

IGaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions

Michael Kipp and Patrick Gebhard

DFKI

Embodied Agents Research Group

michael.kipp@dfki.de, patrick.gebhard@dfki.de

Abstract. We present IGaze, a semi-immersive human-avatar interaction system. Using head tracking and an illusionistic 3D effect we let users interact with a talking avatar in an application interview scenario. The avatar features reactive gaze behavior that adapts to the user position according to exchangeable gaze strategies. In user studies we showed that two gaze strategies successfully convey the intended impression of dominance/submission and that the 3D effect was positively received. We argue that IGaze is a suitable setup for exploring reactive nonverbal behavior synthesis in human-avatar interactions.

1 Introduction

While embodied agents have a wide range of applications, interactive systems with a face-to-face conversation are of particular interest [1]. However, most current HCI systems are turn-based instead of being *reactive*. In reactive systems user actions should trigger an instantaneous response on the agent side which in turn influences the user, resulting in a tightly coupled feedback loop. Such reactive behavior can only be explored with continuous user input, for instance by visually tracking the user. As theater expert Johnstone observed: "the bodies of the actors continually readjusted. As one changed position so all the others altered their postures." [2]. To simulate and study such effects in human-avatar interactions, *reactive* agents are required in an *immersive* setup. In this paper we discuss a minimalistic approach to creating immersiveness and implementing reactive gaze behavior that instantaneously adapts to the user's current position.

Gaze is a powerful interaction modality with many functions like signaling attention, regulating turn-taking or deictic reference [3]. Gaze also serves as an indicator for mood, personality and status. The latter has been explored by social scientists, semioticists and theater professionals alike [4, 2]. Because of its communicative importance gaze is highly relevant for embodied virtual agents [5, 6], in robotics [7] and human-computer interaction (e.g. COGAIN¹).

STEVE was one the first immersive human-avatar interaction systems [8]. Users were instructed by a 3D-situated virtual tutor who displayed a number of

¹ <http://www.cogain.org>

gaze behaviors, including continuous gaze following and gaze aversion [9]. However, no empirical studies on the impact on personality/status were reported. Moreover, the used VR goggles had the possible risk of *VR sickness*. Heylen et al. [5] investigated gaze behavior of a cartoon-style talking head. Results showed that users found the functionally optimized gaze strategy easiest to use, they found the character more friendly and completed the task in less time. Fukayama et al. [10] showed that varying the gaze pattern in terms of amount, mean duration and target points has a significant impact on impression formation. Bente et al. [11] proposed a system for investigating social gaze and found that prolonged gaze led to better evaluation of the interlocutor, a finding that explains a part of our results. Poggi et al. [6] created a formalism for generating gaze using a meaning-signal mapping. They leave open the question how to *react* to the user’s continually changing position. Lee et al. [12] created the *Eyes alive* system where pupil movement (saccades) was generated using statistical models of real people. Their data-driven approach outperformed random gaze. The system is complementary to ours which neglects saccade movement.

Except for [8], most systems have a 2D view of the agent with the user at a fixed position that was not tracked. Immersive systems like STEVE run the risk of VR sickness. IGaze intends to study and apply tightly coupled interactions between user behavior and agent behavior in a semi-immersive setup. While empirical studies traditionally compare the usage of embodied agents to more traditional interfaces [13, 14], more recent studies try to specifically validate the effects of *particular behaviors* [15, 16]. In this paper, we empirically validate the effect of specific gaze behaviors.

2 The IGaze System

The IGaze system is an immersive human-avatar interaction system for studying reactive nonverbal behavior. Immersiveness is established by two factors: an illusionistic 3D effect makes the user feel like s/he is moving in 3D and a continuous gaze adjustment that makes the avatar follow the user with the head.

The setup consists of a 42” display and an IR camera behind the screen, pointing at the user. The user wears glasses with 2 infrared LEDs attached. For the IR camera we use Nintendo’s Wii remote (1024x768 resolution). In the current modular architecture, the input module computes a hypothetical head position from the detected IR lights, based on J. Lee’s *WiiDesktopVR* software². Head position values are used by the behavior control system (a) to position the virtual camera at user’s location oriented toward the avatar and (b) to orient the avatar’s head (e.g. always looking at the user). The animation controller handles facial viseme animation and the combination of procedural animation (head rotation) with keyframe animation (breathing). We use Horde3D³, developed by N. Schulz, for rendering. The 3D character (Figure 1) has a 40-joints skeleton and 4 viseme morph targets. Speech is synthesized using OpenMARY⁴.

