Accuracy of head orientation perception in triadic situations: Experiment in a virtual environment

Ronald Poppe, Rutger Rienks, Dirk Heylen
Human Media Interaction Group, Department of Electrical Engineering, Mathematics and Computer Science, University of Twente, PO Box 217, NL 7500 AE Enschede, The Netherlands; e-mail: poppe@ewi.utwente.nl
Received 24 July 2006, in revised form 24 November 2006; published online 29 June 2007

Abstract. Research has revealed high accuracy in the perception of gaze in dyadic (sender – receiver) situations. Triadic situations differ from these in that an observer has to report where a sender is looking, not relative to himself. This is more difficult owing to the less favourable position of the observer. The effect of the position of the observer on the accuracy of the identification of the sender’s looking direction is relatively unexplored. Here, we investigate this, focusing exclusively on head orientation. We used a virtual environment to ensure good stimulus control. We found a mean angular error close to 5°. A higher observer viewpoint results in more accurate identification. Similarly, a viewpoint with a smaller angle to the sender’s midsagittal plane leads to an improvement in identification performance. Also, we found an effect of underestimation of the error in horizontal direction, similar to findings for dyadic situations.

1 Introduction
Research has revealed high accuracy in the perception of gaze in dyadic (sender – receiver) situations. In these situations, a sender looks at receivers, or slightly next to them. The task of the receivers is to report either whether they are being looked at, or quantitatively assess where the sender is looking. Triadic situations differ from this setting in that an observer has to report where a sender is looking, not relative to himself. This was found to be a more difficult task, owing to the more unfavourable position of the observer (Krüger and Hückstedt 1969).

The effect of the position of the observer on the accuracy of identification of the sender’s looking direction is relatively unexplored. We present here an experiment that investigates this effect. Within perception research, it is of major importance to have good control over the stimuli. With the traditional tools that are used for perception research (live situations, pictures, video), it is difficult to achieve this control, especially over a range of viewpoints. We addressed this limitation by using a 3-D virtual environment (VE), as suggested by Symons et al (2004). This allows full control over all stimuli, which makes it an appropriate tool for research into human perception (Loomis et al 1999). As VE we use the virtual meeting room (figure 4), described in Reidsma et al (2007). The virtual representation of a human sender is termed ‘avatar’.

We discuss previous research into the perception of gaze and head orientation in section 2. The setup of our experiment is described in section 3 and results are presented in section 4.

2 Perception of head orientation
The direction of gaze is determined by a combination of head orientation and eye orientation (Kleineke 1986). Traditional research has addressed human perception for both components, and how they interact. In our research, we did not vary eye orientation. Our avatar’s eyes point straight ahead, which allows us to focus on the perception of head orientation alone. In the following, we summarise research into perception of gaze, and focus on the role of head orientation.
The first assessments of the accuracy of human perception of gaze direction were carried out by Gibson and Pick (1963) and Cline (1967). They observed that the sender’s head direction is important for the perception of being looked at by the receiver in dyadic situations. Maruyama and Endo (1983) found that the contribution of face direction to apparent gaze direction was larger than that of the eyes, for moderate angles. Von Cranach and Ellgring (1973) concluded that, also in triadic situations, observers’ gaze judgments are determined to a great extent by head direction of the sender.

These studies used artificial settings in which a sender had to look directly at, or slightly to the side of, the receiver. In natural settings, people either look directly at, or markedly away from, a person (Argyle and Cook 1976). Given this fact, it is to be expected that observers can identify the sender’s focus of attention more reliably in natural settings. In an experiment in which observers had to determine whether a receiver was being looked at or not, Vine (1971) found a much higher inter-observer agreement for actual interactions compared to artificial ones (approximately 94% instead of 76%). Instead of using a (virtual) human receiver, we used balls as gaze targets. Although we are not aware of any studies of differences in accuracy when using artificial targets instead of humans, we expect the results to be comparable.

