

Where To Look? Automating Attending Behaviors of Virtual Human Characters

Sonu Chopra-Khullar and Norman I. Badler
Center for Human Modeling and Simulation
Computer and Information Science Department
University of Pennsylvania
schopra, badler@graphics.cis.upenn.edu
<http://www.cis.upenn.edu/~schopra/home.html>

Abstract

This research proposes a computational framework for generating visual attending behavior in an embodied simulated human agent. Such behaviors directly control eye and head motions, and guide other actions such as locomotion and reach. The implementation of these concepts, referred to as the *AVA*, draws on empirical and qualitative observations known from psychology, human factors and computer vision. Deliberate behaviors, the analogs of scanpaths in visual psychology, compete with involuntary attention capture and lapses into idling or free viewing. Insights provided by implementing this framework are: a defined set of parameters that impact the *observable effects* of attention, a defined vocabulary of looking behaviors for certain motor and cognitive activity, a defined hierarchy of three levels of eye behavior (endogenous, exogenous and idling) and a proposed method of how these types interact.

1 Introduction

This research proposes a computational framework for generating visual attending behavior in an embodied simulated human agent. Such behaviors directly control eye and head motions, and guide other actions such as locomotion and reach. The implementation of these concepts, referred to as the *AVA*, draws on empirical and qualitative observations known from psychology, human factors and computer vision. Deliberate behaviors, the analogs of scanpaths in visual psychology, compete with involuntary attention capture and lapses into idling or free viewing.

Given a high level script that an agent should follow, how do we animate details of the script with the appropriate behavior? The mapping between motor tasks

and the corresponding motion is clear, but attending behavior is often not specified and is *emergent* (where an agent looks changes due to interactions between simultaneous tasks and in response to the dynamics of the environment). Further, motor actions may be modified by input from the attentional system (e.g., if an agent notices an object bearing down him, he will step out of the way).

Some potential applications of this research are:

- Realistic avatars and participants in cyber-chat communities. When an avatar walks to a goal, or looks for someone in the community, his behavior should reflect actual eye behaviors (corresponding to locomotion, visual search and response to peripheral events).
- Virtual reality immersive games. Human players anticipate that animated players move and behave appropriately to the circumstances of the game. Since game environments are typically changing, characters' responses cannot be scripted in advance.
- Determining the ergonomics of computer simulated environments. This research associates standard frequencies of eye movements for primitive cognitive and motor tasks. Frequencies are adjusted in the implementation reflecting degradation in performance due to increasing cognitive load or interference from exogenous factors in the environment. Also, *relative speed* of eye movements is encoded and adjusted based on interference from exogenous effects. Our model of eye behavior could be used to determine when critical events remain unattended.

2 Psychologically Plausible Design

The *AVA* associates a set of primitive motor activities (walk, reach, lift, manipulate, ...) and cognitive actions (monitor, visually search, visually track...) with predefined *patterns* of looking behavior. Monitoring activities

are additionally associated with *memory uncertainty* thresholds. Patterns are estimated in this system based on empirical and qualitative data from related experiments in human factors as well as simple observation. In the *AVA*, looking behaviors implement patterns of eye movements and compete in a psychologically motivated framework. In a multi-task situation or in the presence of exogenous distractors, performance degrades (performance is measured by speed of eye movements to task targets). Interspersed with deliberate looking patterns are lapses into idling.

Input to the *AVA* may be a script generated from a task planner or a loose outline of activity (e.g., While riding a bus, the agent should watch for his stop as it nears. He should also attend or react to other passengers nearby).

2.1 Relevant Psychology Literature- Inputs to Our Method

The purpose of the *AVA* is to generate looking behavior in a *psychologically plausible* framework. A character’s attention is directed by volitional, goal-directed aims known as endogenous factors that correspond to the current task(s) being performed. Involuntary attentional capture by irrelevant stimuli such as peripheral motion or local feature contrast are said to be exogenous factors [25].

