Proc. of 8th Conference on Computer Generated Forces and Behavioral Representation, Orlando, FL, May 1999

Modeling Perceptual Attention in Virtual Humans

Randall W. Hill, Jr.
University of Southern California
Information Sciences Institute
4676 Admiralty Way, Suite 1001
Marina del Rey, CA 90292-6695
310-822-1511 x783
hill@isi.edu

Keywords:
Perception, Attention

ABSTRACT: *This paper describes our efforts to model perceptual attention in virtual humans for the joint synthetic battlefield. With the exception of the work by Reece et al. [19,20] to develop individual combatants, current computer generated forces represent entities at the platform level, hence, the perceptual model of these entities is typically a composite representing the behaviors of multiple humans (e.g., a tank crew). With the growing interest in developing realistic models of human behavior for military simulations [16], a greater emphasis is placed on understanding and representing the behaviors of individual crew members. A composite model of perception and cognition may not be sufficient to provide the desired level of fidelity in a simulation. To be realistic, a model needs to be both psychologically plausible and capable of generating believable behavior. In this paper we describe our efforts to support such capabilities through the development of a flexible, integrated model of perceptual attention in a synthetic helicopter pilot. The model has been developed in a synthetic helicopter pilot [8,9] implemented in Soar, an integrated cognitive architecture.*

1. Introduction

Perception does not stand alone—it is tightly integrated with cognition and motor behavior. Yet, cognition is often treated as an ancillary component of perception and vice versa, depending on the interests of the researcher. With a growing interest in realistic models of human behavior for simulation environments [16], a greater emphasis must be placed on developing holistic models of perception, cognition, and motor behavior.

Simulation environments offer the ability to develop virtual humans—realistic models of human behavior embodied in a synthetic agent or character. Such agents can be used for training [21], games [26], animated characters [3], entertainment, military simulations for mission rehearsal and tactics evaluation, and are potentially useful for testing computational models of psychological theories. While the uses for virtual humans vary, the need for realism and believability motivates the development of models of perception, cognition, and motor behavior that are both psychologically plausible and functionally executable in a simulated world.

The primary purpose of this paper is to present a model of perceptual attention that has been integrated with the cognitive and motor systems of a virtual human, specifically, a helicopter pilot in a military simulation. Our virtual pilot performs tasks such as flying in team formation while simultaneously following a route plan and searching for enemy vehicles. These tasks require visual behaviors such as search, fixation, tracking, and grouping. Such behaviors are not purely perceptual—they involve coordinating the virtual pilot's perception, cognition, and motor systems. Without coordination among these three components, the virtual pilot cannot function robustly—it is susceptible to perceptual or cognitive overloads under certain conditions. To create realistic, functional virtual pilots, we have taken a number of concepts from what is known about how humans coordinate perception and cognition and applied them in our model. In humans, attention provides a way of limiting perceptual processing by focusing on specific visual objects or regions, and it provides a way for cognition to control perception by shifting the focus from one object or region to another. Furthermore, perceptual attention can be captured involuntarily, giving perception a way of alerting cognition to significant events in the environment.

A secondary purpose of this paper is to demonstrate that simulation environments provide a way to investigate the nexus of perception, cognition, and motor skills without having to solve more basic vision problems first. Vision is greatly simplified in simulation environments. For example, objects in the visual field are automatically resolved and recognized without the need for computer vision algorithms. This enables us to focus on issues relating to the integration of perception, cognition and motor behavior in an effort to create a virtual human with realistic behaviors.

Finally, we attempt throughout this paper to make a distinction between the aspects of our model that are psychologically realistic versus ones that are purely functional in nature, modeling human capabilities at a more abstract level. We begin by describing the psychological motivation that forms the basis of our model.

2. Psychological Motivation

2.1 Why is attention needed?

The answer seems obvious—there is simply too much information in the visual field for the human perceptual system to process. As Wolfe [29,30] points out, there are two ways of dealing with this problem. The first way is to ignore excess information. By its nature, the human eye discards data. Acuity is limited over much of the visual field. The arc of greatest acuity is located in the fovea and is only 1-2 degrees, consequently much of the visual input is simply not sensed. The second way of dealing with too much data is to be selective in processing the information that is sensed—attention does this by focusing the processing capabilities on a specific region. Posner [17] describes attention as the system for controlling the way information is routed and for controlling processing priorities.

2.2 Stages of attention

Visual information is processed in preattentive and attentive stages [10,14,29,30]. Preattentive processing is thought to operate in parallel across the entire visual field; it is the stage where segmentation takes place and textures are formed. The products of preattentive processing are the units or objects toward which attention is subsequently directed. Some researchers hypothesize that Gestalt grouping also takes place during the preattentive stage, although these patterns may not be encoded in memory without attention [13]. Grouping is performed based on proximity, similarity, and motion [6,11].

2.3 How is attention oriented?

The metaphor of a spotlight is frequently used to describe how attention is limited to a particular region (or object) and how it can be moved from place to place. The spotlight should not be equated with the eyes, however. Posner [17] identifies two forms of orienting, covert and overt. Although the fovea provides the greatest degree of acuity, attention is not tied to the fovea and can be devoted to other parts of the visual field. This type of orienting is covert—it moves the spotlight of attention without moving the eyes. The model of attention we use in this paper is covert. Conversely, overt orienting involves moving the eyes, and sometimes the head, to shift attention to a new region or object.

