# A Computational Model of Dynamic Perceptual Attention for Virtual Humans

*Youngjun Kim*
*Randall W. Hill, Jr.*
*David R. Traum*
Institute for Creative Technologies
13274 Fiji Way, Suite 600
Marina del Rey, CA 90292
yjkim@ict.usc.edu, hill@ict.usc.edu, traum@ict.usc.edu

Keywords:
Perception, Attention, Virtual human, Non-verbal communication

**ABSTRACT:** *An important characteristic of a virtual human is the ability to direct its perceptual attention to objects and locations in a virtual environment in a manner that looks believable and serves a functional purpose. We have developed a computational model of perceptual attention that mediates top-down and bottom-up attention processes of virtual humans in virtual environments. In this paper, we propose a perceptual attention model that will integrate perceptual attention toward objects and locations in the environment with the need to look at other parties in a social context.*

## 1. Introduction

Modeling interactive virtual humans has been one of the primary goals of immersive virtual environments for training. An important characteristic of a virtual human is the ability to direct its perceptual attention to objects and locations in a virtual environment in a manner that looks believable and serves a functional purpose. Not only must the virtual human pay attention to objects related to the tasks it is performing, but it must also be able to cope with sudden events that demand attention. It is often the case that the amount of information in the virtual environment far exceeds the processing abilities of the virtual human. In fact, only a small fraction of sensory information can be fully processed and assimilated into the cognitive model. A successful model of perceptual attention provides a way of both pruning the incoming sensory input and choosing the most salient information to focus on during the next step of a decision cycle.

The computational models of perceptual attention that we surveyed fell into one of two camps: top-down and bottom-up. Biologically inspired computational models (Itti, 2001; Courty et al., 2003) typically focus on the bottom-up aspects of attention, while most virtual humans (Traum and Rickel, 2002; Chopra and Badler 2001; Marco and Neil, 2002; Conde and Thalmann 2004) implement a top-down form of attention. Bottom-up attention models only consider the image information (e.g, color, intensity, orientation, and motion) without taking into consideration saliency based on tasks or goals. As a result, the outcome of a purely bottom-up model will not consistently match the beahvior of real humans in certain situations. Models like Itti's (2001) can predict the bottom-up salience of features in an image at any point in time, but such a model is not sufficient to predict where to actually look. Humans are generally task-oriented, and it is safe to say that a great deal of one's time is spent looking at objects related to the current task.

Modeling perceptual attention as a purely top-down process, however, is also not sufficient for implementing a virtual human. A purely top-down model does not take into account the fact that virtual humans need to react to perceptual stimuli vying for attention. For instance, it is reasonable to expect that a loud noise, like gunfire or an explosion, will catch the attention of virtual human and invoke some kind of immediate reaction. Top-down systems typically handle this in an ad hoc manner by encoding special rules to catch certain conditions in the environment. The problem with this approach is that it does not provide a principled way of integrating the ever-present bottom-up perceptual stimuli with top-down control of attention.

One of the distinctions between the work described in this paper and some of the other work on models of perceptual attention is the purpose of the model in the context of a virtual human. In many of the systems we reviewed, the purpose of the perception model was to make the virtual human behave as though it was attending to the surroundings and tasks in a natural way. In contrast, our goal is also to develop virtual humans that can perform tasks, react to contingencies, interact with other agents, both virtual and human, plan, and make decisions about what to do next or at some future time (Hill, 2000). To accomplish this, perceptual attention is a critically important mechanism for restricting the sensory information being processed by the perception module and controlling virtual humans to exhibit goal-directed and reactive behaviors. In the first stage of perceptual attention, there are mechanisms that filter the information that comes through the sensory system. Subsequent processes selectively strengthen or weaken the priority of the information. Directing perceptual attention toward the interests of a particular region in space can be achieved by two distinguishable shifts; covert and overt shifts of perceptual attention. It is well known that covert and overt attention shifts affects gaze direction to locations in space (Wolfe, 1994). The sequences of gaze fixations describe the way in which overt attention deployed, whereas directing attention to location in space without moving gaze describes the way in which covert attention is deployed.

In this paper, we present a computational model of perceptual attention for virtual humans. This model extends a prior model of perceptual resolution (Hill, 2000) based on psychological theories of human perception. This models allows virtual humans to dynamically interact with objects and other individuals, balancing the demands of goal-directed behavior with those of attending to novel stimuli. This model has been implemented and tested with the MRE Project (Swartout et al, 2001).