The demands of a particular task generate a characteristic *pattern* of eye movements. Depending on an observer’s intentions or goals, eye fixations will vary even when directed at the same image. In [27], observers were shown a picture and asked to estimate the ages of figures in the picture. Patterns of fixations were directed at the face of each figure. When asked to estimate the “material circumstances” of participants, fixations were directed at the clothes of each figure. Accordingly, in the *AVA* we associate patterns of eye behavior for broad categories of motor and cognitive activity.

The transitioning between simultaneous tasks is characterized in [3] as “shifting intentional set.” When engaged in more than one task that requires the same sensory modality, performance degrades versus the single task condition (a review of divided attention experiments is found in [9]). We account for this phenomenon in our method by increasing response time to task targets as the number of events vying for an agent’s attention increase.

Attention may be directed *covertly* without explicit shifts of gaze or overtly. The *AVA* seeks to characterize the *observable* effects of attention shifts relevant to character animation. Hence, covert shifts are relevant in so much as they *interfere* with or increase response time to targets [11] in unattended locations.

When attention is not engaged, eye saccades to targets are within the order of 100ms and are known as

express saccades [7]. When a character is attending to a task, however, eye saccade time between relevant sites will increase to 200ms [7]. Voluntary engagement of attention acts as a “hold mechanism” [2] and suppresses express saccades to irrelevant stimuli. The tendency to orient gaze toward irrelevant distractors is found in patients with frontal-parietal brain lesions [14] (reflecting impairment of oculomotor control) and in early infancy [10] (reflecting the underdevelopment of selective attention). This range of behavior is characterized in our method by a *distractability* parameter that allows a probabilistic sampling of irrelevant stimuli.

What sorts of exogenous factors capture attention and with what frequency? A review of the literature suggests that peripheral events [11] and *abrupt onsets*, the introduction of new perceptual objects into a scene, capture attention [25] when attention is in a diffuse or divided mode (i.e. the target may appear anywhere). However, when attention is fully engaged in a particular location, capture by onset does not occur [26].

In the absence of any given task, attention follows patterns of spontaneous looking [12] where areas of high local feature contrast capture interest. Figure 1 shows rays intersecting those locations in an agent’s field of view that are the most locally conspicuous.

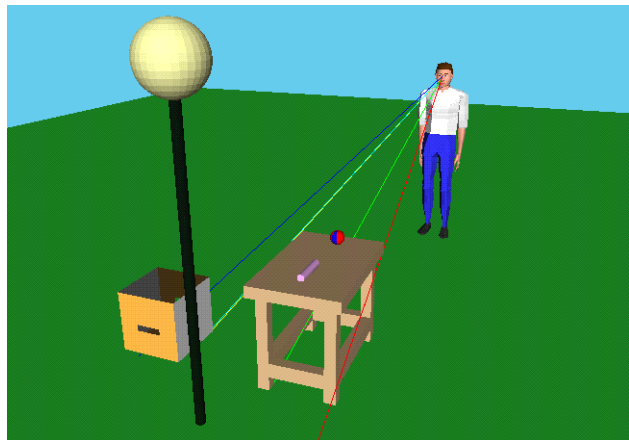


Figure 1: Spontaneous Looking - Rays Intersect Features with Local Contrast

In summary, tasks impose a voluntary pattern of eye movements. As several tasks are simultaneously attempted, performance (in terms of response time to task targets) degrades. Peripheral events capture attention when the agent is engaged in a task which requires diffuse attentiveness (e.g. visual search or divided attention). In the absence of tasks or peripheral stimuli, attention follows patterns of spontaneous looking.

3 System Architecture

Our implementation assigns eye behaviors to broad types of motor and cognitive activities: monitoring and locomotion, reach and grasp, visual search and visual tracking. Behaviors generate a characteristic pattern and frequency of eye movements. Actions are entered by the user of our system as tasks on a queue. A task queue manager process coordinates requested motor and cognitive activities and spawns the appropriate attentional behavior (as well as animating the underlying motion) for an action. Behaviors are implemented in our technique as parallel, executing finite state machines [23].