When using a virtual environment, the situation differs in three aspects from traditional research settings. In the first place, we use a computer screen showing a VE instead of a live setting. In relation to this, Anstis et al (1969) found that there are differences in the perception of gaze from a television screen compared to a live setting. This effect occurred when the television screen was turned to the left or to the right. Due to the non-convexity of the television screen, the ratio of white in the sclera of the eye did not change, which appeared to generate a similar perception as when the screen was not turned. This effect was unlikely to occur in our setting as the observers were seated directly in front of the screen, and the avatar’s appearance was geometrically adjusted to the chosen view in the environment, with its eyes fixed and directed straight forward. Symons et al (2004) noted that, although their results for perception of eye gaze at screen stimuli were qualitatively comparable to those of live stimuli, accuracy was much worse. The possible sources of this discrepancy could be the limited resolution of the stimuli, the lack of depth information, and their use of a different sender. In our experiment, we do not regard eye orientation. Our screen resolution was higher: 1280 × 1024 pixels, compared to 640 × 480 pixels used by Symons et al (2004). Therefore, we expect our results to be comparable to those of live settings.

The second difference between a live or recorded setting and ours is the fact that our avatar representation is an abstraction of a real person. The presented avatar stimuli might be too simplistic to reliably determine their head orientation. However, Sagiv and Bentin (2001) found that schematic faces were capable of producing similar effects to real faces and were an appropriate substitute for photographs. This finding is supported by Wilson et al (2000), who found that perception of head orientation was high, even for low-resolution images. When face outlines only were presented, accuracy was even higher for the schematic stimuli as these were perfectly symmetrical.

A third difference between traditional settings and our VE setting is that our observation angles are significantly different. Past research into triadic gaze has been limited to the situation where sender, receiver, and observer resided in the same horizontal plane (Krüger and Hückstedt 1969), or the angles between sender, receiver, and observer were small (Symons et al 2004). In our experiment, we varied the angles in a large range. Because the observers’ specific situation strongly influences the perception of gaze, Von Cranach and Ellgring (1973) suggested that the observers’ accuracy in discrimination for a specific setting should be measured. Therefore, we assessed in this experiment the observers’ accuracy in distinguishing the focus of interest of the avatar in our VE.
3 Method

3.1 Stimuli

An avatar was positioned in the virtual meeting room at the left side of the table (figure 4). The eyes of the avatar were fixed and pointed straight ahead. 10 balls were placed at a distance corresponding to 1.5 m away from the avatar, at eye height. To ensure good depth estimation, each ball was placed on a stick that intersected with the table. For enhanced discrimination, the balls were numbered and were coloured alternatively red and green. Figure 1 shows a schematic top view of the setup of the experiment.

The balls were placed in the azimuth plane, which is parallel to the floor, in the range \{-45^\circ, 45^\circ\} in front of the avatar. Looking straight ahead corresponds to 0°. Since we used 10 balls, the angular distance between two consecutive balls was 10°. In an informal preliminary experiment, we used angular distances of 30°, 22.5°, and 15°. Even for the 15° condition, we found identification performance of approximately 75%. Since a smaller distance between balls allows for more accurate assessment of perception accuracy, we chose to lower the angular ball spacing to 10°.

In our experiment, we varied the viewpoint along two dimensions. Figure 4 shows an example from each of the 8 views. All viewpoints were located on a sphere, with the centre 0.75 m in front of the avatar, at table height. The centre of projection was always directed at this point. The radius of the sphere corresponded to 3.4 m. The first viewpoint dimension was the rotation in the y direction (\(y\)-angle). The two conditions used were 15° and 45°, where 0° is an exact side view and 90° is an exact top view. We expected that a higher viewpoint would result in a better identification performance since discrimination between balls was higher. We formulate hypothesis 1:

**Hypothesis 1:** A higher viewpoint results in a lower target-identification error.

The second dimension is the rotation in the x direction (\(x\)-angle). Four conditions were used: -45°, 0°, 30°, and 60°. We expected that a larger \(x\)-angle would result in a better identification performance since more of the face is visible. Specifically, we expected the highest identification rates for a front view (\(x\)-angle = 90°). However, these views are not included in our experiment because it was not possible to display all balls without scaling the scene. Hypothesis 2 is formulated as follows:

**Hypothesis 2:** A larger \(x\)-angle results in a lower target-identification error.
Hypothesis 2: An $x$-angle closer to the centre of the arc of balls (in our case 90°) results in a lower target-identification error.

