A Perception and Selective Attention System for Synthetic Creatures

Daniel Torres[1] and Pierre Boulanger[2]

[1] Department of Computing Science, University of Alberta
T6G 2E8 Edmonton, Alberta. Canada
dtorres@cs.ualberta.ca, http://www.cs.ualberta.ca/~dtorres
[2] Advanced Man-Machine Interface Laboratory, Department of Computing Science, University of Alberta
T6G 2E8 Edmonton, Alberta. Canada
pierreb@cs.ualberta.ca

Abstract. This paper describes the selective attention and perception system designed for the synthetic creatures of the ANIMUS[1] project. This system allows agents to evaluate objects, other characters, properties, and events happening in their virtual environment. Perceived data is prioritized to direct the attention of the character toward salient stimuli, providing means for action-reaction and conditioned learning, automatic triggering of reflexive responses to certain stimuli, and further optimization of the cognitive processes by creating a more natural and realistic interface. The system can be customized to fit the artistic and conceptual constraints of each character, facilitating the creation of synthetic creatures with complex behavior and personality.

1 Introduction

Creating "believable" characters for interactive environments is a demanding task both in terms of art and technology. A "believable" character appears to be alive, giving the illusion of having its own thoughts, emotions, intention and personality. Thomas and Johnston[2] say:

> Consider this: a strong situation has been established. The character comes into it with a definite and interesting attitude. Confronted with the problem, he develops his own personality, grows a little, shows who he is, makes his decisions for action...(p. 375)

To support a virtual creature with such life-like behavior, we must maintain a robust and consistent system of emotions, goals and beliefs; make decisions in a changing environment; and take the character to express itself effectively to the user while remaining within the bounds of its role and personality. An acceptable perception system supports the character with all the necessary information about the world and itself in order to facilitate these tasks. Interactivity is greatly hindered without some form of perception. The quality and flexibility of such interactivity, and the performance of the character's inner systems depends on the

A. Butz et al. (Eds.): Smart Graphics 2003, LNCS 2733, pp. 141–150, 2003.

ability of the perception system to convey information in a way that facilitates higher level processes. Evaluating perceived information is an expensive process. It is often needed to measure the impact and significance of the world over the agent's plans and objectives, before a response is generated. Our proposed perception system simplifies this task by means of an *early* discrimination method that prioritizes and highlights interesting stimuli, performs conditioned learning and creates primitive reactive behavior without having to analyze the meaning of sensory data. Broad agents and synthetic characters have become widely used in virtual reality simulations, video games, human training programs[3,4], therapeutic tools, support and help systems[5], virtual drama[6,7,10], and artistic projects. A flexible, customizable and robust perception system provides a solid start for the creation of characters that provoke the illusion of being alive. The rest of this paper is organized as follows: In section 2 we review some previous work. In section 3, we briefly describe the architecture of an ANIMUS system to provide a context for the perception layer. In section 4 we detail our virtual domain and its characteristics. We also define some basic concepts. The detailed description of the perception architecture, its components, and functionality is given in section 5. Section 5.6 illustrates the perception system with an example.

2 Previous Work

Several research groups are working with broad agents. We are interested in those that incorporate some model of personality, goals and emotions (such systems require more sophisticated perception methods than those which only show reactive behavior). An interesting example is C4, developed by Isla[8,9] and the MIT Media Lab: The perception system sees the world through a set of *percepts* (atomical classification and data extraction units) kept in a *percept tree*. When a stimulus is sensed, percepts are used to determine what and how much is known about it. New percepts can be created, and the existing ones refined as needed by the character, each one having its own specialized algorithm to deal with the perceived information. Bates[10] presents the Oz Project, in which agents sense the world through data objects that represent property-value pairs. Sensed data can be incomplete, incorrect, or absent. In Kline[11], perception is done by physical sensors (heat, proximity, pressure), synthetic vision (detecting objects and their features), and direct sensing (by querying the world). Sensors filter the information coming from the world or other sensors, each one reacting to a spectrum of the sensed data, while other systems get the output and construct specific numerical information that can be interpreted by higher rational systems. An example of perception systems for real robots is found in Velázquez[12].

