Highly Realistic 3D Presentation Agents with Visual Attention Capability

Arjen Hoekstra[1,2], Helmut Prendinger[1], Nikolaus Bee[3], Dirk Heylen[2], and Mitsuru Ishizuka[4]

[1] National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
helmut@nii.ac.jp
[2] Computer Science, Human Media Interaction
University of Twente, PO Box 217, 7500 AE Enschede The Netherlands
a.h.hoekstra@student.utwente.nl, heylen@cs.utwente.nl
[3] Institute of Computer Science, University of Augsburg
Eichleitnerstr. 30, D-86135 Augsburg, Germany
[4] Graduate School of Information Science and Technology, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
ishizuka@i.u-tokyo.ac.jp

Abstract. This research proposes 3D graphical agents in the role of virtual presenters with a new type of functionality – the capability to process and respond to visual attention of users communicated by their eye movements. Eye gaze is an excellent clue to users' attention, visual interest, and visual preference. Using state-of-the-art non-contact eye tracking technology, eye movements can be assessed in a unobtrusive way. By analyzing and interpreting eye behavior in real-time, our proposed system can adapt to the current (visual) interest state of the user, and thus provide a more personalized, context-aware, and 'attentive' experience of the presentation. The system implements a virtual presentation room, where research content of our institute is presented by a team of two highly realistic 3D agents in a dynamic and interactive way. A small preliminary study was conducted to investigate users' gaze behavior with a non-interactive version of the system. A demo video based on our system was awarded as the best application of life-like agents at the GALA event in 2006.[1]

1 Introduction

The challenge of giving a good presentation is to provide relevant and interesting content in an easily accessible way while keeping the attention of the audience during the entire presentation time. Human presenters often obtain feedback from listeners regarding their level of attention by simply looking at their behavior, specifically whether they are looking at the currently presented material, typically visualized on slides, at the presenter, or somewhere else. If a presenter,

[1] http://hmi.ewi.utwente.nl/gala/

A. Butz et al. (Eds.): SG 2007, LNCS 4569, pp. 73–84, 2007.

e.g. a museum guide, observes that the attention of the spectators is diverted by other objects, he or she will try to adapt the presentation by taking the interest shift of the audience into account.

Although speech conveys the richest information in human-computer interaction, it is not the preferred input modality for scenarios such as presentation settings, which, as monologues, typically do not assume verbal expressions of interest from the audience. To determine the user's current focus of attention and interest, we therefore propose a system that is based on human eye movements. As an input modality, eye gaze has the advantage of being an involuntary signal that reflects the user's visual interest [16], and its signal is robust and can be assessed accurately [5].

As an interactive system, our proposed setup can be conceived as reviving the 'self-disclosing display' concept introduced in [21], where eye gaze is utilized as an input modality to recognize and respond to a user's interest. Their system would zoom in to areas of user interest and provide explanations via synthesized speech. Our work extends this concept by detecting both user interest and preference between two (visual) alternatives to continue the presentation, and by embodied life-like characters rather than a disembodied voice.

The remainder of this paper is structured as follows. Section 2 discusses related work. Section 3 describes our methods to assess (visual) interest and (visual) preference. Section 4 provides details about the application scenario and the gaze-contingent responses of the agents. In Section 5, we report on the main findings of our preliminary study based on a non-interactive version of the agent application. Section 6 discusses and Section 7 concludes the paper.

2 Related Work

Life-like characters or virtual animated agents are intended to provide the illusion of life or 'suspend disbelief' [3], such that users interacting with those agents will apply social interaction protocols and respond to them as they would to other humans, e.g. by listening to their story and attending to their face and gestures through eye gaze [15]. Life-like characters have been shown to serve multiple purposes successfully; besides presenters, they can act as tutors, actors, personal communication partners, or information experts [14]. An excellent overview of the evolution of virtual presenters – starting from a non-interactive single presenter to multiple presenters with the capability of processing natural input from users – can be found in [18].

