COMPUTER ANIMATION AND VIRTUAL WORLDS
Comp. Anim. Virtual Worlds 2011; **22**:361–369
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cav.425

SPECIAL ISSUE ARTICLE

Modelling selective visual attention for autonomous virtual characters

Elena Kokkinara*, Oyewole Oyekoya and Anthony Steed

Virtual Environments and Computer Graphics Group - University College London, Department of Computer Science, Gower Street, London, WC1E 6BT, UK

ABSTRACT

Autonomous virtual characters (AVCs) are becoming more prevalent both for real-time interaction and also as digital actors in film and TV production. AVCs require believable virtual human animations, accompanied by natural attention generation, and thus the software that controls the AVCs needs to model when and how to interact with the objects and other characters that exist in the virtual environment. This paper models automatic attention behaviour using a saliency model that generates plausible targets for combined gaze and head motions. The model was compared with the default behaviour of the Second Life (SL) system in an object observation scenario while it was compared with real actors' behaviour in a conversation scenario. Results from a study run within the SL system demonstrate a promising attention model that is not just believable and realistic but also adaptable to varying task, without any prior knowledge of the virtual scene. Copyright © 2011 John Wiley & Sons, Ltd.

KEYWORDS
animation; attention; autonomous virtual characters

***Correspondence**
Elena Kokkinara, Synthetic Perceptive, Emotive and Cognitive Systems - Mixed Reality Laboratory, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain.
E-mail: redarahel@gmail.com

1. INTRODUCTION

One of the main requirements for the generation of multimedia such as video production and games is the automation of the behaviour of virtual characters. For example, a virtual character should react in a sensible way whenever a task-relevant object or another character enters its field of view (fov), or it should be able to explore in a virtual environment. However, at the moment, character animations are programmed offline or use very simple procedural animation. The interaction between different virtual characters and the objects that populate these 3D worlds can be limited because of the amount of time and effort that has to be spent by animation experts to animate scripted story lines. Characters can look wooden because they have few behaviours outside of pre-scripted situation, and they may not look aware of events in their surroundings. Ideally, we would like to be given the impression that characters are interacting with their environment in the same way as living, thinking creatures. Automating the animation of the eyes and head of virtual characters will have an important impact in achieving this goal.

This work explores a task demand model also referred to as top-down, voluntary attention based on endogenous factors. This model takes into consideration agent–object interactions based on interaction information of various kinds: intrinsic object properties, information on how to interact with it, object behaviours and expected agent behaviours [1]. Additionally, goal-directed motion is when the avatar direct its attention to focus on objects relevant to a specified task. In this paper, we study this approach in terms of gaze and head motion of the humanoid, as well as in terms of social behaviour. Initially designed as a gaze model, this paper extends a previous model of visual attention [2] to animate the head. The model takes as input the intrinsic object properties and generates eye and head reactions to objects within the environment. Furthermore, an attempt to adjust the attention model to the avatar's social behaviour has been done, with the intent of producing realistic attention towards the avatar that a virtual character shares a conversation with or just when it comes across other avatars in a virtual scene.

The model was integrated into the Second Life (SL) application platform [3], a widely used application providing a set of valuable features for designing and controlling a virtual character. It also provides a variety of existing virtual environments to explore, as well as open source code for the client viewer. Hence, this platform was an appropriate base for the integration of the attention model, enabling evaluation of the model's generalisation to new environments and robustness in different virtual scenarios.

In order to investigate the effectiveness of the methods used, two different experiments were designed. The experiments aimed to examine the believability and realism of the attention model-driven SL characters in two scenarios: object observation and conversation. In the conversation scenario, avatar responses to social situations were compared with videos of similar situations in the real world. Results prove the robustness of a model that provides virtual humans with realistic attention features.

2. RELATED WORK

Human attention, and thus their head and eye gaze, moves to different target locations at a time. The spatial aspects of visual attention, that is, the focus of gaze and the constraint on the head position, are driven by the need to focus on objects of interest. We are more likely to be aware of events that occur at our focus of attention than away from it [4]. In a proposal of a computational framework for generating visual attending behaviour in a virtual human [5], eye behaviour was assigned to broad types of motor and cognitive activities (monitoring and locomotion, reach and grasp, visual search and visual tracking), combining goal-directed, exogenous and idle factors for attention. Peters and O'Sullivan also generated bottom-up visual attention behaviour in virtual humans [6] on the basis of the interactions of multiple components: a synthetic vision system, a model of bottom-up attention processor, a memory system and a gaze controller/generator. Using different mechanisms, Gilles' model consists of an attention manager based on vision and attention [7]. The focus of attention in this case is dependent on other agents' looking behaviour, and each agent receives requests for attention shifts from other agents' gaze targets. It then arbitrates between them and executes them, changing the focus of attention, which can then be used by other agents.