An arbitrating process (called a *Gazenet*) determines where an agent looks by selecting from three levels of behavior: deliberate, exogenous and idling. Two queues are maintained: an **IntentionList** that stores sites or objects that need to be attended due to the demands of current activities and a **Plist** that indicates objects in agent’s peripheral field of view that are moving. When both queues are empty, a spontaneous looking or idling behavior is activated.

Figure 2 illustrates the *AVA*’s architecture. Users enter task requests as text input. The task queue manager for each agent consumes such requests and generates the appropriate eye gaze or looking behaviors for an action (some activities such as walking and monitoring may be requested in parallel). The motions which correspond to motor tasks are also generated. When the memory uncertainty threshold for an activity is reached, the corresponding eye behavior adds relevant sites to an **Intentionlist** (e.g. The locomotion eye behavior will add the goal destination or ground at particular intervals indicating that those locations should be attended). A peripheral motion sensor behavior is active for each agent and updates the **Plist** as needed.

Behaviors of the same *type* compete equally for an agent’s attention. Task related eye behaviors have the highest precedence. As the number of concurrent task eye behaviors increase, response time to targets increases. A probability factor is used to determine overt orienting toward peripheral stimuli. If the agent is engaged in visual search or in a series of parallel tasks (requiring divided attention), the presence and number of peripheral events will increase response time to task-related targets. Spontaneous looking has the lowest precedence and can be interrupted by any other type of behavior.

3.1 Monitoring and Locomotion

Monitoring tasks (locomotion and visual tracking being a general case) use uncertainty thresholds [16] that relate how often a signal, event, or goal should be glanced at in order to maintain an accurate view of the signal’s state in memory. When the uncertainty threshold for

a given monitoring task is reached in our system, the relevant site is added to **IntentionList**.

While walking, for example, an agent in our system looks toward the horizon or destination and occasionally glances at the ground [21]. This is an example of a monitoring task with high uncertainty thresholds. If the state of the terrain changes, becoming slippery or uneven, for example, the uncertainty threshold associated with the ground plane is reduced, causing the agent to glance more frequently at the ground in front of his feet.

3.1.1 Limit Monitoring

Monitoring may also be associated with limit conditions [16]. As a signal’s state approaches a critical or cautionary level, it will occasion more frequent eye fixations. For example, when crossing the road, an agent will more frequently glance at the light or crossing signal if it is yellow rather than green.

3.2 Reaching and Grasp

Traditional experiments indicate that eye movements precede hand movements and since eye saccades are extremely fast [1], the eye arrives before the hand motion is started.

When initiating a reach and grasp motion, we generate eye movement toward the relevant grasp site by adding it to **IntentionList**. If an agent is picking up a cup, we look at the cup handle. If the agent is lifting a box, we generate a sequence of eye motions to the box grips [4]. Clearly, the eye establishes targeting for the hand [1].

3.3 Visual Search

We model visual search by first determining the angle between the center of fixation and the target. We generate a sequence of intermediate positions that move the eye from its current position to the target location. Ray casting is used to determine target visibility. If the target is not present in the environment or occluded by another object, a sweep of the field of view is performed. Each position to be searched is placed, in order, on **IntentionList**. This eye behavior corresponds to experiments and a computational model proposed in [17]. Visual processing proceeds in a low to high accuracy manner. When asked to locate a specific object in a scene, subjects in [17] performed a *series* of (progressively more accurate) eye saccades toward the object rather than immediately fixating it.

3.4 Motion in the Periphery

A peripheral motion sensor determines (using geometric reasoning) those objects in an agent’s periphery that are

