Controlling the Focus of Perceptual Attention in Embodied Conversational Agents

Youngjun Kim
Institute for Creative Technologies
13274 Fiji Way, Suite 600
Marina del Rey, CA 90292
yjkim@ict.usc.edu

Randall W. Hill, Jr.
Institute for Creative Technologies
13274 Fiji Way, Suite 600
Marina del Rey. CA 90292
hill@ict.usc.edu

David R. Traum
Institute for Creative Technologies
13274 Fiji Way, Suite 600
Marina del Rey. CA 90292
traum@ict.usc.edu

1. INTRODUCTION

In this paper, we present a computational model of dynamic perceptual attention for virtual humans. The computational models of perceptual attention that we surveyed fell into one of two camps: top-down and bottom-up. Biologically inspired computational models [2] typically focus on the bottom-up aspects of attention, while most virtual humans [1,3,7] implement a top-down form of attention. Bottom-up attention models only consider the sensory information without taking into consideration the saliency based on tasks or goals. As a result, the outcome of a purely bottom-up model will not consistently match the behavior of real humans in certain situations. 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. 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. This model extends the prior model [7] with perceptual resolution based on psychological theories of human perception [4]. This model 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 [5].

2. COMPUTATIONAL MODEL OF DYNAMIC PERCEPTUAL ATTENTION

We have developed a model called Dynamic Perceptual Attention (DPA) to compute object salience and to control gaze behaviors. 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 about the selected object. Our embodied agent dynamically decides where to look, which object to look for, and how long to attend to the object.

2.1 Decision-Theoretic Control

One of the consequences of modeling perception with limited sensory inputs is that it creates uncertainty about 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. Top-down and bottom-up perceptual processes provide information to the DPA module in the form of tuples composed as follows:

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

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 $tuple_i$

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). 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 and social costs 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.

2.2 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* [3], 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 1.

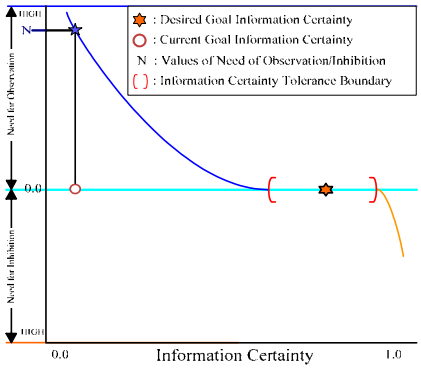

Figure 1. Information Certainty and Need

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 about the target from top-down and bottom-up processes, it sets objCGI to 0. After the values for objCGI and ICTB are set, the virtual human computes the NEED for observation or for inhibition on each tuple as follows:

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

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

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$_i$)*.

2.3 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).

2.4 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 attention, the virtual human will maintain its 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.

3. CONCLUSION AND DISCUSSION

The proposed computational model of controlling the focus of perceptual attention for embodied agents 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 [7], we have demonstrated that embodied agents can respond dynamically to events that are not even relevant to the tasks and shift their attention among objects in the environment.

4. ACKNOWLEDGMENT

The project 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.

5. REFERENCES

[1] 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

[2] 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.

[3] 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.

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

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

[6] Klein, R. Inhibition of return. *Trends in Cognitive Sciences*, 4:138–147, 2000.

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