Eyes are an intriguing part of the human face, and are sometimes even seen as 'windows to the soul'. In verbal communication, the major functions of eye gaze include paying and signaling attention, conversation regulation, and demonstration of intimacy [1,8]. In dyadic communication, two types of core functions of gaze direction can be distinguished [10]: (1) *Monitoring functions*. The speaker observes the interlocutor to gather information about the listener's attending behavior (e.g. at the end of long utterances) or about signals indicating that the listener will 'take floor'. (2) *Regulatory and expressive functions*. The speaker manages turn taking, e.g. by looking away from the listener to signal that he or she is going to speak. The speaker may also express an emotion of embarrassment or high arousal by averting the eyes from the listener. An investigation of gaze direction among multiple conversational partners can be found in [22]. It should be noted that the usage of our system differs from the interaction situation described above in that users cannot converse with agents verbally. In our case, the agents will monitor a user's state of visual interest (or lack of interest) in the presentation, and react accordingly.

Recent attempts to integrate eye behavior into interactive systems are reported in [23,19], who discuss the use of eye tracking in various applications - so-called 'attentive user interfaces' or 'visual attentive interfaces'. *Magic Pointing* is based on the user's conscious eye behavior in order to improve the positioning of a mouse pointer [23]. An eye-pointer approach is implemented in our augmented reality videoconferencing system called *iPick* [2]. *InVision* exploits involuntary gaze movements to estimate a user's plan or needs [19]. Similar *InVision*, our system exploits the non-conscious nature of eye movements in a non-command fashion.

Approaches that rely on eye gaze as input and life-like characters as communication media are currently rare. In the context of computer mediated communication, [6] conducted a study where different conditions (video, audio-only, avatar with random gaze, avatar with informed gaze) are compared for their effectiveness. Here, the virtual avatar based on a gaze model derived from [10] scored highest in terms of naturalness and involvement (experience) except for the video condition. Another study assessed the level of engagement and attentiveness in a situation where two humans converse with a virtual agents [17]. The authors were able to confirm findings from human communication, e.g. that persons look more at their conversational partners as listeners than as speakers. An agent-specific (and possibly surprising) result is that users attend more to an agent talking to them than to the human speaker.

3 Interest Estimation

The focus of interest is determined by a modified version of the algorithm described in [16]. These authors implemented an intelligent virtual tourist information environment (*iTourist*), for which they propose a new interest algorithm based on eye gaze. Two interest metrics were developed: (1) the Interest Score (IScore) and (2) the Focus of Interest Score (FIScore). IScore refers to the object 'arousal' level, i.e. the likelihood that the user is interested in a (visual) object. When the IScore metric passes a certain threshold, the object is said to become 'active'. The FIScore calculates the amount of interest in an active object over time. Since we were mainly interested in whether a user's attention is currently on a particular object, a simplified version of the IScore metric was sufficient for our purpose. The basic component for IScore is $p = T_{ISon}/T_{IS}$, where T_{ISon} refers to the accumulated gaze duration within a time window of size T_{IS} (in our application, 1000 milliseconds). In order to account for factors that may enhance or inhibit interest, [16] characterize the IScore as $p_{is} = p(1 + \alpha(1 - p))$. Here, α encodes a set of parameters that increase the accuracy of interest estimation.

The modification factors are modelled as follows [16]:

$$\alpha = \frac{c_f\alpha_f + c_c\alpha_c + c_s\alpha_s + c_a\alpha_a}{c_f + c_c + c_s + c_a}$$

The terms in this formula are defined as:

- α_f is the frequency of the user's eye gaze 'entering' and 'leaving' the object ($0 \le \alpha_f \le 1$),
- α_c is the categorical relationship with the previous active object ($\alpha_c = -1|0|1$),
- α_s is the average size of all possible interest objects compared to the size of the currently computed object ($-1 \le \alpha_s \le 1$),
- α_a encodes whether the object was previously activated ($\alpha_a = -1|0$), and
- c_0, c_f, c_c, c_s, and c_a represent empirically derived constant values of the corresponding factors. Some of these factors are domain dependent and are thus not applicable in all contexts.

The factors α_c and α_a were not (yet) integrated to our system. α_c concerns (semantic) relations between objects; α_a can be used to make the system respond in a different way when an object is activated multiple times.

We continue by explaining α_f and α_s, the two remaining factors. α_f is represented as $\alpha_f = \frac{N_{sw}}{N_f}$, where N_{sw} denotes the number of times eye gaze enters and leaves the object and N_f denotes the maximum possible N_{sw} in the preset time window. When the user's gaze switches to some object many times, the value of the modification factor will increase and hence there will be a higher chance on excitation. α_s is represented by $\alpha_s = \frac{S_b - S}{S}$, whereby S_b represents the average size of all objects, S denotes the size of the currently computed object, and the smallest object is never more than twice as small as the average object. This modification is intended to compensate for the differences between the size of the potential interest objects. Due to some noise in the eye movement signal, larger objects could have a higher chance of being 'hit' than smaller ones, which should be avoided.

