

# On Alignment of Eye Behaviour in Human-Agent interaction

*Huub Priüst*

*Robbert-Jan Beun*

*Rogier M. van Eijk*

Technical Report 2008-039

November 2008

Department of Information and Computing Sciences

Utrecht University, Utrecht, The Netherlands

[www.cs.uu.nl](http://www.cs.uu.nl)

ISSN: 0924-3275

Department of Information and Computing Sciences  
Utrecht University  
P.O. Box 80.089  
3508 TB Utrecht  
The Netherlands

# ON ALIGNMENT OF EYE BEHAVIOUR IN HUMAN-AGENT INTERACTION

*Huub Priüst, Robbert-Jan Beun, Rogier M. van Eijk*

*Utrecht University, Department of Information and Computing Science*

*P.O. Box 80089, NL-3508 TB Utrecht, The Netherlands*

*Email: {huub,rj,rogier}@cs.uu.nl*

**Paper presented at 1<sup>st</sup> International Conference on Human-Robot Personal Relationships, June 12-13, Maastricht University, the Netherlands.**

## **ABSTRACT**

Intelligent user interfaces provide smooth interaction with the user, possibly by employing an embodied conversational agent. This paper argues that human-agent interaction improves by provoking alignment of coordination devices. Thereby we concentrate on eye behaviour. We show that automatic alignment of eye behaviour, described for human-human interaction (Pickering & Garrod, 2004), carries over to human-agent interaction. We experimentally investigate the role of alignment of eye behaviour and attention on the fluency of interaction and user-perception in human-agent interaction. A pilot study of interactions between humans and the embodied conversational agent iCat ( (Philips Research Technologies)) indicates that an agent that simulates eye contact and attention, provokes more eye contact from the user which increases fluency of interaction and perceived alertness of the agent.

## **INTRODUCTION**

In order to realize intelligent interaction between humans and computers, human-human interaction may be mimicked, and an embodied conversational agent (ECA) may be employed in the interface. In this conversational metaphor for intelligent user interfaces, the conversational agent may be more or less human-like.<sup>1</sup>

The question how human-human interaction comes about has many answers. Starting point here is the view of Clark ( (Clark, 1996)) that dialogue is a joint activity, which is carried out in coordination by the dialogue participants. Participants are coordinated in a cooperative dialogue just as participants in any successful joint activity, such as dancing, are coordinated. Some aspects of human-human interaction have been shown to carry over to human-agent communication (cf. (Bartneck, 2003), (Zanbaka, Goolkasian, & Hodges, 2006)). Our goal here is to investigate alignment of eye behaviour in this respect.

## **HUMAN-HUMAN INTERACTION**

Natural human-human interaction is composed of communicative acts and physical actions. Communicative acts may be verbal (cf. ( (Searle, 1969))) or non-verbal acts. One may distinguish between dialogue management acts versus content acts. Content acts are directly related to the

---

<sup>1</sup> See (Hutchkins, 1989) for other possible metaphors for interface design.

content of the interaction, e.g. information exchange or, in general, the task at hand, whereas dialogue management acts are concerned with the interaction itself, and directed to managing the flow of interaction. Examples of dialogue management acts are greetings, signals providing feedback on attention and processing, and error signaling. See (Bunt, 2000) for an extensive taxonomy of dialogue management acts.

### **The flow of interaction**

The flow of interaction in a dialogue is determined by the interplay between the participants by means of their communicative acts and actions. Thereby dialogue management acts play an important role. The main subjects of managing the flow of interaction are turn-taking, timing, feedback, perceptual contact, dialogue structuring, and social obligations management. For instance, it is common in dialogue that one of the participants has the floor (by speaking or acting), whereas the other participant's contributions are limited to feedback by dialogue management acts, preferably expressed in such a way that they do not interfere with the speaker / actor. Employing these 'backchannel cues', i.e. listener responses that (dis)confirm interest and understanding without interrupting the flow of dialogue, multiple modalities are advantageous, such as nodding (or shaking), eye contact, short vocal utterances such as 'yes', 'uhuh', etc. Viewing human dialogue as an ongoing joint activity, dialogue management acts are acts by which participants coordinate the next step in their ongoing joint activity. Dialogue management acts are coordination devices: they are employed to coordinate the interaction. See (Clark, 1996) for other coordination devices.