² <http://www.cs.cmu.edu/~johnny/projects/wii>

³ <http://www.horde3d.org>

⁴ <http://mary.dfki.de>

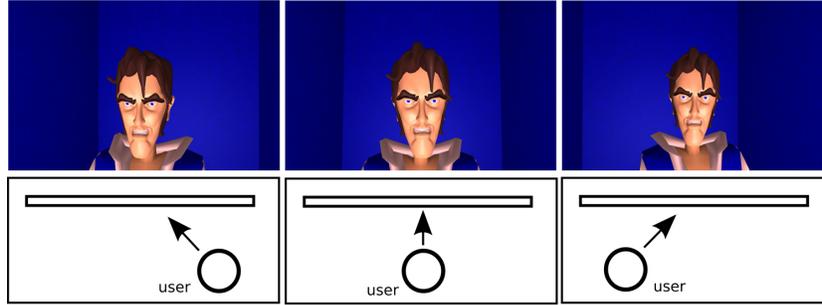


Fig. 1. The illusionistic 3D effect consists of adapting the virtual camera according to the user’s head position (distance, height, sideways position).

The 3D illusion is created by positioning the camera according to the position of the user in front of the screen. The user can “look into” the room by moving his/her head (see Figure 1). To avoid jumps in the camera movement due to misreadings or flare, we smooth incoming values v_{in} using a change factor η to obtain the new value $v_{new} = (1 - \eta) v_{old} + \eta v_{in}$. Especially the z-value (distance), estimated from the distance of the two IR blobs, is quite unstable at distances of $> 1m$; we therefore applied stronger smoothing to the distance z-update ($\eta = .1$) than to the x/y-update ($\eta = .6$).

2.1 Gaze strategies

We implemented 3 gaze strategies: the *Mona Lisa* strategy (= continuous gaze following), dominant and submissive strategy. The gaze aversion behavior was different for dominant and submissive. Strategies were modeled using timed finite state automata depicted in Fig. 2.

The **Mona Lisa strategy (ML+, ML-)** consists of following the user’s position with the eyes all the time. Two variants should check on the impact of the 3D effect: With the 3D effect switched on (ML+) the avatar looks at the position of the virtual camera. When switched off (ML-) it looks at the hypothesized position in the real world (usually *not* the camera). The Mona Lisa gaze is related to Poggi’s *magnetic eyes*, hypothesized to mean dominance [4]. It also fits *stare* (request for attention), *look in the face* and *look straight into someone’s eyes* (expression of dominance, defy), or even *cold anger*.

Dominant strategy (Dom): High status, according to Johnstone, is gained by outstaring your interlocutor [2]. Moreover, he observed that if *A* breaks eye contact and does not look back, *A* is higher. Also, according to the Visual Dominance Ratio (VDR) measure, the higher status person gazes roughly the same amount while listening and while speaking, whereas the lower status person spends more time gazing while listening [17]. Our dominant strategy consists of maintaining eye contact while speaking and randomly changing from gazing to averting while listening. More precisely, the avatar establishes and holds eye contact when speaking, and after speaking, immediately looks away. When listening, the avatar establishes eye contact after 0–3 sec., then holds it for 4.5–7.5

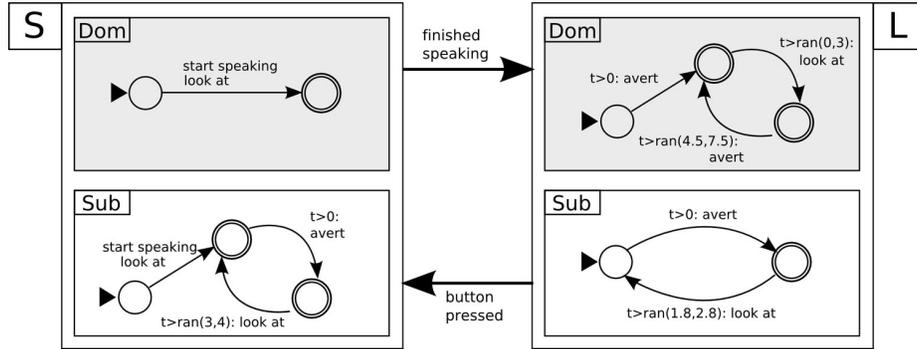


Fig. 2. Timed automata were used to model the gaze behaviors of dominant (upper states) and submissive (lower). S/L refer to speaking/listening modes.

sec. before looking away. The dominant avert behavior consists of a movement 12° *away* from the user and 5° *upward* from the default up-down angle.