The simulation of visual perception is vital for the simulation of behaviour. Gaze models have been developed for the generation of natural eye movement for virtual characters [8]. Grillon and Thalmann simulated gaze for crowds using an automatic interest point detection algorithm based on bottom-up attention behaviors [9]. The main approach in existing models is also the derivation of statistical models from eye and head tracking data [10]. In the Eyes-Alive model [11], videos of eye movements are segmented and classified into talking and listening modes so that they can construct a saccade model for each mode. Other works use various machine vision algorithms to extract critical parameters from video [12]. These parameters are then transmitted to the rendering site and used for avatar animation. The last approach consists of using machine vision to attempt to locate targets of interest in virtual or real scenes, where Itti *et al*. [8] developed a computational model that predicts the spatiotemporal deployment of gaze onto any incoming visual scene.

Multimedia such as video production and games often include computer graphics animations in which virtual actors engage in conversations. An essential factor for the believability of an embodied conversational agent is the generation of eye and head movements. In Masuko's method [13], the head and eye movements of the virtual actor were generated and synchronised with the conversation on the basis of the utterance. Utterance is computed by the voice and the text sentences with 3D information to be spoken. Turn-allocation strategies in multiparty conversation and associated gaze behaviours were also considered by Gu and Badler [14]. Gaze models can be generated in conjunction with statistical models in order to generate the appropriate expressions for conversation scenarios. Bilvi's model predicts what should be the value of gaze in order to have a given meaning in a given conversational context [15]. Using a stochastic model [11] as an engine for generating realistic shifts in gaze (saccades), Poggi's gaze lexicon was created for mapping signals and meanings of an agent's internal state, where the appropriate gaze expression is generated according to the context of the conversation [16].

3. METHODOLOGY

In this paper, we extend a previous work on gaze modelling [2] in two ways: we develop the attention model to drive both head and eye gaze and we integrate this model into SL. This provides an excellent platform for testing in general avatar encounters, and in the Experiments section, we discuss user experiments based on this platform.

3.1. Attention Model

The attention model [2] is designed to adapt to the complex interaction within the scene and is based on data-driven modelling of gaze behaviour.

The trigger for the eye and head movement generation in the attention model are the objects or other characters surrounding the avatar. Consequently, a fundamental first step was to ensure that the characteristics (i.e. position, orientation) of these SL objects and avatars are considered by the model. Crucially, as the SL world changes (i.e. by new avatars coming in or avatars leaving and objects being rendered or destroyed), the attention model adapts. The next step was to load information about the model-controlled avatar's eyes and head current location in each frame. Then, the model determines the target with the highest saliency. An object from within the fov can be selected as the target depending on its intrinsic saliency, as depicted in (Figure 1) and the following equations. The four criteria were weighted and summed to compute saliency scores for each possible targets in the database. The horizontal fov was set to 70° for the eye (i.e. a maximum angle of 35° towards the left or right), whereas the vertical fov was set to

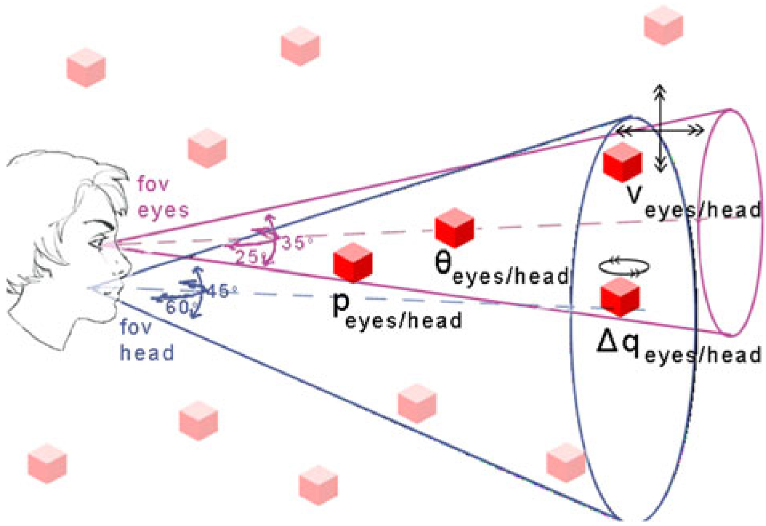


Figure 1. Visual depiction of saliency criteria. Objects are selected based on the highest saliency score for both eye and head.