### **Automatic alignment**

(Dijksterhuis & Bargh, 2001) argue that social behaviour is based for an important part upon direct links between perception and behaviour: much social behaviour is automatically triggered by perception of actions of others. The majority of routine social behaviour follows the 'perception-behaviour expressway' ( (Dijksterhuis & Bargh, 2001)), i.e. a direct linking between perception and action. These findings are based on research on mirror neurons in the neuropsychological literature. There is evidence for automatic links controlling speech, facial expressions, gestures, posture and other nonverbal behaviour. For instance, in the literature on speech it has been established that participants in a cooperative dialogue align with regard to their dialect, speaking rate and pausing frequency (cf. (Street, 1984), (Cappella & Planalp, 1981)); furthermore dialogue participants may mimic foot shaking and nose rubbing carried out by a person with whom they interact ( (Chartrand & Bargh, 1999))<sup>2</sup>; and it has recently been established that dialogue participants involved in a cooperative task automatically converge their posture ( (Shockley, Santana, & Fowler, 2003)). Pickering & Garrod ( (Garrod & Pickering, 2004), (Pickering & Garrod, 2004)) have applied the findings on the perception-behaviour expressway to their theory of human dialogue. They agree with Clark ( (Clark, 1996)), that dialogue is a joint activity which involves cooperation between dialogue participants in a way that establishes a joint meaning of the dialogue as a whole, but they add automatic alignment, a device that facilitates language processing in dialogues. To come to a joint understanding, participants align their representational models at various levels: words, syntactic, semantic, situational alignment.

So, the flow of interaction in human dialogue is managed largely by dialogue management acts, and is partly determined by non-conscious automatic processes. It has, as far as we know, not been investigated whether automatic alignment carries over to agent-human interaction.

### **Eye behaviour**

Eye behaviour serves many functions in human interaction (cf. ( (Argyle & Cook, 1976), (Leathers, 1997))). Like other communicative acts, eye signals may be used as content acts and as dialogue management acts. To contribute to the content of the interaction, a speaker may communicate beliefs, intentions or affective state by means of eye behaviour, e.g. express (un)certainly on a topic or express emotional state. A hearer may indicate degree of attentiveness (paying attention/ interest / arousal / intimacy) by means of his/her eye behaviour. In addition, eye behaviour may indicate the relationship between participants (power / status / impression management). See (Poggi, Pelachaud, & Rosis, 2000) for an extensive typology of meanings of

---

<sup>2</sup> (Chartrand & Bargh, 1999) refers to the non-conscious mimicry of behaviour in human interaction as the *chameleon effect*.

eye behaviour. As for managing the flow of interaction, eye behaviour is efficient in structuring the dialogue (e.g. stress certain information by eye behaviour) and in turntaking ( (Vertegaal, Shell, Chen, & Mamuji, 2006), (Bunt, 2000)): e.g. looking away is a means for a speaker to keep the floor. Resuming eye contact, sometimes referred to as gaze, with a dialogue partner is a way to pass the turn to the hearer. Asking for a turn may be done by widely opening the eyes, comparable to taking breath for starting to speak. Eye behaviour may also be employed to 'ask for' feedback. Like other behaviour, eye behaviour, especially dialogue management acts, seems partly determined by automatic alignment.

## **HUMAN-AGENT INTERACTION**

Several studies suggest that users in interaction with embodied conversational agents behave as they normally do in social interaction, and apply 'social heuristics' (cf. (Reeves & Nass, 1998) and (Rickenberg & Reeves, 2000)). ECAs may inhibit different embodiments with different degrees of anthropomorphisation. As for embodiment, (Bartneck, 2003) studies the difference between robotic versus screen embodiment, and shows effects of social facilitation in that embodied robotic characters seem to have a stronger social facilitation effect than embodied screen characters. When ECAs show human-like appearances and behaviour, users tend to ascribe human characteristics to them or 'anthropomorphise' them. The question how the degree of anthropomorphisation of an embodied agent influences interaction is still open. (Beun, Vos, & Witteman, 2003) shows that the mere presence of an ECA in the interface, independent of the human-like character, has positive influence on retainability of information by the user.

### **Carry over**

In recent studies, it has been shown that several aspects of human interaction carry over to the interaction between humans and human-like ECAs. For instance, (Zanbaka, Goolkasian, & Hodges, 2006) shows effects of gender with regard to persuasiveness in that subjects may be more persuaded by agents of the opposite sex. Rickenberg & Reeves ( (Rickenberg & Reeves, 2000)) indicate that employing an embodied agent in the interface may increase the perceived presence of the user.