## 2. Modeling Perception in Virtual Humans

Our virtual humans are implemented in Soar, a general architecture for building intelligent agents (Newell, 1990) and build on the STEVE Architecture (Rickel and Johnson, 1999) in the immersive environment called the Mission Rehearsal Exercise (MRE) (Swartout et al, 2001). The virtual humans' behavior is not scripted; rather, it is driven by a set of general, domain-independent capabilities. The virtual humans perceive events in the scenario, by interacting with the simulator, reason about the tasks they are performing, and they control the bodies and faces of the PeopleShop™ animated bodies to which they have been assigned.

We have developed a model of perceptual resolution based on psychological theories of human perception (Hill, 1999 and 2000). Hill's model predicts the level of details at which an agent will perceive objects and their properties in the virtual world. He applied his model to synthetic helicopter pilots in simulated military exercise. We extended the model to simulate many of the limitations of human perception, both visual and aural.

### 2.1 Visual Perception

The virtual human perceives dynamic objects, under the control of the simulator, by filtering updates (e.g., body location and orientation, gaze location and orientation, velocity, size, and color) that the simulator periodically broadcasts. As shown in figure 1, we limited the virtual human's simulated visual perception to 190 horizontal degrees and 90 vertical degrees so that the virtual human only gets updates that he is currently sensing through the field of view (FOV). When the virtual human senses the objects in the FOV, it first processes how salient each object is in the respect of size, distance, and color. We consider the computational model (Nothegger et al., 2004) to compute the visual salience of each object that is measured by observing individual visual attributes (e.g., size, shape, and color).

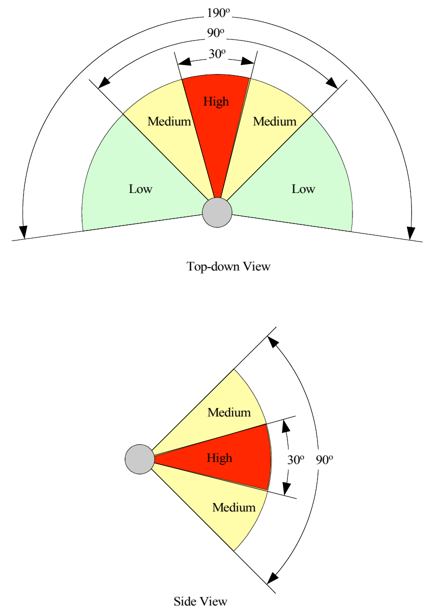



Figure 1. Virtual Human's Functional Visual Field (FVF)

After computing the visual saliencies of the perceived objects, we applied a sigmoid function as a utility function that reduces the degree of salience of an object in the respect of angle disparities between the virtual human and the object. Then we classified the levels of saliencies on those objects as high, medium, or low, depending on where the objects is in the virtual human's field of view and whether attention is being focused on it.

### 2.2 Aural Perception

To model aural perception, we estimate the sound pressure levels of objects in the environment and compute their individual and cumulative effects on each listener based on the distances and directions of the sources. This enables the virtual humans to perceive aural events involving objects not in the visual field of view. For example, when a virtual human hears a vehicle is approaching from behind, he can choose to look over his shoulder to see who is coming. Another effect of modeling aural perception is that some sound events can mask others. A helicopter flying overhead can make it impossible to hear someone speaking in normal tones a few feet away. The noise could then prompt the virtual human to shout and could also prompt the addressee to cup his ear to indicate that he cannot hear.

Given a set of visually or aurally perceived objects, the agent's perceptual model updates the attributes of objects that fall in the limited sensory range. At any point in time, the virtual human must recognize which object is the most salient among those objects and draw his focus of attention on the object. The next section describes our approach to computing the salience of the objects in the field of view and the subsequent behaviors associated with shifting the agent's gaze.

## 3. Computational Model of Perceptual Attention

To compute object salience and to control gaze behaviors, we have developed a model called Dynamic Perceptual Attention (DPA). Internally, DPA combines objects selected by bottom-up and top-down perceptual processes with a decision-theoretic perspective and then selects the most salient object. Externally, DPA controls an embodied agent's gaze not only to exhibit its current focus of attention but also to update beliefs (e.g., position) of the selected object. That is, the embodied agent dynamically decides where to look, which object to look for, and how long to attend to the object.

Figure 2. A snapshot of the MRE simulation

### 3.1 Decision-Theoretic Control

One of the consequences of modeling perception with limited sensory inputs is that it creates uncertainty on each perceived object. For instance, if an object that is being tracked moves out of an agent's field of view, the perceptual attention model increases the uncertainty level of the target information of the object that a virtual human tries to observe.