50° (i.e. a maximum angle of 25° upwards or downwards). The horizontal fov for the head was set to 120°, whereas the vertical fov was set to 90°. The steps and equations used in computing the scores are described as follows:

1. *Proximity* p relates to the Euclidean distance between the user's eye from the object. Closer objects are allocated higher scores than farther objects. Given the position of the character's eye (or head), $E = (e_x, e_y, e_z)$, and each object, $O_i = (o_x, o_y, o_z)$, *proximity* p is computed from the Euclidean distance between the two 3D points as

$$p = \sqrt{(e_x - o_x)^2 + (e_y - o_y)^2 + (e_z - o_z)^2} \quad (1)$$

A Gaussian curve fit of the proximity is computed from

$$y = f(x) = \sum_{i=1}^{n} a_i e^{\left[-\left(\frac{x-b_i}{c_i}\right)^2\right]} \quad (2)$$

The Gaussian model is used for fitting peaks and is given by Equation (2) where a_i are the peak amplitudes, b_i are the peak centroids (locations), c_i are related to the peak widths, n is the number of peaks to fit and $1 \leq n \leq 8$. Proximity p is fitted with the values $a_1 = 19.11$, $a_2 = 6.68$, $b_1 = 1.83$, $b_2 = 3.27$, $c_1 = 0.87$ and $c_2 = 1.7$. The saliency score, S_p, of the object's proximity is also computed from Equation (2) where $x = p$ and is normalised by dividing by a_1 (i.e. peak amplitude) to keep the range between 0 and 1.

2. *Eccentricity* θ is based on the angular distance of objects from the centre of gaze (head-centric vector), allocating attention to objects under direct scrutiny than objects in the user's peripheral vision. θ is defined as the magnitude of the dot product is computed as

$$\theta = \arccos\left(\frac{u \cdot v}{|u||v|}\right) \quad (3)$$

where $u = (u_x, u_y, u_z)$ is the head-centric vector and $v = (v_x, v_y, v_z)$ is the direction vector of the eye (or head) to the object, $(e_x, e_y, e_z) - (o_x, o_y, o_z)$. A Gaussian curve fit of the eccentricity is computed from Equation (2). Eccentricity θ is fitted with the values $a_1 = 40.13$, $a_2 = 8.09$, $b_1 = 14.39$, $b_2 = -14.05$, $c_1 = 4.18$ and $c_2 = 40.5$. The saliency score, S_θ, of the object's eccentricity is computed from Equation (2) where $x = \theta$ and is normalised by dividing by a_1 (i.e. peak amplitude).

3. *Velocity* v is based on the object's speed across the user's visual field, allocating attention to objects moving quickly across the user's gaze than slow-moving or still objects. It is defined as the rate of change of the object's location. v is defined as the rate of change of the object's location and is computed as

$$v = \frac{\Delta O_i}{\Delta t} \quad (4)$$

where ΔO_i is the Euclidean distance between an object's location at time t_1 and its location at time t_2, and Δt is the time interval of the frame duration. The normalised saliency score, S_v of the object's velocity is given by $v/20$ (i.e. a reasonable maximum speed of 20 feet per second).

4. *Orientation* Δq is based on objects' differences in rotation with attention allocated to objects with higher rotation speed. It is defined as the change in object's angular position over time. Δq is defined as the

change in object's angular position over time and is computed as

$$\Delta q = 2 \arccos(q_1{}^{-1}.q_2) \tag{5}$$

where quaternions q_1 and q_2 represent two orientations at time t_1 and t_2, respectively. The normalised saliency score, $S_{\Delta q}$, of the object's orientation is given by $\Delta q/180$ (i.e. a reasonable maximum change in orientation of 180°).

