The Rickel Gaze Model: A Window on the Mind of a Virtual Human

Jina Lee[1], Stacy Marsella[1], David Traum[2], Jonathan Gratch[2], and Brent Lance[1]

[1] Information Sciences Institute, University of Southern California, USA
{jinal,marsella,brent}@isi.edu
[2] Institute for Creative Technologies, University of Southern California, USA
{traum,gratch}@ict.usc.edu

Abstract. Gaze plays a large number of cognitive, communicative and affective roles in face-to-face human interaction. To build a believable virtual human, it is imperative to construct a gaze model that generates realistic gaze behaviors. However, it is not enough to merely imitate a person's eye movements. The gaze behaviors should reflect the internal states of the virtual human and users should be able to derive them by observing the behaviors. In this paper, we present a gaze model driven by the cognitive operations; the model processes the virtual human's reasoning, dialog management, and goals to generate behaviors that reflect the agent's inner thoughts. It has been implemented in our virtual human system and operates in real-time. The gaze model introduced in this paper was originally designed and developed by Jeff Rickel but has since been extended by the authors.

1 Introduction

Research on gaze has shown that it plays a large number of cognitive, communicative and affective roles in face-to-face human interaction [1] [2] [3]. Gaze is of course central to attentional mechanisms, helping to provide information to and regulate cognitive processes accordingly, but also can be an intentional or unintentional signal to others about these processes. It similarly informs, reflects and conveys underlying emotional processes and attitudes. And as a powerful nonverbal signal, it plays a critical role in regulating dialog and social processes in general. Given its myriad roles, it is not surprising that gaze has been called the window on the soul of a human.

We envision a similar role for gaze in virtual humans, as a window on the “mind” of a virtual human. The gaze model introduced in this paper was originally designed and developed by Jeff Rickel. It has evolved through the work of the authors but clearly within the structure developed by Rickel. The model is driven by a virtual human architecture [4] [5] that interleaves behaviors related to planning and execution of tasks and attention capture. Task-related behaviors (e.g., checking the status of a goal or monitoring for an expected effect or action) trigger a corresponding gaze shift, as does attention capture (e.g., hearing a new sound in the environment). Gaze during social interactions is driven by the dialogue state and the state of the virtual human's own processing, including gaze at an interlocutor who is speaking, gaze aversion during utterance planning (to claim or hold the turn), gaze at an addressee when speaking, and gaze when expecting someone to speak.

C. Pelachaud et al. (Eds.): IVA 2007, LNAI 4722, pp. 296–303, 2007.

The tight integration of gaze behaviors to our underlying cognitive model ensures that the outward attention of the virtual humans is synchronized with their inner thoughts. Thus the gaze behavior reveals virtual human processes as opposed to human cognitive processes and may differ in timing from average human gaze behavior. Nevertheless, those differences in gaze, if properly realized, should ideally help a human adjust or entrain. This is in contrast to approaches that focus more on mimicking the physical properties of human gaze [6] [7] [8]. When our virtual human takes longer to understand speech, the gaze reflects that. Ideally, the virtual human is perceived as slower in that regard than a native (human) speaker and the human user will ideally adjust.

A key aspect of our approach is that it is part of a virtual human with highly-detailed models of socio-cognitive processes. This in turn supports myriad connections between those processes and gaze, allowing the gaze to play a large number of cognitive, communicative and affective roles, as it does in people. This essentially makes gaze a sparse resource and raises the question of how to regulate or prioritize those connections so that contentions for the virtual human's gaze are resolved.

In this paper, we describe our approach for a gaze model for virtual humans. The gaze behavior generated by our model is realized through the SASO research prototype (Stabilization and Support Operations) [9], which grew out of the Mission Rehearsal Environment [10], to teach leadership and negotiation skills under high stress situations. In this system, the trainees interact and negotiate with a life-size virtual human that resides in a virtual environment.