Some aspects of eye behaviour were shown to carry over to human-agent interaction: (Garau, Slater, Bee, & Sasse, 2001) shows that adding eye gaze to an ECA improves the 'communication experience' of subjects, but only when eye gaze is related to the conversational flow (e.g. turn taking) and inferred from the audio stream. (Garau, Vinayagamoorthy, Brogni, Steed, & Sasse, 2003) found that 'low-anthropomorphic' agents are adversely affected by adding eye gaze, whereas adding eye gaze to 'high-anthropomorphic' agents shows significant interaction effects on the perceived quality of interaction. So, more eye gaze control only leads to higher perceived quality of communication if the anthropomorphic degree is high.

## **FOCUS OF THIS STUDY**

The focus of this study is on alignment of eye behaviour as dialogue control acts in human-agent interaction. Since we can control the behaviour of the agent, we pose the following research questions: 1. Does simulation of eye contact by a monitoring agent provoke eye contact from the user? 2. Does simulation of attentive eye behaviour by a monitoring agent enhance the flow of interaction? and 3. Does it have any effect on the perceived quality of interaction?

## **EXPERIMENT**

To answer these research questions, the iCat platform ( (Philips Research Technologies)) was extended with an object tracker to be able to monitor / attend to the subjects, simulating eye contact, and with a speech module that provides prerecorded utterances. A pilot experiment was run at the CKE Usability Lab at Utrecht University. The sample size doesn't allow any firm conclusions, but gives some indication.

### **Task**

To evoke an interaction consisting of an interplay between communicative acts and physical actions, a simple task was created. After a short introduction, subjects were asked by iCat to build

a block tower in several steps. To enforce movement, subjects had to gather building blocks from a table that was at 1.5 meters distance from the table at which the tower was built. The distance between iCat and subject is 1.5 meters; a distance at which iCat's head movement in combination with its gaze in the direction of the subject is perceived as eye contact. Interaction between subject and iCat is conducted by iCat through questions and instructions. Questions and instructions were prerecorded and these dialogue building blocks were triggered manually by the experiment leader, who monitors the interaction, so the (number of) dialogue acts is controlled. On the part of the subject, only few verbal responses were provoked, most reactions were physical (taking blocks, walking towards some table, building). Both agent and subject had visual access to each other and to the task domain, but only the subject could manipulate objects in the domain.

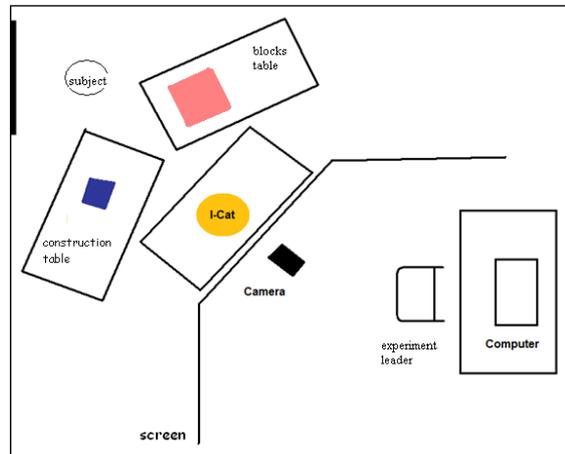


Fig 1. Experimental setting

### Conditions

We considered two conditions: a dynamic versus a static condition: The dynamic condition incorporates a dynamic character in the interface i.e. a screen character iCat that monitors its dialogue partner by means of head movement and eye contact. In the static condition, the embodied agent is motionless, staring aside.

### Implementation

To establish the dynamic condition, an object tracker was implemented and added to the screen version of the iCat. The Philips OPPR platform was extended with an object tracker based on the Camshift Algorithm. By movement of the head, the iCat can follow the subject in where s/he goes and what s/he does so that the subject is tracked continuously. During verbal interaction in the dynamic condition, the iCat's gaze is in the direction of the subject, thus implementing primitive gaze control. (A more sophisticated implementation of gaze is reported in (Poel, Heylen, Nijholt, Meulemans, & van Breemen, 2007).)

### Questionnaire

After a short introduction by iCat, subjects were asked to build a blocks tower in several steps. After the experiment, subject were asked to fill in a questionnaire. This questionnaire was developed to measure the perceived quality of the interaction. The 7-point scale questionnaire covered 20 items on 3 topics: the perceived social skills of the embodied agent (kindness, alertness, trustfulness etc.), the interaction (effectiveness, pleasantness, etc.) and the task itself (difficulty, enjoyability, etc.).