To illustrate this idea, consider the screen snapshot of the MRE simulation shown in figure 2. An injured boy is being attended to by his mother and a medic. A sergeant is conversing with a human participant. Since the mother, the boy, and the medic are out of the visual field of view of the sergeant while the sergeant is conversing with the human, the sergeant's uncertainty levels about each of these characters will increase with time.

The information flow of the DPA module is shown in figure 3. Top-down and bottom-up processes provide information to the DPA module in the form of tuples composed as follows:

$$tuple_i = \left\langle objP_i, objC_i, objDGI_i, objCGI_i, k_i \right\rangle$$

where, $objP_i$: priority of the $tuple_i$
$objC_i$: concern of the $tuple_i$
$objDGI_i$: desired goal information of the $tuple_i$
$objCGI_i$: current goal information of the $tuple_i$
$k_i$: constant for the $tuple_i$

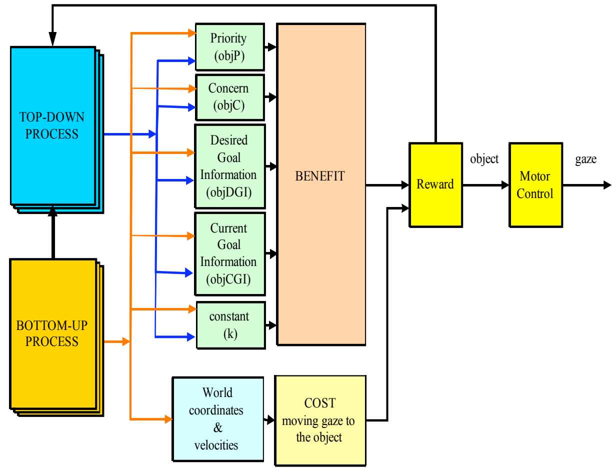



Figure 3. The information flow of the DPA module

The priority attribute, objP, is used to indicate the absolute importance of an object, whereas the concern attribute, objC, is used to indicate a conflict between the desired goal information (objDGI) and the current certainty of information (objCGI). For instance, even if a person is given a high priority task, he may not be concerned about monitoring objects associated with the task if the task is going well, resulting in less frequent observations. If the task runs into some difficulties, he will increase his concern for the task, resulting in more frequent observations.

By considering both attributes (i.e., priority and concern), our virtual humans compute the benefits of attending to objects. Information certainty is one of factors that help the virtual human decide which object it has to focus on. To deal with certainties of the perceived objects, we have chosen to take a decision theoretic approach to computing the perceptual costs and benefits of shifting the focus of perceptual attention of the perceived objects. In the next two sections, we will describe how to compute the perceptual costs and benefits of shifting the focus of perceptual attention. The expected cost is computed by calculating the perceptual cost of shifting the gaze to the selected object. The expected benefit is computed by considering the value of acquiring accurate information about the selected object. Once a decision has been made, DPA shifts the virtual human's gaze to focus his perceptual attention on the object that has the highest reward.

### 3.2 Computing the Benefit

To compute the benefit of focusing perceptual attention on an object requires the estimated values of object-based information certainty. We consider object-based information certainty as a key factor in computing the benefit of shifting the focus of attention to the object. The term, *object-based information certainty*, is used here to describe the level of information certainty of an object rendered in the agent's mental image of a virtual world.

Humans determine the desired goal information certainty of perceived objects (*objDGI*) based on their subjective preferences or prediction and then make efforts to maintain the current certainty of information (*objCGI)* within a specific range of objDGI, defined as the information certainty tolerance boundary (*ICTB*).

Information certainty is dynamic both in space and time. If (*objCGI*) is out of *ICTB*, we activate one of two kinds of NEEDs: the NEED for observation or the NEED for inhibition. The NEED for observation is activated if objCGI goes below $ICTB_{lower}$. The NEED of inhibition is activated as objCGI goes over $ICTB_{upper}$. According to Klein's account of the *inhibition of return* (Klein, 2000), too much information can be a bad thing. By modeling the inhibition of return, perceptual attention will not permanently focus on the most active salient information but will increase the chances of diverting perceptual attention to less salient information.

The orthogonal process model between information certainty and the NEEDs of observation and inhibition is shown in figure 4.