The next section summarizes the different functions of gaze and various gaze models implemented in other virtual human systems. Section three provides the details of our gaze model and its implementation. We end by discussing issues in our model and future directions, including extensions to the model and experiments we are preparing to conduct.

2 Related Work

There has been extensive psychological study of the functions of gaze behaviors. Argyle and Cook [11] provides an overview of the various movements and functions of gaze. The following summarizes a few of those functions.

Gaze is used to exchange social signals. Even when two people are not interacting, if one is being looked at by another, he/she expects something to happen or an interaction to start [11, p.85]. A request for attention may also be signaled through gaze. After making mutual gaze, one may shift gaze to a third object and return to mutual gaze to draw the other person’s attention to the third object silently.

Argyle and Cook also identified a number of important functions and patterns of gaze during conversation [11, p.114-124]. Conversational gaze serves to send social signals, open a channel to receive visual non-verbal messages, and control the synchronization of speech. Gaze aversion occurs at the beginning of utterances, while speaking, when asked a question, and during hesitant pieces of speech. In general, gaze aversion can serve to avoid overload of information and external distraction. Gaze is also used to regulate turn-taking between the speaker and the addressee. As a speaker ends his utterance, he makes a prolonged eye gaze at the listener, at which point the listener makes a gaze aversion and starts speaking.

There have been many implementations of gaze behaviors in virtual agents. One of the first of these was Animated Conversation [12], which implemented a real-time interaction between two virtual agents. The gaze model was based on conversational behaviors such as turn taking. Many gaze implementations in virtual humans are similarly based on communicative signals, such as REA, the Real Estate Agent [13]. Pelachaud et al. [8] use Bayesian belief nets to determine when to gaze in conversation, based on frequency data collected from human interaction. Our approach models both conversational gaze and environmental interaction with an emphasis on revealing the cognitive state of the virtual human.

3 The Gaze Model

The approach to modeling gaze in the Rickel model assumes that gaze is closely tied to the agent's cognitive operations that are at any time vying for processing time. These operations may include perceptions of events, the update of beliefs, understanding speech, planning, and taking actions in the world, of which the selected operation serves to determine both the type of gaze as well as its physical manner.

In this section, we begin with describing the various types of gazes in the model. Then we describe the different cognitive operations and how they determine gaze. Finally, we provide an example to demonstrate the generation of gaze behaviors.

3.1 Different Types of Gaze and Their Properties

Our gaze model produces a wide range of gaze behaviors. Different gazes specified through a set of properties that describe the type, style, speed, as well as the agent's rationale behind the gaze behaviors. Figure 1 summarizes the various gaze properties that can be specified.

- **Gaze-type:** A symbol describing the type of gaze at the target. It can be one of avert, cursory, look, focus, or weak-focus. Focus requires having the body oriented towards the target and may cause stepping whereas weak-focus avoids stepping towards the target.
- **Target:** The name of an object that the agent is gazing at or shifting gaze to, or averting in the case of gaze aversion.
- **Priority:** A symbol describing the priority of the cognitive operation that triggered this gaze command.
- **Speed:** The desired speed of the gaze shift. It can be one of slower, slow, normal, fast, or default.
- **Track:** If the gaze type is glance, look, or focus, this slot specifies whether the object should be continuously tracked, or looked at once but not tracked. If the gaze type is avert, this slot holds a symbol that describes the type of aversion (offset from eyes, down, sideways-down, up, sideways-up).
- **Reason:** A token that represents the rationale behind why we are doing the gaze. This specifies the cognitive operations or the sub-phases of the operations associated with gaze.

Fig. 1. Properties of Gaze Commands

Gaze-type, target, and track define the physical properties of gaze. In addition, reasons specify the underlying rationale for the gaze. Examples of reason are *planning_speech_hold_turn*, *speaking*, *monitor_expected_action*, etc. In the original implementation, reason did not play a functional role; the manner of the gaze was specified in the gaze properties as noted. In a more recent work, we have begun to pass the "reason" for the gaze to the animation system that realizes the body. This provides the animation system with more information to specialize the manner. However, the basic idea behind the implementations remains. The model of gaze should have a large space of gaze types with varying physical manner. With such a model, the inner intricacies of the agent's reasoning can be revealed by different gaze manners.