The saliency of each object within the fov is computed from a summation of the normalised saliency scores.

$$S_O = S_\theta + S_p + S_v + S_{\Delta q} \tag{6}$$

The summation S_O was then used to guide attention, as the selected target in each frame will become the object with the highest overall saliency score. An eye or head fixation towards the particular object will then occur. The fixation duration for the eye was limited to 300 ms, whereas the typical head fixation duration was extended to 1000 ms as long as the target object remained within each respective fov.

The eyeball was interpolated over six frames by fitting to an exponential velocity curve to produce a smooth movement during saccades as presented in Lee *et al.* [11,17]

$$y = 14e^{[-\pi/4(x-3)^2]} \tag{7}$$

where $x = \text{frame}\{1, 2, 3, 4, 5, 6\}$.

In order to correlate the eye and head animations, the eyes saccade in the same direction as the head during the head's saccadic transition. The head's saccadic transition was fitted to the following model [18].

$$\theta_h(t) = e^{0.027t}\left(\sin\frac{\pi}{2}t\right)^{1.90} \tag{8}$$

As in previous implementation, in order to decrease the probability of centre bias in the model, a random selection algorithm computes the target object on 25% of the time, hence the probability that the saliency scoring algorithm uses is given by $P(salience) = 3/4$. Using this method, the objects in the periphery can also participate in the target competition but with less probability than the highly salient objects. It also means that the eyes and head are animated even when the virtual character is idle.

3.2. Modelling Social Behaviour

A virtual scene could contain a large collection of objects. Avatars could also be part of the scene object database and hence possible targets selection. Aiming to test the eye–head saliency model realism in a conversation between two avatars, we expected that an object such as another avatar in the scene would have increased the chance of attracting attention. Therefore, the contents of the objects' database were populated in favour of the avatars. Hence, each time a new avatar was encountered in the scene, the following body parts were added to the system: head, left eye, right eye, chest and torso. Having the character chat with another avatar in a virtual scene, the four criteria should yield relatively higher saliency scores for the avatar's body parts.

3.3. Second Life

Second Life is part of the growing family of online social spaces supporting a massive number of users in a shared virtual environment. The SL platform has been used before for a wide range of research, such as emulating the input device for massively multiplayer online games using eye control [19]. Furthermore, the SL platform has been used to study the spatial social behaviour of the users in a virtual world [20] and to monitor 'social expectations' in multi-agent systems [21].

The attention model was integrated as an external library into Snowglobe version 2.0.0, an open source SL-compatible viewer, built jointly by the open source community and Linden Lab.

3.3.1. Default Second Life Model.

The default SL model derives its implementation from typical social behaviour such as the turn-allocation strategies employed by Gu and Badler [14]. The eyes are driven by a model where the eyes either look straight ahead or away at a random angle based on preset time constraints (similar to Eyes Alive [11]). Head-turn priority is allocated to mouse cursor movement and the actions of other avatars within its vicinity. Head turns are allocated to other avatars when they walk into close proximity of the character, with increased attention being allocated to avatars that are talking or typing into chat. The SL character's head is fixed to look straight ahead during an idle posture. Hence, the default model computes the eye and head animation on each frame based on these competing social signals.

4. EXPERIMENTS

In order to investigate the effectiveness of the eye and head animation model, two different experiments were designed. The main objective was to prove the realism of the model in different modes: observation of objects and social behaviour within conversation scenarios.

4.1. "Look At" Experiment

The first experiment compares the attention model-controlled SL character with the default model-controlled SL character. The aim is to detect if the attention model-controlled virtual character is perceived to be more realistic than the default SL implementation.

4.1.1. Experimental Design.

For this experiment, the online video evaluation approach has been used. Participants were asked to compare pairs of videos in terms of perceived realism. Each pair contained a video captured while using the automatic attention model configuration and a video captured while using the default SL implementation (Figure 2). The videos were positioned randomly at the top or bottom of the experiment interface. The camera was manually zoomed in and out and rotated closely around the avatar, thus providing a varying and clear view. The videos in a pair depict the same scenario, and the participant watched the videos simultaneously. There were in total five different SL scenes (five pairs of videos):

1. *Examine a bar* - the avatar is observing the objects of the scene (glasses, bottles, chairs, etc.).
2. *Sit in a bar* - the avatar is sitting by the bar examining the objects and other avatars around the scene.
3. *Watching dancers* - the avatar is looking at other avatars dancing on the dance floor.
4. *Examine paintings* - the avatar is placed in an art gallery scene, observing six paintings.
5. *Garden* - the avatar stands in front of a gazebo and scrutinises the objects underneath it.