3 The ANIMUS Project

Our perception model is part of a bigger project called ANIMUS. Here we briefly mention its basic architecture to illustrate the context in which perception is done and how it interacts with other parts of our synthetic characters. An ANIMUS creature is a synthetic character composed of three main layers, namely **Perception** layer (described in this paper), **Cognition** layer, and **Expression** layer. The Cognition layer is a set of high-level systems that allow the character to have goals, emotions, personality, and social and rational behavior. This layer processes all information coming from the perception layer and orchestrates the actions of the expression layer. The Expression layer is a **Pose-based** animation engine similar to the ones in Isla[8] and Downie[13]. It renders a 3D graphic representation of the character and creates animations in real time to express the inner states of the creature. The Perception layer has direct connections to both the Cognition and Expression layers. An ANIMUS character can be human, animal, or otherwise.

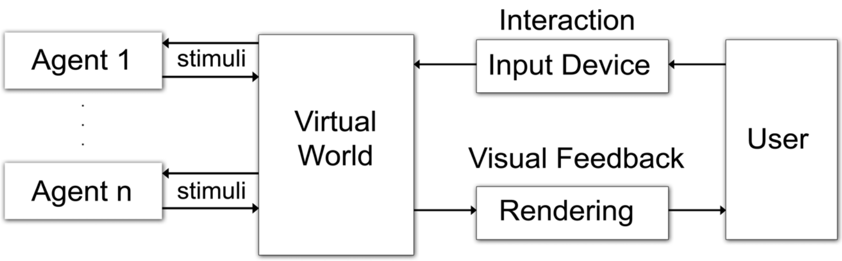

Fig. 1. The basic layout of the Virtual Environment

4 The Virtual Environment

Basic Definitions. The basic layout of the Virtual Environment can be seen in Fig. 1. It contains a number of *Virtual Characters*, the *Virtual World Agent* and an optional avatar representation of the *human user(s)*. Virtual characters are ANIMUS creatures as described in Section 3. The basic information primitive is called *feature*, a concept-value pair similar to what we find in Bates[10], representing something that can be sensed by a character. Features range from physical properties like color, weight, temperature and volume intensity, to more complex concepts like pain, hunger, and fatigue. When a feature is produced by a character it is called a *stimulus*. A *stimuli* is the set of stimulus expressed by a character at a given time.

Information Exchange. The *Virtual World* is an agent that keeps track of all existing characters and administrates a *blackboard* with the set of all stimuli at time t. It also synchronizes the activity of the characters by calling them to update their perception system a certain number of times per second. The Virtual Environment has any number of characters interacting in real time between themselves and the human users. When a character needs to perceive its environment, stimuli is obtained from the blackboard or by directly querying other characters. The perceived information usually contains noise or transformations induced by the Virtual World in order to make it *relative* to the character[1]. One stimuli is updated by each character and posted in the blackboard for other characters to see. This perception-reaction-expression information cycle repeats for the duration of the simulation.

5 A Selective Attention and Perception Layer

The objectives of the Perception Layer are:

1. To serve as an entry-point for the world, thus letting the character perceive its environment.
2. To segment, discriminate, and prioritize information to bias the attention of the creature toward salient stimuli.
3. To implement reflexes as an automatic response to certain stimuli
4. To perform low-level conditioned learning

The possibility of creating a selective attention model without having to analyze the meaning of perceived information was inspired from the concepts of *early* and *late* selection of human perceptual processing[18]. Early selection suggests that attention can be directed before the perceived information is semantically processed, by blocking some stimuli and preserving others. Late selection states that attention comes after the correct evaluation of information by high-level cognition systems. It was later proposed that early selection would not completely block certain stimuli, but only degrade or attenuate it before it would reach higher levels of cognition. Our perception layer is an early selection system. This creates faster response times and resource optimization since higher AI evaluation algorithms only process important information, while the rest of the stimuli remains available for late selection.

5.1 System Architecture

The modules of the Perception System are illustrated in Fig. 2. At each perception frame, the input to the Perception layer is a set of stimuli from other agents and the world. The output is a set of *prioritized* stimuli to the high level cognition system of the character. The *Reflex Agents* module can also trigger automatic reflex responses as a reaction to the perceived information.

