a team
4. I believe the sender* to be helpful

* The sender is the agent who gazes at a specific point on the grid

Randomizing

<u>Sets</u>

There are 3 sets of 8 points selected on the grid for the sender the target. Each set of points was present in a fixed order. The sets are numbered: T1 T2 T3.

Scenarios
Scenario A: The static neck, the robot's tablet will display a video of a face that looks towards a specific point while the tablet is kept in a default position by the robot neck.
Scenario B: The dynamic neck, the robot's tablet will display a video of a face that looks straight ahead. The physical movements of the robot neck will be the only indication of its gaze direction.
Scenario C: A human will direct its gaze towards a specific point on the grid using his neck while looking in a straight direction.

Tests
Each participant get randomly assign to one of the 9 tests that has not yet been assigned to another participant. Each test consists of 3 subtests consisting of a scenario and a set of target point for that scenario. Test where constructed in the following way to balance out the scenarios and sets.

	Scenario 1	Set 1	Scenario 2	Set 2	Scenario 3	Set 3
Test 1	A	T1	C	T2	B	T3
Test 2	A	T2	C	T3	B	T1
Test 3	A	T3	C	T1	B	T2
Test 4	B	T1	A	T2	C	T3
Test 5	B	T2	A	T3	C	T1
Test 6	B	T3	A	T1	C	T2
Test 7	C	T1	B	T2	A	T3
Test 8	C	T2	B	T3	A	T1
Test 9	C	T3	B	T1	A	T2

(Table 1: Assignment of tests)

The data
During the experiment a participant was assign one of the test combinations explained earlier. The data directly obtained from a test contained the coordinates the participant had chosen in a certain combination of set and scenario. In table 2 you can see part of the data collected on participant with test 7.

Test	Test1	
Set 1	T1	
Scenario 1	C	
	Assessments	
	Y	X
1	d	17
2	j	28
3	o	26
4	c	24
5	l	14
6	j	19
7	u	8
8	t	25

(Table 2: Test data participant 7)

The data on the assessed target locations of the sender where rated with an error value in order to compare them. The error value was based on the shortest distance between the center of the assessed grid location and that of the correct location in centimeter. We used 2 different methods to establish the correct gaze location for the participants resulting in two different analyses. The first method was looking at the distance between the assessed grid location and the targeted grid location by our sender resulting in an error value labeled as the "direct error". The second method was based on the constancy of the participant's assessments, looking at the distance between the assessed grid location and the average of all assessments of the corresponding target location in similar settings. The error value labeled by the consistency method was labeled as the "relative error". The direct error shows how well our setup preformed in the given condition, where our relative error shows the best obtainable result in our experiment if everything would have been mapped to be as consisted as possible with the participants in order to create a best case scenario.

Calculating the direct and relative error
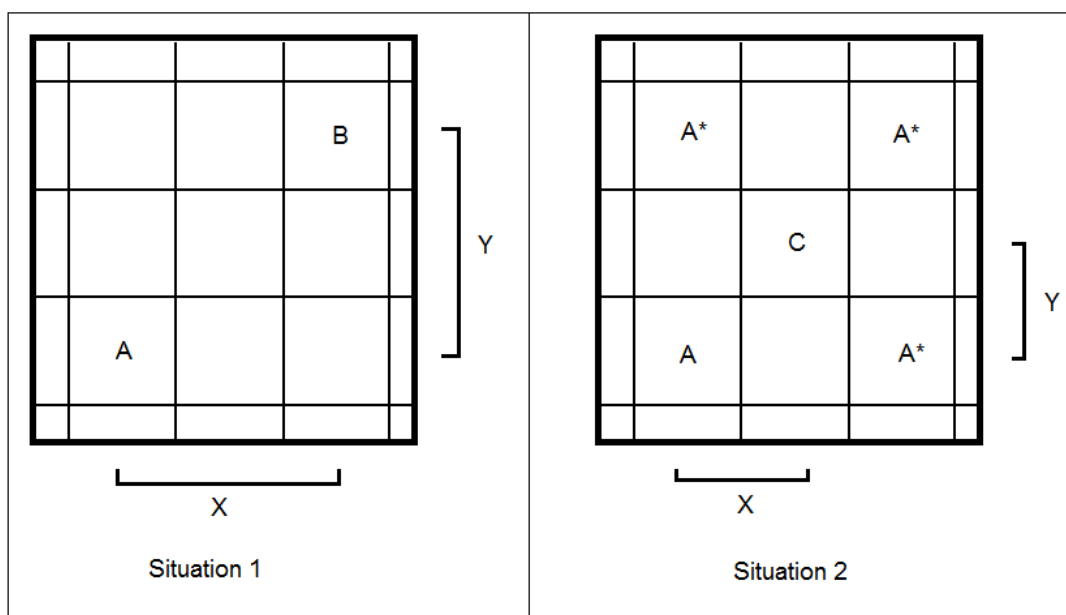
(Figure 9: Displays the buildup of direct error left and relative error right)