All five different scenarios were accompanied by three questions (see Figure 2) related to the avatar's attention, eye and head motion, respectively.

1. *Overall* - Where does the attention to the scene look more natural?
2. *Eye* - In which video does the eye movement look more realistic?
3. *Head* - In which video does the head movement look more realistic?

The subject had the choice of Top, Equal or Bottom (Figure 2) as an answer.

4.1.2. Results and Analysis.

A pilot trial was undertaken by five supervised subjects, and because the reactions were positive, another 51 volunteers performed the experiment online at their own machine and in their own time. After recording the 56 participants' votes on the five pairs, a preference matrix was computed for the default SL and the attention model mode on the basis of the voting results of the subjects (Table I). The number in each cell denotes the selection frequency of a specific method (model, defaults or equal) when answering one of the three questions (overall, eye and head realism), adding the answers for all the five videos, with one point given for each choice. From a plain first analysis of Table I, we can notice that participants clearly preferred the attention model videos over the default SL videos, for all of the three questions. In more detail, a chi-square analysis was conducted for each of the three questions. Because the data are not independent for each question, chi square could only evaluate the significance of the preference in each question separately. Hence, for the question about the overall believability of the avatar, the preference on the model was statistically significant with $\chi^2(1) = 46.667$, $p < 0.001$. The pairwise comparison between the three answers (model, default, equal) also revealed significant differences as can be seen in Table I. The analysis for the question about the eye motion's believability of our avatar had similar results. The preference on the model was statistically significant with $\chi^2(1) = 37.089$, $p < 0.001$. The

Figure 2. "Look At" experiment interface.

Table I. Participants preferences for "Look At" experiment over the five pairs of videos, chi-square analysis for the three questions and pairwise comparisons of the questions.

Question	Preferences			Comparison							
	Attention model	Default SL model	Equal	Model-default-equal		Model-default		Model-equal		Default-equal	
				χ^2	p	χ^2	p	χ^2	p	χ^2	p
Overall	140	80	50	46.667	< 0.001	16.364	< 0.001	42.632	< 0.001	6.923	< 0.01
Eyes	137	63	70	37.089	< 0.001	27.38	< 0.001	21.686	< 0.001	0.368	$= 0.544$
Head	162	58	50	86.756	< 0.001	49.164	< 0.001	59.17	< 0.001	0.593	$= 0.441$

Note: SL, Second Life.

last question about the head motion realism of the avatar also gave us statistically significant differences between the preferences on the behavioural model ($\chi^2(1) = 86.756$, $p < 0.001$).

4.2. Conversation Experiment

The second experiment attempts to evaluate the realism of the attention model associated with social behaviour. The aim was to test the effect of the model, when the avatar encounters other avatars, rather than simple objects as in the previous experiment. By using scenarios, which are based on small chats, the subject's opinions about the realism of the automatic eye and head motion and the overall attention generation were recorded, comparing with real people's behaviour.

4.2.1. Experimental Design.

For the conversation experiment, we used the online video evaluation approach again. This time participants were asked to watch pairs of videos, where the first of each pair contained real people (amateur actors) having a chat, whereas the second one depicted the same chat animated in an SL scene. For the main animated character, the enhanced attention model for social behaviour (see *Modelling Social Behaviour* section) was attached to the SL viewer, whereas for the speech animation, SL's lip synch feature was used through the microphone. A second SL client (using the default SL application) was also used for the conversation scenes in order to control the reactions of the second avatar in the chat. The task for the subjects was to rate (from 1 to 5) the avatars' behaviour in terms of overall believability, attention realism and eye and head motion compared with the real people. The experiment contained three different conversation scenarios in total, and the exact dialogues were reproduced in both real and animation videos (Figure 3):

1. *Asking information* - a woman is looking for a grocery store and she is asking a passer-by for information.
2. *Talking about a new bar* - a man suggests a new bar to his friend in the neighbourhood.
3. *Lost dog* - a woman is looking for her lost dog in the park.

