Do You See What Eyes See? Implementing Inattentional Blindness

Erdan Gu [1], Catherine Stocker [2], Norman I. Badler [1]

[1] Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, 19104-6389
{erdan, badler}@seas.upenn.edu
[2] Department of Psychology, University of Pennsylvania, Philadelphia, PA, 19104
{cstocker}@ sas.upenn.edu

Abstract. This paper presents a computational model of visual attention incorporating a cognitive imperfection known as inattentional blindness. We begin by presenting four factors that determine successful attention allocation: conspicuity, mental workload, expectation and capacity. We then propose a framework to study the effects of those factors on an unexpected object and conduct an experiment to measure the corresponding subjective awareness level. Finally, we discuss the application of a visual attention model for conversational agents.

1 Introduction

If an embodied (virtual) agent is expected to interact with humans in a shared real or virtual environment, it must have the cognitive ability to understand human visual attention and its limitations. Likewise, an embodied agent should possess human attention attributes so that its eyes and resultant body movements convey appropriate and humanly understandable behaviors. Suppressed or inappropriate eye movements can by themselves damage the communicative effectiveness of an embodied agent. Thus, in order to build convincing computational models of human behavior, one should have a thorough understanding of communication and interaction patterns of real people. Attention models may be the key to leading animated agents out of the "uncanny valley" where increasing visual accuracy, combined with lifeless eyes, results in a "ghoulish" appearance when animated.

As a first step to making the appearance of virtual agents more realistic, we are creating a model of human visual attention. The visual attention system has been proposed to employ two filters – bottom-up [10] [18] and top-down [9][4] – to limit visual processing to the most important information of the world. In our early work [7], we suggested a computational model that was unique because not only did it integrate both of these filters, but also combined 2D snapshots of the scene with 3D structural information. However, after extensive examination of the Psychology literature, we became aware of the many intricate shortcomings of human cognition, and recognized

the importance of incorporating *inadequacies* in processing as a means of making a simulated human agent more realistic.

Inattentional blindness [21], as the name implies, occurs when objects that are physically capable of being seen in fact go unnoticed. Inattentional blindness was chosen as the primary phenomenon to include in our framework for two reasons. First, evidence suggests that it mainly involves the attention system, rather than other cognitive structures such as memory or language [1]. Other prominent attentional deficits, such as change blindness, appear to be tied much closer to these additional cognitive structures [17]. Second, inattentional blindness is a robust feature of multi-modal attention and analogous paradigms, such as the "cocktail party effect", have been well documented in auditory attention [19]. Therefore, once this model is complete its future applications will not be restricted to the visual system, but can be extended into other realms of cognitive processing.

While it is commonly believed that an object requires only perceptible physical properties to be noticed in a scene, recent studies have found that people often miss very visible objects when they are preoccupied with an attentionally demanding task [20]. Mack and Rock coined the term inattentional blindness, and concluded that conscious perception is not possible without attention [12]. Green [6] attempted to classify all of the prominent features of the phenomenon, and suggested that there are four categories that these features fall into: *conspicuity*, *mental workload*, *expectation* and *capacity*. Through experimental testing, Most *et al.* [13] forged a link between attention capture and inattentional blindness, and revealed the single most important factor affecting the phenomenon, the attentional set. They also introduced the concept of different levels of attentional processing, which, in our work, is categorized as four stages of subject awareness [22]: *unnoticed*, *subliminal*, *non-reflective* and *semantic*.

In order to formulate a realistic attentional framework, we will examine attentional deficiency and inattentional blindness, while attempting to answer three questions:

1. What kinds of stimulus properties will influence the likelihood of missing the unexpected object or event?
2. What kinds of perceiver-controlled mechanisms decide what should be permitted into consciousness and what should be rejected?
3. How much, if any, of a scene do we perceive when we are not attending to it?