In picture 9 you can see an example of how the direct and relative error is calculated. The direct error is calculated in situation 1 by looking at the guessed location on the grid marked A and the targeted location on the grid marked B. The error calculated by taking the distance between the center of point A and B on the X and Y resulting in a direct X and direct Y error. The direct error is then calculated by taking the absolute value provided by $\sqrt{(A^2 + B^2)}$ using Pythagoras theorem. [x]. In situation 2 the relative error is calculated in a similar way, only here point B is replaced by point C. Point C is the point resembling the average of all guessed made by participants in that specific condition. In Situation 2 the guesses of other participants are shown by a A*, together with point A they influence the location of point C. Notice that point C is unlike the other point on the grid not bound to be in the center of a square. Once point C is calculated it will be used with point A in the same way point A and B where used in situation 1. Table 2 shows the new data table after adding the directs en relative error to table 1. The X and Y values of errors can be either negative or positive showing the direction of the error on that axis.

Test	Test1							
Set 1	T1							
Scenario 1	C							
point	Assessments		direct error in cm			relative error in cm		
	Y	X	X	Y	total	Y	X	total
1	d	17	2,5	2,5	3,5	0,8	-0,8	1,2
2	j	28	5,0	2,5	5,6	0,0	-0,8	0,8
3	o	26	-5,0	0,0	5,0	-5,8	-0,8	5,9
4	c	24	5,0	2,5	5,6	2,5	0,0	2,5
5	l	14	0,0	-2,5	2,5	5,0	-9,2	10,4
6	j	19	2,5	7,5	7,9	0,0	4,2	4,2
7	u	8	-7,5	2,5	7,9	-0,8	-1,7	1,9
8	t	25	0,0	-5,0	5,0	-5,8	0,0	5,8
			average		5,38		average	4,09

(Table 2: Assessments measured in terms of the direct and relative error)

Calculating the angles

The relative and direct errors are calculated as a distance between locations on the grid. The disadvantage of using this distance is that its scale depended on how far an agent is away from the grid. Making an error of 20 centimeter when assessing a target point located one meter away can be considered less accurate than when making this same error while located 10 meter away from the target point. In order to take the distance of an agent into account we used a measure based on the angle between an assessment of a participant and the corresponding target point. We calculated both a horizontal and vertical angle error between a point and our agent and combined those into a total angle error.
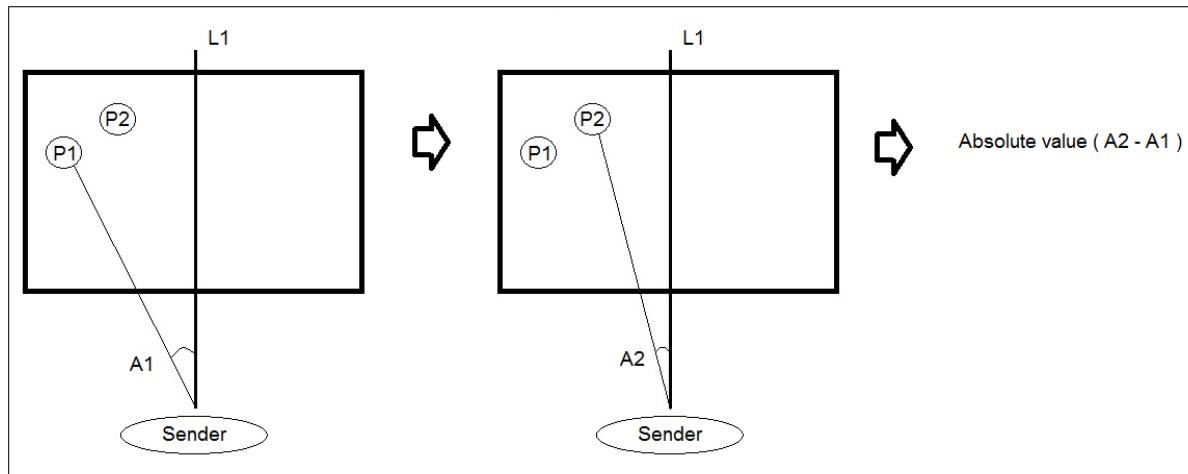
(Figure 10: Calculating the horizontal angle error)