3.2 Cognitive Operations Associated with Gaze Commands

In our model, we assume that a set of cognitive operations in turn produce a variety of gaze behaviors. Gaze requests are associated with and are made by these operations. For instance, there are different gaze behaviors for the sub-phases of outputting speech depending on whether the agent is about to speak, intends to hold turn, etc. Table 1 provides a partial overview of the mapping between cognitive operations and their impact on gaze. Some processes are not listed for space reasons. In particular, appraisals and coping operations also impact gaze (e.g., there are 13 different types of coping strategies, such as shifting responsibility to other person or resigning from achieving the goal, with various manners of gaze associated).

The cognitive operations can be largely grouped into categories that are based on their functions. There are operations to manage the conversation such as *planning speech*, *listen to speaker*, and *interpret speaker's utterance*. These operations describe the different phases of conversational interaction and may show a pattern of sequence. For example, the agent may listen to someone, interpret the speech, plan and execute speech, and then wait for grounding. There are also cognitive operations that are tied to updating of the agent's beliefs, desires, and intentions, as well as operations associated with perceptual processes such as monitoring for events and attending to sound in the environment.

An important point to note about this model is that there is a large set of distinct cognitive operations. The role of the gaze model is to reflect and convey what cognitive operations the agent is engaged in. Therefore, the properties of the gazes will differ according to the current cognitive operation.

Gaze behaviors in category 3 reflect the gathering of visual information about the world. These include both top-down processes such as monitoring objects/events for changes as well as bottom-up processes such as orienting towards the source of sound [15].

In addition to gaze behaviors associated with different cognitive operations, there is a priority scheme among them to allow one operation to interrupt another. For instance, the agent might be delivering an utterance when there is an explosion. The agent then needs to choose whether to respond to the unexpected event or continue with the current operation. The priorities among operations will resolve the contention and the model will generate gaze behaviors associated with the selected operation.

Table 1. Association between cognitive operations and gaze behaviors

CATEGORY 1: CONVERSATION REGULATION

Cognitive Operation	Behavior	Quality	Reference
Planning speech	Gaze aversion	Slower, offset from eyes	[2]
Start an utterance	Look at hearer	Focus, track	[2]
During speech	Look at hearer	Slow, focus, track	[11, p.99]
Utterance is a rejection or counter-proposal	Gaze aversion (Avoid threat)	Slow, Sideways-down	[11, p.92-99]
Utterance is reluctant acceptance	Gaze aversion	Slow, Sideways-down	[2]
Utterance is about past event	Gaze aversion	Slow, Sideways-up	[2]
Done speaking	Look at hearer	Slow	[2]
Hold turn	Gaze aversion	Slow, offset from eyes	[2]
Listen to speaker	Look at speaker	Weak-focus, track	[11, p.101]
Interpret speaker's utterance	Look at speaker	Weak-focus, track	[11, p.121]
Expect speech from the other	Look at speaker	Weak-focus, track	[11, p.121]
Wait for grounding (acknowledgement or repair)	Look at other	Weak-focus, track	[2]

CATEGORY 2: UPDATE INTERAL COGNITIVE STATE

Cognitive Operation	Behavior	Quality	Reference
Update desire, relevance, intention while planning	Gaze aversion (Cognitive load)	Slower, offset from eyes	[14]
Update belief while planning	Look at object (Gather info.)	Look, track	[1]

