"Look!" – Using the Gaze Direction of Embodied Agents

Johann Schrammel* **Arjan Geven*** **Reinhard Sefelin*** **Manfred Tscheligi*‡**
*CURE - Center for Usability Research and Engineering ‡ICT&S Center, University of Salzburg
Hauffgasse 3-5, 1110 Wien, Austria S.-Haffner-Gasse 18, 5020 Salzburg, Austria
{schrammel; geven; sefelin; tscheligi}@cure.at

ABSTRACT
This paper describes the results of three studies investigating an embodied agent that supports its interaction with the user by gazing at corresponding objects within its close environment. Three experiments were conducted in order to research whether users can detect an agent's line of sight, whether the agent's gaze direction can help to guide the users' attention towards designated locations and whether such a setup can be used to improve realistic interaction situations. The results show that a) users can detect the agent's gaze direction quickly (within 200 ms) but not very exactly, b) the use of the agent's gaze direction can speed up but also slow down the detection of objects in dependence on their location and c) that the agent's gaze towards corresponding objects during the interaction can have counterproductive effects in realistic settings.

Author Keywords
Embodied Agent, Gaze Direction, Computer Vision

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces – Interaction Styles.

INTRODUCTION

Embodied agents are used and researched more and more as promising means of future interaction. Embodied agents have a more or less human look and feel and typically are capable of synthesizing spoken language. People are known to attribute emotions and feelings to computers and interact socially with them even if they do not look human [3, 4]. Using human-like representations increases these tendencies. For example it was found that a user can even build and maintain a relationship with an agent [1].

An important element of human-human interaction is gaze-behaviour. Therefore, it is not surprising that research has shown that an agent's gaze patterns have a significant impact on its perceived naturalness and also on the user's reactions. Speakers, for example, shift their gaze towards a listener as they get to the end of a thought. Colburn et al. [2] could show that the implementation of this and other patterns elicits changes in the viewers' eye gaze patterns. [1] and [5] showed that an avatar whose gaze behaviour is related to the conversation does not only change the viewer's behaviour but that it leads to a significant improvement of the perceived quality of the conversation.

With the increasing maturity of computer vision technology it was possible to equip computer systems and embodied agents with "eyes" that can perceive objects and activities in the environment and use this knowledge to improve the interaction with their users. This technology connects the real physical world with the virtual computational world and allows for systems that can detect, locate, recognise and understand objects and situations in the real world [6].

In our study we investigated the combination of these two technologies: Cognitive vision & Embodied agents. We wanted to now whether an agent can help its users if it looks actively and consciously at corresponding objects. In other words: Can an agent's line of sight help users to perceive and to recognize objects inside their close environment? We started from the observation that humans tend to look at certain objects when they become the topic of a conversation. A person who asks another person to pass a cup will first look at the person and then at the cup.

We wanted to investigate whether an embodied agent can and shall do the same: Are users able to locate the area at which the agent is looking? Can users recognise objects faster when the agent looks at them? And can this behaviour be applied to support daily memory tasks?

In order to answer these questions we conducted three experiments. Experiments one and three were conducted with the same 16 participants (8 men, 8 women, av. age 25.4, max. 33, min. 21, 9 persons used corrective lenses). The second experiment was conducted with 10 other participants (5 men, 5 women, av. age 27.7, max. 37, min. 19, 5 persons used corrective lenses, no colour-blinds).

We used the avatar character system named KATE developed by Haptek [7]. KATE comes along with pre-defined poses for looking directions that can be addressed in 7 steps for the directions left, right, up and down. The resulting poses use the eyes, head and torso orientation in different degrees depending on the targeted direction. To provide a more realistic impression KATE autonomously

CHI 2007, April 28–May 3, 2007, San Jose, California, USA.

performs tiny movements. KATE uses a SAPI text-to-speech engine and provides synchronized lip movements.

EXPERIMENT ONE: DETECTION OF GAZE DIRECTION

The first experiment was designed to see how strongly users agree or diverge on the perceived fixation points of the agent on a desk. Only if users agree strongly on the interpretation of the agent's' gaze direction it makes sense to use this means for interaction purposes.