The horizontal angle error is calculated as the difference between the angles's produced by the target point and the point made in a participants assessment with the horizontal location of the sender gaze. In figure 10 you can see a by a participant estimated point called P1. Line L1 is a straight line from middle of the sender through the center of the grid illustrating an angle of zero. The angle produced by point P1 is called A1 and is calculated as the angle between P1 and line L1. The location targeted by our agents gaze is called P2 which produce an angle A2 in the same way as point P1. The absolute difference between angle A1 and A2 is then taken as the horizontal angle error. We didn't calculate the angle between point P1 and P2 directly but in this way in order to sort errors depending on how far points where located from the center of the grid in order to look whether the distance from the center influenced the accuracy in which participants could locate our agents gaze.
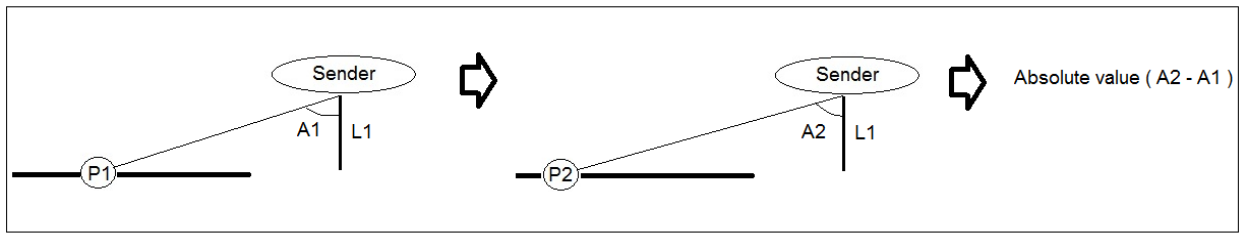
(Figure 11: Calculating the vertical angle error)

The vertical angle error is calculated as the difference between the angles produced by the target point and the point made in a participant's assessment with the horizontal location of the sender gaze. In figure 11 you can see the grid represented by a thick horizontal line with point P1 as the location of a participant's estimation. Line L1 is the vertical line indicating the height between location of the sender's eyes and the grid. The angle between the sender eyes and point P1 is labeled here as A1. The location of the senders gaze on the grid is labeled P2 which is used in the same way as P1 in order to calculate angle A2. The vertical angle error is than calculated by taking the absolute value of the difference between angle A2 and A1.

      The total angle error is calculated by combining the value's of the horizontal angle error with the corresponding vertical angle error in a three dimensional space. The total angle error is equal to $\tan^{-1}(\sqrt{(\tan(HE)^2 + \tan(VE)^2}\,))$ where HE is the horizontal angle error and VE the corresponding vertical angle error. In the calculation of the total angle error are errors on the horizontal and vertical angle of equal value. The total angle error is used on both the relative and absolute target points creating two new error values we called Relative angle and Absolute angle in order to avoid confusion with the Direct and Relative error which refer to an error as a distance.

Results

Description

We used four measures to evaluate the performance of our participant in estimating a gaze location. These four measures can be divided into two dimensions, distance opposed to angle and direct opposed to relative. To look at the errors in distance we used the direct and relative error, for the errors in angle's we used the direct and relative angle. In table 3 you can see a descriptive overview of our result per scenario labeled A, B and C. N is the number of error values used in the analysis, for each of the three conditions there are 72 values provided by 9 participants who each guessed 8 locations in one scenario. The mean, standard deviation and standard error are smaller for the relative error then for the direct error in each of the scenarios. The reason that participants score lower relative errors than direct errors is because the relative errors are constructed as a best case scenario for the participants, where the direct error is calculated independent of the participant's performance as a group. The Direct and Relative angle have a relatively small standard errors compared to their mean value's. This can be explained by the fact that distance of the sender from the grid was larger than the size of grid, making the effect of the differences in location on the grid more consistent than with the Relative and Direct error.

		N	Mean	Std. Deviation	Std. Error
Direct error	A, Static neck	72	16,09	11,43	1,35
(in cm)	B, Dynamic neck	72	16,77	12,18	1,44
	C, Person	72	9,22	6,11	,72
	Total	216	14,03	10,78	,73
Relative error	A, Static neck	72	9,17	6,35	,75
(in cm)	B, Dynamic neck	72	7,62	5,63	,66
	C, Person	72	5,80	4,02	,47
	Total	216	7,53	5,57	,38
Direct angle	A, Static neck	72	1,05	,38	,04
(in degree)	B, Dynamic neck	72	,96	,43	,05
	C, Person	72	1,08	,36	,04
	Total	216	1,03	,39	,03
Relative angle	A, Static neck	72	1,05	,42	,05
(in degree)	B, Dynamic neck	72	1,08	,38	,05
	C, Person	72	,97	,38	,05
	Total	216	1,03	,40	,03