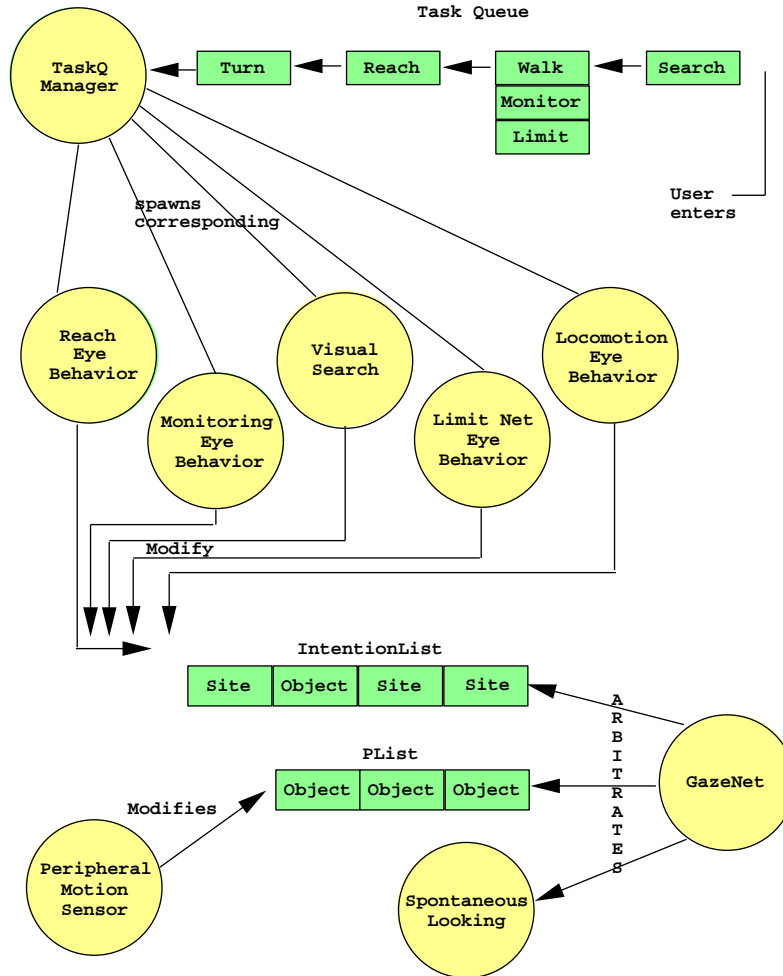


Figure 2: Method Architecture

moving. Such objects are added to **Plist**. All moving objects will not necessarily be attended (when the agent does look at such an object, the behavior embodied is attentional capture by exogenous, peripheral motion). Appearance changes, such as flashing, are not sensed as motion. Such changes are a form of abrupt onsets (see section 2.1). With a minimal computational overhead, the motion sensor behavior in the AVA can check for appearance changes (by querying the display status of objects in the graphics database).

3.5 Spontaneous Looking

A spontaneous looking, or free viewing eye behavior, is activated in those instants when there are no deliberate or exogenous events vying for attention. Attention is drawn to items in the environment that are likely to be informative or significant. Psychologists argue this is due to a need to reduce uncertainty about our surroundings [12].

Novel or complex items are considered significant.

Novelty may be measured by motion, color, isolation, or complexity of shape. Image processing approaches in [24, 13] look for areas in the field of view that are locally conspicuous. Luminance is considered salient in [24] while color and orientation of edges are the measure of conspicuousness in [13].

Since we wish to generate real-time eye behavior, we use a simplified novelty measure. The system copies a snapshot of the agent’s field of view into a pixel buffer. We select those pixels whose color values are the furthest from their neighbors in RGB space. We convert the location of these pixels back into 3D world by inverting and applying the graphics pipeline rendering transforms.

3.6 Interleaving and Confidence Levels

The interleaving of an agent’s attention will happen as a natural consequence of competing behaviors in our system. In contrast, given a set of sequential motor activities, our system must determine when to abandon

the current eye behavior and initiate a subsequent one. A boolean variable is maintained in each net that implements eye behavior based on a reach or locomotion. This variable indicates an expectation that the current activity will complete successfully. Normally, such a variable is set when the hand is in close proximity to the relevant grasp site or the agent is close to his destination. If an agent is confident or expert, however, this variable will be set earlier in the execution of the reach motion reflecting greater confidence in the agent's skill. Setting this boolean variable thus allows attention to be directed to the next activity while the motor system completes the motor task. Notice that if this variable is set at the beginning of the task, the interpretation is consistent with human behavior: it means the agent knows where to reach or walk even without looking at the object or goal.