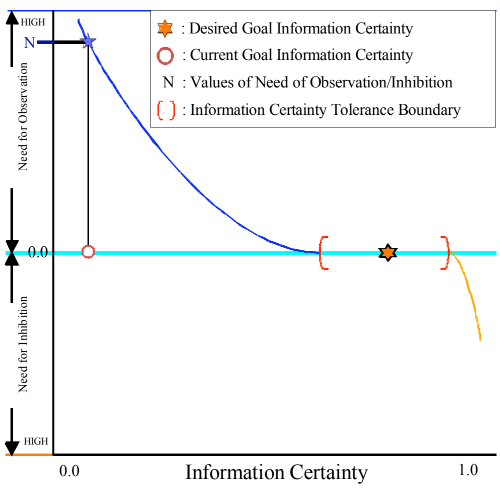



Figure 4. The interrelation of Information Certainty and Need

The desired goal information certainty (objDGI) is determined by the priority attribute (objP). The information certainty tolerance boundary is set by the concern attribute (objC). The higher the concern attribute is, the narrower the length of the boundary is. The current goal information certainty of the target object (objCGI) is set by top-down and bottom-up processes. If a virtual human cannot retrieve any information certainty of the target from top-down and bottom-up processes, it sets objCGI to 0. After the values for objCGI and information certainty tolerance boundary are set, the virtual human computes the NEED for observation or for inhibition on each tuple as follows:

$$\mathrm{NEED}(\mathrm{tuple}_i) = \begin{cases} -1.0 \times objP_i \times \exp^{\alpha} & \text{if } \mathrm{objCGI}_i > \mathrm{ICTB}_{upper} \\ 0 & \text{if } \mathrm{ICTB}_{lower} \leq \mathrm{objCGI}_i \leq \mathrm{ICTB}_{upper} \\ objP_i \times \exp^{\beta} & \text{if } \mathrm{objCGI}_i < \mathrm{ICTB}_{lower} \end{cases}$$

$$\text{where, } \alpha = \mathrm{objCGI}_i - \mathrm{ICTB}_{upper} \text{ and } \beta = \mathrm{ICTB}_{lower} - \mathrm{objCGI}_i$$

The NEED on $tuple_i$ is used as a force that produces a benefit of diverting perceptual attention into $tuple_i$. The benefit is computed as follows:

$$BENEFIT(tuple_i) = \frac{NEED(tuple_i)^2}{2}$$

Once *BENEFIT(tuple$_i$)* is computed, it will used with *COST(tuple$_i$)* to compute the *REWARD(tuple$_i$)*.

### 3.3 Computing the Cost

Even if the benefit of drawing attention to one object is higher than the benefits of attending to others, the virtual human should not automatically select that object as the best one since the cost of shifting the focus of attention must also be considered. To compute the cost of shifting perceptual attention from one object to another, we consider two sets of factors: physical and social. Physical factors include the degrees of head and eye movements and distance efficiency. Social factors indicate the relative costs of perceptual gaze shifts in social interaction. For instance, it may be rude to look away when someone is speaking (high cost of shift), yet it may be very important to attend to an unexpected or potentially dangerous event (high cost not to shift).

### 3.4 Shifting Perceptual Attention

With the benefit and two sets of cost factors of each tuple, we compute *REWARD(tuple$_i$)* as follows:

$$REWARD(tuple_i) = BENEFIT(tuple_i) - COST(tuple_i)$$

After calculating REWARD(tuple) of all tuples, the virtual human selects a tuple that has the highest REWARD. If the selected tuple is holding the current focus of perceptual attention, the virtual human will keep focus on it. If not, it will divert its perceptual attention to the tuple having the highest REWARD.

The duration of a gaze at an object affects the information certainty level. While a virtual human gazes at an object obj (i.e., overt monitoring), objCGI increases. Likewise, while obj is monitored only in the virtual human's memory and projection (i.e., covert monitoring), objCGI decreases. Covert monitoring will cause the certainty of information to decay over time.

## 4. Implementation in MRE Scenario

When our scenario starts, a simulated army vehicle carries a human participant (lieutenant) to an accident site where an Army vehicle has crashed into a civilian car, injuring a boy. The participant then takes on the task of directing the troops to rescue the boy by interacting with our three embodied conversational agents—the sergeant (SGT), the mother, and the medic. We controlled the sergeant's gaze movements with DPA. The sergeant is initially looking at the boy to update the boy's health status with information.

In a typical example of interaction with the system, the lieutenant starts with a general inquiry as to what is going on, "Sergeant, what happened here?" Since this inquiry is given as an aural event, the aural perception filters the aural event and then gives a tuple for the event to DPA.