Theories and Experiment

First we define the four factors critical to inattentional blindness and describe how they are used in our experiment to study their effects on subjective awareness level. By questioning subjects who participated in our experiment, we hoped to determine quantitative descriptions of each parameter's individual and combined importance in attention allocation.

The Four Factors Model

Because cognitive resources are limited, attention acts as a filter to quickly examine sensory input and allow only a small subset of it through for complete processing. The rest of the input never reaches consciousness, so is left unnoticed and unremembered. It has been suggested that the attentional filter is affected by four factors [6]: conspicuity, mental workload, expectation and capacity.

Conspicuity

Conspicuity refers to an object's ability to grab attention, and can be divided into two distinct groups: sensory and cognitive conspicuity [20]. Sensory conspicuity refers to the physical or bottom-up properties of an object, such as contrast, size, location and movement. Cognitive conspicuity, on the other hand, reflects the personal and social relevance that an object contains. Face pop-out – the phenomenon where faces that are meaningful to a person are more likely to capture attention – is an example of cognitive conspicuity in visual attention capture.

Mental Workload

There is only a finite amount of attention available to be rationed to objects and events. Thus, items that require more attention decrease one's ability to allocate this limited resource to other objects. As tasks become more difficult they increase the mental workload of the subject and require more attention, increasing the likelihood that an unexpected event will go unnoticed. Similarly, as tasks become less difficult, they require less attention. An object requiring less mental processing with time is said to be habituated [6]. This will cause workload to decrease and allow for other objects in the scene to be attended to more readily. An example of habituation is learning to drive a car. While driving may begin as a very difficult task, as it becomes more ingrained in one's repertoire of abilities, it becomes less mentally taxing.

Expectation

While the habituation process slowly decreases workload levels for the entire scene with time, expectation quickly causes specific stimuli to gain more weight over time and trials. According to the Contingent-Capture Hypothesis [20], as items and properties of items become more expected they become part of an attentional set. This attentional set then informs a person what is important and relevant in a scene. Inattentional blindness occurs when certain items are expected so much that people ignore any others. The Contingent-Capture Hypothesis, and the attentional set's involvement in inattentional blindness, will be described in detail in the next section.

Capacity

Attentional capacity refers to the number of items and information that a person can attend to at a time. Variations in capacity are a result of the individual differences between people, but are also affected by a person's current mental state (fatigue), cognitive processes (habituation), and physiological state (drugs and alcohol) [6].

Our experiment and its parameters

Our study was based on a famous demonstration of inattentional blindness, "Gorillas in our midst" [16], which asked participants to count the number of times a basketball was passed among a group of people. During this activity, a individual in a gorilla costume walked into and through the scene. Rather remarkably, many subjects do not recall seeing anything unusual! In our variation (Fig. 1), subjects were assigned the task of counting the number of ball passes between images of human-like characters that we created in a virtual environment. During this time, an unexpected image passed through the scene and the event continued, undisturbed.

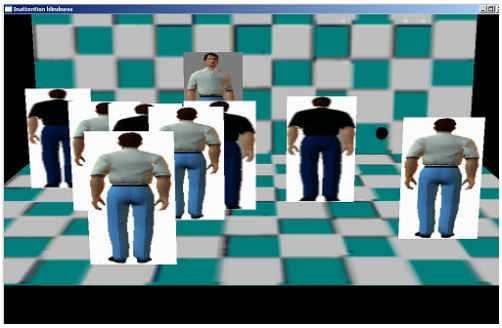


Fig.1 Example Frame of Animation Demo in the Experiment: Eight players (four in black T-shirts, four in white) move around the screen randomly while two 'balls': (one white, one black) bounce between them. Subjects were responsible for counting the number of passes made to the black T-shirt team using the black ball. A pass was considered to be completed when the ball hit the image, and the image 'jumped' Fifty seconds into the task an unexpected, face-forward, gray boxed character (the unexpected object) passed through the scene, but the players continued as normal. The task lasted a total of 90 seconds.