2.4 How is attention controlled?

This question differs from the last in that it deals with what prompts the spotlight of attention to move. Attention can be controlled top-down or bottom-up [4,14,17,29,30]. Top-down control is endogenous—it is directed by a decision of the cognitive system and performed in service of a task. This type of control is used to support visual behaviors such as search and tracking. Bottom-up control is exogenous—in this case, attention is captured by an external stimulus. Stimuli that have been shown to capture attention include luminance, color, motion, and abrupt onset [4].

2.5 Cognition and perception

Attention forms the nexus of perception and cognition. The cognitive system needs what the perceptual system offers—visual objects with features such as location, color, orientation, and motion. Without these percepts, the cognitive system would not know about the external world. At the same time, the human perceptual system requires the focus and control provided by attention to allocate limited processing resources to a selected region of the visual field. Without a focus, the perceptual system can be overloaded, with unpredictable consequences for the cognitive system. The cognitive system steers attention, focusing the perceptual resources according to goals and tasks. Conversely, attention can be involuntarily captured by certain types of stimuli in the visual field, providing inputs to the cognitive system that were not sought after but which may have profound effects on subsequent decisions.

What is not clear from this background is how perception and cognition actually interface as a coherent whole via the attention mechanism. While there are theories about what is produced at each stage of perception (preattentive versus attentive), what can be said about limits or constraints on the metaphorical spotlight? At what point does the cognitive system actually attend to a visual object? Moreover, does attention begin after a visual object is fully processed and added to working memory, or does attention cause the final stages of processing to occur? Finally, what are the mechanisms available to cognition for steering attention, and how is attention involuntarily captured? In the next section, we will address these questions in the context of a virtual human, a helicopter pilot in a military simulation.

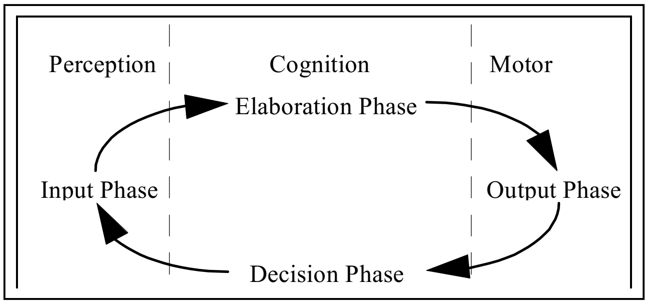

Figure 1: Soar decision cycle

3. Modeling Attention in a Virtual Pilot

3.1 Soar cognitive architecture

The virtual helicopter pilot was implemented in Soar, which is both an architecture for constructing intelligent systems and a unified theory of cognition [12,15,22,23]. All tasks in Soar are formulated as attempts to achieve goals in problem spaces. Each problem space consists of a set of states and a set of operators, states representing situations and operators representing actions. Operators perform the basic deliberate acts of the system—they can perform simple, primitive actions that modify the internal state and generate primitive external actions, or they can perform arbitrarily complex actions, such as executing a mission. The basic processing cycle is to repeatedly propose, select, and apply operators of a given problem space to a state. Operator selection, application, and termination are all dynamically determined by the system's knowledge, stored in the form of productions (condition-action rules). Any changes in goals, states, and perceptions can cause these productions to fire. There is no conflict resolution, so all applicable productions are allowed to fire in parallel.

Whenever there is insufficient or conflicting information on what activity to do next, Soar reaches what is called an impasse. This triggers the architecture to create a subgoal in which to resolve the conflict. Productions devoted to the subgoal can then fire, possibly leading to new subgoals, before the conflict is finally resolved. These subgoals disappear whenever the impasse is resolved. Subgoals serve three important functions. First, they dynamically make explicit the lack of knowledge in problem solving, making it possible to deliberately reason about how to resolve an impasse at a higher level. Second, they serve an organizing role. Much as traditional programming languages modularize code into separate procedures, Soar productions are organized based on the subgoals to which they are relevant. Third, subgoals facilitate Soar's ability to learn. As an impasse is resolved, Soar records the dependencies underlying this resolution and creates one or more new productions that prevent similar impasses from arising in the future. This process is known as chunking.

3.2 Perception in Soar

Soar agents execute the decision cycle shown in Figure 1. In the Soar framework, perception occurs during the input phase—this is when percepts are processed and the results are placed in the agent's working memory. During the elaboration phase, productions are matched with the contents of working memory and fire in parallel until quiescence is reached, meaning that no more rules fire. The rules that fire during the elaboration phase do not actually change the contents of working memory, rather, they create preferences for changes to working memory and they produce motor commands. These commands are issued to the motor system during the output phase. During the decision phase, a procedure evaluates the preferences that were generated during the elaboration phase; it decides what changes to make in working memory, and it chooses an operator to apply in the current context. Once the decision phase ends, the decision cycle begins again.

To begin answering the questions posed in the last section about the role of attention in a virtual human, the interface between cognition and perception in Soar occurs in working memory, where the percepts are placed during the input phase. Once in working memory, percepts are matched and used by operators to perform tasks associated with the current goal hierarchy. Sensors are controlled by issuing commands to the motor system, thus completing the connection between perception, cognition, and motor behavior. The problem, however, is that Soar does not place constraints on the number of percepts that can be placed in working memory at one time. Nor does the theory behind Soar tell us how to control the level of detail produced by the perceptual system, nor how to control the amount of information processed during the input phase. What is lacking is a way of focusing and controlling perceptual attention.