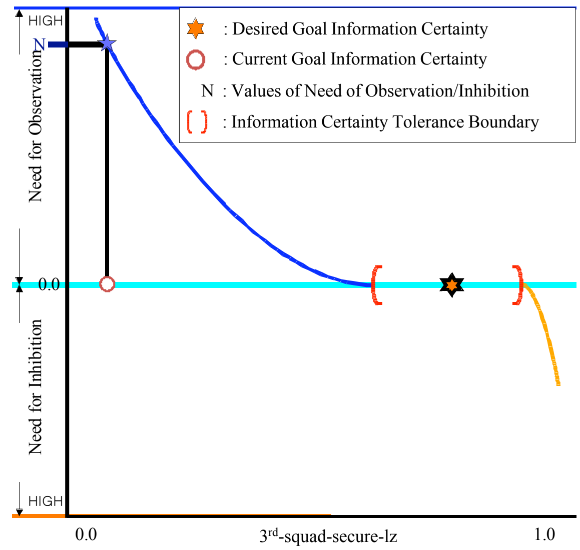
When DPA gets this tuple from the aural perception, DPA shifts the sergeant's perceptual attention, which currently attends to the boy, to react to the aural event. As the result of the shift of perceptual attention, the sergeant recognizes that the lieutenant made an inquiry. Then, the sergeant internally processes the inquiry.

As a result of considering both perceptual objects (the boy and the lieutenant) the sergeant turns from the boy and faces the lieutenant, answering, "There was an accident. This woman and her son came from the side street and our driver didn't see them." The Lieutenant continues by asking "Who is hurt?", and the sergeant replies "The boy and our driver." Now when the Lieutenant asks "How bad is the boy hurt?," rather than answering directly, the sergeant defers to the medic, who has better knowledge of such things, and directs him to answer, by looking at him and calling his name "Tucci?"

Looking up the medic answers. "The boy has critical injuries. Sir we need to get a medevac in here ASAP."

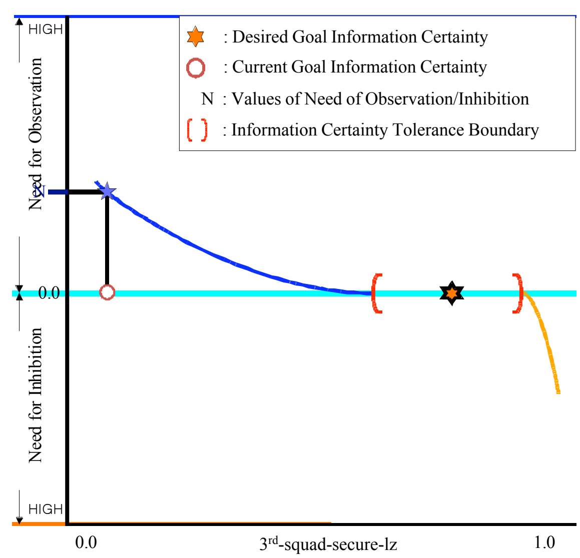
The lieutenant decides to call for the medical evacuation helicopter as requested and secures the local area. Then the lieutenant commands the sergeant to execute the task of setting up a landing zone (LZ) so that the helicopter can safely land.

When the sergeant starts executing the secure-landing zone task, the sergeant contacts the 3rd squad in order to dispatch the squad to the LZ. The interaction between the information certainty of position of the 3rd squad and the benefit of observing the information is shown later with graphs and trees. The sergeant initially knows the position of the 3rd squad and the location of the LZ. When the sergeant contemplates execution of the task (secure-lz), he tries to gain high-level information certainty of the spatial information of the 3rd squad, who will be dispatched to the LZ to secure it. The information graph on the task is shown as follows:

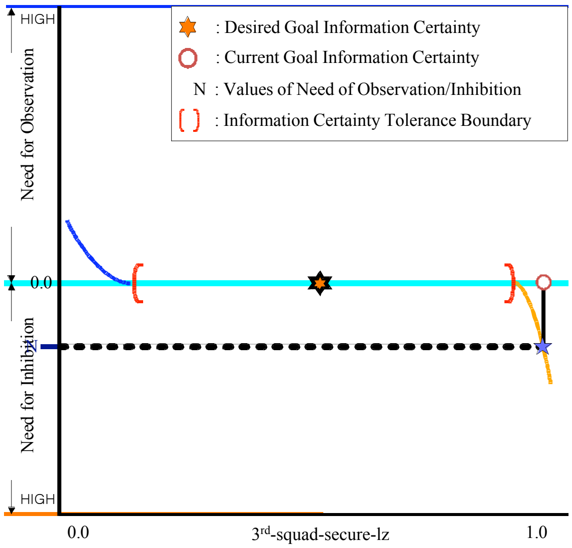



Since the squad is not in the area for securing the LZ, the sergeant's DPA module determines there is a benefit derived by observing the current spatial information of the squad.