The four factors of inattentional blindness were measured by adjusting various parameters during the experiment. The appearance and movement of the objects contained in the scene, as well as the scene itself, were varied in order to affect the cognitive workload, sensory conspicuity and attentional set.

The first variation, the mental workload of the subject, could be high, medium, or low, determined by the speed that the balls moved and the amount of background clutter. A subject in a high mental workload group observed very fast moving balls and a cluttered (green and white checkered) background; the medium mental workload group saw medium speed moving balls and a cluttered background; the low mental workload group watched a slow moving ball and an uncluttered (all gray) background.

The sensory conspicuity of the unexpected object could also be varied: high, medium, or low, determined by the inherent physical salience of the unexpected object. Here, the saliency was dependent on the speed, as well as the trajectory that the unexpected object took. High sensory conspicuity groups were presented with an unexpected image that appeared and disappeared while moving quickly along the background of the scene. The unexpected object of the medium sensory conspicuity

group moved at a medium speed, in an irregular manner (beginning in the background, moving back-and-forth towards the foreground) across the screen. The low sensory conspicuity group received an unexpected object that moved at a slow speed in a straight line across the background of the scene.

Finally, the attentional set held by our subjects always contained the color black because they were attending to the black T-shirt group and tracking a black ball. What varied in the attentional set parameter is how similar the unexpected object's features were to the attentional set held by the subject, so the values were: matched, neither matched nor unmatched, or unmatched, according to the color of the unexpected object's T-shirt (black, maroon or white respectively). In Table 1, we list the variables in the experiment and their corresponding factors.

Table 1: Summary of the relationship between the four factors and the experimental parameters. It shows how the four factors interact with shown the attentional set and object properties

Factors		Definition	Parameters
Conspicuity	Sensory	Pop-out due to an object's inherent physical saliency in a scene.	Color & Intensity ·Contrast ·Opaqueness ·Environment ·Clutter ·Illumination Size Movement ·Velocity ·Trajectory
	Cognitive	Pop-out due to the perceiver's mental state and task relevance.	Personal Relevance ·Meaningful Face Pop-Out ·Familiarity
Workload		The amount of attention that the current item requires. Reduces probability of attention shift.	Difficulty Environment Habituation ·Time ·Trial
Expectation		The amount of attention an object receives varies according to a perceiver's beliefs about its relevance in the scene, due to past experience.	Attentional Set ·Task-specific features
Capacity		The total amount of attention available varies by individual	Individual differences Mental State

Computational Framework

Green's four-factor model specifies a theoretical set of parameters involved in inattentional blindness, while Most *et al.* provide the evidence for a detailed

progression from "ignored" to "part of consciousness." Our model integrates the two theories – attempting to retain the individual contribution of each – into a comprehensive theory of attention allocation (Fig. 2).

Dynamic internal representation of the world

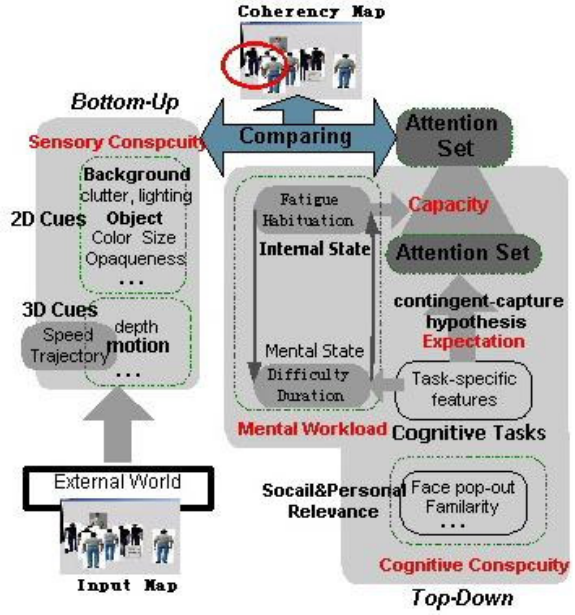


Fig.2 Block Diagram for computational framework. It illustrates the computational model of visual attention incorporating the four factors model and the contingent capture hypothesis.