3.3 Soar pilot in a joint synthetic battlespace

The Soar virtual pilot is implemented in a distributed, interactive simulation called ModSAF. The Soar pilot flies a synthetic helicopter and performs tactical operations with a team of other Soar pilots. Teams of Soar pilots are deployed along with thousands of other entities (i.e., tanks, trucks, individual combatants, airplanes, etc.) in what is known as a joint synthetic battlespace.

Figure 2 shows the relationship between the Soar pilot and the simulation environment. The perceptual and motor systems are implemented as a set of C routines that provide the interface between the simulator and the Soar pilot.

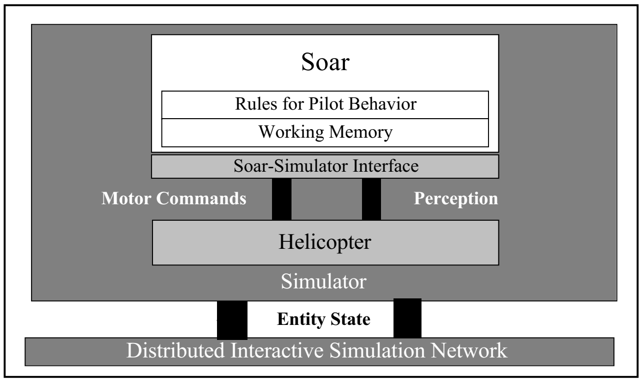

Figure 2: Soar Pilot Architecture

Perception in a joint synthetic battlespace

Perception in the joint synthetic battlespace involves four distinct problems: perception of terrain, perception of messages, perception of cockpit instruments, and the perception of entities. The first problem, the perception of terrain, is critical for tasks such as flying, especially at low levels where the possibility of colliding with the ground requires the pilot to continually monitor for the presence of obstacles in the flight path. The terrain is available on demand in the form of an un-interpreted polygonal surface. Currently, the virtual pilot perceives the terrain via a look-ahead sensor that samples points along the flight path for their altitude and the pilot adjusts the flight parameters accordingly [24]. This approach to terrain perception provides sufficient functionality for the pilot to fly safely, but it does not provide the level of realism needed for tasks such as flying along a contour line around a hill instead of simply flying over it. Such behaviors require a more general visual processing capability than we have yet addressed.

The second problem, the perception of messages, involves receiving communications in a structured format provided by the Command and Control Simulation Interface Language (CCSIL) [7], using simulated radios. This fills a functional need for passing information between agents, and it is realistic to the extent that it uses standard military messages (e.g., the Operations Order, Situation Report, etc.)

The third problem, the perception of cockpit instruments, is what enables the virtual pilot to keep track of the helicopter's current state (i.e., air speed, altitude above ground level, heading, and so on). Based on these vehicle parameters, the virtual pilot modifies the parameters for flying the helicopter. Finally, the perception of entities is necessary for tasks like tracking, formation flying, and targeting. The virtual pilot perceives other entities using a simulated visual sensor, which is specifically designed to model human visual perception of entities.

Each of the four types of perception is handled independently of the others, hence, there is no direct interaction among these types of perception in the virtual pilot. While this is probably not realistic at the level of modeling how human pilots move their eyes from the cockpit instrument panel, to the terrain, to other entities, it does satisfy a number of key functional requirements in the virtual helicopter pilot, while also modeling the human ability to deal with competing perceptual demands when flying. Given the independent nature of how each form of perception is implemented in the virtual pilot, we can focus on the issue of integrating perception and cognition via attention in one area at a time before addressing the bigger issue of how they should all be integrated. In this paper we focus on the issue of attention in the context of the perception of entities, so no more will be said about terrain, messages, and cockpit instrument perception here.

Modeling entity perception

Entity perception is driven by the arrival of a stream of entity-state updates. Each update characterizes the momentary state of an entity: it provides information about the identity, location, and velocity of an entity, such as a tank. These updates are filtered through models of the pilot's visual sensors to determine what information is potentially perceptible. Entities that are too far away will be imperceptible. Entities within the perceptible range of the model may still be rendered imperceptible if they are occluded by a terrain feature or an environmental factor such as smoke or dust. The sensor models also determine the resolution of the percept based on factors like distance, dwell time, and visibility. Hence, an entity may initially be recognized only as a vehicle when perceived at a great distance, but it may be identifiable as a specific tank model at a closer range.

Given that the state information of the perceptible entities is directly available to the virtual pilot, many of the standard vision problems can be finessed. For instance, understanding whether a new update refers to a known entity, or whether it represents a newly perceived object is not a problem. Each entity has a unique identifier in the simulation that can be used to associate the entity state information with visual objects in the pilot's memory. Thus, it is simple to resolve new percepts with previously observed entities. While this is probably not a realistic model of how humans resolve visual percepts, it does provide the human functional capability to recognize entities it has seen before.