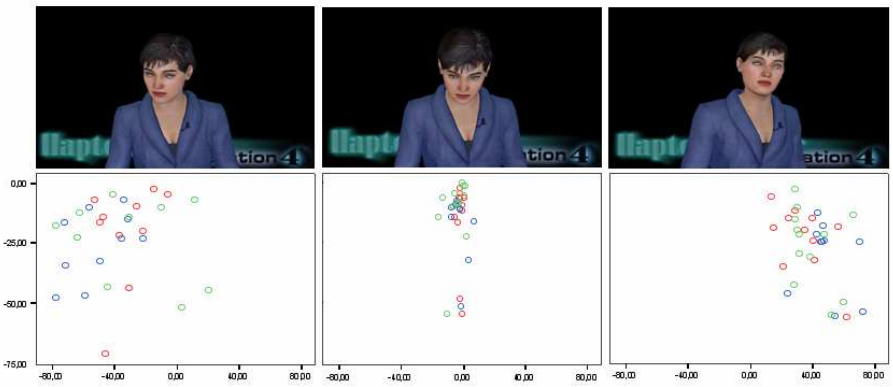

Figure 1: Scatter plots showing the identified locations of three example poses

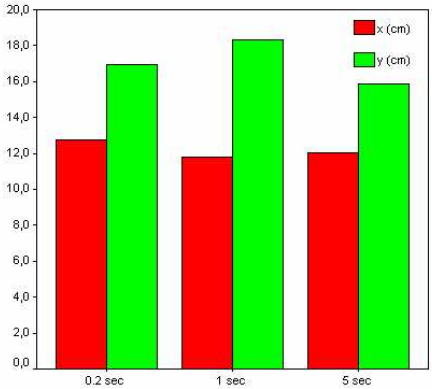

Figure 2: Standard deviation of X and Y coordinates

Method

The experimental setup consisted of an agent looking down on a grid on the desktop in front of the subject (see Figure 3). The subject could see the agent and the grid comfortably when seated. We varied the amount of time the subject could see the agent between three conditions: 200 ms, 1 second and 5 seconds. After that the agent disappeared from the screen and the subject was asked to point with his or her finger to the location at which the agent was looking. Then the subject could see a graphic of the grid on the screen and was asked to click on the position of her or his finger i.e. the perceived agent's fixation point on the desktop. We used 16 different poses of the agent with gaze directions spread across the desktop according to the authors' opinions.

Results

The experiment showed that subjects agreed only vaguely on the location at which the agent was looking. Figure 1 shows three examples of the used poses of the agent and scatter plots of the corresponding positions indicated by the subjects. The figure shows that subjects were able to follow the agent's gaze more accurately when it was directed straight towards the subjects. When the agent looked to its right or left hand side the accuracy decreased. The standard deviation values on the x-axis (across the subjects viewing direction) varied between 1.75 cm (middle pose in Figure 1, shown for 0.2 s) and 34.19 cm (left pose in Figure 1, shown for 5 s). The standard deviations on the y-axis (straight ahead from the subject) were larger for all poses with the smallest value being 3.17 cm for a pose where the agent was looking to the side (only the profile is visible, shown for 5 s) and a maximum of 35.53 cm for a pose where the agent's head was oriented straight ahead but its eyes were directed to the side.

Also the qualitative interviews showed that it is harder for the subjects to determine the distance from the display to the agent's fixation point (y-axis) than to detect the gaze's direction (x-axis).

A detailed analysis of the data shows that the amount of agreement depends on the actual pose of the agent. Poses where both head and eyes are directed towards a target location show more agreement than poses that make only use of the eyes to indicate the direction. Deviations between body-head directions and gaze directions were due to the limited possibilities to manipulate details of the avatar's body such as eyelids and brows.

Surprisingly the amount of time that the agent was presented to the subjects did not have a significant impact on the accuracy of the gaze detection. ($F = .213$, $p > .1$). Apparently, even the very short period of time of 200 ms was enough to follow the agent's eye gaze as accurately as with a 5 seconds display (see Figure 2).

Concluding, we can say that gaze can give subjects a sense of direction, but it is difficult for subjects to pinpoint it to an exact location.

EXPERIMENT TWO: REACTION SPEED

In the second experiment we investigated whether the agent's gaze direction can help to guide the subjects' attention towards designated locations.

Method

Subjects were seated in front of the computer and were asked to press a button out of five identified by its colour as fast as possible (see Figure 3). The time was logged by the system and used as dependent variable.

Five buttons of different colours were placed on the desktop: One button in the centre, two halfway to the side and two at the very border (Factor A: Location of target).