[1] This principle is called *sensory honesty*. If two characters are on the same side of a wall they are able to see each other, but if one is placed at the other side, they will not be able to know their position, even though it is contained on the blackboard.

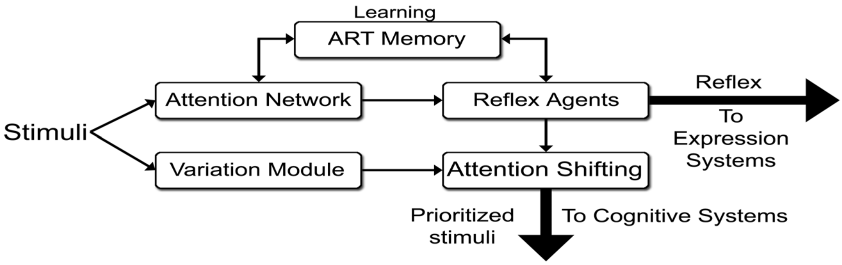

Fig. 2. The architecture of the Perception Layer

5.2 The Attention Network

The functionality of the Attention Network is inspired by Koch[16]. Each neuron contains an input weight w_i where $0 \leq w \leq 1$, and a threshold θ_i initially set to 0, and responds to a particular stimulus that may or may not be contained in the stimuli set. When a stimuli enters the network, the net input for a matching neuron is $v_{i_t} = x_{i_t} w_i$, where x_{i_t} is the magnitude of stimulus i at time t, and the output is an excitation level determined by $I_i = (v_{i_t}, \theta_i)$ (Usually a sigmoid function) if $v_{i_t} < \theta_i$ and $I_i = 1$ otherwise. There are features that can only be produced as a combination of other features (like pain, for example), thus the excitation output of some neurons is linked to other neurons, creating the network. Both the input weight and the threshold of each individual neuron changes as a function of entrant stimuli and time, giving the neuron short-term memory characteristics. A learning rate $\eta_i, 0 \leq \eta \leq 1$ controls how fast the threshold adapts to the magnitude of the stimulus, and the function $\gamma(w_{i_t}, \sigma_i)$, usually a parabolic or sigmoid decay, makes the input weight converge to 0 in σ_i frames if the input remains constant or less than the current threshold. The output of the Attention Network is an **attention stimuli** composed of the excitation magnitude of each triggered neuron. The careful selection of attention neurons and network design determines the perception capabilities of the creature. One basically controls *what* can be sensed and *how* sensible this information is. Like in real biological systems, some creatures will be able to note features ignored by others, creating a more realistic behavior and a more flexible tool for building original characters.

5.3 The Variation Module

The Variation Module is a set of variation units based on the *Inhibition of Return*[18] effect observed in human perception, which suggests that recently attended stimuli is inhibited over time. The perception of a given stimulus is inhibited when its variation is greater than a certain frequency factor[2]. We say that a stimulus produced change from one perception frame to the next one

[2] Such factor must be greater than 1 and smaller than the current perception rate.

if the absolute difference of their magnitudes is less than a threshold value. The output of a variation unit is 1 if variation is detected for its corresponding stimuli and the current variation rate is less than the frequency factor, and 0 otherwise. This inhibits interest in strongly repetitive phenomena like an object oscillating with constant frequency, while biasing attention to more unique events like a lightbulb that suddenly goes off. The output of the Variation Module is a **variation stimuli** composed of the magnitudes of all variation units triggered by the perceived stimuli.

5.4 Reflex Triggering

The Reflex Module is a set of *Reflex agents*. Each reflex agent is an independent unit containing two elements: a reflex stimuli and a reflex function. The reflex stimuli is a purposely chosen combination of features with attention magnitudes, and the reflex function is a unique procedure in terms of the features in the reflex stimuli. When a reflex agent finds a close correspondence between its reflex stimuli and some attention stimuli coming from the Attention Network, the original magnitudes of the perceived stimuli become the parameters of its reflex function, generating a carefuly designed reflex. Each agent handles the creation and behavior of such reflex in its own particular way, allowing flexibility and originallity between characters. In Fig. 3, an ANIMUS character perceives bullets fired by the user. When a bullet is close to its body, an evasive reflex agent appraises speed and proximity to move the body away from the bullet. This animation demonstrates that by simply using reflex agents we can create characters with interesting reactive behavior.