Perceptual overload

Recall that the Soar decision cycle (refer to Figure 1) has an input phase, which is when the perceptual processor receives sensory inputs and transforms them into working memory elements that can be matched during the elaboration phase. What happens when there are no limits set on perceptual processing? Soar does not impose a limit on perceptual processing, so the input phase will not end until every percept has been processed. In the case of entity perception, this primarily involves computing a variety of geometric relationships between the pilot and each entity. It turns out that the cost of these computations can be expensive, especially when there are many entities in the visual field. Without a way of limiting perceptual processing, overload conditions can occur, where there are hundreds of entities in the field of view and the pilot perceptually attends to all of them, every decision cycle. When this occurs, the Soar pilot begins to lose control of the aircraft and sometimes crashes into the terrain because it cannot keep up with the perceptual processing demands while simultaneously flying the helicopter (and performing other tasks) in real-time. While this may seem to argue in general for better resource allocation strategies, the problem of perceptual overload is the primary culprit in the virtual pilot. Without a mechanism to focus and limit perceptual processing, the Soar decision cycle can be dominated by input. Before describing how we used attention mechanisms to address the perceptual overload problem, we will briefly discuss another related issue.

Perceptual requirements

There are a number of pilot behaviors requiring entity perception. For instance, virtual pilots need to be able to perceive individual enemy vehicles in order to bypass, scout, hide, or engage them as targets. For these behaviors it is important to perceive the entity's state, position, orientation, and geometric relationship to the virtual pilot—these attributes are automatically computed for entities. But there are also behaviors where entity-level perception does not suffice. For instance, the escort behavior involves leading or following a group of helicopters to a landing zone. The escort must keep a prescribed distance and angle of orientation between itself and the group of helicopters being escorted. The way this behavior was initially modeled was to have the escort choose one of the escorted helicopters as a base of orientation and then continually adjust its own flight parameters to maintain the proper position with respect to the individual. This approach only works, however, as long as the helicopter used as a base of orientation remains with the escorted group—if it is shot down, crashes, or otherwise deviates from the group's formation, then the escort pilot becomes confused. Instead of orienting on an individual helicopter, the escort pilot should orient on the group as a whole. But if the pilot only perceives entities, then what is the basis for orientation? The pilot clearly needs perceptual support for tracking groups.

Perception needs attention mechanisms

We have shown two functional shortcomings in the initial model of perception and cognition. First, the perceptual system was susceptible to overload whenever there were too many entities in the visual field. There were no constraints on the amount of perceptual processing done during a decision cycle. This is both psychologically unrealistic and functionally disastrous for the virtual pilot under the right circumstances. Second, some pilot behaviors were not supported by the level of resolution afforded by entity perception—it was difficult to reason about groups of entities when all that is perceived are individual entities. Again, this lacks psychological realism—the Gestalt principles of grouping by proximity and similarity is predicated on the ability to perceive coherent groups—and it points to a need for the virtual pilot to perceive groups as well as individuals. For example, in tasks such as targeting the pilot needs to focus on a single entity at high acuity, but tasks such as scouting and situation assessment require the pilot to perceive groups of objects and their approximate location.

While these shortcomings were functional in nature, they also pointed to some psychologically unrealistic aspects of the virtual pilot's perceptual system. To overcome these shortcomings, we sought to address both the functional and psychological concerns with the model by applying principles of human perceptual attention to the virtual pilot. Recall from the earlier discussion that human perceptual attention limits what is perceived and selects the priorities for information routing and visual processing. By employing attention, humans do not typically experience the kind of perceptual overloading that occurs in the Soar pilot—attention focuses the processing resources on a subset of the total available information. By adding a model of attention to the Soar pilot, the problem of perceptual overloading should be eliminated. Furthermore, Eriksen and Yeh hypothesize that attention acts more like a zoom lens [5] than a spotlight. According to this metaphor, a zoom lens has a reciprocal relation between the magnification or level of detail and the size of the viewing field. When the zoom lens is set to a low level of magnification, the field of view is greater, albeit with a low level of detail. Conversely, as the power of the lens increases, the amount of detail increases, while the field of view decreases. According to Eriksen and Yeh, an ideal lens produces a constant amount of information across all power settings.

Applying the zoom lens metaphor of attention to the Soar pilot, the ability to perceive at different resolutions is provided by enabling the pilot to perceive groups when attention is set at low resolution and to perceive individual entities at high resolution. With an attention mechanism, the pilot exercises greater control over the amount of information that is processed, ameliorating the overload problems experienced previously. At the same time, the groups perceived at low acuity provide the right level of abstraction for tracking and reasoning about groups of entities. Furthermore, groups play an important role in controlling and shifting the focus of attention.

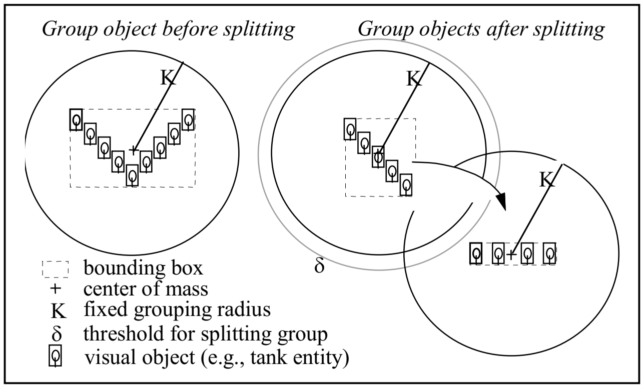

Figure 3: Grouping and splitting

3.4 Perceptual Grouping

Automatic grouping

Group objects are formed by the virtual pilot's perceptual system in one of two ways: automatically or voluntarily. Automatic grouping is initiated by the perceptual system during the preattentive stage of processing, which takes place during the Soar input phase (Figure 1). Groups are formed based on proximity and similarity (e.g., same vehicle type), which are commonly used Gestalt principles for grouping [6,11]. Each decision cycle, newly sensed visual objects are compared to existing groups—if a visual object is within 500 meters of the center of mass of a group, and its attributes are similar to the group's members, then it is clustered with that group. Otherwise, a new group is formed. New group objects are added to working memory and pre-existing group objects are updated at the end of the input phase.