Our attention capture framework relies on the cooperation of an internally-driven top-down setting and external bottom-up input. The bottom-up setting uses the "saliency" (sensory conspicuity features) of objects in the scene to filter perceptual information and compute an objective saliency map. Primary visual features such as color, contrast and motion are the features examined by this filter. Simultaneously, top-down settings, such as expectation and face pop-out determine the set of items that are contextually important, such as the attentional set, which is a subjective feature pool of task-prominent properties maintained in memory. At any moment, focused attention only provides a spatio-temporal coherence map for one object [15]. This coherence map highlights the object that has been calculated to be the most important at that moment in the scene, and can thus be used to drive the gaze of an embodied agent.

The final coherency map is created in three steps. First, a spatial coherency map is created, then it is augmented by temporal coherency and finally moderated by the attentional set. The spatial coherency map is computed by transforming a snapshot of the scene to the retinal field by a retinal filter. It is generally believed that the internal

mental image is built through non-uniform coding of the scene image. This coding is determined by the anatomical structure of the human retina, causing the image to appear very clear wherever the center of the retina is located, and increasingly blurry as distance from the center increases. In other words, whatever a person looks directly at will appear the most clear in their mental image, and objects will appear less clear the further they are from the in-focus object. Log-polar sampling [2] is employed as an approximation to the foveated representation of the visual system. The processing occurs rapidly (i.e., within a few hundred milliseconds) and in parallel across a 2D snapshot image of the scene. To allow real-time computation, interpolation between the partitions of receptive fields is implemented [8]. For each trial of our experiment, the size of the fixation field (the patch with the highest resolution) remained approximately constant since the distance from the subject to the screen, as well as the resolution of the animated demo, were fixed.

Once the spatial map is created, a temporal mapping highlights the direction of important movement. A final coherency map is generated by integrating these two maps and filtering the objects of interest using the attentional set.

The Contingent-Capture Hypothesis and the Attentional set

The attentional set, determined by subjective expectation, will further tune the generated spatio-temporal coherency map. The Contingent-Capture Hypothesis states that the only time that an object receives attention is when it, or properties of it, is contained in the attentional set held by the subject [5]. Most *et al*. expand on this theory, revealing that before an object can even be considered for attention, and thus compared to the attentional set, a transient orienting response to the object must occur. Consequently, the likelihood of noticing an unexpected object increases with the object's similarity to the currently attended object. In our animation demo, since the

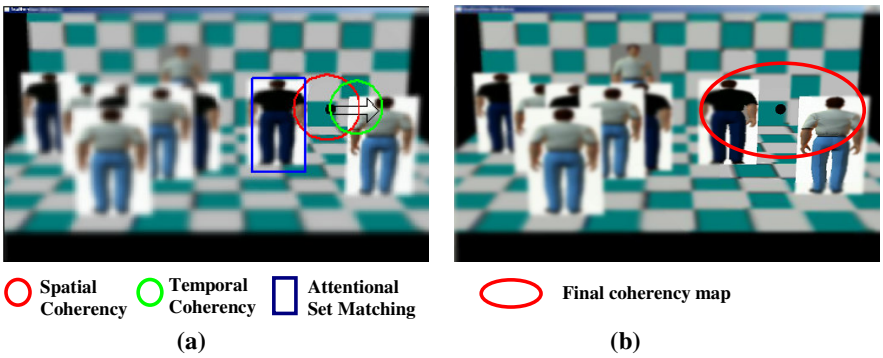


Fig. 3 Generation of Coherency Map. **(a):** Three influences of attention capture: spatial, temporal and attentional set. **(b):** The final coherency map, resulting from the combined effect of the three influences.

task was to count the number of times that the black ball hit the black T-shirt players, attentional set={black T-shirt people, black ball} would be warranted by the

Contingent-Capture Hypothesis. Fig. 3(a), demonstrates the three influences on the final coherency map. The red circle represents the spatial coherency map, the green circle denotes the temporal coherency, and the blue square reveals the object that matches the black color as well as the black T-shirt people held as a property of the attentional set. The red ellipse in Fig. 3(b) illustrates the readjusted coherency map that incorporates all three influences.