(Table 3: Descriptive about the direct error, relative error, direct angle and relative angle)

All four measurements give different means between the scenarios. In order to look at whether the difference between means is significant we used a Repeated measures ANOVA. The Mauchly's Test of Sphericity showed a significance larger than 0.05 for all variables making it acceptable to assume Sphericity. When we look at the Univariate Tests displayed in table 4, we can see that only the Direct error has a significant effect ($p<0,001$). Because both the angle based measures Direct angle and Relative angle showed no significance of lower then 0.1 we will only go into the distance based measures Direct error and Relative error, though Relative error did not prove to be significant.

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.
Scenarios	Direct error	Sphericity Assumed	313,989	2	156,994	12,190	,001
	Relative error	Sphericity Assumed	51,342	2	25,671	2,309	,131
	Direct angle	Sphericity Assumed	,072	2	,036	2,401	,122
	Relative angle	Sphericity Assumed	,063	2	,031	1,189	,330
Error(scenarios)	Direct error	Sphericity Assumed	206,068	16	12,879		
	Relative error	Sphericity Assumed	177,865	16	11,117		
	Direct angle	Sphericity Assumed	,241	16	,015		
	Relative angle	Sphericity Assumed	,421	16	,026		

(Table 4: Univariate Tests, showing significance for the direct error)

Direct errors
There is a significant difference between the Direct error mean values created in the three different scenarios. How the mean value's compare to their standard deviant in illustrated for each of the 3 scenarios in box plot 1. Each mean is built up of 72 Direct error values with outliers displayed as a circle and extreme values as a star. Most participants made smaller direct errors in scenario C than when they were in scenario A or B. These smaller direct errors mean that participants made more accurate estimations of the target point's location when this point was produced by a person sitting in front of them rather than being displayed through a tablet. The mean value's of the collected direct errors also shows a difference between scenario A and B where the mean value in scenario A is lower than that in scenario B. Participants made on average better estimations of a tablets gaze location when it was held still opposed to having a dynamic setting.
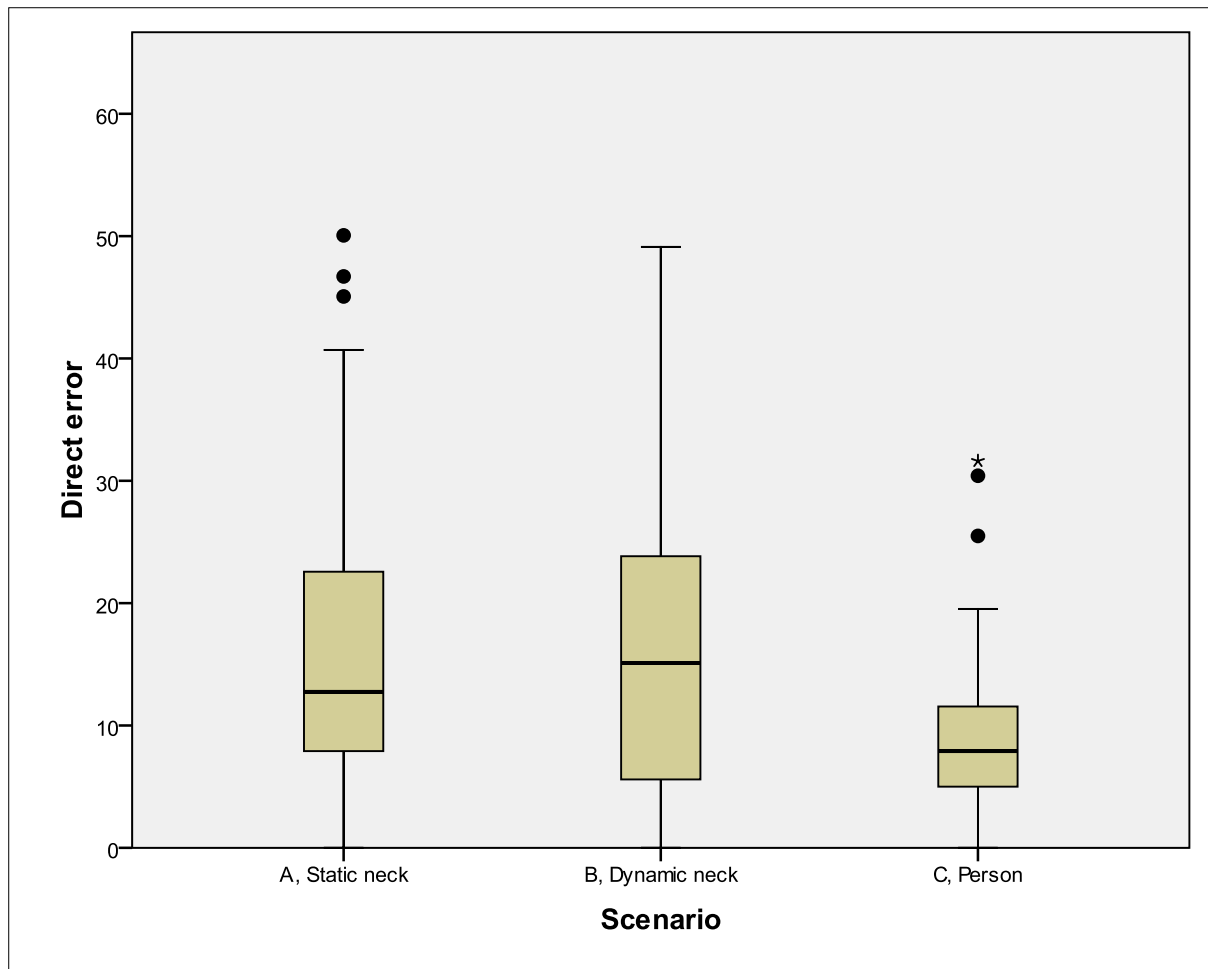
(Box plot 1: Direct error)