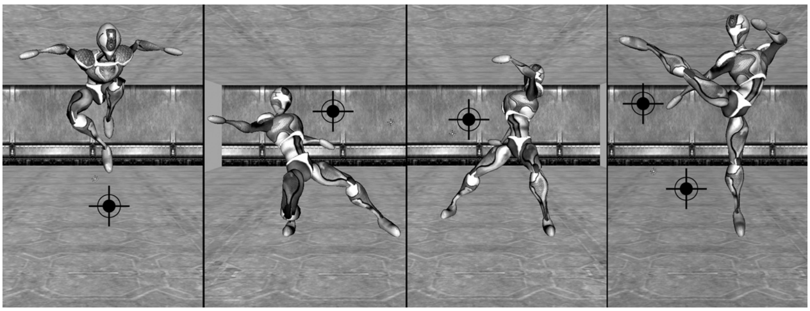

Fig. 3. An ANIMUS character with a simple reflex agent that moves the body away from bullets

5.5 Learning

The Conditioned Memory module is a set of ART units. For each reflex agent that has triggered a response, there exists a unique ART model that synthesizes a sample of the stimuli at the time the reflex was produced. This module is based on work by Mumford[14] and follows the suggestion that events in the environment are not arbitrarily produced but are due to hidden variables. Synthesizing coded signals for the creation of comparison points in a feed-forward, feed-backward structure helps to create knowledge about such hidden variables and hopefully teach the character to react in response to certain stimuli.

To explain the functionality of this module consider the case of *Pavlov's dog.* Pavlov discovered that dogs could be trained to salivate in response to the sound of a bell. The training consisted on simultaneously presenting the stimuli of a bell and a piece of meat to the dog. After some time, the dog relates the sound of the bell to the appearance of the meat and salivates in preparation for the feast. Similarly, our virtual dog perceives the meat and its features, triggering the salivation reflex. The conditioned memory keeps a synthetic version of the stimuli (containing both the signals from the food and the bell). After some reinforcement (done by repeating the scenario to the dog) the weight of the synthesized stimuli is enough to elicit the salivation reflex with only the presence of the bell. This happens because the bell's stimuli matches, to some extent, the template contained in the ART model of the salivation agent. In other words, the virtual dog "suspects" they are related and therefore triggers the agent. When the food appears to confirm this relation, the ART model gets updated and the bell's stimuli creates a closer match next time it is perceived.

The Attention and Variation stimuli are analyzed in the **Attention Shifting** module in order to prioritize the perceived stimuli before it is presented to the high lever cognition systems.

5.6 Alebrije: A Working Example

To demostrate the simplicity and functionality of the Perception Layer we have created a simulation called **Alebrije**. It consists of a main character (a lizard-like creature named *Alebrije*[3]) and a number of *Noise Insects*, whose only function is to emit noise as they are moved by the user or the computer. Alebrije has a perception network that responds to noise, and a reflex agent that makes him turn and face whatever stimulates his attention. The only behavior policy is that Alebrije stays awake as long as something interesting is happening and otherwise goes to sleep. Once the scenario is set the user can create any number of noise insects and move them in the virtual world. Alebrije is free to react to those who best stimulate his attention. For this experiment we use 2 noise insects: one oscillates in a cyclic fashion by the computer, while the other is controlled by the user. The perception rate is set to 20 samples per second, variation threshold is 4 samples/sec, and the decay for the noise-sensible neuron is set to 30 samples. Fig. 4 shows the noise magnitude along 350 frames as it is created by both insects. The dotted line belongs to the automated noise insect and the continuous line is the user-controlled insect. We will refer to the computer insect as $C1$, and the

[3] Alebrije is a term used by some Mexican artisans to describe a particular kind of imaginary creature.

human controlled insect as $C2$. At this time, the ART memory module is not completely implemented so this experiment will not reflect its features.