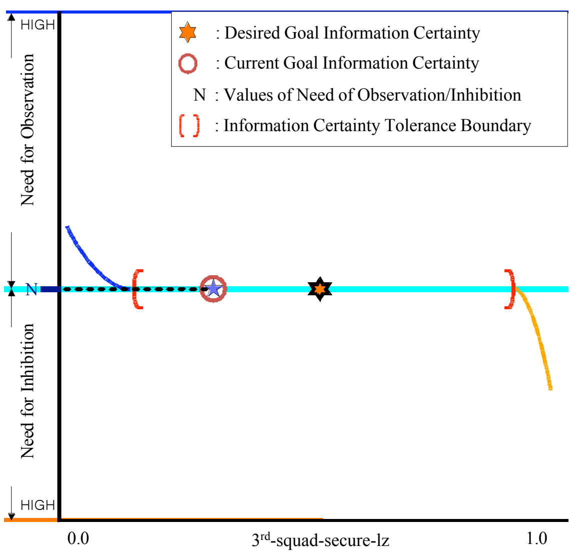
Next, the sergeant contacts and then commands the squad forward to the LZ. After he observes that the squad is moving toward the LZ, he reduces the slant of the curve since he gets hopes of achieving the task that may be given from the emotion module. The changed information graph on the task is shown as follows:



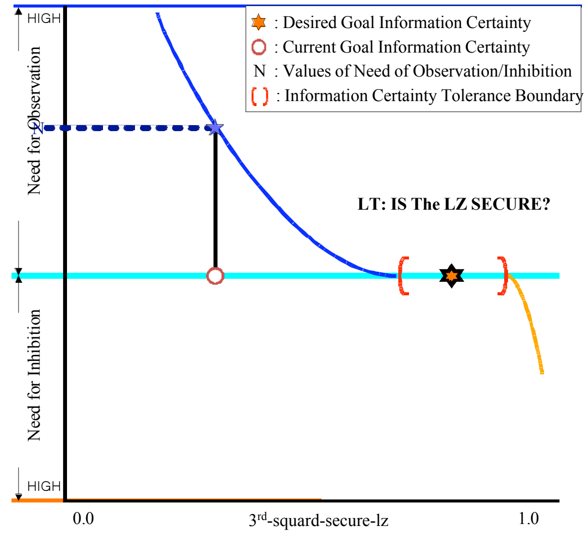

After the LZ is secured by the squad whom the sergeant highly trusts, he does not need to maintain the information status that the LZ is secure with one hundred percent certainty but can lower the priority of the information (i.e., lower the desired certainty of information and expand the tolerance boundary). With this shift in priority, allows the sergeant to observe, search, or track other information. The changed information graph on the task is shown as follows:




While the sergeant observes, searches, or tracks other objects, the certainty of the information of the security of the LZ will gradually decrease. The changed information tree on the task is shown as follows:




Next, the LT asks the sergeant, "Is the LZ secure?" This speech event increases the desired certainty of information and makes the tolerance boundary narrow since the sergeant wants to be very sure of the information that he will convey to his superior officer, the LT. The changed information graph and tree on the task is shown as follows:




This speech event changes the benefit level by the changing the attributes of the tuple(3$^{rd}$-squad-secure-lz). This, in turn, affects the sergeant's emotional state by increasing the degree of distress, suggests that he should update the belief of the status of security of the LZ. As the result, the sergeant gazes at the landing zone to determine whether it is still secured by the dispatched squad members, and he responds with the status of the landing zone to the lieutenant.

This example illustrates the importance of gaze in acquiring perceptual information and monitoring task performance while embedded in the social context of conversation. Our aim is to have the sergeant's behavior seem appropriate within this context, both in terms of behaving human-like and using perceptual gaze to mediate between costs and benefits of information updating actions.

## 5. Related Work

There are few models or frameworks that address the issue of where and what an observer should look at in a given time. Findlay and Walker (Findlay and Walker, 1999) present a comprehensive psychological model of the information flow routes and competitive pathways in saccade generation. Their model has not been implemented as a computational system yet, but it served as a source of inspiration for aspects of the work described in this paper.

There are a number of comprehensive computational models of perceptual attention for virtual humans. Chopra-Khullar and Badler (Chopra-Khullar and Badler, 2001) built one of the most extensive models to date, a psychologically motivated framework for generating the visual attending behaviors of an animated human figure. Their implementation generates believable animation behaviors for a virtual human performing a fairly scripted set of tasks, but it is not clear how the model would fare in a much more dynamic environment where the need to react to events in the world is much higher than the virtual world they describe. The model appears to fall into the top-down attention category, where gaze behaviors are scheduled and placed in a queue.