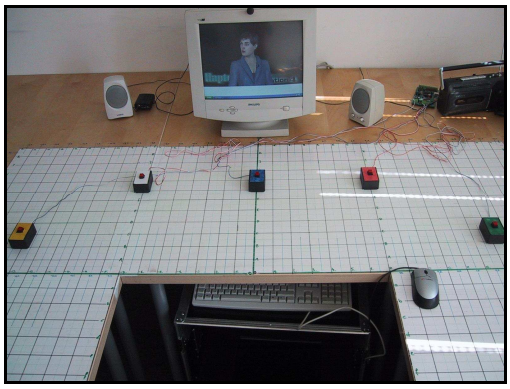

Figure 3: Desk Layout of the reaction speed experiment

When the agent gave the instructions its gaze-behaviour was varied between two conditions: The agent kept looking directly at the subject, or the agent looked at the corresponding button (Factor B: Agent gaze behaviour).

The information regarding the colour of the button was either indicated visually by a coloured field below the display of the agent or by synthesized spoken text of the agent (e.g. "blue button") (Factor C: Instruction modality).

We assumed that the subjects would perform better if the agent looks at the target. Furthermore we assumed that this effect would be more significant for the visual instruction modality because only this setting forced the subjects to look at the screen and at the agent.

Results

A three-way repeated measure ANOVA found a significant main effect for Factor C: instruction modality. The visual modality was significantly faster than the auditory modality (mean of 1.52s versus 2.11s, $F = 103.7$, $p < .001$). This effect can be explained by the fact that subjects can almost immediately see the colour while it takes more time to understand spoken text. In addition, no cognitive translation of language to colour is needed in the visual condition.

The ANOVA also showed a significant main effect for Factor A: location of target ($F = 18.2$, $p < .001$). This also can be explained quite easily, as the subjects' hands typically initially were placed in the centre of the table. Therefore it took them less time to reach the buttons in the centre than the ones on the edges of the table. Furthermore it can be assumed that subjects learned the colour of the centre-buttons better than the colours of the peripheral buttons because the centre-buttons were more salient.

The ANOVA did not find a main effect for Factor B: agent's gaze behaviour but showed a significant interaction ($F = 7.6$, $p = .004$, means see Table 1) between agent's gaze behaviour (Factor B) and location of target (Factor A).

Although we saw that the button in the centre was pressed much faster than the other buttons, the response time was actually slower when the agent's gaze was directed at this button (by about 0,1s) than when the gaze was directed at the participant. On the other hand, we found faster response times for the peripheral buttons when the agent looked at these buttons. These findings might be based on two confounding effects that have their roots in trust and salience:

Factor B: Agent…	Factor A: Location	Mean (s)
…looks at the target	centre	1,835
	halfway to side	1,822
	Very border	1,928
…looks straight	centre	1,731
	halfway to side	1,811
	Very border	1,956

Table 1: Reaction times for different factor combinations

The reaction-speed was typically faster for the buttons right in front of the user. Users could hit the blue button in the centre faster than the others simply because they did not have to look for it; they have learned its position.

On the other hand, eye gaze towards the peripheral buttons increased user performance. This can be explained by the reduced saliency of these buttons: whereas the blue button was salient through its centeredness, the other buttons could only be seen by turning to look at them.

But why did users need more time when the agent looked at the button in the centre. Our hypothesis based on observations is that users always double-check before they hit the button. When the agent supports them, they always tend to look at the agent like someone who asks: Shall I really hit this button?" So we have two effects: The agent's gaze increases the detection of the buttons in the periphery, but it also increases the time between detection and button-press because users always want to double-check.

In the periphery both effects almost offset each other. In the centre the user's double-checking led to a clear difference between the two conditions.

EXPERIMENT THREE: MEMORY ASSISTANCE WITH GAZE

The third experiment dealt with the possibility to support everyday tasks in a "real-life" personal assistance scenario. We were interested in the effects of eye gaze support on the memorisation of tasks related to objects on the desktop.

Method

The subjects were seated in front of the system and were told that they should imagine that they just entered their office on a typical working day. Furthermore they should imagine that their personal electronic agent summarises today's tasks. Next they participated in two other experiments (Experiment one as described above, and another unrelated study). After that (approximately 30 minutes) we asked the subjects to recall the tasks, which were described by the agent.