When using the target-identification error as a measure, no attention was paid to the direction of the error. In studies of dyadic gaze an underestimation of the gaze target for heads turned to either side has been reported (Gibson and Pick 1963; Cline 1967; Vine 1971). This is the effect of identification of targets being biased towards the viewpoint. We expected to see the same systematic error in triadic gaze situations. Our third hypothesis follows from this:

Hypothesis 3: There is an underestimation effect for head orientation in $x$ direction.

The setup of our experiment resembles that of Krüger and Hückstedt’s (1969). Specifically, we used the same distance between sender and receiver as in one of their conditions, and placed our centre of projection at the same location. However, Krüger and Hückstedt (1969) varied the distance between sender and receiver, and used different values for the sender’s head orientation and distance between observer and centre. Moreover, we also varied our viewpoints in the $y$ direction.

Use of the centre of projection as described above has an additional advantage over the use of avatar’s head as centre: the scene did not have to be scaled down (or, alternatively, the virtual distance between viewpoint and centre of projection enlarged) to make sure all balls would fit on the screen. By shifting the centre of projection from the avatar’s head, the distance between viewpoint and head varies between viewpoint conditions. For example, when the $y$-angle is 15°, the distances for an $x$-angle of $-45°$ and 60° are 3.40 m and 4.55 m, respectively. Another point to mention is the introduction of a small roll angle due to the subsequent rotations. However, this has only an effect on the rotation of the 2-D projection.

Participants were seated in front of a 19 inch TFT screen (resolution 1280 × 1024 pixels) that was placed on a desk in an office environment. The VE was scaled such that the height of the head of the avatar measured exactly 3.3 cm on the screen. We assumed that the eyes of the participants were at a distance of 60 cm from the screen. In this situation, the perceived height of the avatar’s head was similar to a condition where a live sender with a head height of 22.0 cm is observed from a distance of 4.0 m.

3.2 Procedure

Participants were asked to state towards which ball the avatar’s head was directed. For each of the 8 viewpoints, the avatar looked at each of the 10 balls once, in a random order. To ensure more accurate results, we repeated these 80 samples 5 times, resulting in 400 answered samples per participant. To eliminate the need to test all orders of the 8 viewpoints, we used a Latin square design. In summary, this gives us one between-participants variable (order of viewpoints) and two within-participants variables (viewpoint and ball), measured at 5 different times. Note that the viewpoint is actually a combination of the $x$-angle and $y$-angle variables.

Participants could indicate their answer by pressing the corresponding button at the bottom of the screen. Then the experiment proceeded with the next sample. There was no time constraint and participants were free to have short breaks when needed. They could view their progress in the experiment but did not receive any feedback on their judgment.

3.3 Participants

A total of thirty-two persons (seven female) participated in the experiment. The participants were students and employees of our department aged between 22 and 54 years (mean age was 32 years). All participants had normal or corrected-to-normal vision; none of them was a trained observer.
4 Results
The total number of samples that have been judged was 12,800. No pruning of these data has taken place, although some participants reported occasional ‘missclicks’.

4.1 Learning and fatigue
First, we checked whether learning or fatigue effects occurred. We looked at the performance within each of the 5 repetitions. The mean angular distance between target ball and identified ball was calculated for all 80 samples within each of the repetitions. Homogeneity of variance for the repetition variable was confirmed with Levene’s test, and an ANOVA with repeated measures was performed subsequently. The repetition variable proved to be significant ($F_{4,124} = 2.950, p < 0.05$). In figure 2a, the angular ball error is plotted as a function of the repetition. As can be seen, the graph is monotonically decreasing, which indicates that a learning effect has occurred.

Since the duration of the experiments was reasonably short, fatigue probably did not have an effect on the performance. The duration of an experiment was between 12.6 and 35.4 min (mean duration was 19.9 min). This is the sum of all answer times, but answer times longer than 30 s were regarded as breaks and were omitted from the total time. Given this moderate duration, and the fact that the graph in figure 2a is monotonically decreasing, we assume that fatigue did not influence the performance.

Since the effect of repetition on the performance was small, in the further analysis of our data we grouped samples with similar viewpoint and ball over the 5 repetitions.

4.2 Viewpoint order
Next, we investigated whether the order of the viewpoints had any effect on the identification performance. We performed an ANOVA with order as a between-participants variable and $x$-angle, $y$-angle, and ball as within-participants variables ($8 \times 4 \times 2 \times 10$). The mean (absolute) angular ball error over the 5 repetitions was used as the dependent variable. Note that, at this point, we only looked at the magnitude of the error, not at the direction. We discuss this later.