Cassell and Vilhjalmsson (Cassell and Vilhjalmsson, 1999) have used gaze as an important communicative behavior in their animated characters. Their animated characters have several limitations: (1) the model does not operate in real time, (2) the model only includes conversational gaze, and (3) the model does not include variability due to emotional state or individual differences.

Rickel and Johnson (Rickel and Johnson, 1999) also employ gaze in their tutoring agent, STEVE, who looks at the student during conversational interaction, and looks at objects in the environment when performing tasks or monitoring the student. Their main purpose of adopting eye movements into agents is to generate eye movements for non-verbal communication (e.g. turn-taking) that are controlled by top-down attention. The general limitation of STEVE is that a gaze command typically comes at the beginning of a cognitive activity, but is not updated during that activity. So, for example, if STEVE starts talking to a person, he gazes at them. Then, if his attention is drawn to an action in the environment, he will remain gazing at that action until something else causes a gaze command.

Hill (Hill, 1999, 2000) applied a simulation of attention for a virtual helicopter pilot. The virtual helicopter pilot selectively draw attention to an object(s)/area(s) based on features of objects and their priority to tasks, and perceptual grouping of objects. However, the helicopter pilot has no animation of head and eye movements. We extended Hill's model of perceptual resolution based on psychological theories of human perception.

## 6. Relationship to Social Attention

While the model of perceptual attention presented above handles many aspects of gaze behavior, there is another factor in the broader scope of attention. Information certainty is just one of the motivations for gaze, but information can be acquired through other means than gaze, and gaze can be used for more than acquiring information. In this section, we describe how these features can be added to the perceptual attention model, presented above, for a more complete model of gaze and attention in virtual humans.

In a social setting, it is often important to use gaze to regulate the flow of conversation, including signals of turn-taking, and feedback. Some of this can be modeled directly as a 'concern' for information certainty, such as needing to look at an addressee while speaking to get information about whether that addressee is listening, understands, and agrees. Likewise, looking away from an addressee while planning speech could perhaps be modeled as inhibition of this feedback information when more cognitive facilities are needed for planning the utterance. Some other factors are less easily modeled as relating to information, however. An alternative reason for gaze aversion by a speaker is that it makes it harder for an addressee to take the turn by speaking. Gaze also can be used as a form of non-verbal communication, e.g., to direct the gaze of others to an object, even when one does not need more information oneself. Another issue is that inappropriate gaze or aversion can send undesired signals about the attention and respect of the speaker or addressee – this will need to be figured in to the cost model.

In a social situation, perceptual attention may sometimes interact with social and conversational attention. In the example given in Section 4, above, we already described how a question about a proposition can change the desired certainty of information. Conversation can also be used to affect the actual certainty. For example, rather than looking at the landing zone, the Sgt might instead radio to the squad and ask them about the security. We then have three means of monitoring: covert monitoring though memory and inference about future projection, overt perceptual gaze, and social monitoring through (perhaps prompted) reports of other agents. It may be difficult to arbitrate between these sources of information when they conflict. For instance, one may remember the landing zone as secure and have no reason for thinking it will change. On the other hand, a verbal status report may conflict with this. If trust is sufficiently high in the report's certainty (and the trustworthiness of the reporter), one may choose to override the covert monitoring with this information. Another option is to motivate a new gaze to arbitrate.

Social means may also change the relative costs of perceptual gaze shifts. For instance, it may be very rude to look away when someone is speaking (high cost of shift), yet it may be very important to attend to an unexpected or potentially dangerous event (high benefit to shift). With high utility on either end, the choice may be difficult, and moreover potentially very costly either way. One way around this is to reduce the cost of the shift with a social action, such as apologizing, or using a non-verbal gesture indicating that the speaker should wait a moment.

## 7. Discussion and Future Work

The proposed computational model for controlling the focus of perceptual attention for virtual humans provides the potential to support multi-party dialogues in a virtual world. As we begin to integrate perceptual attention into multi-party, multi-conversational dialogue layers (Traum and Rickel, 2002), we have demonstrated that virtual humans can respond dynamically to events that are not relevant to the tasks and shift their attention among objects in the environment and have gotten positive feedback to informal demonstrations. The model we have described here is still a prototype that has to be tuned and tested in a wider range circumstances. In particular, by integrating more robust and deliberate language tasks with the model we have described in this paper, we believe we have made progress toward natural gaze behaviors in embodied conversational agents. In addition, by integrating the concept of measuring the salience of a specific class of spatial features with the model, we believe that this model will provide a large potential for generating more reactive and realistic bottom-up attention.