Three different test conditions were used: a) six tasks associated with an object on display on the table and the

agent looking at it (e.g. calling someone with an telephone as the corresponding object). b) six tasks with an corresponding object on display on the table but the agent is not looking at it and c) three tasks without a corresponding object on the table (e.g. to buy flowers).

In condition a) the agent looked directly at the objects on the table when it described the task to be remembered whereas in the other conditions (b+c) it maintained eye contact with the subject. For conditions a) and b) the presentation was counterbalanced between the subjects i.e. the tasks related to an object on the table were presented alternately in both conditions with and without using the agent's gaze direction. Tasks were always presented in the same order, with condition c) irregularly intermingled. Figure 4 shows the desk layout of the experiment.

Results

Interviews revealed that the objects on the table were highly discriminable for the subjects mainly due to the spoken text, 8 of the subjects mentioned that they concentrated mainly on the spoken text and did not pay attention to the agent's gaze. The other 8 subjects said that the gaze was assisting them.

This made the final results even more surprising. The repeated measures ANOVA ($F=4.269$; $p= .023$) showed significant differences between the conditions. Post hoc comparisons with t-tests for paired samples using Bonferroni-corrected alpha levels (.025) showed significant lower recall rates of condition a) both in comparison to b) ($t = -2.529$; $p = .023$) and c) ($t = -2.629$; $p = .019$). A comparison of b) and c) does not show any differences ($t = -.865$; $p = .401$). For objects that had "gaze support" from the agent, the recall rate was actually lower than for objects where the agent maintained eye contact with the participant. This effect was consistent for 10 of the 12 test objects.

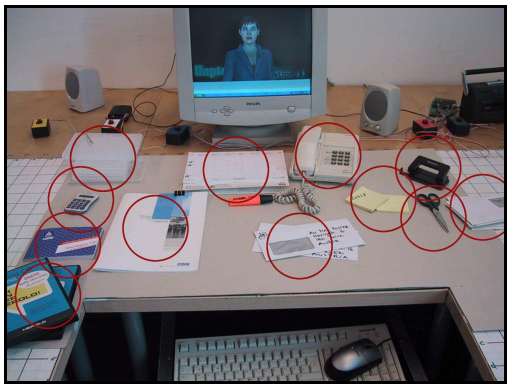

Figure 4: Desk layout of the memory assistance experiment

An explanation of this result might be that as the agent has to break eye contact with the subject to focus on an object on the table, subjects pay less attention to what the agent is saying. If we take into account that breaking and re-establishing eye contact is a very normal behaviour in daily life conversations, these results imply that interaction with an agent is maybe not as human-like as expected.

DISCUSSION

Our experiments led to three main results:

1. Users can fast but only roughly estimate the position at which an agent is looking.
2. There is evidence that an agent's gaze direction can help to guide the users' attention faster to the peripheral parts of the workspace.
3. "Pointing" towards corresponding objects during the interaction can have counterproductive effects. The users' attention might be more distracted by the changing movements than focused onto the targeted object.

Gaze patterns that are directed on real life objects in an agent-human interaction have to be implemented very carefully. Although the agent's gaze direction might help users to detect the objects under discussion we will also have to consider that users might misinterpret tiny movements and feel irritated when the agent breaks eye contact.

Further experiments should study real humans' gaze-behaviour in such situations. If we understand this in deep we might be able to develop agents whose gazes cannot only look natural but also help their users to follow and remember a conversation.

REFERENCES

1. Bickmore, T.W. (2003). Relational Agents: Effecting Change through Human-Computer Relationships, Ph.D. Thesis, MIT.
2. Colburn, R.A., Cohen, M.F. and Drucker, S.M. (2000). The Role of Eye Gaze in Avatar Mediated Conversational Interfaces. Technical Report, MSR-TR-2000-81. Microsoft Research
3. Nass, C., Steuer, J. and Tauber, E.R., (1994). Computers are Social Actors, CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems
4. Reeves, B. and Nass, C. (1996). The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, Cambridge, UK
5. van Es, I.; Heylen, D.; van Dijk, B. & Nijholt, A. (2002). Gaze behavior of talking faces makes a difference. CHI '02: Extended abstracts on Human factors in computing system
6. Vernon, D. (2004). Cognitive Vision – The Development of a Discipline, Proc. IST 2004 Event 'Participate in your future'.
7. www.haptek.com