The effect of the order was found to be insignificant ($F_{7,24} = 0.881$, ns) and no interaction effects between order and any of the other variables was found. In the following, we group samples with similar conditions, regardless of the order in which the viewpoints were presented.

The average angular ball error over all samples was 5.2$^\circ$: $x$-angle, $y$-angle, and ball all appeared to have a significant effect on the identification performance. Also, interaction effects were reported for all combinations of these variables. We discuss these effects below.

![Figure 2. Mean angular ball error plotted as a function of (a) the repetition variable, and (b) the $x$-angle condition. Vertical bars show standard deviations.](image-url)
Recall that our viewpoint was tilted either 15° or 45°. The latter situation allowed for better distinction between balls since the on-screen distance between the balls was larger (see also figure 4). This resulted in a lower angular ball error for the 45° condition (3.78°) than the 15° condition (6.53°). These findings are significant ($F_{1,24} = 96.297, p < 0.001$) and consistent with hypothesis 1.

4.4 x-angle
A decrease in angular ball error was found for increasing x-angles, ie as more of the face became visible. Figure 2b shows this trend, which is significant ($F_{3,72} = 18.446, p < 0.001$, with Huynh–Feldt correction). This confirms hypothesis 2.

4.5 y-angle, x-angle, and ball
The angle under which the avatar’s head is viewed is not solely determined by the x-angle. The ball that is being looked at also affects the x component of the angle. The y-angle influences both the x component of the angle, as the y component. Indeed, we see a significant interaction effect for x-angle and ball ($F_{27,648} = 18.189, p < 0.001$), for y-angle and ball ($F_{9,216} = 3.748, p < 0.001$, with Huynh–Feldt correction), and for the three variables together ($F_{27,648} = 6.459, p < 0.001$). Furthermore, a marginal interaction effect was found for x-angle and y-angle ($F_{3,72} = 2.619, p = 0.057$).

We observed that the y-angle, x-angle, and ball variables together determine the angle under which the avatar’s head is viewed. This angle can be defined as the angle between viewpoint and avatar, and avatar and ball (VAB angle). This angle could alternatively be decomposed into an x component (angle when projected onto the azimuth plane) and a y component (angle when projected onto the plane orthogonal to the azimuth plane and through either the line viewpoint–avatar, or avatar–ball). However, since we only varied the location of the balls in the azimuth plane, the y-angles that were used in this experiment were within only a small range. Therefore, we restricted our investigation to the x component. We regarded the viewpoints where the y-angle was 15°. For each of the combinations of viewpoint and ball, we calculated the VAB angle and the mean angular ball error over all repetitions and participants. Figure 3a shows a scatter plot of the two variables. There is a strong correlation between the two variables ($R = 0.497, p < 0.001$). Future work should aim at investigating the role of the y-angle in this effect.

One possibility is the limited resolution of the screen. An increase in the angle between the observer’s viewpoint and the midsagittal plane results in a decreased lateral shift in facial features. A limited resolution strengthens the effect that small far-off shifts are displayed similarly. This effect might explain the larger estimation errors for larger
Figure 4. Different views on the scene. (a–h) Correspond to viewpoint 1–8. (a) $x$-angle $= -45^\circ$, $y$-angle $= 15^\circ$; (b) $x$-angle $= -45^\circ$, $y$-angle $= 45^\circ$; (c) $x$-angle $= 0^\circ$, $y$-angle $= 15^\circ$; (d) $x$-angle $= 0^\circ$, $y$-angle $= 45^\circ$; (e) $x$-angle $= 30^\circ$, $y$-angle $= 15^\circ$; (f) $x$-angle $= 30^\circ$, $y$-angle $= 45^\circ$; (g) $x$-angle $= 60^\circ$, $y$-angle $= 15^\circ$; (h) $x$-angle $= 60^\circ$, $y$-angle $= 45^\circ$. A colour version of this figure is shown on the Perception website at http://www.perceptionweb.com/misc/p5753/.)
VAB angles. However, this effect alone would result in the largest errors for situations where the VAB angle is 90°, which is not confirmed by our data (see figure 3a).