Subjective Awareness Level

Following completion of the task, participants filled out a questionnaire to determine if they noticed an unexpected object. To discover the level of processing that the object received, questions probed how well they perceived the object. Questions began by vaguely asking about anything unusual, and increased in specificity until subjects were asked to choose the unexpected image out of a line-up of eight.

We now introduce the concept of awareness level to describe the degrees of perceptual organization achieved by the visual system. At the lowest extreme is complete inattentional blindness – attentional resources failed to be allocated to the object resulting in a failure to notice it. At the opposite end is the highest level of consciousness, the semantic level, where the object is perceived as a figure-ground discrimination with meaning. In between the two extremes are the subliminal level and the non-reflective level. The subliminal level is represented by a subject's acknowledgement of the presence of the unexpected object, but no conscious awareness of any of its physical characteristics. Hence, important subliminal messages were transmitted for further processing because they were salient enough to cause a transient orienting response, but were prevented from reaching higher levels. With a little more attentional investment, objects could have been processed at the non-reflective level. At the non-reflective level the object receives enough attention to allow the subject to retain some, but not all, of its features in memory. At this level, the subject has not yet developed a figure or ground structure. Thus, a partial

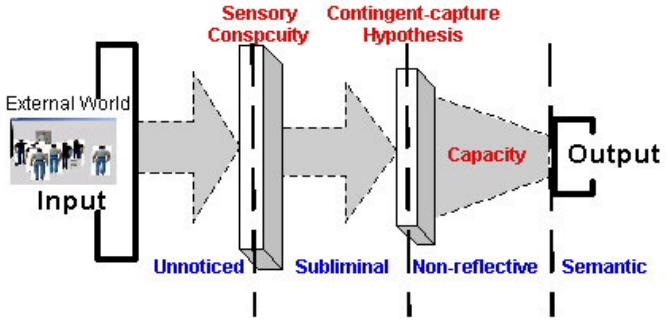


Fig. 4 Workflow of three filters. It demonstrates how three filters work to determine different level of process.

description of the object can be expected, but some details will be missed. Fig. 4 shows a block diagram of these processing levels.

The amount of attention devoted to the processing of an object can also be explained by how several filters work. When an object is not physically salient enough to catch attention, it is discarded by the sensory conspicuity filter, resulting in no processing and, consequently, no conscious awareness of it. An object has passed the sensory conspicuity comparison when it was eye-catching enough to induce an unconscious transient shift of attention. If the properties of this object do not match those held in the attentional set, it falls out of current coherence map, having received only minimal attention. But even if the object was physically salient and held many properties that matched the attentional set, it can still be discarded due to the capacity bottleneck. At this level, the object has been processed quite a bit, but not completely, so a subject's description of the object would contain some partial or even incorrect details. Finally, the object approaches the semantic level and is fully processed in conscious perception. For people who allowed the unexpected item to be sustained in attention, a detailed description is not difficult.

Experiment Results and Discussion

Thirty-six participants were randomly assigned to one of 27 groups that varied according to three parameters: mental workload, sensory conspicuity and attentional set. The data from six participants was discarded because of previous experience with inattentional blindness, or incorrect performance on the task. The results are summarized in Table 2 and illustrated in Fig. 5. The awareness level is assigned as a score from 1 to 4, corresponding to the processing levels from unnoticed to semantic, respectively. Each group included 10 subjects. The average score for the matched, unmatched, and neither matched nor unmatched attentional set groups was 2.5, 2.1 and 3.0, respectively.