In order to determine the user's preference in situations involving a two-alternative forced choice (2AFC), i.e. "how the presentation should continue", we exploited the so-called 'gaze cascade' effect. This effect was discovered in a study where users had to choose the more attractive face from two faces [20]. It could be demonstrated that there was a distinct gaze bias towards the chosen stimulus in the last one and a half seconds before the decision was made.

Our system integrates a recently developed real-time component for automatic visual preference detection, the *AutoSelect* system, which is based on the gaze cascade phenomenon [4]. *AutoSelect* was tested in a study where users were instructed to choose their preferred necktie from two presented neckties, i.e. in a 2AFC setting. There was no input modality available except the subjects' eye gaze. After the decision of *AutoSelect*, subjects were asked to confirm (or reject) the result of the system. Starting from an initial set of thirty-two pairs of neckties, subjects repeatedly indicated their preference, amounting to sixty-two decisions. The system achieved an accuracy of 81%. In order to avoid 'polite' answers from subjects, we conducted a follow-up study, where system and subject decisions were assessed independently. The accuracy of the re-designed system was slightly lower (72%).

Examples of the exploitation of the gaze cascade effect and of the use of the interest algorithm will be given in the next section.

4 Responding to User Interest and Preference

Our implemented system involves a team of two presentation agents that introduce the user to research at the National Institute of Informatics (NII), Tokyo (see Fig. 1 and video clip[2]).

The two agents were designed based on the appearance of two famous Japanese actors. In order to support their life-likeness, the agents are highly realistic and expressive. They can perform various gestures, such as greeting and counting, or 'beat' and deictic gestures. In addition to body gestures, mimics for "joy", "surprise", and "sadness" are available. High-quality synthesized speech using the Text-to-Speech (TTS) engine from Loquendo[3] is combined with proper lip synchronization, and the head of the agents can be adjusted to any (natural) direction, e.g. to the direction of the other agent when giving turn, or to the virtual slide. (When listening to a presentation, paying attention to its visualized content is of key importance. However, the audience will also focus on the presenter's face to increase comprehension via perception of lip movements in addition to speech, especially when listeners are not native speakers of English, as in our case.) A discussion about the animation techniques employed for our agents can be found in [12].

Both the agents and the environment are controlled by MPML3D [13], a reactive framework that supports anytime interaction, such as real-time interpreted input from the eye tracker. The agents will adapt their performance based on user eye gaze in two ways:

- If the user shows interest in a particular interface object (an 'interest object') not currently discussed (e.g. the view), or non-interest in a currently discussed object (e.g. a presentation slide), the agents will interrupt their presentation and react accordingly.
- At decision points in the presentation flow, the user's preference determines the subsequent topic.

4.1 Adapting to User Interest

In the system, the following interest objects are defined (see Fig. 1; from left to right): (a) NII logo; (b) male agent; (c) left part of the slide; (d) right part of

[2] A demo video can be found at `http://research.nii.ac.jp/~prendinger/GALA2006/`