4 Example Simulation

Consider a scenario where an agent is asked to walk to a destination: in order to reach the destination, he must cross a road, watch out for oncoming traffic and monitor the appropriate traffic signal. We animate such a scenario by entering those three task requests into our system.

A task queue manager net for our agent, a *Jack* virtual human model, will consume these actions requests. A walking eye behavior net will be spawned that periodically adds relevant sites to **IntentionList**: the destination (a table on the other side of the road) and, infrequently, the ground in front of *Jack's* feet. Also, the walking motor activity will be spawned (the corresponding eye behavior will remain active as long as the motor activity is not complete). Figures 3(a)-(c) show several snapshots from the animation where our agent looks out for and responds to (by visually tracking) oncoming traffic (a line is rendered indicating viewpoint or line of sight).

A monitoring eye behavior will be spawned that periodically adds the traffic light as a figure to be monitored on **IntentionList**. If the light turns yellow, the frequency with which the monitoring behavior adds the traffic light to **IntentionList** will increase. The monitoring behavior will only remain active while the agent is crossing the street.

A monitoring eye behavior will also be spawned to check for oncoming traffic on the right side of the road. This behavior will also remain active until *Jack* crosses the street.

Behaviors modeling exogenous factors (involuntary attention capture by task unrelated events) will be a peripheral motion sensor and spontaneous looking. Figure 4(a) shows *Jack* glancing at his destination. In figure 4(b) *Jack* lapses into idling and notices the edge

of the box figure (the most locally conspicuous region). In figure 4(c) *Jack* tracks a ball that flies into his field of view (other task demands from deliberate activity are not vying for attention in that instant).

5 Related Work

Work in the areas of robotics, computer vision, intelligent tutors and facial animation provide some complementary efforts in researching visual attention.

M.I.T.'s Cog Project [5, 19, 15] is developing a humanoid robot which learns or acquires skills during its interactions with its environment. Determining the focus of attention aids in reducing complexity of processing (attention acts as a filter that selects which regions of interest to process in camera images).

Terzopoulos's artificial fish project [22] implements a vision module that determines the identity and location of nearby fish (by querying the graphics database). Feedback from this sensor is used to manage schooling and avoidance behaviors.

Rickel and Johnson's virtual intelligent tutor [18], Steve, has a perception module which is used to monitor changes in the virtual world. The perception mechanism is used to monitor events in an actual student's field of view and can feedback changes to the tutor's planning system.

Our method differs from the preceding projects since we are concerned with predicting human looking behavior due to the demands of simultaneously executing tasks and exogenous effects. Similar to the intent of the preceding projects, our method also allows for feedback from the attention system to our agent's motor capabilities.

Image processing techniques have been developed that attempt to model where humans look in the absence of deliberate activity. Our method incorporates a *much simplified* version of such approaches [24, 13] for spontaneous looking behavior in real-time.

Research in animation has explored issues of eye engagement during social interactions or discourse between virtual agents [6]. Similarly, visual cues of attention between a robot and a human instructor are explored in [19]. The AVA may be used to extend systems that deal with issues of facial animation and social interaction of virtual agents.

6 Conclusion

Believable virtual actors need to exhibit the appropriate attending behaviors in order to be suitably convincing and human-like. Gaze is a significant and often subtle indication of intent and cognitive process. Automating the generation of looking behaviors is an important endeavor since such behaviors are *emergent* and often