Submissive strategy (Sub): Low status, according to Johnstone, means being outstared by your interlocutor. Moreover, if *A* breaks eye contact and looks back, *A* is lower. Our submissive strategy makes the avatar only look briefly every now and then and immediately avert the gaze again. In the submissive strategy, the avatar establishes eye contact when starting to talk but averts his gaze immediately after eye contact. His gaze remains averted for 3–4 sec. He then establishes eye contact again and looks away immediately. During listening, the pattern is the same with the difference that the avatar holds eye contact for 1.8–2.8 sec. The submissive avert behavior consists of a movement *away* from the user (5° while speaking, 8° while listening) and 15° *downward*.

3 Experiment

In our experiment subjects played applicants in a virtual application interview. 14 subjects (aged 21–36, 5 female, 9 male, German native speakers) participated. Subjects interacted with the avatar (interviewer) in a private cabin and wore a headset with microphone to make them believe that speech input is understood. The avatar was remote controlled by the experimenter (wizard of oz) who triggered the utterances. We had the following *hypotheses* regarding the outcome of our experiment: **(H1)** The 3D effect is not uncomfortable, **(H2)** the 3D effect helps people to immerse, **(H3)** dominant gaze behavior is perceived as dominant, **(H4)** submissive gaze behavior is perceived as submissive.

3.1 Pilot Study

In a pilot study we asked the 10 subjects to take the application interview as seriously as possible and to answer truthfully. Many subjects displayed a high degree of stress similar to a real application setting. This had three negative side-effects: (1) the subjects hardly moved, thus not noticing the 3D effect, (2)

they were so focused on their answers that little attention was given to the avatar’s behavior, and (3) the avatar was judged by the content of the interview questions (subjects found the avatar getting ”too personal” or sometimes being ”more relaxed”). When we found no effects in the analysis we modified the design in various ways: (a) we demonstrated the 3D effect prior to the interview, (b) we asked subjects to pay attention to the avatar’s gaze behavior, (c) we told subjects not to take the interview too seriously (e.g. invent answers), (d) we changed the answer scale from 5 points to 7 points because only few subjects had used the extreme points.

3.2 Main Study

Procedure The subjects were told to participate in an experiment about a “virtual application interview training”. They should act as if in an application interview for an academic position. However, they were asked to pay attention to the avatar’s gaze and not take the application answers themselves too seriously. Moreover, we demonstrated the 3D effect before the interview.

<i>condition</i>	<i>gaze behavior</i>	<i>3D effect</i>
ML-	continuous ”Mona Lisa” gaze following	inactive
ML+	continuous ”Mona Lisa” gaze following	active
Dom	dominant gaze behavior	active
Sub	submissive gaze behavior	active

Table 1. The four conditions of our experiment.

During one session the avatar asked 32 interview questions. Each question was 2–3 sentences long to give room for avatar head movement. The subject had to answer each question and after 4 questions the screen was turned blank and the subject filled in a paper/pencil *in-session questionnaire* to rate the past experience. The subject’s saying ”ready” triggered the next 4-question session. Thus, we had 8 4-question sessions. In each session the condition was changed: ML-, ML+, Dom or Sub (Table 3.2). The order of conditions was random and balanced across subjects. They virtual character performed gaze behavior both while speaking and listening. In order for the system to know that an answer was finished, a human operator had to press a button when the subject finished his/her answer. The whole interaction lasted between 15–25 minutes, and was followed by a post-questionnaire.

Each *in-session questionnaire* (paper and pencil) asked for 5 ratings on a 7-point scale⁵. The subject was asked whether s/he found the avatar (1) likable, (2) dominant, (3) extrovert, (4) natural, and (5) how stressed the subject him/herself felt. In the *post-questionnaire* we first asked the subjects to describe (free form) any differences between the session’s segments. We then explained the 3D effect to the subject and asked in 3 questions (7-point scale) whether (1) the 3D effect was uncomfortable, (2) the 3D effect was enjoyable, and (3) whether any differences in gaze behavior were discernible.

⁵ Extreme values were labeled *not at all* and *very much*, middle value was labeled *neutral*.

Results Fig. 3 (left) shows the mean values of our four questions (likable, dominant, extrovert, natural) over the four conditions (ML-, ML+, Dom, Sub). We first checked whether and how the conditions differed with regard to the questions using ANOVA which yielded significant main effects for condition ($F(3,39)=3.70, p < .05$), question ($F(3,39)=3.60, p < .05$) and condition-question interaction ($F(9,117)=2.59, p < .01$), therefore the four conditions had different answer patterns.