[3] `http://www.loquendo.com/`

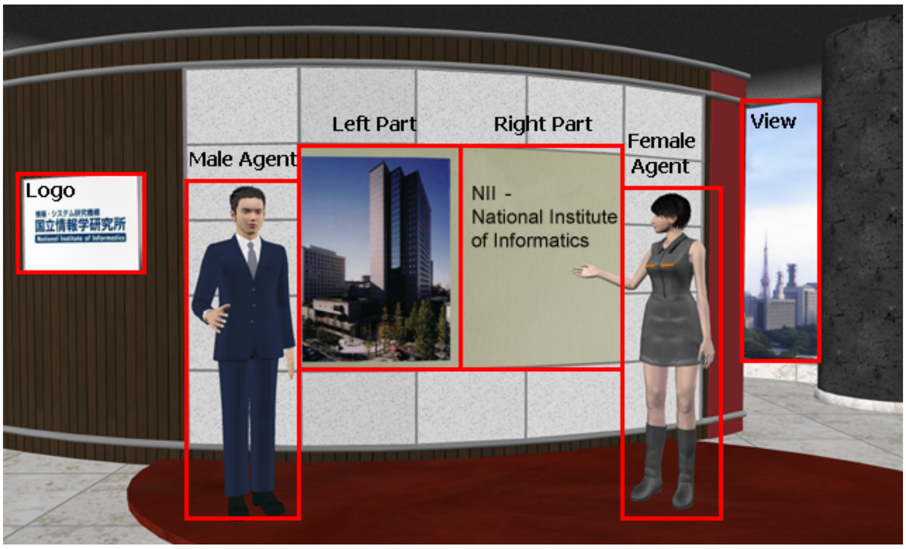

Fig. 1. Interest objects in the virtual environment

the slide; (e) female agent; (f) the view out of the window to the right. For each interest object, the IScore is calculated every second. When the score exceeds the threshold, the object becomes 'activated' and the agent(s) will react (if a reaction is defined). Agent responses (or non-responses) are defined for three types of situations:

1. *Continuation of presentation*: If the user attends to the currently explained (part of a) presentation slide (which is desired), the agent will continue with the presentation. Fig. 2 depicts a situation where the user attends to the explanation of the male agent by gazing at the slide content.
2. *Interruption of presentation*: If the user is detected to be interested in an interface object that is not currently discussed, the system chooses between two responses:
   (a) Suspension: If e.g. the user looks out of the virtual window (at the "View" object) rather than attending to the presentation content explained by the male agent, the female co-presenter agent asks her colleague to suspend the research presentation and continues with a description of the view (see Fig. 3).
   (b) Redirecting user attention: Here, the presenter agents do not suspend the presentation to comply with the user's interest. Instead, the co-presenter alerts the user to focus on the presentation content.

The existing implementation of our presentation system handles interruptions in a simple way. If a user's interest object is not the currently explained object (typically a slide) the presentation will be suspended at first by providing information about that object, and subsequently, the co-presenter agent will try to redirect the user to the presentation content.

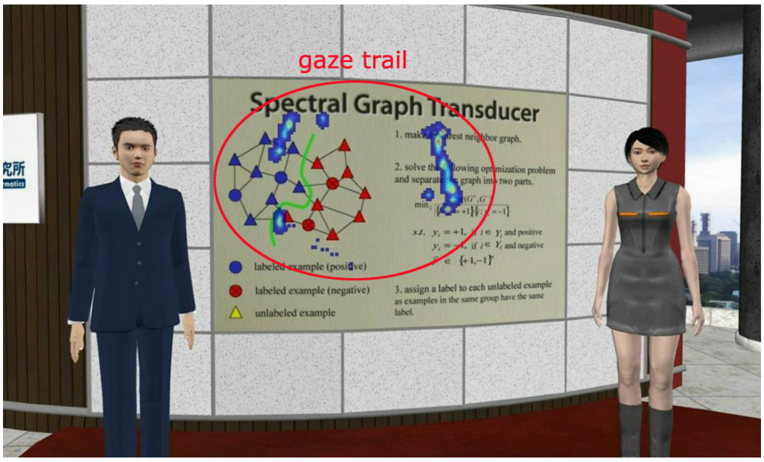

Fig. 2. User is interested in slide content. The corresponding gaze trail is visualized by 'heat trails'.

Ken:	The transductive learning makes up for the smallness of user feedback. The transducer assigns labels from which the relevance is unknown with the same label as neighboring terms with known relevance.
User:	*[IScore exceeds threshold of the "View" object and gets activated.]*
Yuuki:	Ken, hold on a second... I couldn't help noticing that you are admiring the view we have from NII at the city. You can see our building is very close to the Imperial Palace. All the greenery belongs to the park of the palace. Well, so much about our neighborhood. Let's go back to our presentation, but please concentrate this time.

Fig. 3. Sample agent dialogue when the user is interested in the outside view

4.2 Following User Preference

At predefined points during the presentation, the agents ask the user to choose the next presentation topic, while a slide depicting two options is displayed. The gaze cascade phenomenon will occur naturally in this situation. Users alternately look at the left part and the right part of the slide, and eventually exhibit a bias for one part. The decision process occurs within seven seconds. Thereafter, the presentation continues with the selected topic.

5 Exploratory Study

A small study was conducted to assess users' eye behavior when watching a non-interactive version of the research presentation by the agent team, i.e., although eye gaze was recorded, the agents did not adapt to user interest or preference.