Groups are dynamic in nature: they form, split, merge, move, and change shape. For this reason, perceptual grouping must also be dynamic. The attributes of group objects are updated every input phase. If members of a group have changed their positions, then the center of mass has to be re-computed along with all the other geometric relationships. Splitting occurs when a member of a group moves beyond a distance of K + δ from the group's center of mass (Figure 3). K, which is 500 meters in our model, is the threshold defining the radius of the group and δ is a small increment of K used for making the split. Once a visual object is split from a group, it is not orphaned for long, however. The clustering algorithm checks the orphan's distance to the center of mass of each of the other proximity-based groups—it is added to the first group within K meters. If there are no groups within K meters of the orphan, a new group is formed with the orphan as its only member, and other members may subsequently be added to it.

In operational terms, splitting occurs when a sub-unit of vehicles, which the pilot perceives as visual objects, maneuvers away from its parent. For example, a scout platoon may break away and move ahead of the main body of a force in order to scout the front. Perceptually, the split may occur over multiple decision cycles. Initially one vehicle will exceed the distance threshold, forcing a split from the group. Assuming no other groups are nearby, a new group is formed containing the single vehicle. Several decision cycles later a second vehicle may split from the original group. This vehicle will be added to the new group if is close enough, otherwise a new group is formed. In the same way other members of the platoon are split and added to the new group. Thus splitting a group usually occurs gradually and does not require clustering all of the visual objects from scratch.

Currently we do not handle the merging of groups. Such a capability would be useful, however, since there are times when the pilot needs to perceive that several groups have massed together and a new, larger group has been formed. For instance, a scout on a reconnaissance mission would need to recognize that the enemy force is massing its vehicles, which could be an indicator of a change in its tactical intentions. We will address this issue in future work.

For automatic grouping, we chose a radius of 500 meters because it tends to cluster vehicles from platoon and company-size units together. Although we fixed the radius for our initial model at 500 meters, it is actually adjustable parameter of the clustering algorithm. To implement a zoom lens with multiple degrees of resolution, the clustering radius could be dynamically adjusted. At low resolutions a large clustering radius would yield larger groups, and vice versa. Another approach to grouping, which may be more realistic, is to cluster visual objects according to their angular proximity and distance in the pilot's visual field instead of by their positional proximity. This moves away from a world-centered coordinate system towards a more viewer-centered approach to visual processing. This would also tend automatically vary the size of the groups, depending on their proximity to the viewer. Visual objects closer to the viewer will tend to be spread out across a wider angle, resulting in more groups with fewer members. Groups far away from the viewer would potentially contain more members. This technique for automatic grouping potentially provides a more natural way of perceiving and reasoning about groups.

Voluntary grouping

The second method of grouping is voluntary. It is invoked by the virtual pilot for tasks such as escorting and tracking of groups. The pilot identifies the individual visual objects to be tracked and groups them with a command to the perceptual system. Once the pilot makes the decision to track a set of visual objects as a group, the perceptual system creates a group object in working memory that can be used for further cognitive processing. For instance, when the task is to escort a division of transport helicopters, it is more effective for the virtual pilot to orient itself on the center of mass of the group than on a particular helicopter. To accomplish this the pilot visually identifies the helicopters it intends to escort and then forms a group object, which can then be used for the escort task. The group objects formed voluntarily are exclusive—the automatic grouping mechanism cannot add members to a voluntary group, even when they are proximally located and have similar attributes.

Groups have perceivable properties

Group objects have many of the same attributes as entities, but there are a few distinctions: (1) A group object's location is computed as the center of mass of its members—the center of mass provides the basis for computing the geometric relationships between the pilot and the group. (2) Group objects have a bounding box, which circumscribes the group's members. (3) Group objects also keep track of the number, types (e.g., air defense, tank, helicopter, etc.) and status (i.e. destroyed or normal) of the visual objects contained within them. This information provides important cues to the pilot for guiding attention. (4) When a group member fires a weapon, the muzzle flash is recorded as an attribute of the group. Since flashes of light are the type of exogenous event that can capture attention, the sudden onset of the group's flash attribute can trigger productions to fire that move attention toward the group.

Automatic versus voluntary grouping

Proximity-based groups are useful for tasks like evading and targeting, where the group's location, force (i.e., friendly or opposing), and composition are exactly what the pilot needs to know in deciding what to do next. Proximity-based groups are not, however, sufficient for tasks like escorting. This is because the escorting task involves tracking a specific set of visual objects (e.g., a division of transport helicopters), and it is likely that this set will not be exclusively contained in one proximity-based group. In practice, two situations arise that make proximity-based groups inadequate for escorting. In the first case, the set of visual objects being tracked are not located close enough to one another to be automatically clustered by proximity, so they are clustered into disparate groups. The second case occurs when task-irrelevant visual objects are proximally mixed with the visual objects being tracked, which can lead to confusion. These two situations raised the need for a voluntary, or top-down, method of grouping, which provides the pilot with a method of intentionally forming groups of visual objects to perform specific tasks.