CATEGORY 3: MONITOR FOR EVENTS / GOAL STATUS

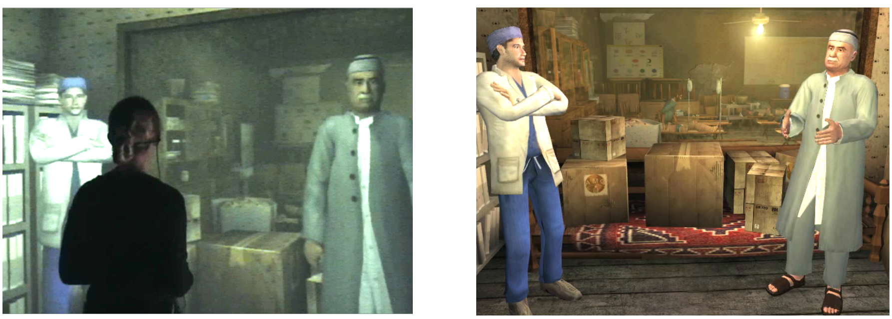
Cognitive Operation	Behavior	Quality	Reference
Check status of precondition object when planning is blocked	Look at object (Gather info.)	Look, track	[15]
Monitor surroundings for unexpected changes	Look at object (Gather info.)	Look, track	[15]
Monitor for expected changes in the environment or actions by others	Look at other (Gather info.)	Look, track	[15]
Attend to a loud or unusual sound	Look at object (Gather info.)	Slow, look, track	[15]

3.3 Example

Let us go through an example from SASO to demonstrate how gaze behaviors are generated. In this scenario, a human user plays the role of a captain whose mission is to move a local clinic to a safer location. To complete the mission, she needs to negotiate with the doctor and the mayor of the city. Note that this is a multi-party setting and the gaze model is able to produce gaze shifts among different participants.

The following contains the transcript of user and virtual humans along with the cognitive operations that drive the gaze behaviors of each agent. Figure 2 shows the screenshots of the behaviors (the doctor is the left character with crossed arms and the mayor is the right character).

(1) CAPTAIN: **Hello, I am captain Kirk.**
DOCTOR: *[planning-speech-look-at-hearer: Look at hearer (captain)]*
(2) DOCTOR: **Hello, captain**.
DOCTOR: *[expect-speech: Look at the speaker (captain)]*
…
(11) MAYOR: **(to captain) We have many things to attend to.**
DOCTOR: *[listen-to-speaker: Look at speaker (mayor)]*
(12) CAPTAIN: **It is imperative that we move the clinic out of this area.**
…
(18) DOCTOR: (to elder) **This conflict is madness. It is killing people.**
DOCTOR: *[monitor-expected-action: Look at agent (captain)]*
(19) CAPTAIN: **Doctor would you be willing to move the clinic downtown? We can give you supplies.**
…



(a) Line 2: Doctor and the mayor are looking at the captain

(b) Line 11 Doctor looks at the mayor while mayor is speaking to the captain

Fig. 2. Gaze behaviors from SASO

4 Conclusion

In this paper, we have described the Rickel gaze model for embodied conversational agents. The basic vision behind the model is that gaze should reflect the inner state of the agent's cognitive processing. The model achieves this by having a large set of distinct gaze behaviors whose physical characteristics reflect or serve cognitive operations. The model has successfully been implemented within our virtual human system and drives the agent to change its gaze as both the situation and its own internal cognitive processing evolve. For example, it exchanges grounding with other agents, monitors for expected events, or attends to unusual sounds.

There are a range of possible improvements we envision for the model. As the model was originally developed, gaze manner was specified within the model and that provided parameters to a procedural animation of gaze. We are currently testing an approach that passes the *reason* parameter to the styles that generates nonverbal behaviors [16] [17] so that the animation system is not tied to specific parameterization but can explore more expressive variations that may also be tied to other aspects of the body's state as well as the capabilities of the animation system.

Cultural and individual variation is another aspect we hope to model. The amount of mutual gaze, duration, and target of gaze are influenced by individual's personality or cultural background. For example, prolonged gaze during face-to-face interaction could be considered as showing interest in one culture while in another, it could be interpreted as being hostile. Rich case-by-case studies highlighting the cultural variation will be required to model the differences.