In order to test which scenarios significantly differ in mean value from one and other we use a pair wise comparisons test as shown in table 5. There is no significant difference between the two videoconferencing scenarios Static neck and Dynamic neck. There is a significant difference between the Person scenario and the two videoconferencing scenarios. Participant produced significantly smaller direct errors in the Person scenario then in the Dynamic neck and Static neck scenario. Participants

| Measure	(I) scenarios	(J) scenarios	Mean Difference (I-J)	Std. Error	Sig.
Direct error	A, Static neck	B, Dynamic neck	-,685	2,042	1,000
		C, Person	6,867	1,274	,002
	B, Dynamic neck	A, Static neck	,685	2,042	1,000
		C, Person	7,552	1,672	,006
	C, Person	A, Static neck	-6,867	1,274	,002
		B, Dynamic neck	-7,552	1,672	,006

(Table 5: pair wise comparisons)

Relative errors
We did not find a significant effect between the mean values of the Relative error's produced by our participants in the three given scenarios. Finding no significant effect for the Relative error is counter intuitive when considering that there were significant differences with the Direct errors, suggesting then there is no difference in the consistency in which participants locate a sender's gaze location though they are more accurate in some scenarios. When looking in more detail to the Relative error's there can be seen some difference between the scenarios as shown in box plot 2. In each scenario you can see outliers up to 3 times the mean value, though this is most common for scenario A in which there is no physical head movements. The mean value is as with the direct errors the lowest for scenario C, but shows a smaller difference with scenario A and B with regard to the relative error. When looking at the relative errors we see that in contrast to with the direct errors scenario B now shows a lower mean than scenario A.
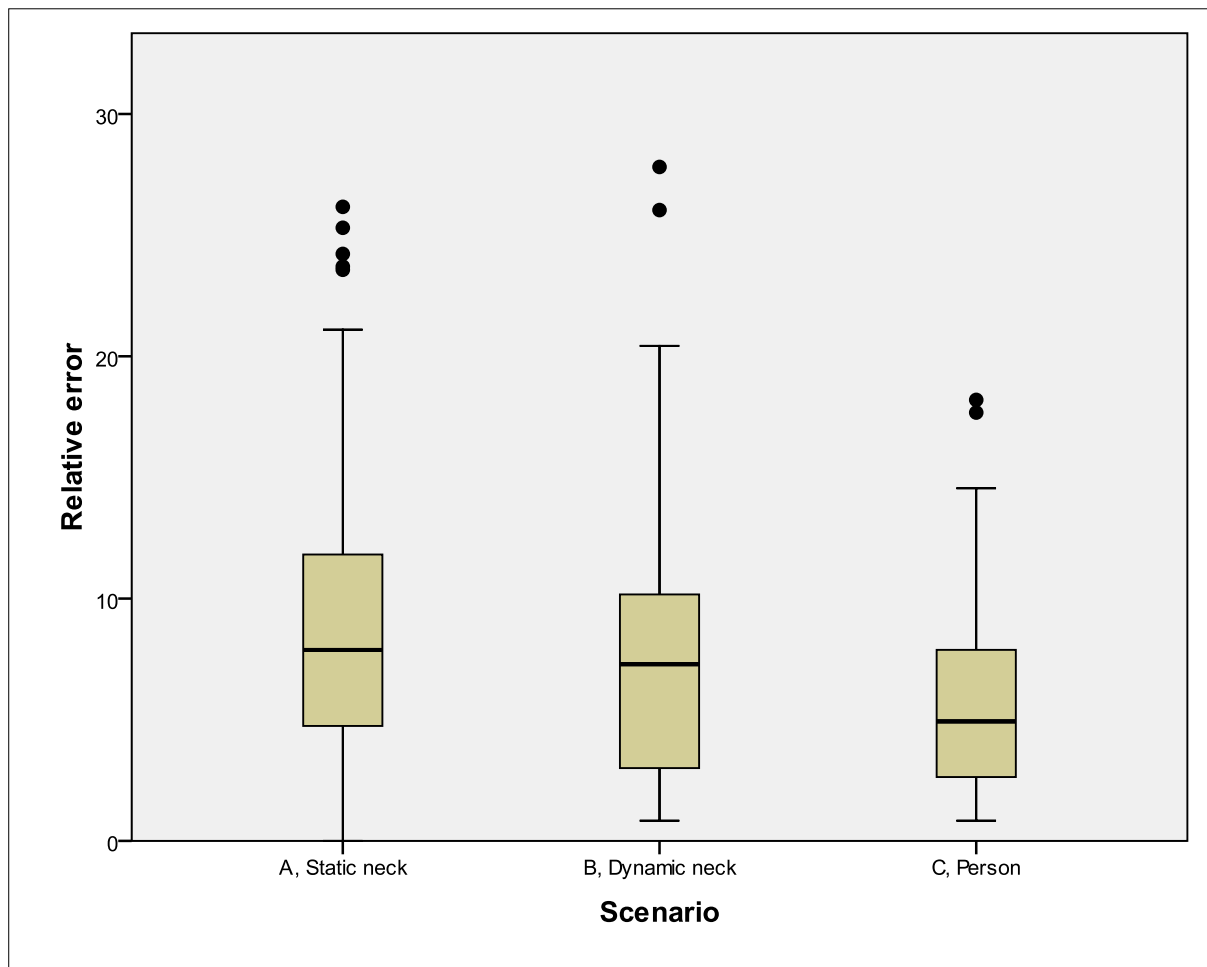
(Box plot 2: Relative error)

The Relative errors are constructed as an absolute value of a distance in order to avoid errors canceling each other out when looking at the mean values. When displaying the Relative error's without using the absolute value's you get a visualization of how far from one and other participants estimated a target gaze location to be. Figure 12 shows three scatter plots of the non-absolute Relative errors in each scenario, where point (0, 0) corresponds to the target gaze location. Scenario C in which the sender was face to face with participant produced less scattered estimation then in scenario A in which a videoconferencing setting was used.
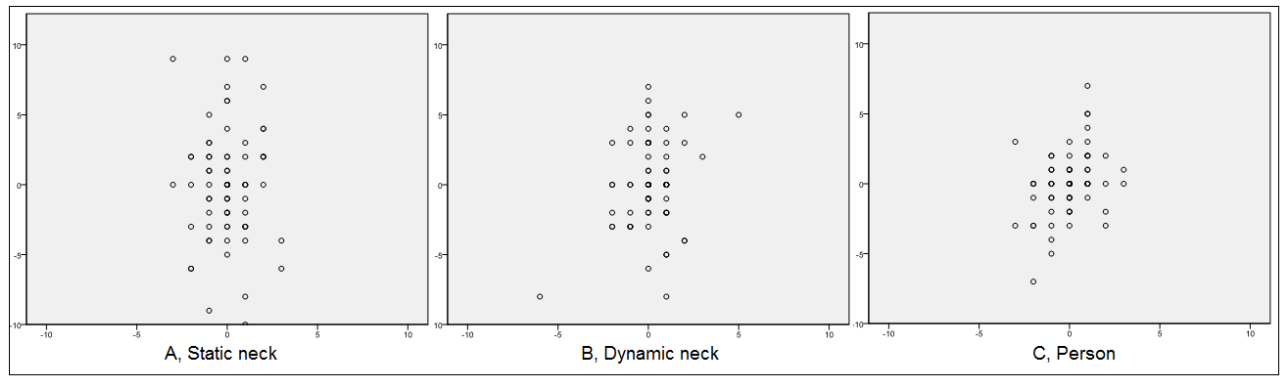
(Figure 12: Scatter plots relative error)