3.5 Controlling attention

As in humans, the virtual pilot's attention can be controlled in two ways, endogenously and exogenously. To control attention endogenously, which is a top-down, goal-driven form of control, the pilot normally chooses to perceive only groups, which is the lower acuity mode of perception that saves the cost of processing visual objects at high resolution. Groups provide cues such as location, size, and some details about the membership that can be used for searching for specific visual objects. The pilot may be interested in focusing on visual objects with a particular set of features. For example, one of the highest priorities in the pilot's visual search is to identify enemy air defense vehicles—this is driven by a goal of survival. Lower priorities for visual search include enemy tanks and armored vehicles, particularly those that are in firing range. To shift attention to a visual object involves specifying a set of features to the perceptual system that will cause the visual object, if it is present in the visual field, to be processed and added to working memory. One or more of these features may be selected: group membership, the distance between the visual object and the viewer, vehicle class (e.g., helicopter, tank, or truck), vehicle type (e.g., T-80, AH-64, and so on), and force orientation (opposing versus friendly force). Although these do not exactly match the basic visual features found in humans, there is a correspondence. Wolfe [30] defines a basic feature as something that supports efficient visual search and effortless segmentation. He identifies ten basic visual features: color, orientation, curvature, vernier offset, size, motion, shape, pictorial depth cues, stereoscopic depth, and gloss. These basic features are preattentively processed and used for visual search. In the same way, we use groups as a way of segmenting a scene, and individual visual objects have features corresponding to depth, size, and shape that can be used for searching and attending.

Attention may also be captured exogenously or involuntarily. This occurs when a feature reaches a threshold, such as when there is a flash of light from a muzzle blast. The sudden onset of this stimulus causes the visual object to automatically be placed in the pilot's working memory. Once in working memory, the pilot's productions fire and react to the event. In this way, visual objects can capture attention by exogenous events. Other events that could possibly be used to capture attention in this way are motion, luminance, and the sudden appearance of smoke or dust.

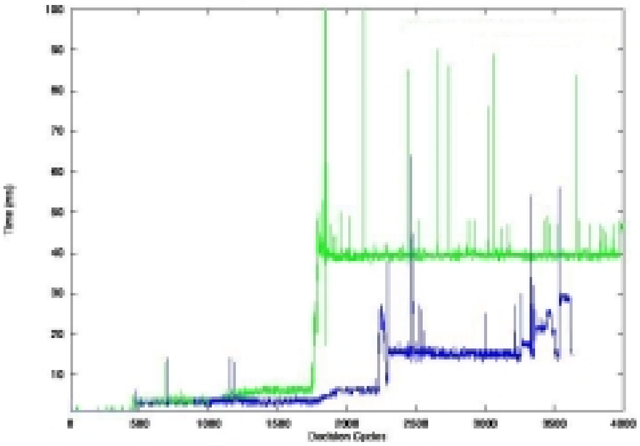

Figure 4: Test Results. Compares the time (ms) for perceptual processing per decision cycle of two identical missions, one without perceptual attention (light gray) and the other with it (dark). Perceptual attention saved computation on average and in peak demands.

4. Experimental Results

One way of evaluating the effects of adding attention to the pilot's perceptual system is to compare the perceptual loads on the pilot with and without attention. To gather data, the pilot was sent on the same mission (a standard deep attack) two times. The mission was such that for approximately 1500 decision cycles there were only a few vehicles in the pilot's visual field at any given time while the pilot traveled to a destination. Once at the destination, the pilot would periodically pop-up and peek over a hill at 120 other vehicles, then mask itself again behind the hill so that the vehicles were no longer in view. The amount of time spent in perceptual processing per decision cycle was measured over the course of the mission. The results are shown in Figure 4. Note that the average time spent on perceptual processing per decision cycle was approximately 15 milliseconds when attention was used, with peaks around 40 milliseconds. Without attention, the average perceptual processing time was around 50 milliseconds, with peaks above 500 milliseconds (not shown due to scale of graph). From experience, we found that peaks much above 100 milliseconds will sometimes result in helicopter crashes and that an average rate of 50 milliseconds puts a high load on the pilot.

5. Discussion

One of fundamental aspects of attention in our model is that perceptual processing occurs in preattentive and attentive stages. The effect of this approach to controlling attention is that it reduces computation. Preattentive processing identifies new visual objects and updates the basic state information on previously perceived ones, but it does not perform detailed geometric reasoning or add information to working memory. Groups are also formed and updated preattentively. When a visual object is selected for attention, all of its attributes are computed and placed in working memory—only attended visual objects make it into working memory. Computation is saved by focusing attention on a small number of visual objects at any given time.

One of the weaknesses of our current approach is that there is not a hard limit placed on the number of visual objects that can be attended simultaneously. Although we employ strategies that effectively limit the focus of attention, this is not the same as Eriksen and Yeh's zoom lens hypothesis, where there is a constant amount of information available at any given setting. In our current model, we cannot guarantee that the processing peaks will always be below 100 milliseconds, or that the average load will stay at a reasonable level. Although a more powerful computer could improve these statistics, this solution would not address the more fundamental problem of how to balance perception and cognition. The visual processing in the joint synthetic battlespace is relatively simple and does not require the more complex forms of computation used in computer vision. As the fidelity of the joint synthetic battlespace (or other virtual worlds) increases, so will the complexity of the visual processing, which could easily put us back in the same position of needing to balance the needs of perception and cognition.