## 8. Acknowledgment

The project or effort described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

## 9. References

J. Cassell and H. Vilhjalmsson: "Fully Conversational Avatars: Making Communicative Behaviors" *Autonomous Agents and Multi-Agent Systems,* 2:45-64, Kluwer Academic Publishers, 1999.

S. Chopra-Khullar and N. Badler: "Where to Look? Automating Attending Behaviors of Visual Human Characters" *Autonomous Agents and Multi-Agent Systems*, 4(1-2), pp.9-23, 2001.

T. Conde and D. Thalmann, An Artificial Life Environment for Autonomous Virtual Agents with multi-sensorial and multi-perceptive features , *Computer Animation and Virtual Worlds* , Volume 15, Issue 3-4, John Wiley, 2004

N. Courty, E. Marchand, and B. Arnaldi: "A New Application for Saliency Maps: Synthetic Vision of Autonomous Actors", *IEEE Int. Conf. on Image Processing, ICIP'03*, Barcelona, Spain, Sep. 2003.

M. Garau, M. Slater, S. Bee, and M. A. Sasse, "The impact of eye gaze on communication using humanoid avatars", *ACM SIGCHI*, 2001.

M. Gillies and D. Neil: "Eye Movements and Attention for Behavioural Animation", in *The Journal of Visualization and Computer Animation*. 13: pp 287-300 2002

R. Hill: "Modeling Attention in Virtual Humans" *Proceedings of the 8th Conference on Computer Generated Forces and Behavioral Representation*, SISO, Orlando, Fla., 1999

R. Hill: "Perceptual Attention in Virtual Humans: Toward Realistic and Believable Gaze Behaviors" *Proceedings of the AAAI Fall Symposium on Simulating Human Agents*, pp.46-52, AAAI Press, Menlo Park, Calif., 2000.

R. Hill, J. Gratch, S. Marsella, J. Rickel, W. Swartout, and D. Traum: Virtual Humans in the Mission Rehearsal Exercise System. *Künstliche Intelligenz* (KI Journal). Special issue on Embodied Conversational Agents, 2003.

L. Itti and C. Koch, "Computational Modeling of Visual Attention", *Nature Reviews Neuroscience,* Vol. **2**, No. 3, pp. 194-203, Mar 2001

C. Nothegger, S. Winter, and M. Raubal, "Selection of Salient Features for Route Directions". *Spatial Cognition and Computation* 4(2): 113-136, 2004.

R. Klein, "Inhibition of return", *Trends in Cognitive Sciences*, 4, 138–147. 2000.

N. Moray: "Designing for attention" *Attention: Selection, Awareness, & Control, Oxford Press*, 1993.

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

J. Rickel and W. Lewis Johnson: "Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control" *Applied Artificial Intelligence*, 13:343-382, 1999

W. Swartout, R. Hill, J. Gratch, W.L. Johnson, C. Kyriakakis, C. LaBore, R. Lindheim, S. Marsella, D. Miraglia, B. Moore, J. Morie, J. Rickel, M. ThiÚbaux, L. Tuch, R. Whitney and J. Douglas: "Toward the holodeck: Integrating graphics, sound, character and story" *In Proceedings of 5th International Conference on Autonomous Agents*, 2001.

D. Traum and J. Rickel: "Embodied Agents for Multi-party Dialogue in Immersive Virtual Worlds", *AAMAS'02*, July 15-19, 2002, Bologna, Italy.

J. Wolfe: "Guided Search 2.0: A revised model of visual search" *Psychonomic Bulletin & Review*, 1 (2), pp.202-238, 1994.

## Author Biographies

**Youngjun Kim** is a Ph.D. candidacy at University of Southern California and works with Dr. Randall W. Hill, Jr. at ICT. His research focuses on modeling perceptual attention for virtual humans

**Randall W. Hill, Jr.** is the Director of Applied Research and Transition at ICT and a research assistant professor of Computer Science at University of Southern California. He earned his M.S. and Ph.D. degrees in computer science from the University of Southern California in 1987 and 1993, respectively. His doctoral thesis showed how intelligent tutoring systems can be used in interactive simulations to assist students in learning how to operate complex equipment.

**David R. Traum** is a Research Scientist at ICT and a research assistant professor of Computer Science at University of Southern California. He completed his PhD in Computer science at University of Rochester in 1994. His research focuses on collaboration and dialogue communication between agents, including both human and artificial agents.