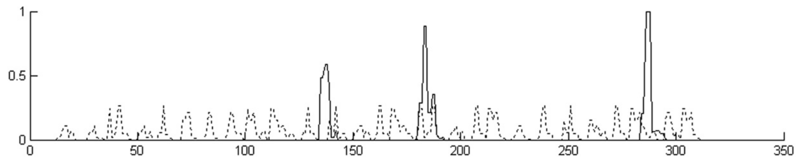

Fig. 4. The unmodified noise stimuli created by the insects.

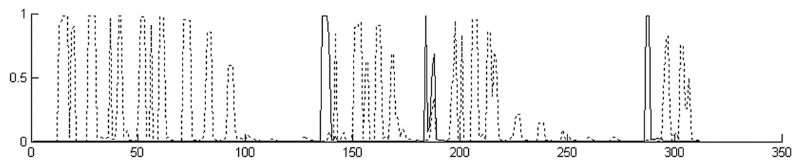

Fig. 5. Reaction to both stimuli by Alebrije's Attention Network

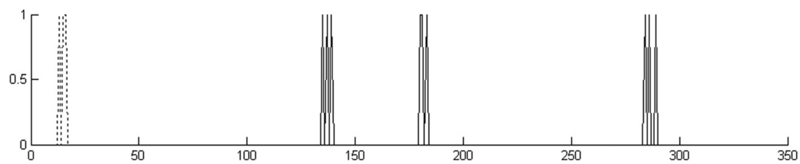

Fig. 6. Reaction to both stimli by Alebrije's Variation Module

When $C1$ begins to move, a stimuli with its noise magnitude is perceived by Alebrije's Attention Network. The noise neuron is excited with an output of aproximately 1 since its initial threshold is 0. The Variation Module detects the change and also produces 1. The reflex stimuli of the "stare" reflex agent is set to trigger when $noise_t > 0.7$, so Alebrije wakes up and stares at $C1$. At this point, w_{noise} begins to decay and θ_{noise} becomes equal to the perceived noise. As $C1$ keeps changing more than 4 times per second for the rest of the simulation, the Variation Module output quickly becomes 0 for this creature and never finds occasion to change. The cyclic peaks of noise produced by $C1$ take Alebrije's noise neuron to produce the output we observe in Fig. 5 until the weight decay causes the neuron to *adapt* to this pattern and its output converges to 0 after frame 100. At this point Alebrije looses interest and goes back to sleep. $C2$ joins the simulation at frame 135 with a noise magnitude two times greater than θ_{noise} (which had adapted to the peak noise of $C1$). The noise neuron reacts and outputs 1, adjusting its w and θ values. The Variation Module also recognizes the change and outputs 1 for $C2$, but remains in 0 for $C1$. Alebrije wakes up and turns to face $C2$, which immediately stops moving. Alebrije then turns to $C1$ since its noise magnitude is higher in comparison to $C2$ (which is now static) but around frame 170 the noise neuron adapts again and Alebrije goes to sleep. This same event repeats two more times at frame 178 and 284. Since $C2$ has a lower variation rate, the Variation Module outputs 1 every time it moves. If at frame 284 (when $C2$ moves for the last time) $C1$ and $C2$ had produced the same magnitude, the variation stimuli would have broken the tie favoring $C2$, taking Alebrije to consider it *more interesting*.

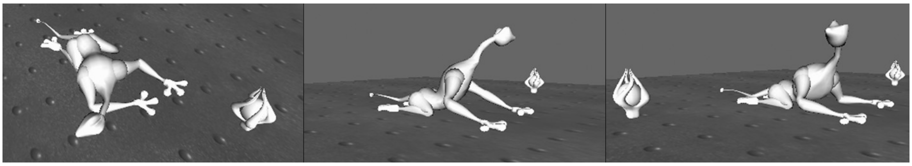

Fig. 7. The Alebrije character while sleeping (left), looking at computer's insect (center) and turning to face the user's insect(right)