We plan to address this more general issue in a couple of ways. First, the visual sensor will be modeled more realistically by adding foveal and peripheral fields to the virtual pilot's visual cone. By creating high and low fields of acuity, the problem of attending simultaneously to many visual objects is addressed by limiting how many can be perceived. This is the approach advocated by researchers in active vision [2,27]. With a fovea, however, less of the visual field will be accessible to the pilot. As a consequence, the pilot will require visual behaviors for controlling its gaze. Ballard [2] and others [18,27] suggest indexing schemes for remembering where things are located, which is necessary in deciding where to look.

Second, we would like to investigate how to limit the amount of information that is processed and placed in working memory by the perceptual system. Should there be an absolute limit on the amount of information? Or should there be a method for scaling the processing to the amount of perceptual information and time available? The outcome should be that both the perceptual and cognitive systems have a sufficient opportunity to do their work, and the interface between the two is better understood.

6. Related work

There are a number of researchers interested in generating believable behavior in virtual humans. Rickel and Johnson [21] developed a virtual tutor that teaches the student how to operate equipment aboard a Navy ship. Their tutor, also built with Soar, shifts its gaze alternately between the student and the equipment in a human-like fashion, which helps to direct the student's attention and creates a more credible tutor-student interaction. The difference from our work is that the tutor's behavior is generated to create believable external behavior as opposed to actually changing the way it perceives the world—it perceives everything regardless of the gaze direction. Similarly, Chopra and Badler [3] generate believable gaze (overt) behaviors in an animated figure. However, as with Rickel and Johnson's tutor, it does not appear that the gaze behaviors are based on a realistic model of perceptual attention and cognition—the gaze behaviors do not affect what the agent actually perceives, nor are the generated to service a cognitive need for information.

Aasman [1] built a virtual car driver that maneuvers through traffic in a simulation world. It employs a detailed cognitive model for controlling overt attention via gaze and head movements. Aasman's work is potentially useful in extending our pilot's ability to control gaze, but like the others, he does not address the issue of balancing perceptual and cognitive processing. Wiesmeyer [28] built a model of covert attention using Soar, but modeled attention with operators rather than addressing the issue of what is processed by the perceptual system. While Wiesmeyer's model did produce performance data comparable to humans on a number of well known psychological tests, his model does not address the issue of controlling the amount of information processed, hence it would be susceptible to overloading under the right conditions.

With respect to perceptual grouping, Flinchbaugh and Chandrasekaran [6] did some of the early work in AI on perceptual grouping using Gestalt principles of proximity, similarity and motion. While this work addressed grouping, the system was not integrated with a cognitive system, nor was the processing performed in a real-time system.

Numerous AI researchers have addressed the importance of attention in vision. For example, Tsotsos et al. [25] argue that attention acts to optimize the search procedure in solving a vision problem, whether in the brain or in a computer, and they have implemented an artificial neural network to solve vision problems where attention selects spatial regions and features of interest. It is not clear whether they have employed their model in a dynamic environment yet, nor how they would integrate it with goal-driven behavior. Reece and Shafer [18] employed a model of attention in their work on (simulated) robot driving. Task-based perceptual priorities and the use of visual routines drove the search for visual objects in a scene. They showed that by making assumptions about world constancy, they could avoid re-sensing relatively static aspects of a scene, thus saving computation. Their work and other AI research on active perception [2,27] address many of the efficiency issues that we are concerned with in this paper. We are attempting to extend this body of research by integrating active perception with a cognitive architecture like Soar, which is capable of goal-directed behavior and general problem solving. Without a strong cognitive component, realistic and believable behavior will not be possible.

7. Conclusions

Attention plays a significant role in defining the interface between perception and cognition in humans. Without attention, the perceptual and cognitive systems can become unbalanced, resulting in unacceptable behaviors, both functionally and in terms of realism. To create realistic virtual humans we have proposed an approach to attention based on a psychological framework. This model has been implemented and tested in a virtual helicopter pilot that performs complex tasks in a military simulation. The model described here is not complete—it is a step toward a more comprehensive model of attention, which must include other effects such as overt orienting of attention and gaze control. Moreover, we would like to understand the interplay between top-down and bottom-up attention, particularly in the context of gaze control and knowing where to look and how to the information to assess the situation. These topics will be addressed in future research.

8. Acknowledgements

We gratefully acknowledge the support of the Defense Advanced Research Projects Agency (DARPA) and the Office of Naval Research (ONR). DARPA's Advanced Simulation Technology Thrust (ASTT) provided funding via subcontract L74158 with the University of Michigan, under prime contract N61339-97-K-008. ONR's support was provided under award N00014-98-1-0132.

9. References

[1] J. Aasman. "Modeling Driver Behaviour in Soar," Ph.D. Thesis, Rijksuniversiteit Groningen, Leiden, 1995.

[2] D. Ballard. "Animate Vision," *Artificial Intelligence* 48, (1991), 57-86.

[3] S. Chopra and N. Badler. "Where to Look? Automating Visual Attending Behaviors of Virtual Human Characters," *Proceedings of the Third International Conference on Autonomous Agents*. May 1-5, 1999. Seattle, WA.

[4] H. Egeth and S. Yantis. "Visual Attention: Control, Representation, and Time Course," *Annual Review of Psychology*, 1997, 48:269-97

[5] C. Eriksen and Y. Yeh. "Allocation of Attention in the Visual Field," *Journal of Experimental Psychology, Human Perception and Performance*, 1985, Vol. 11, No. 5, 583-597.