To evaluate the model, we also plan to conduct a number of experiments with human users. We are particularly interested in the user's responses to the behaviors and what they infer from the behaviors. We are also interested to find out how effective the gaze behaviors are in improving the quality of interaction between virtual humans and human users.

Acknowledgements. This work was sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM), and the content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

1. Argyle, M., Ingham, R., Mccallin, M.: The different functions of gaze. Semiotica, 19–32 (1973)
2. Kendon, A.: Some functions of gaze direction in social interaction. Acta Psychologica 26, 22–63 (1967)
3. Abele, A.: Functions of gaze in social interaction: Communication and monitoring. Journal of Nonverbal Behavior 10, 83–101 (1986)
4. Rickel, J., Johnson, L.L.: Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. Applied Artificial Intelligence 13, 343–382 (1999)
5. Marsella, S., Gratch, J., Rickel, J.: Expressive Behaviors for Virtual Worlds. In: Prendinger, H., Ishizuka, M. (eds.) Life-Like Characters Tools, Affective Functions, and Applications, Springer, Heidelberg (2003)
6. Peters, C., Pelachaud, C., Bevacqua, E., Mancini, M., Poggi, I.: A model of attention and interest using gaze behavior. In: Panayiotopoulos, T., Gratch, J., Aylett, R., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 229–240. Springer, Heidelberg (2005)

7. Fukayama, A., Ohno, T., Mukawa, N., Sawaki, M., Hagita, N.: Messages embedded in gaze of interface agents: Impression management with agent's gaze. In: Proc. of SIGCHI conf. on Human factors in computing systems, pp. 41–48. ACM Press, New York (2002)
8. Pelachaud, C., Bilvi, M.: Modelling gaze behavior for conversational agents. In: Rist, T., Aylett, R., Ballin, D., Rickel, J. (eds.) IVA 2003. LNCS (LNAI), vol. 2792, pp. 93–100. Springer, Heidelberg (2003)
9. Traum, D., Swartout, W., Marsella, S., Gratch, J.: Fight, Flight, or Negotiate: Believable Strategies for Conversing under Crisis. In: Panayiotopoulos, T., Gratch, J., Aylett, R., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, Springer, Heidelberg (2005)
10. Swartout, W., Gratch, J., Hill, R.W., Hovy, E., Marsella, S., Rickel, J., Traum, D.: Toward virtual humans. AI Magazine 27, 96–108 (2006)
11. Argyle, M., Cook, M.: Gaze and Mutual Gaze. Cambridge University Press, Cambridge (1976)
12. Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., Stone, M.: Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. Computer Graphics 28, 413–420 (1994)
13. Bickmore, T., Cassell, J.: Social Dialogue with Embodied Conversational Agents. In: Bernsen, N. (ed.) Natural, Intelligent and Effective Interaction with Multimodal Dialogue Systems, pp. 23–54. Kluwer Academic Publishers, Dordrecht (2004)
14. Doherty-Sneddon, G., Phelps, F.G.: Gaze Aversion: A Response to Cognitive or Social Difficulty? Memory & Cognition 33, 727–733 (2005)
15. Chopra-Kullar, S., Badler, N.: Where To Look? Automating Attending Behaviors of Virtual Human Charactors. In: Proc. of 3rd Annual Conf. on Autonomous Agents, pp. 9–23 (1999)
16. Kallmann, M., Marsella, S.: Hierarchical Motion Controllers for Real-Time Autonomous Virtual Humans. In: Panayiotopoulos, T., Gratch, J., Aylett, R., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, Springer, Heidelberg (2005)
17. Lee, J., Marsella, S.: Nonverbal Behavior Generator for Embodied Conversational Agents. In: Gratch, J., Young, M., Aylett, R., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, Springer, Heidelberg (2006)