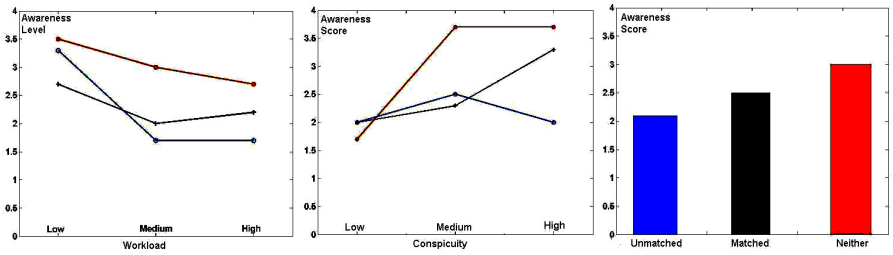
Table. 2: Summary of the levels of processing averaged by the subjects in each group.

Attentional set	Average	Workload			Conspicuity		
		Low	Med	High	Low	Med	High
Match (subj : 10)	2.5	2.7	2.0	2.2	2.0	2.3	3.3
Unmatch (subj: 10)	2.1	3.3	1.7	1.7	2.0	2.5	2.0
Neither (subj: 10)	3.0	3.5	3.0	2.7	1.7	3.7	3.7
Average		3.2	2.2	2.2	1.9	2.8	3.0

Thus, we can consider the results favorable since they agree with the four-factor model and our computational framework. This validates our model's assumption on these three very important factors of inattentional blindness. There are a few interesting findings to note.

1. We found the neither matched nor unmatched object is generally the most easily noticed one of the three attentional set groups. While counterintuitive, this finding is supported by our model. The model allows for the possibility that objects that perfectly match the attentional set will be discarded in level one if they are not physically salient enough. It would be reasonable to believe that the black and white T-shirt unexpected images (matched and unmatched,

respectively) were not physically salient in the scene, and could have been discarded in level one. The maroon T-shirt unexpected object (neither matched nor unmatched), could have been inherently salient enough to pass through the first bottom-up filter and then made its way into awareness because of its similarity to the attentional set in pant color and body shape as well as the T-shirt which is darker than it is light. (That is, it was more black than white – so more likely to be in the attentional set than in the inhibition set). More work should be done to illuminate the causal features in this situation.



(a) **(b)** **(c)**

Fig. 5 For all charts, red corresponds to the neither matched nor unmatched attentional set, black corresponds to the matched set and blue corresponds to the unmatched set. (a) Awareness Score vs. Workload. The unexpected object becomes more noticeable as the workload is reduced for all three attentional set groups. (b) Awareness Score vs. Conspicuity. The unexpected object receives greater processing when sensory conspicuity increases though there is some noise in the unmatched group. (c) Awareness Scores vs. Attentional set. The unexpected object receives the most processing when it is neither matched nor unmatched and the least when it is unmatched.

2. Additionally, there are two interesting findings about workload. Not only does it show the largest difference between its largest variations, suggesting that workload is the most important feature of attention capture and inattention blindness, but it also shows its largest variation between its *medium* and *low* settings (as opposed to the expected high and low settings). The only difference between the high and medium setting is the ball speed, but the ball with high speed was extremely fast. It is possible that the high setting was too difficult, and that people were more easily distracted because they had actually given up on the task. The medium speed may have been just difficult enough. This is another important parameter to investigate.

Application

The importance of a flawed attention model is considerable. Communication, especially face-to-face conversational interaction [3], is affected not only by the individuals involved, but also by what is taking place in the external environment [14]. To improve the naturalness of conversations, we are attempting to use the attentional

framework to create embodied agents that are aware of a perceived world. While attention to the conversational partner is the most basic form of signaling understanding by the agent, a listener whose eyes never waver from her partner, despite background events, appears lifeless.