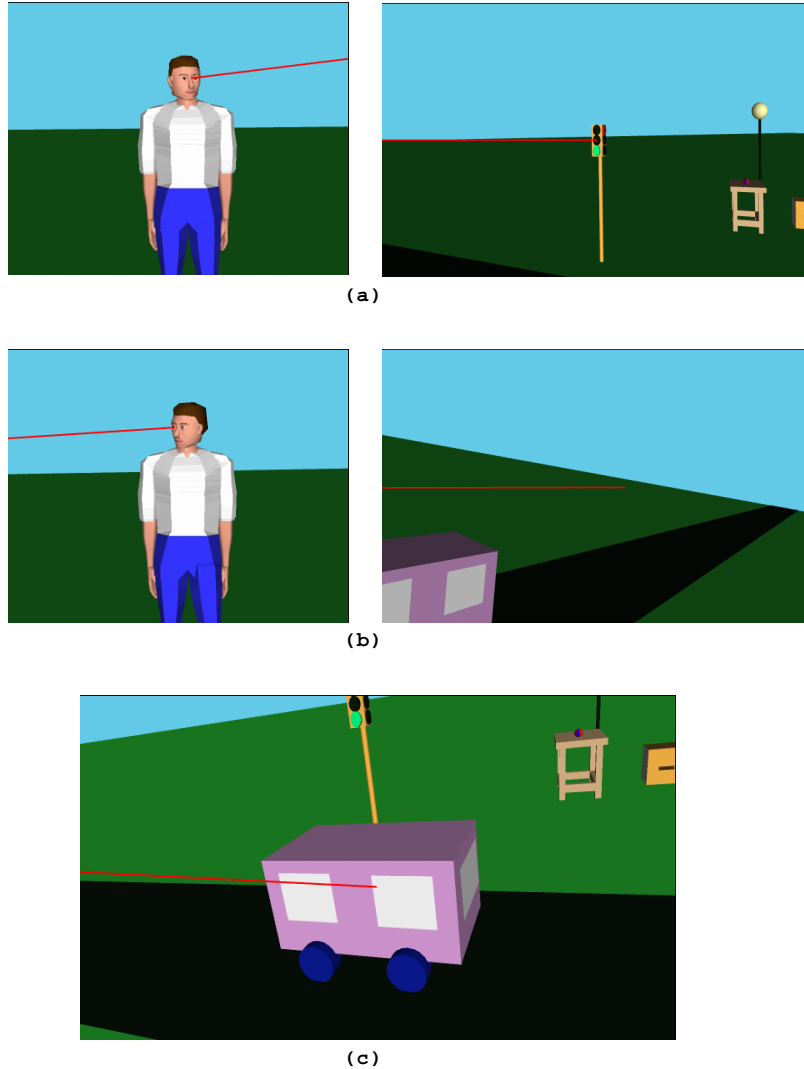


Figure 3: Jack monitors light and avoids traffic

cannot be predicted by a manual animation process. Further, synthetic actors in dynamic virtual environments must respond to changing circumstances and exogenous events. Scripted behavior is inadequate in such scenarios.

The contribution of the *AVA* method is a unified, psychologically-motivated framework that generates a character’s visual attention at interactive rates for a given set of primitives. Deliberate behaviors, the analogs of scanpaths in the psychology literature [27, 20] compete with involuntary attention capture [25, 11, 8] and lapses into idling or free viewing [12, 20]. When information about a task is known, the scene graph is queried for efficiency purposes. When an agent lapses into free viewing or idling, no task constraints are active so a simplified image processing technique is employed. Monitoring tasks (such as locomotion, tracking) are as-

sociated with memory uncertainty thresholds (a concept coined in the study of the ergonomics of avionics cockpits). Uncertainty thresholds allow the interleaving of simultaneously executing tasks and idling (e.g., although a task such as locomotion is ongoing, attending to task sites continuously is not required).

Motor activity itself may adapt in our system due to feedback from the attention system. For example, when the queue lengths of objects requiring deliberate or peripheral attention exceed a combined threshold, we will slow all currently active motor tasks (e.g., locomotion or reach). Further, when moving objects appear to be on collision course with the agent, he will modify his locomotion accordingly (by stopping, speeding up or altering course).