[6] B. Flinchbaugh and B. Chandrasekaran. "A theory of spatial-temporal aggregation for vision," *Artificial Intelligence*, 17 (1981), 387-407.

[7] S. Hartzog, M. Salisbury. "Command Forces (CFOR) Program Status Report," Proceedings of the Sixth Conference on Computer Generated Forces and Behavioral Representation, Orlando, Florida, July, 1996.

[8] R. Hill, J. Chen, J. Gratch, P. Rosenbloom, M. Tambe, "Soar-RWA: Planning, Teamwork, and Intelligent Behavior for Synthetic Rotary Wing Aircraft," *Proceedings of the 7th Conference on Computer Generated Forces & Behavioral Representation*, Orlando, FL., May 12-14, 1998.

[9] R. Hill, J. Chen, J. Gratch, P. Rosenbloom, M. Tambe, "Intelligent Agents for the Synthetic Battlefield: A Company of Rotary Wing Aircraft," *Proceedings of Innovative Applications of Artificial Intelligence* (IAAI-97), Providence, RI, July 1997.

[10] B. Julesz and J.R. Bergen. "Textons, The Fundamental Elements in Preattentive Vision and Perception of Textures," *The Bell System Technical Journal*, Volume 62, No. 6, July-August 1983.

[11] M. Kubovy and A. Holcombe. "On the Lawfulness of Grouping by Proximity," *Cognitive Psychology*, 35, 71-98, 1998.

[12] J. Laird, A. Newell, P. Rosenbloom, "Soar: An architecture for general intelligence," Artificial Intelligence 33(1):1-64, 1987.

[13] C. Moore, and H. Egeth. "Perception without attention: Evidence of grouping under conditions of inattention," *Journal of Experimental Psychology: Human Perception and Performance*, 1997, Vol. 23, No. 2, 339-352.

[14] U. Neisser. *Cognitive Psychology*. NY: Appleton-Century-Crofts, 1967.

[15] A. Newell. *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press, 1990.

[16] R. W. Pew and A. S. Mavor, editors. *Modeling Human and Organizational Behavior: Application to Military Simulations.* National Academy Press, Wash., D.C, 1998.

[17] M. Posner. "Orienting of attention," *Quarterly Journal of Experimental Psychology*, (1980) 32, 3-25.

[18] D. Reece and S. Shafer. "Control of Perceptual Attention in Robot Driving," *Artificial Intelligence*, 78 (1995) 461-505.

[19] D. Reece and R. Wirthlan. "Detection Models for Computer Generated Individual Combatants," *Proceedings of the Sixth Conference on Computer Generated Forces and Behavioral Representation*. Orlando, FL, 1996.

[20] D. Reece, D. Ourston, M. Kraus, and I. Carbia. "Updating ModSAF for Individual Combatants: the DI SAF Program," Proceedings of the Seventh Conference on Computer Generated Forces and Behavioral Representation. Orlando, FL, May 12-14, 1998.

[21] J. Rickel and W. Lewis Johnson. "Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control." To appear in *Applied Artificial Intelligence*.

[22] P. Rosenbloom, J. Laird, A. Newell, and R. McCarl, "A preliminary analysis of the Soar architecture as a basis for general intelligence," Artificial Intelligence 47(1-3):289-325, 1991.

[23] P. Rosenbloom, J. Laird, A. Newell (editors)., The Soar Papers: Research on Integrated Intelligence. MIT Press, Cambridge, MA, 1993.

[24] M. Tambe, K. Schwamb, and P. Rosenbloom. “Building Intelligent Pilots for Simulated Rotary Wing Aircraft,” Proceedings of the Fifth Conference on Computer Generated Forces and Behavioral Representation, 1995.

[25] J. Tsotsos, S. Culhane, W. Yan Kei Wai, Y. Lai, N. Davis, and F. Nuflo. “Modeling visual attention via selective tuning,” *Artificial Intelligence* 78 (1995) 507-545.

[26] M. van Lent and J. Laird. “Developing an artificial intelligence engine.” Proceedings of the 1999 Game Developers’ Conference, San Jose, CA.

[27] G. Wasson, D. Kortencamp, and E. Huber. “Integrating active perception with an autonomous robot architecture.” *Proceedings of the Second International Conference on Autonomous Agents.* May 1998

[28] M. Wiesmeyer. “An Operator-Based Model of Human Covert Visual Attention. CSE-TR-123-92.” Ph.D. thesis, Department of Engineering Division and Computer Science., University of Michigan, Ann Arbor, Michigan, 48109-2122.

[29] J. Wolfe. “Guided Search 2.0: A revised model of visual search,” *Psychonomic Bulletin & Review*, 1994, 1 (2), 202-238.

[30] J. Wolfe. “Visual Search,” In H. Pashler (editor), *Attention*. Hove, East Sussex, UK: Psychology Press Limited, 1998.

Author Biography

RANDALL W. HILL, JR. is a project leader at the University of Southern California Information Sciences Institute (USC-ISI) and a research assistant professor in the computer science department at USC. He received his B.S. degree from the United States Military Academy at West Point in 1978 and his M.S. and Ph.D. degrees in computer science from USC in 1987 and 1993, respectively. His research interests are in the areas of integrated intelligent systems, cognitive modeling, perception, and intelligent tutoring systems.