### RESULTS

We ran a modest pilot experiment with 9 subjects, each subject performing the task in either the static or the dynamic condition. We video recorded the interactions, hand-coded the data and

analysed the quantitative data on frequency of eye behaviour and the qualitative data from the questionnaires.

#### **Frequencies**

We measured the number of times that a subject looked at the screen character during the interaction. We found a significant difference between the two groups: the group of subjects that interacted with the monitoring iCat seeking eye contact versus subjects interacting with a static character. Subjects in the dynamic character condition sought eye contact more often (average = 46.0) than subjects in the static character condition (36.75). The difference is significant ( $t(7) = -2.391$ ,  $p = .032$ , one-tailed). This means that the absolute frequencies of eye contact differ and also the frequencies of eye contact, relative to the number of dialogue acts, since the number of dialogue acts was kept constant throughout the different dialogues. We also found a significant difference between the groups regarding frequency of eye contact relative to the absolute duration of the interaction ( $t(7) = 4.867$ ,  $p = .012$ ). This means that the answer to the first research question is: yes, subjects tune their eye contact to the eye behaviour of the agent, i.e. users align their eye behaviour, resulting in more eye contact. This is in line with the view on dialogue as joint activity.

#### **The flow of interaction**

Does more attention and eye contact influence the flow of interaction with the iCat? We know from human-human interaction that eye contact enhances fluency. To answer this question we have analysed when, i.e. at which moments in the dialogue, subjects looked at the iCat: while listening, while performing an action (e.g. picking blocks / moving from one table to the other / building the tower), or directly after some action (releasing the turn, waiting for further instruction). There was a tendency for the subjects in the dynamic condition to make eye contact with iCat relatively more often while moving and after having performed an action than subjects in the static condition, but not while listening. However, the difference was not clear cut.

#### **Perceived alertness and enjoyability**

We found no difference between the two groups as far as the overall concepts of social skills of the agent, interaction and task were concerned. On individual items, we found two significant differences in that subjects in the dynamic condition perceived higher alertness of the iCat ( $t(7) = -3.656$ ,  $p = .035$ , nonequal variances) and experienced the interaction more enjoyable ( $t(7) = -3.657$ ,  $p < .008$ ).

### **CONCLUSION AND DISCUSSION**

Automatic alignment in human-agent interaction has hitherto not explicitly been taken into account in the design of intelligent user interfaces, though it plays an important role in human dialogue. We have focused on alignment of eye behaviour. The pilot study indicates that an agent in the interface that simulates attention and eye contact, results in different behaviour of the subjects: it provokes eye contact from the user. This is an automatic alignment process; in the pilot, it enhanced the perceived alertness of the agent and the enjoyability of the interaction. We are currently running experiments at Utrecht University with more subjects and with Philips physically embodied iCat to see whether the results above can be confirmed and worked out in more detail (employing Observer XT for annotation).

Since more backchannel behaviour increases fluency in human-human interaction, we have argued that it is important for the design of successful intelligent user interfaces that incorporate autonomous agents, to take alignment of coordination devices into account. Results in (Garau, Vinayagamoorthy, Brogni, Steed, & Sasse, 2003) suggest that alignment of a user to an agent may depend on how human-like the embodied agent is: A user may align strongly with a human-like embodied agent, whereas a non-anthropomorphic agent may invoke less alignment. This may be due to user expectations (Reeves & Nass, 1998), and raises the question which factors should be considered to determine whether an embodied agent should represent the intelligent interface and which degree of anthropomorphism an agent should incorporate (e.g. type of user (Rickenberg & Reeves, 2000)), task at hand, desired degree of equalness of the interlocutors).

It is important to stress that this paper has, obviously, highlighted only one side of the medal: alignment of the user with the agent. Modern research on intelligent user interfaces is heading towards perceptive agents, both in technical and in conceptual sense (cf. (de Croon, Postma, & van

den Herik, 2006), (Vertegaal, Shell, Chen, & Mamuji, 2006)). The current study indicates that the user (unconsciously) adapts to an agent that provokes alignment of nonverbal aspects such as eye behaviour, which enhances the flow of interaction. In future, perceptive agents may take care of their 'part of the bargain' and adapt to the (eye) behaviour of the user, resulting in balanced alignment.