Questionnaire

Each participant filled in a questionnaire with questions about his or her feelings toward the preformed exercises. The questions could be answered by strongly agree, agree, uncertain, disagree and strongly disagree. We used a repeated measures ANOVA in order to compare the given answers for which we encoded these answers as value's from 0 to 4, having an equal distance between each answers. We did not find any significant difference between the means in each scenario, as shown in table 6 where the measures Difficulty, Accuracy, Team and Helpful refer to the question in from questionnaire as described in the instructions and are shown in appendix A. In each scenario participants where on average uncertain about the accuracy of their guesses and their feeling towards the sender. The degree in which participants found it difficult to locate the targeted points differed between the three scenarios. We also looked at a more general interpretation in which we classified both strongly agree and agree as a form of agreeing. We combined the strongly agree with the agree answers and the strongly disagree with the disagree answers in order to created three categories agree, uncertain and disagree. With our more general interpretation method we still didn't found any significant result, though this seems plausible for a larger data set.

None of participants reported that it was not difficult to locate the targeted point in the Static neck scenario, where for the Person scenario six out of the nine participants reported this. Participants found it easier to estimate the gaze location of person in a face to face setting then when having to evaluate the gaze location of an agent shown on a tablet.

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Scenario	1,Difficulty	Sphericity Assumed	5,63	2	2,815	2,643	0,102	0,248
	2,Accuarcy	Sphericity Assumed	0,667	2	0,333	0,471	0,633	0,056
	3,Team	Sphericity Assumed	0,222	2	0,111	0,123	0,885	0,015
	4,Helpful	Sphericity Assumed	0,296	2	0,148	0,215	0,809	0,026

(Table 6: Repeated measures ANOVA of the questionnaire)

Conclusion and Discussion

Our experiment showed that participants where significantly more precise in detecting the gaze location of another person when that person was sitting directly in front of them, then when that persons gaze was embodied by video feed on a tablet. We used 2 different tablet condition, one static in which a recording of face would make head moments in order to target a gaze location, and one dynamic condition in which the tablet would play a video of a non moving face while using a robotic neck to physically turn the tablet toward a gaze location.

When comparing the two tablet scenarios we found a mean value of 16.09 for the direct error produced in the tablets static condition and a mean value of 16.77 for the direct error produced in the dynamic neck scenario. The different mean values between two tablet scenarios shows a difference in the precision in which participants guessed a target location. The difference in the precision in which participants tried to located a target gaze location where not significant with a P value of 1. The reason that no the significant difference was found might be caused by gaze offsets created by the camera locations in our hardware. We set the parameters for our robotic neck by holding a pilot experiment in which we asked trained subjects for feedback. Because we were aware of the risk of possible offset in our target gaze location due to hardware adjustments we used the Relative error as a best case scenario in which a possible offset would always favor the average participant. The mean values of the Relative errors corresponded with our expectations that Scenario C, Person (Relative mean 5.80) would perform better then scenario B, Dynamic neck (Relative mean 7.62) which would perform better then scenario A, Static neck (Relative mean 9.17). The mean values of the Relative errors did not differ significantly according to the results from the repeated measures ANOVA. There is a change that the Relative errors showed no significance due to the size of our sample set, this would be a plausible explanation taking into account that we only used nine participants in our experiments.

The angle measures Relative angle and Direct angle gave no significantly different mean values between our scenarios, showing a P value of .330 for the direct angle and a P value of .122 for the relative angle. Having no significant effect on the Direct angle while the Direct error shows a significant effect is remarkable because both measures are based on the exact same grid coordinates. The difference in significance between the direct angle and direct error might be caused by having a to small sample set. We could not prove a significant difference between dynamic and static video conferencing settings in regard to gaze direction detection. According to D.A. Kenney there is a minimum number of 42 samples needed when performing a statically analysis with a power of 0.95 (D.A.Kenny, 1987). The values produced by our participants only counted as 9 samples in the repeated measures ANOVA because they where averaged, this would suggest that our number of participants was to low be likely to find significant effects.

Though we couldn't find any significant difference between the dynamic and static scenarios the Relative error means are conform our expectations that a dynamic scenario can be evaluated more accurately than a static one. Follow up research into the effects of a movable tablet on gaze direction detection with more participants is needed in order to draw more conclusive conclusions about the subject. For follow up research we recommend to go more into the consistency between participants, because we believe this to be more reliable than the precision of the participant's estimations due to the difficult nature of adjusting hardware correctly. Previous research of Schrammel et al showed that participants where less accurate in locating a static senders gaze location when that sender was targeting positions located more to sides of the senders location (J.Schrammel, 2007). In continuing research it could be interested to compare how the location of a targeted point influence the detection of that point in both a static and dynamic tablet scenario. A static scenario might prove to be less effected by a point's location because of its ability to physically turn. In order to investigate the effects of the location of a target we recommend to use a larger field that our, which was 90 by 65 centimeter, in order to amplify a possible effect.