6 Conclusions and Future Work

We have briefly described the ANIMUS architecture in order to provide a context for the Selective Attention and Perception System. We defined the basic units of information involved in the Virtual Environment, as well as the dynamics of the data flow and the different modules used to allow communication between characters and the user. The general architecture of the Perception System was explained and each one of its components analyzed in depth. A sample scenario was described along with its configuration and dynamics to show the system in a working case. We are pleased with the obtained results from our Perception System and believe it successfuly accomplishes its original requirements. As the Conditioned Learning Module is still not completely implemented, future work will include finishing this module and preparing test scenarios to stress the features of the system. There are many configuration variables that have to be manually adjusted at this time to get the optimal results in different characters, a group of templates will greatly save time and provide a starting point for the design of new synthetic creatures.

References

1. The ANIMUS Project: A Framework for the Creation of Emotional and Rational Social Agents. `http://www.cs.ualberta.ca/~dtorres/projects/animus` Computing Science Department, University of Alberta.
2. Thomas, F. and Johnston, O. Disney Animation: The Illusion of Life. Abbeville Press, New York. 1981.

3. Marsella, S. and Gratch, J. Modeling the Interplay of Emotions and Plans in Multi-agent Simulations. In Proceedings of the 23rd Annual Conference of the Cognitive Science Society, 2001.
4. Gratch, J. Modeling the Interplay Between Emotion and Decision-Making. In the 9th Conference on Computer Generated Forces and Behavioral Representation, 2000.
5. Andre, E. et al. Integrating Models of Personality and Emotions into Lifelike Characters. Proceedings of the workshop on Affect in Interactions - Towards a new Generation of Interfaces in conjunction with the 3rd i3 Annual Conference, pages 136–149, Siena, Italy, October 1999.
6. Mateas, M. An Oz-centric Review of Interactive Drama and Believable Agents. AI Today: Recent Trends and Developments. Lecture Notes in AI 1600. Berlin, NY: Springer. 1999.
7. Neal Reilly, W. Believable Social and Emotional Agents. Ph.D. Thesis. School of Computer Science, Carnegie Mellon University. USA. June 1996.
8. Isla, D. et al. A Layered Brain Architecture for Synthetic Creatures. IJCAI, Seattle, WA, August 2001
9. Isla, D. and Blumberg, B. New Challenges for Character-Based AI for Games. AAAI Spring Symposium on AI and Interactive Entertainment, Palo Alto, CA, March 2002
10. Bates, J. et al. Integrating Reactivity, Goals and Emotions in a Broad Agent. In 14th Annual Conference of the Cognitive Science Society, 1992.
11. Kline, C. and Blumberg B. The Art and Science of Synthetic Character Design. Proceedings of the AISB1999 Symposium on AI and Creativity in Entertainment and Visual Art, Edinburgh, Scotland, 1999.
12. Velázquez, J. When Robots Weep: Emotional Memories and Decision-Making. In AAAI Proceedings, 1998.
13. Downie, M. Behavior, Animation, Music: The Music and Movement of Synthetic Characters. M.Sc. Thesis, Media Arts and Sciences, Massachusetts Institute of Technology, USA. February 2001.
14. Mumford, D. Neuronal Architectures for Pattern-theoretic Problems. In Large-Scale Theories of the Brain, p. 125–152 MIT Press, Cambridge Massachusetts. USA. 1995.
15. Damasio, A. and Damasio H. Cortical Systems for Retrieval of Concrete Knowledge: The Convergence Zone Framework. In Large-Scale Theories of the Brain, p. 61–74 MIT Press, Cambridge Massachusetts. USA. 1995.
16. Koch, C. and Crick F. Some Further Ideas Regarding the Neuronal Basis of Awareness. In In Large-Scale Theories of the Brain, p. 93–109 MIT Press, Cambridge Massachusetts. USA. 1995.
17. Carpenter, G. and Grossberg, S. Adaptive Resonance Theory (ART). In The Handbook of Brain Theory and Neural Networks, p. 79–85. MIT Press, Cambridge, Massachusetts. USA. 1995.
18. Gazzaniga, M. et al. Cognitive Neuroscience, The Biology of The Mind. Norton, New York. 1998.