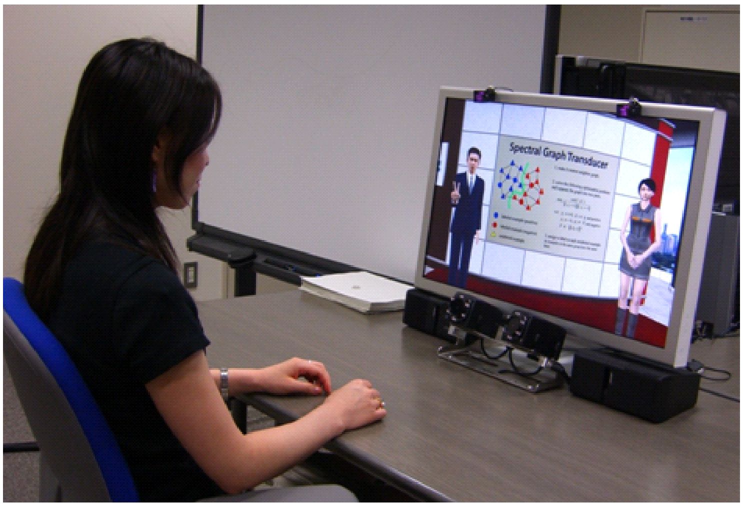

Fig. 4. Experimental setup

This approach seemed justified as a first step, given the lack of experience with attentive behavior of human spectators of a presentation performed by two animated agents. Hence, the aim of the study was to assess likely types of gaze behaviors. This information can then be used to refine the functionality of the interactive system, which will be followed by an extensive study.

5.1 Method

Subjects: The data of four subjects (average 30 years) were analyzed. Subjects received a monetary compensation for participation (1,000 Yen).

Apparatus and Procedure: Subjects were seated in front of a 30 inch screen (distance 80 cm) and stereo cameras of the faceLAB eye tracker from Seeing Machines.[4] The cameras and speakers were located below the screen. Two infrared pods were attached at the upper part of the display for illumination of the eyes (see Fig. 4). Then calibration of each subject was performed. Subjects were given no instruction other than watching the presentation.

In the presentation prepared for the study, the agents first introduce themselves, and then explain the researches of three professors of NII. The total length of the presentation is 14:49 min.

Data Analysis: The faceLAB software allowed us to extract the coordinates of gaze points on the screen. The recorded data was then processed and analyzed with MATLAB. 'Heat trails' were used for visualization, as they present the amount of fixations over time as a continuous movie. The heat trails were made transparent with the chroma key (Bluescreen) effect and merged with the

[4] `http://www.seeingmachines.com/`

captured video of the presentation. Animations and (virtual) environment changes were analyzed with the ANVIL annotation tool [11].

5.2 Results

The most distinctive result of the study could be found for situations where the agents ask the subject to select the subsequent topic. All of the subjects showed the gaze pattern characteristic of the 'gaze cascade' effect in both occurrences of a decision situation. This outcome generalizes the results of [20,4] to a setting featuring two agents referring to slide content depicting two choices (displayed left and right on the slide). It indicates that the cascade phenomenon can be reliably used to let users select the continuation of the presentation in a non-command fashion.[5]. It should be noted, however, that in the non-interactive presentation shown in the study, the subjects' preference had no effect on the continuation of the presentation.

Deictic arm gestures of embodied agents and agents' head-turning to slide content are an effective way to redirect the attention of users to the (virtual) slides [15]. We were interested in differences in the effect of deictic gestures depending on whether a new slide is shown, or some textual content of a displayed slide is changed, e.g. a new item is added to a given slide content. In the study, every subject had noticed a new slide within 2 sec (19 new slides presented). On the other hand, changes on slides (18 occurrences) were noticed with some delay, with 97% redirected attention within 3 sec. Although we expected more occasions where an attentive agent would have to alert the user, a 15 min presentation is probably too short to observe a user's diminishing attention.

The functionality of the interactive system also provides for the possibility that users attend to interface objects not related to the presentation content, such as the NII logo or the view outside the building (see Fig. 1). In the study, however, subjects spent 99% of the total time on the agents or slides. Since the actual view of the subjects was essentially limited to those interface objects (see Fig. 4), there was little room for attending to anything else. Other results regarding cumulative gaze distribution include attention to speaking agent (53%), attention to presented slides (43%), and attention to non-speaking agent (3%).