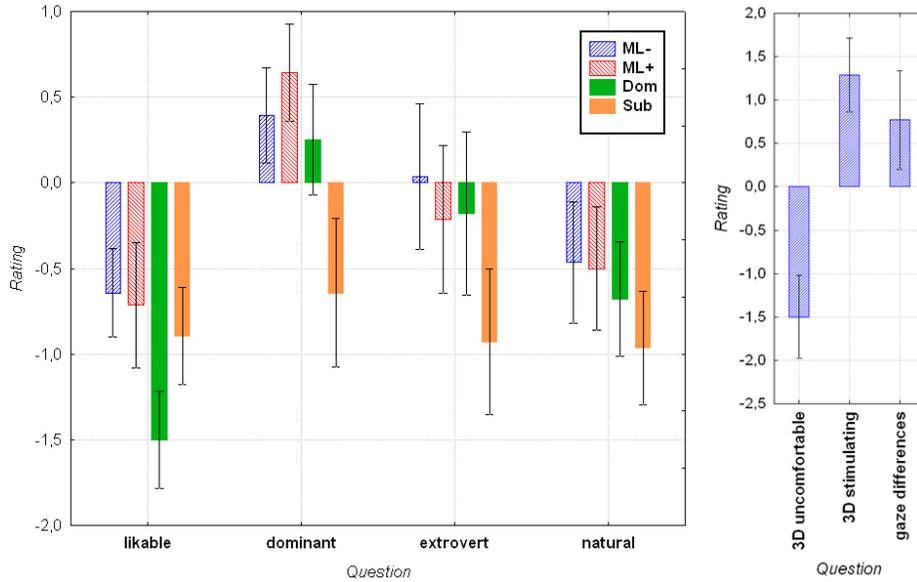


Fig. 3. (a) Left figure shows mean values and standard error of the 4 questions vs. 4 conditions. (b) Right figure shows mean and std. err. for 3 debriefing questions.

We computed ANOVAs for each question to find out whether specific questions differed with respect to condition, using the Fisher LSD test over all condition pairs to single out the exact differences. For question *likable*, we found a main effect ($F(3,39)=3.12, p < .05$) in the ANOVA with significant difference between ML- vs. Dom ($p < .01$) and ML+ vs. Dom ($p < .05$) using Fisher LSD. Question *dominant* had a main effect ($F(3,39)=4.01, p < .05$). Significant differences were ML- vs. Sub ($p < .05$), ML+ vs. Sub ($p < .01$) and Dom vs. Sub ($p < .05$). We found similar results for question *extrovert*: a main effect ($F(3,39)=3.56, p < .05$) and differences between ML- vs. Sub ($p < .01$), ML+ vs. Sub ($p < .05$) and Dom vs. Sub ($p < .05$). To our surprise, the question *natural* yielded no main effect ($F(3,39)=1.22, p < .32$) which means that subjects found all conditions natural to the same degree. The question *stress* did not result in a main effect either ($F(3,39)=1.27, p = .30$).

Fig. 3 (right) shows the means of the 3 post-questionnaire questions asked after the session (7 point scale: -3 to 3). The first two questions showed a significant difference from zero (neutral value). The answers to "do you find the 3D effect uncomfortable?" were significantly below zero, -3 being *not at all* ($t(13)=-3.14$, $p < .01$). The answers to "do you find the 3D effect stimulating?" were significantly above zero ($t(13)=3.03$, $p < .01$).

4 Discussion

The study successfully validated our hypotheses that our encoded dominant/submissive behaviors are perceived as dominant/submissive (**H3,H4**). What is interesting is that conditions ML-, ML+ and dominance are so close to each other. However, the conditions can be divided along the dimension of *liking*. Here, the dominant behavior is significantly perceived less likable than ML. This indicates that instead of implementing purely dominant behavior, we implemented dominant+negative behavior (Dom) and dominant behavior (ML-, ML+). The former can also be called arrogance, a key word that also emerged in debriefing interviews. The latter conforms with findings that continuous gaze leads to more positive evaluation [11]. We found that stress obviously did not impact the judgement of dominance as most subjects did not find the situation stressful.

As for the 3D effect, we wanted to know whether subjects would experience irritation similar to the VR sickness. However, our analysis showed that subjects did not find it uncomfortable (**H1**) but actually found it stimulating (both significant). Many subjects told us afterwards that they liked both the 3D effect and the fact that the agent was actually following them with his gaze (the Mona Lisa effect). However, we were surprised it did not seem to matter whether the 3D effect was switched on or off. So while the effect did generate excitement it did not affect the perception of avatar personality and did not raise stress level or comfort (**not H2**).