References

B. Biguer, C. P. (1984). The contribution of coordinated eye and head movements. *Exp. Brain Res.* , pp. 462–469.

B.Biguer, C. a. (1984). The contribution of coordinated eye and head movements in hand pointing accuracy. *Experimental Brain Research* , 462-469.

D.A.Kenny. (1987). *Statistics for the social and behavioral sciences.* Boston: Little, Brown .

D.Hogan. (2008, 11 3). *Videoconferencing More Confusing For Decision-Makers Than Face-To-Face Meetings.* Retrieved 5 27, 2012, from ScienceDaily: http://www.sciencedaily.com/releases/2008/10/081028184748.htm

D.T.Moore. (2008, 11 10). *The Advantages and Disadvantages of Video Conferencing.* Retrieved 5 27, 2012, from Ezine Articles: http://ezinearticles.com/?The-Advantages-and-Disadvantages-of-Video-Conferencing&id=1676217

DI.Perrett, J. M. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philos Trans R Soc Lond B Biol Sci. 335* , 23-30.

G. Butterworth, N. J. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology* , 55–72.

H.Nakanishi, Y. M. (2008). Minimum movement matters. *In Proc. CSCW* , 303-312.

J. Gemmell, K. T. (2000). Gaze awereness for video conferencing a software approach. *IEEE Multimedia* , 7-15.

J. Wainer, D. F.-S. (2008). Embodiment and humanrobot. *In Proceedings of the International* , 6.

J.N.Bailenson, A. J. (2002). Gaze and taskperformance in shared virtual environments. *The Journal of Visualization and Computer Animation, Issue 13* , 313-320.

J.Schrammel, A. M. (2007). "Look!" – Using the Gaze Direction of Embodied Agents. *CHI 2007 Proceedings • People, Looking at People* .

L.W. Barsalou, P. N. (2003). Social Embodiment. *Psychology of Learning and Motivation* , 43–92.

M.Morales, P. J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development, Volume 21* , 373–377.

Moore, D. T. (2008, 11 10). *http://ezinearticles.com.* Retrieved 5 27, 2012, from Ezine Articles: http://ezinearticles.com/?The-Advantages-and-Disadvantages-of-Video-Conferencing&id=1676217

R.Yang, Z. (2001). Eye Gaze Correction with Stereovision for Video-Teleconferencing. *Microsoft Research Technical Report 2001-19* .

Skype grows FY revenues 20%, reaches 663 mln users. (2011, 3 8). Retrieved 7 2012, 15, from telecompaper: http://www.telecompaper.com/news/skype-grows-fy-revenues-20-reaches-663-mln-users

T. Farroni, G. C. (2002). Eye contact detection in humans from birth.Eye. *Proceedings of the National Academy of Sciences* , 9602–9605.

Appendix A

The sender is the agent who gazes to a specific point on the grid.
The receiver is the one who tries to locate the gaze location of the sender on the grid.

Spots on the grid where targeted in 3 differed situations:

Scenario A = The robotic neck moved
Scenario B = The person shown on the tablets screen moved
Scenario C = The person sitting in front of you moved.

Answer the following questions for the **scenario A.**	strongly agree	agree	uncertain	disagree	strongly disagree
1. I found it difficult to locate the targeted points.	☐	☐	☐	☐	☐
2. I believe my guesses where accurate	☐	☐	☐	☐	☐
3. I had the feeling the sender and I where a team	☐	☐	☐	☐	☐
4. I believe the sender to be helpful	☐	☐	☐	☐	☐

Answer the following questions for the **scenario B**.	strongly agree	agree	uncertain	disagree	strongly disagree
1. I found it difficult to locate the targeted points.	☐	☐	☐	☐	☐
2. I believe my guesses where accurate	☐	☐	☐	☐	☐
3. I had the feeling the sender and I where a team	☐	☐	☐	☐	☐
4. I believe the sender to be helpful	☐	☐	☐	☐	☐

Answer the following questions for the **scenario C**.	strongly agree	agree	uncertain	disagree	strongly disagree
1. I found it difficult to locate the targeted points.	☐	☐	☐	☐	☐
2. I believe my guesses where accurate	☐	☐	☐	☐	☐
3. I had the feeling the sender and I where a team	☐	☐	☐	☐	☐
4. I believe the sender to be helpful	☐	☐	☐	☐	☐