6 Discussion

While gaze-contingent interfaces are getting increasingly popular [5], it remains an open question how 'responsive' an interface that relies on eye gaze as an input modality should be. The problem of distinguishing between eye movements that are just explorative and those that are meant as an input is known as the 'Midas Touch' problem: "Everywhere you look, another command is activated; you cannot look anywhere without issuing a command." [9, p. 156]. Our presentation system avoids the Midas touch problem by (i) strictly confining the

[5] Given the small sample size, our results should always be seen as preliminary.

screen areas that could yield an agent response (the interest objects), and (ii) calculating user interest based on a well-established metric [16].

While we can determine visual interest and preference with high accuracy, the system is not perfect and might misinterpret the user's intention. The current implementation does not allow the user to 'undo' incorrect system decisions. Conceivable remedies include: (1) the user explicitly confirms a decision of the system, e.g. by simple verbal input ("Yes", "No"), and (2) the system detects the user's acknowledgement (head nodding) or negative signal (head shaking).

Another question concerns the manner in which an interface should be manipulated by gaze behavior. [7] propose the following types of interactivity:

1. *Change in information.* Objects of user interest are explained in more detail.
2. *Change in point-of-view.* The camera position changes to the interest object of the user.
3. *Change in storyline.* The sequence of story events is dependent on where the user attends to.

Our current system supports all of those possibilities (to some extent). For instance, when the user is interested in the virtual outside view, the agent provides additional explanation of the view and the (virtual) camera shifts to show a full screen image of the view. Interest objects can also be conceived as 'hyper-links' to particular scenes of a story. Our gaze cascade based selection method can be seen as a (restricted) implementation to decide the progress of the presentation.

7 Conclusions

The use of eye gaze offers a powerful method to adapt a presentation to the current interest of a user, i.e. make the presentation contingent to user interest. Eye gaze as an input modality is particularly beneficial when verbal feedback is either not assumed or difficult to provide. Most presentations given by lecturers or museum guides are one-way communications that can nevertheless be adaptive to the audience if a presenter is sensitive to the behavior of the audience, such as their exhibition of visual interest or non-interest. Furthermore, while the audience certainly has interest in specific presentation topics or parts of the presentation, it is unusual (or at least impolite) to verbally point out objects of interest repeatedly during the presentation. The online analysis of eye behavior thus provides an unobtrusive method to estimate user interest continuously.

In this paper, we describe a presentation system that features two virtual 3D presentation agents capable of responding to a user's focus of attention and interest in a natural way. The agent system [13] supports synchronized speech and lip movements, timed gestures, mimics, and head movements. In order to estimate user interest, the presentation system uses a previously developed algorithm [16], and has the presentation agents respond in an appropriate way. The gaze cascade effect [20,4] is exploited at decision points in the presentation in order to determine with which presentation topic the user would like to continue.

In our future work, we will proceed along two research avenues. First, we plan to extend the interest estimation algorithm to cover relationships between interest objects in order to unveil e.g. a user's interest in comparing visual objects rather than choosing between them. Second, we intend to improve the presentation system by integrating narrative principles. This is important since currently, agent response to user input (visual interest) mostly 'interrupts' the presentation flow, which is thereafter simply resumed following the pre-defined storyline. It would be desirable to utilize user attention as a means to control the presentation in a natural and non-conscious way while preserving the narrative cohesion and persuasiveness of the presentation flow.

Acknowledgements

The research was supported by the Research Grant (FY1999–FY2003) for the Future Program of the Japan Society for the Promotion of Science (JSPS), by a JSPS Encouragement of Young Scientists Grant (FY2005–FY2007), and an NII Joint Research Grant with the Univ. of Tokyo (FY2006). The first author was supported by the JSPS Encouragement Grant. The third author was supported by an International Internship Grant from NII under a Memorandum of Understanding with the Faculty of Applied Informatics at the Univ. of Augsburg. We are also indebted to Dr. Ulrich Apel (NII) for scripting the dialogues.