## REFERENCES

- Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. London: Cambridge University Press.
- Bartneck, C. (2003). Interacting with an Embodied Emotional Character. *Proceedings of DPPI'03*. Pittsburgh, Pennsylvania, USA.
- Beun, R., Vos, E., & Witteman, C. (2003). Embodied Conversational Agents: effects on memory performance and anthropomorphisation. *4th Int. Workshop IVA*. Kloster Irsee.
- Bunt, H. (2000). Dynamic Interpretation and Dialogue Theory. In H. D. 1. Bunt, *The Structure of Multimodal Dialogue II* (pp. 139-166). Amsterdam: North Holland.
- Cappella, J., & Planalp, S. (1981). Talk and silence sequences in informal conversations: III Interspeaker influence. *Human Communication Research*, (1) Pages 117-132.
- Chartrand, T., & Bargh, J. (1999). The Chameleon Effect: The Perception-Behaviour Link and Social Interaction. *Journal of Personality and Social Psychology*, Vol.76 No.6, 893-910.
- Clark, H. (1996). *Using language*. New York: Cambridge University Press.
- de Croon, G., Postma, E., & van den Herik, H. (2006). A situated model for sensory-motor coordination in gaze control. *Pattern Recognition Letters*, Vol. 27, Issue 11, Pages 1181-1190.
- Dijksterhuis, A., & Bargh, J. (2001). The perception-behaviour expressway: automatic effects of social perception on social behaviour. *Advances in Experimental Social Psychology*, Academic Press, Vol. 33 (2001), Pages 1-40.
- Garau, M. S., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. (2003). The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment. *CHI 2003*, (pp. Vol.5, No.1, Pages 529-536).
- Garau, M., Slater, M., Bee, S., & Sasse, M. (2001). The impact of eye gaze on communication using humanoid avatars. *CHI 2001*, (pp. 309-316).
- Garrod, S., & Pickering, M. (2004). Why is conversation so easy? . *TRENDS in Cognitive Sciences*, Vol.8 No.1 Pages 33-39.
- Hutchkins, E. (1989). Metaphors for Interface Design. In M. e. Taylor, *The Structure of Multimodal Dialogue* (pp. 11-28). Amsterdam: North Holland.
- Leathers, D. (1997). Chap.3 Eye Behaviours. In D. Leathers, *Successful Nonverbal Communication*. Boston: Allyn and Bacon.
- Philips Research Technologies*. (n.d.). Retrieved March 28th, 2008, from Philips Research Technologies: [http://www.research.philips.com/technologies/syst\\_softw/robotics/](http://www.research.philips.com/technologies/syst_softw/robotics/)
- Pickering, M., & Garrod, S. (2004). Toward a mechanistic Psychology of Dialogue. *Behavioral and Brain Sciences*, (27) 169-190.
- Poel, M., Heylen, D., Nijholt, A., Meulemans, M., & van Breemen, A. (2007). Gaze Behavior, Believability, Likability and the iCat. *Proceedings 6th Workshop on Social Intelligence Design* (pp. 109-124). Nijholt, A., Stock, O., Nishida, T. (eds).
- Poggi, I., Pelachaud, C., & Rosis, F. d. (2000). Eye Communication in a Conversational 3D Synthetic Agent. *Special Issue of Artificial Intelligence Communications, IOS Press*, Vol. 13 (3), Pages 169-181.
- Reeves, B., & Nass, C. (1998). *The Media Equation*. Cambridge: Cambridge University Press, CSLI.

Rickenberg, R., & Reeves, B. (2000). The Effects of Animated Characters on Anxiety, Task Performance, and Evaluations of User Interfaces. *CHI Letters (2000)*, (pp. Vol. 2, Issue 1, Pages 49-56).

Searle, J. (1969). *Speech Acts*. London: Cambridge University Press.

Shockley, K., Santana, M.-V., & Fowler, C. (2003). Mutual Interpersonal Postural Constraints are Involved in Cooperative Conversation. *Journal of Experimental Psychology: Human Perception and Performance* , Vol. 29 No.2, Pages 326-332.

Street, R. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research* , Vol. 11:2, Pages 139-169.

Vertegaal, R., Shell, J., Chen, D., & Mamuji, A. (2006). Designing for augmented attention: Towards a framework for attentive user interfaces. *Computers in Human Behaviour* , Vol. 22, Pages 771-789.

Zanbaka, C., Goolkasian, P., & Hodges, L. (2006). Can a Virtual Cat persuade You? The Role of Gender and Realism in Speaker Persuasiveness. *CHI Proceedings (2006)*, (pp. 1153-1161).