4.6 Direction of error
In the following, we use the data from the viewpoint where y-angle is 15° and x-angle is 60°. This is the only condition where the y-angle is small and where data are gathered for head turns in both directions away from the viewpoint. As can be seen in figure 4g, the head of the avatar is positioned more or less behind ball number 8. When we analyze the mean angular direction, we obtain the graph in figure 3b. The mean error for ball 8 is close to zero. We notice a negative estimation error for balls to the right of ball 8 (as seen from the viewpoint) and positive errors for balls to the left. A negative error corresponds to a systematic error to the right (as seen from the viewpoint). This implies that our results show an underestimation effect for head orientation, similar to those found for dyadic situations. This confirms hypothesis 3.

5 Conclusions
A virtual environment was used to assess the effect of the location of the observer on the accuracy of head orientation perception in triadic situations. The virtual environment allowed good stimulus control. The sender was a virtual representation of a human (avatar). Instead of using a (virtual) human receiver, we used balls as targets. The balls were placed in the horizontal plane in a 90° range from the avatar at a distance corresponding to 1.50 m. We used 8 viewpoints for the observer, with 2 different values for the y component, and 4 values for the x component. Participants were asked to identify towards which of the 10 target balls the sender’s head was oriented. The average angular ball error over all samples was 5.2°. We found that a higher viewpoint allowed better discrimination of the balls, which led to lower identification-error scores. The x component of the viewpoint had an important effect on the identification performance. For this x component, we regarded the angle viewpoint–avatar–target, projected onto a horizontal plane. We found that small values for this angle, which correspond to a viewpoint where the target was between the viewpoint and the head of the avatar, resulted in more accurate identification. An increase in this angle led to an increase in the identification error. Furthermore, an underestimation effect was found for the direction of the error in the x component. These findings are in line with earlier research into dyadic situations.

The role of the y component and the interaction with the x component on the accuracy of head orientation remain to be investigated. Also, the distances between viewpoint and avatar, and between avatar and target need to be varied, to see how they influence identification performance. Our current experiment focused on targets that resided in the same horizontal plane. Future work will also take into account variations in the position of the targets in the vertical direction.

Acknowledgments. This work was partly supported by the European Union 6th FWP IST Integrated Project AMI (Augmented Multi-party Interaction, FP6-506811, publication AMI-205), and is part of the ICIS program. ICIS is sponsored by the Dutch government under contract BSIK03024. We wish to thank Matthijs Poppe for his useful comments regarding the design and analysis of the experiment.

References
Anstis S M, Mayhew J W, Morley T, 1969 “The perception of where a face or television ‘portrait’ is looking” American Journal of Psychology 82 474 – 489
Cline M G, 1967 “The perception of where a person is looking” American Journal of Psychology 80 41 – 50
Kleinke C L, 1986 “Gaze and eye contact: a research review” Psychological Bulletin 100 78 – 100
und Angewandte Psychologie 16 452 – 472
Loomis J M, Blascovich J J, Beall A C, 1999 “Immersive virtual environment technology as a
basic research tool in psychology” Behavior Research Methods, Instruments and Computers 31
557 – 564
Maruyama K, Endo M, 1983 “The effect of face orientation upon apparent direction of gaze”
Tohoku Psychologica Folia 42 126 – 138
meeting rooms: from observation to simulation” AI & Society 21 in press
Sagiv N, Bentin S, 2001 “Structural encoding of human and schematic faces: holistic and part-based
processes” Journal of Cognitive Neuroscience 13 937 – 951
Symons L A, Lee K, Cedrone C C, Nishimura M, 2004 “What are you looking at? Acuity for
triadic eye gaze” Journal of General Psychology 131 451 – 469
Vine I, 1971 “Judgement of direction of gaze: an interpretation of discrepant results” British Journal
of Social and Clinical Psychology 10 320 – 331
in Social Communications and Movement: Studies of Interactions and Expression in Man and
Research 40 459 – 472
This article is an advance online publication. It will not change in content under normal circumstances but will be given full volume, issue, and page numbers in the final PDF version, which will be made available shortly before production of the printed version.

Conditions of use. This article may be downloaded from the Perception website for personal research by members of subscribing organisations. Authors are entitled to distribute their own article (in printed form or by e-mail) to up to 50 people. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.