References

1. Argyle, M., Cook, M.: Gaze and Mutual Gaze. Cambridge University Press, Cambridge (1976)
2. Barakonyi, I., Prendinger, H., Schmalstieg, D., Ishizuka, M.: Cascading hand and eye movement for augmented reality videoconferencing. In: Proceedings 2nd IEEE Symposium on 3D User Interfaces 2007 (3DUI-07), pp. 71–78. IEEE Computer Society Press, Los Alamitos (2007)
3. Bates, J.: The role of emotion in believable agents. Communications of the ACM 37(7), 122–125 (1994)
4. Bee, N., Prendinger, H., Nakasone, A., André, E., Ishizuka, M.: AutoSelect: What You Want Is What You Get. Real-time processing of visual attention and affect. In: Tutorial and Research Workshop on Perception and Interactive Technologies (PIT-06) LNCS (LNAI), vol. 4021, pp. 40–52. Springer, Heidelberg (2006)
5. Duchowski, A.T.: Eye Tracking Methodology: Theory and Practice. Springer, London, UK (2003)
6. Garau, M., Slater, M., Bee, S., Sasse, M.A.: The impact of eye gaze on communication using humanoid avatars. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI-01), pp. 309–316. ACM Press, New York (2001)
7. Hansen, J.P.,Engell.-Nielson, T., Glenstrup, A.J.: Eye-gaze interaction: A new media – not just a fast mouse. In: The Second Swedish Symposium on Multimodal Communication (1998)
8. Heylen, D.: Head gestures, gaze and the principles of conversational structure. International Journal of Humanoid Robotics 3(3), 241–267 (2006)

9. Jacob, R.J.K.: The use of eye movements in human-computer interaction techniques: What You Look At is What You Get. ACM Transactions on Information Systems 9(3) 152–169 (1991)
10. Kendon, A.: Some functions of gaze-direction in social interaction. Acta Psychologica 26, 22–63 (1967)
11. Kipp, M.: Gesture Generation by Immitation – From Human Behavior to Computer Character Animation. PhD thesis, Saarland University, Disseration.com, Boca Raton, Florida (2004)
12. Nischt, M., Prendinger, H., André, E., Ishizuka, M.: Creating three-dimensional animated characters: An experience report and recommendations of good practice. Upgrade. The European Journal for the Informatics Professional VII(2), 36–41 (2006)
13. Nischt, M., Prendinger, H., André, E., Ishizuka, M.: MPML3D: a reactive framework for the Multimodal Presentation Markup Language. In: Proceedings 6th International Conference on Intelligent Virtual Agents (IVA-06), Springer, LNAI,vol. 4133, pp. 218–229 (2006)
14. Prendinger, H., Ishizuka, M.(eds.): Life-Like Characters. Tools, Affective Functions, and Applications. In: Cognitive Technologies, Springer Verlag, Berlin Heidelberg (2004)
15. Prendinger, H., Ma, C., Ishizuka, M.: Eye movements as indices for the utility of life-like interface agents: A pilot study. Interacting with Computers 19(2), 281–292 (2007)
16. Qvarfordt, P., Zhai, S.: Conversing with the user based on eye-gaze patterns. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI-05), pp. 221–230. ACM Press, New York (2005)
17. Rehm, M., André, E.: Where do they look? Gaze behaviors of multiple users interacting with an embodied conversational agent. In: Proceedings of International Conference on Intelligent Agents (IVA-05), LNCS (LNAI), vol. 3661, pp. 241–252. Springer, Heidelberg (2005)
18. Rist, T., André, E., Baldes, S., Gebhard, P., Klesen, M., Kipp, M., Rist, M.,Schmitt, M.: A review of the development of embodied presentation agents and their appication fields. In: Prendinger, Ishizuka, vol. 14, pp. 377–404
19. Selker, T.: Visual attentive interfaces. BT Technology Journal 22(4), 146–150 (2004)
20. Shimojo, S., Simion, C., Shimojo, E., Scheier, C.: Gaze bias both reflects and influences preference. Nature Neuroscience 6(12), 1317–1322 (2003)
21. Starker, I., Bolt, R.A.: A gaze-responsive self-disclosing display. In: Proceedings CHI-90, pp. 3–9. ACM Press, New York (1990)
22. Vertegaal, R., Slagter, R., van der Veer, G., Nijholt, A.: Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. In: Proceedings of CHI-01, pp. 301–308. ACM Press, New York (2001)
23. Zhai, S.: What's in the eyes for attentive input. Communications of the ACM 46(3), 34–39 (2003)