An agent with a realistic attentional system also has the ability to use the perceptual information it gains from the external world to enhance its engagement during a conversation. Engagement is defined here as the process by which two (or more) participants establish and maintain their perceived connection during interactions they jointly undertake [19][20]. Three types of engagement cues are categorized: those

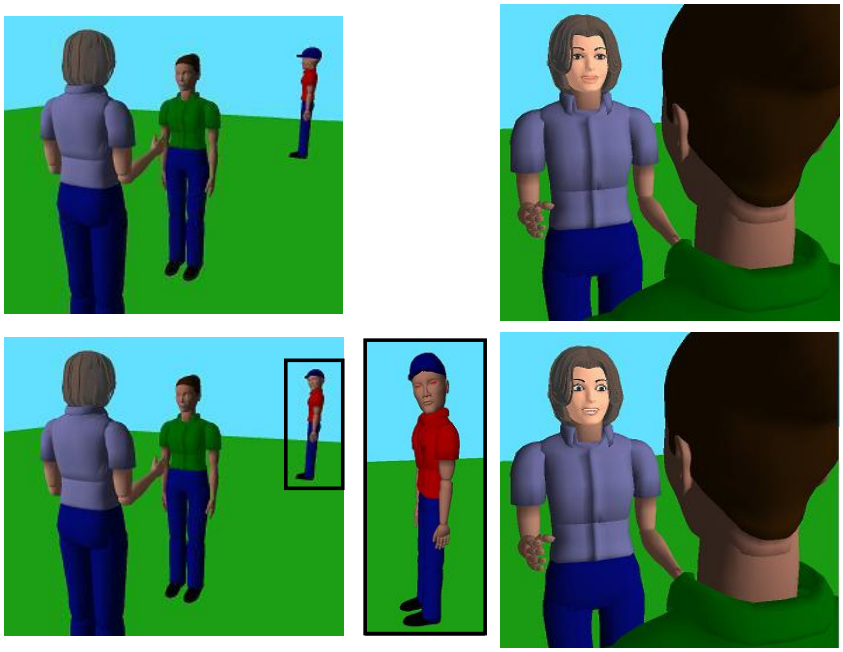


Fig. 6: Snapshots of two conversational agents interacting. During the conversation, a man with red eyes walks through the background. In the first case (top), the red T-shirt man walks off and does not turn his face towards the speaker. Thus, the speaker continues to talk, paying no attention to the man, even though he has fallen into her line of vision. In this situation, the perceptual information of the man is discarded by the visual attention model of the speaker. In the second case (bottom), when the man turns his head and shows his red eyes, the speaker is shocked. The face pop-out and physical saliency of the man causes the engagement of the speaker to shift from the listener to the external world stimuli.

with oneself, those with a conversational partner, and those with the environment. Our inattention blindness framework can improve the engagement behaviors of an embodied agent, particularly for the transition from self/partner to the environment. Therefore, in conjunction with an eye-movement model [11], the attentional model will increase the realism of an agent's engagement behaviors, as demonstrated in Fig. 6.

Future Work and Conclusion

As embodied agents become more commonplace elements of interpersonal interactions, adequate computational frameworks for cognitive processes are essential. Not only must the framework replicate normal human functioning, it should also demonstrate abnormal and imperfect human functioning, or else the agent will never be able to assimilate into a human-interactive environment. We have presented current theories of inattentional blindness and demonstrated how to integrate them into one model of visual attention. We attempted to justify our model with an experiment that examined three of the most important parameters, and discovered that the results agree with our proposed computational framework.

Future work for the model will include: further exploration of the parameters of habituation and capacity level, as well as more experimentally supported quantification. In addition, it is important to have models that can predict attention failure in order to decide how to compensate for, as well as reduce, human errors in perception in critical situations such as operating machinery or security monitoring. We hope that future work on our model can help contribute to these challenging problems.