Attention is a process that utilizes a allocatable, steerable resource (the eyes) and as such requires a con-

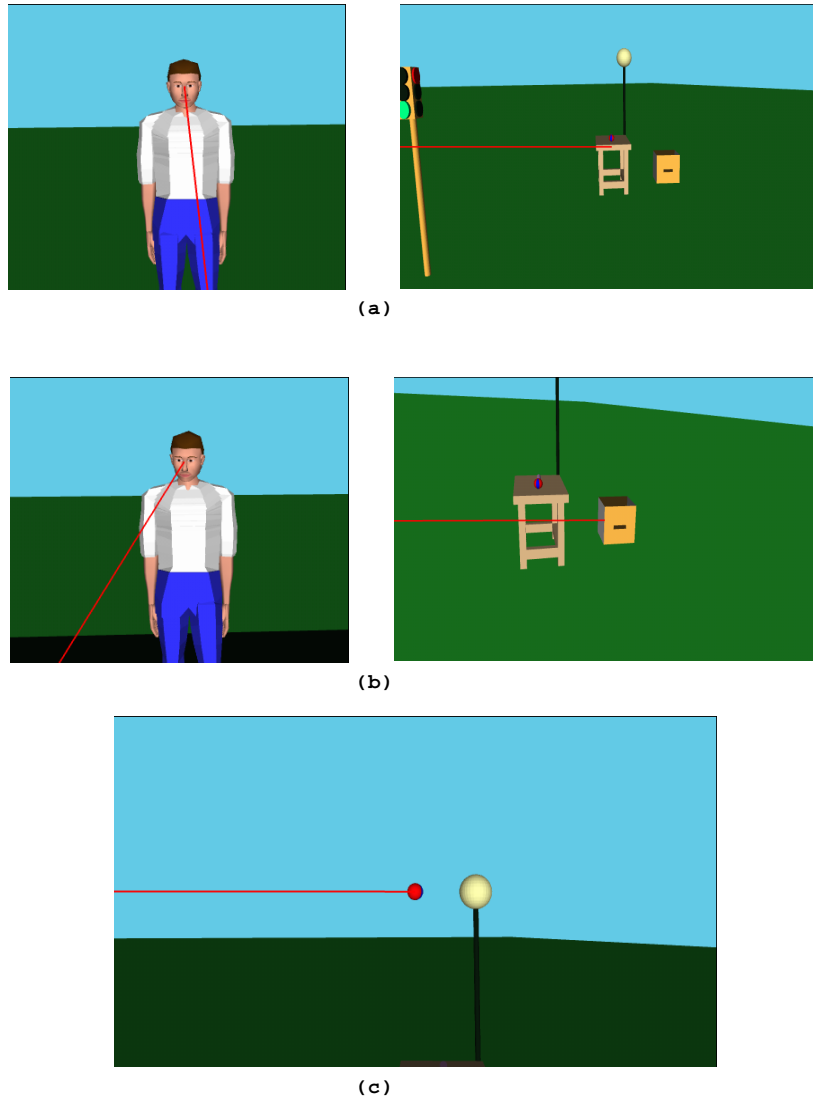


Figure 4: Jack glances at destination and responds to exogenous events

trol algorithm and a time budget for movement and sensing. Competing behaviors require prioritization and arbitration. Visual perception is a significant component of the human behavior repertoire. Through our methodology we have shown that automatic attention control is both feasible and useful for animated human-like characters.

7 Acknowledgments

This research is partially supported by U.S. Air Force through Delivery Order #8 on F41624-97-D-5002; Office of Naval Research (through Univ. of Houston) K-5-55043/3916-1552793; DARPA SB-MDA-97-2951001; NSF IRI95-04372; NASA NRA NAG 5-3990; and Just-System Japan.

References

- [1] R.A. Abrams, D.E.M. Meyer, and S. Kornblum. Eye-hand coordination: Oculomotor control in rapid aimed limb movements. *Journal of Experimental Psychology: Human Perception and Performance*, 16:248–267, 1990.
- [2] A. Allport. Attention and control: Have we been asking the wrong questions? a critical review of 25 years. *Attention and Performance*, 14:183–218, 1993.
- [3] A. Allport, E. Styles, and S. Hsieh. Shifting intentional set: Exploring the dynamic control of tasks. *Attention and Performance*, 15:421–452, 1994.
- [4] D. Ballard, M. Hayhoe, F. Li, and S. White-