5 Conclusion

We presented IGaze, a semi-immersive system for reactive human-avatar interactions. We use head tracking, an illusionistic 3D effect and a life-size display of the avatar's upper body to create immersiveness. Different gaze strategies (dominant/submissive) were implemented using timed automata and successfully validated in a user study.

Many prior systems have neglected the questions that arise when continuous input data from the user is available. To build truly reactive systems, we have to devise tools and systems to model the tightly coupled feedback that is characteristic for human interactions. IGaze takes a minimalistic approach to the setup, employs timed automata for modeling reactive behavior and will be extended in the future with new I/O modules. For output we envisage realtime procedural animation of gesture that adapts to user actions [18]. New input modalities include speech or accelerometer-based input devices.

Acknowledgments. This research has been carried out within the framework of the Excellence Cluster Multimodal Computing and Interaction (MMCI), sponsored by the German Research Foundation (DFG).

References

1. Cassell, J., Sullivan, J., Prevost, S., Churchill, E.: Embodied Conversational Agents. MIT Press, Cambridge, MA (2000)
2. Johnstone, K.: Impro. Improvisation and the Theatre. Routledge/Theatre Arts Books, New York (1979) (Corrected reprint 1981).
3. Heylen, D.: A closer look at gaze. In: Proc. of the workshop "Creating Bonds with Embodied Conversational Agents". (2005)
4. Poggi, I.: Mind, Hands, Face and Body: A Goal and Belief View of Multimodal Communication. Weidler Buchverlag, Berlin (2007)
5. Heylen, D., van Es, I., Nijholt, A., van Dijk, B.: Controlling the gaze of conversational agents. In: CLASS Workshop. (2003)
6. Poggi, I., Pelachaud, C., de Rosi, F.: Eye Communication in a Conversational 3D Synthetic Agent. *AI Communications* **13**(3) (2000) 169–181
7. Mutlu, B., Hodgins, J.K., Forlizzi, J.: A storytelling robot: Modeling and evaluation of human-like gaze behavior. In: Proceedings of HUMANOIDS'06, 2006 IEEE-RAS International Conference on Humanoid Robots, IEEE (December 2006)
8. Rickel, J., Johnson, W.L.: Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence* **13** (1999) 343–382
9. Lee, J., Marsella, S., Traum, D., Gratch, J., Lance, B.: The rickel gaze model: A window on the mind of a virtual human. In: Proc. of the 7th International Conference on Intelligent Virtual Agents (IVA-07). (2007)
10. Fukayama, A., Takehiko, O., Mukawa, N., Sawaki, M., Hagita, N.: Messages Embedded in gaze of Interface Agents - Impression management with agent's gaze. In: Proceedings of SGICHI. (2002) 41–48
11. Bente, G., Eschenburg, F., Aelker, L.: Effects of simulated gaze on social presence, person perception and personality attribution in avatar-mediated communication. In: PRESENCE. (2007)
12. Lee, S., Badler, J., Badler, N.: Eyes alive. In: TOG / Proc. of SIGGRAPH, San Antonio, TX, ACM Press (2002) 637–644
13. Krämer, N.C., Tietz, B., Bente, G.: Effects of embodied interface agents and their gestural activity. In: Proc. of the 4th International Conference on Intelligent Virtual Agents, Springer (2003)
14. Lester, J.C., Converse, S.A., Stone, B.A., Kahler, S.E., Barlow, S.T.: Animated pedagogical agents and problem-solving effectiveness: A large-scale empirical evaluation. In: Proceedings of the Eighth World Conference on Artificial Intelligence in Education, Amsterdam, IOS Press (1997) 23–30
15. Kipp, M., Neff, M., Kipp, K.H., Albrecht, I.: Toward Natural Gesture Synthesis: Evaluating gesture units in a data-driven approach. In: Proc. of the 7th International Conference on Intelligent Virtual Agents (IVA-07), Springer (2007) 15–28
16. Foster, M.E., Oberlander, J.: Corpus-based generation of head and eyebrow motion for an embodied conversational agent. *Journal on Language Resources and Evaluation - Special Issue on Multimodal Corpora* **41**(3-4) (2007) 305–323
17. Dovidio, J.F., Ellyson, S.L.: Decoding visual dominance: Attributions of power based on relative percentages of looking while speaking and looking while listening. *Social Psychology Quarterly* **45**(2) (June 1982) 106–113
18. Neff, M., Kipp, M., Albrecht, I., Seidel, H.P.: Gesture Modeling and Animation Based on a Probabilistic Recreation of Speaker Style. *Transactions on Graphics* (2008) to appear.