Acknowledgements

This work is partially supported by the ONR VITRE project under grant N000140410259, and NSF IIS-0200983. Opinions expressed here are those of the authors and not of the sponsoring agencies. The authors are deeply grateful to Jan M. Allbeck for the human model. Also the authors thank all the participants in our IRB-approved experiments.

References

1. Becklen, R. and Cervone, D. (1983). Selective looking and the noticing of unexpected events. *Memory and Cognition, 11*, 601-608.
2. Bernardino, A. and Santos-Victor, J. (2002). A binocular stereo algorithm for log-polar foveated systems. *In Proc. 2nd International Workshop on Biologically Motivated Computer Vision*, pages 127–136. Springer-Verlag.
3. Cassell, J., Bickmore, T., Campbell, L., Vilhjalmsson, H. and Yan, H. (2001). More Than Just a Pretty Face: Conversational Protocols and the Affordances of Embodiment. *Knowledge-Based Systems, 14 (2001)*, pp. 55-64.
4. Chopra-Khullar, S. and Badler, N. (2001) Where to look? Automating attending behaviors of virtual human characters. *Autonomous Agents and Multi-agent Systems* 4, 9-23.
5. Folk, C. L., Remington, R. W. and Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology*: *Human Perception and Performance*, 18, 1030-1044.
6. Green, G. (2004), Inattentional blindness and conspicuity. Retrieved November 10, http://www.visualexpert.com/Resources/inattentionalblindness.html

7. Gu, E. (2004), Attention Model in Autonomous Agents, Technical Report, CIS, University of Pennsylvania.
8. Gu, E., Wang, J. and Badler, N. (2005). Generating Sequence of Eye Fixations Using Decision Theoretic Bottom-Up Attention Model. *3rd International Workshop on Attention and Performance in Computational Vision*..
9. Itti, L. (2003), Visual attention. *The Handbook of Brain Theory and Neural Networks*, pages 1196–1201.
10. Itti, L., Koch, C. and Niebur, E. (1998), A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259.
11. Lee, S. P., Badler, J. and Badler, N. (2002). Eyes Alive. ACM Transactions on Graphics 21(3):637-644.
12. Mack, A. and Rock, I. (1998). *In Inattentional Blindness*. 1998. Cambridge, MA: MIT Press.
13. Most, S. B., Scholl, B. J., Clifford, E. R. and Simons, D. J. (2005). What you see is what you set: Sustained inattentional blindness and the capture of awareness. *Psychological Review*, 112, 217-242.
14. Nakano, Y. I. and Nishida, T. (2005). Awareness of Perceived World and Conversational Engagement by Conversational Agents, AISB 2005 Symposium: Conversational Informatics for Supporting Social Intelligence & Interaction, England.
15. Rensink, R. (2002). Internal vs. external information in visual perception. *Proceedings of the 2nd International Symposium on Smart Graphics*.
16. Simons, D. J. and Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28,* 1059-1074.
17. Simons, D. J. and Rensink, R. A. (2005). Change blindness: past, present, and future. *Trends in Cognitive Sciences*, 9, 16-20.
18. Sun, Y. and Fisher, R. (2003). Object-based visual attention for computer vision. *Artificial Intelligent*, 146:77–123.
19. Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, 3, 449-459.
20. Ward, T. A., An Overview and Some Applications of Inattentional Blindness Research, research paper for PSY 440 (Perception), Stephen F. Austin State University.http://hubel.sfasu.edu/courseinfo/SL03/inattentional_blindness.htm
21. Wolfe J. M. (1999). "Inattentional amnesia", in Fleeting Memories. *In Cognition of Brief Visual Stimuli.* Cambridge, MA: MIT Press. 71-94.
22. Woodman, G. F. and Luck, S. J. (2003). Dissociations among attention, perception, and awareness during object-substitution masking. *Psychological Science*, 14, 605–611.