- head. Hand-eye coordination during complex tasks. *Investigative Ophthalmology and Visual Science*, 33(4):1355, March 1992.
- [5] R. Brooks, C. Breazeal, R. Irie, C. Kemp, M. Marjanovic, B. Scassellati, and M. Williamson. Alternative essences of intelligence. In *AAAI98*, 1998.
- [6] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *ACM SIGGRAPH Annual Conference Series*, pages 413–420, July 1994.
- [7] B. Fisher. The role of attention in visually guided eye movements in monkey and man. *Psychological Research*, 48:251–257, 1986.
- [8] A. Hillstrom and S. Yantis. Visual motion and attentional capture. *Perception and Psychophysics*, 55(4):399–411, 1994.
- [9] W. Hirst. The psychology of attention. In *Mind and Brain: Dialogues in Cognitive Neuroscience*, pages 105–141, 1986.
- [10] M. Johnson. Visual attention and the control of eye movements in early infancy. *Attention and Performance*, 15:291–310, 1994.
- [11] J. Jonides. Voluntary versus automatic control over the mind’s eye movement. *Attention and Performance*, 9:187–203, 1981.
- [12] D. Kahneman. *Attention and Effort*. Prentice-Hall, 1973.
- [13] C. Koch and S. Ullman. Shifts in selective visual attention: Toward the underlying neural circuitry. *Human Neurobiology*, 4:219–227, 1985.
- [14] E. Ladavas, G. Zeloni, G. Zaccara, and P. Gangeni. Eye movements and orienting of attention in patients with visual neglect. *Journal of Cognitive Neuroscience*, 9(1):67–75, 1997.
- [15] M. Marjanovic, B. Scassellati, and M. Williamson. Self-taught visually-guided pointing for a humanoid robot. In *Fourth International Conference on Simulation of Adaptive Behavior*, Cape Cod, Massachusetts, 1996.
- [16] N. Moray. Designing for attention. In *Attention, Selection, Awareness, and Control: A Tribute to Donald Broadbent*, pages 53–72. Clarendon Press, Oxford, 1993.
- [17] R. Rao, G. Zelinsky, M. Hayhoe, and D. Ballard. Modeling saccadic targeting in visual search. In D. Touretzky, M. Mozer, and M. Hasselmo, editors, *Advances in Neural Information Processing Systems*. MIT Press, 1996.
- [18] J. Rickel and W. L. Johnson. Integrating pedagogical capabilities in a virtual environment agent. In *Proceedings Autonomous Agents 1997*, 1997.
- [19] B. Scassellati. Mechanisms of shared attention for a humanoid robot. In *AAAI Fall Symposium on Embodied Cognition and Action*, 1996.
- [20] L. Stark and Y. Choi. Experimental metaphysics: the scanpath as an epistemological mechanism. In W.H. Zangemeister, H.S. Stiehl, and C. Freksa, editors, *Advances in Psychology: Visual Attention and Cognition*, chapter 1. North-Holland, 1996.
- [21] M. Swain and M. Stricker. Promising directions in active vision. *International Journal of Computer Vision*, 11:109–126, 1993.
- [22] D. Terzopoulos, X. Tu, and R. Grzeszczuk. Artificial fishes: Autonomous locomotion, perception, behavior and learning in a simulated physical world. *Artificial Life*, 1(4):327–351, 1994.
- [23] T. Trias, S. Chopra, B. Reich, M. Moore, N. Badler, B. Webber, and C. Geib. Decision networks for integrating the behaviors of virtual agents and avatars. In *Proc. IEEE Virtual Reality Annual International Symposium*, pages 156–162, 1996.
- [24] J.K. Tsotsos, S.M. Culhane, W.Y.K. Wai, Y. Lai, and F. Nufflo. Modeling visual attention via selective tuning. *Artificial Intelligence*, 78:507–545, 1995.
- [25] S. Yantis. Stimulus-driven attentional capture and attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 19(3):676–681, 1993.
- [26] S. Yantis and J. Jonides. Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*, 16(1):121–134, 1990.
- [27] A. L. Yarbus. *Eye Movements and Vision*. Plenum Press, 1967.