Participants were asked to rate four questions (see Figure 3) on a Likert-scale rating of 1 to 5, assuming that the acted real video is 5. The questions relate to the avatar's overall behaviour, attention and head and eye motion, respectively.

1. *Overall* - Please rate the believability of the virtual scene in general.
2. *Attention* - Please rate the attention of the avatar to the avatar it refers.
3. *Eye* - Please rate the realism of the eye movement of the avatar.
4. *Head* - Please rate the realism of the head movement of the avatar.

This time, the pilot trial was conducted by five supervised subjects and then other 46 volunteers performed the experiment online at their own machine and in their own time. The analysis from the 51 in total responses is presented below.

4.2.2. Results and Analysis.

The Likert-scale ratings were entered into a single factor ANOVA with three levels (overall, attention, head, eye). The Likert-scale ratings of 1 to 5 for each question (overall, attention, eyes, head) were added for the three pairs of videos. Thus, the maximum rate for each of the four questions was 15 (5×3 scenarios). The main analysis revealed a main effect of question ($F(3,150) = 5.52$, $p < .01$), and the *post hoc* tests showed a significant difference between rating of eyes' realism and overall realism ($p = 0.014$) and between eyes and attention ($p = 0.003$). In Figure 4, we can see the differences in the four means and standard deviation of the rating of each question. Overall, the mean values of each question ranged from 7.5 (equates to 2.5 when normalised to the Likert scale of 5) for the eye to 8.6 (equates to 2.87) for attention with similar standard deviation.

5. DISCUSSION

In this paper, we proposed an attention model for animating the eyes and head of an avatar autonomously. Results demonstrate a realistic model for eye–head attention and

Figure 3. Conversation experiment interface.

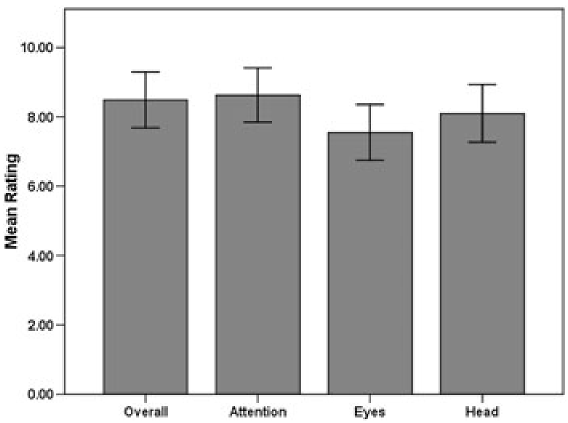


Figure 4. Mean rating for conversation experiment. The rating for each of the four questions (overall, attention, eyes, head) has been added for the three pairs of videos.

social behaviour. The attention model was shown to be flexible and adaptable to various tasks.

The hypothesis that the perceived realism of the avatar operating with the attention model would be preferred over the default SL model was tested with the “Look At” experiment. Moreover, we tested whether there is any perceived difference between the two models and found that rankings across questions indicate the superiority of the attention model. In detail, the overall attention of the attention model-controlled avatar was considered more realistic than the default SL model. The eyes and head movements’ generation by the attention model also had a significant preference. This affirms the success of our model’s approach for scene analysis towards realistic attention generation.

The automatic attention generation was also evaluated in terms of realism of the avatar’s social behaviour. The model was extended in order to append more importance to other avatars in the scene so as not to simply perceive them as objects. The analysis of the conversation experiment revealed that the subjects highly rated the perceived realism of the avatar’s overall behaviour and attention (Figure 4). However, the analysis also exposed a preference for the believability of the performance of the head motion over the believability of the eyes motion. This finding, along with participants’ feedback (general comments), indicates possible loss of realism as participants preferred if our avatar had longer eye contact with the other avatar. It must be noted that longer eye contact could have been achieved with a per-object weight, which would be higher for the eyes.

6. CONCLUSION

This paper focussed on identifying relevant targets for the eyes and head of virtual characters in real time. The attention model drew support from the fact that in the real world, targets (objects or other humans) compete for human visual attention. This work aimed to serve as a proof of concept in the viability of driving virtual characters with believable models. The attention model has been shown to enable an SL avatar to appear to perceive its environment and engage other avatars and objects in a believable manner. The feasibility of driving autonomous virtual characters has been established, although further studies need to be carried out for complete virtual autonomy, that is, where virtual characters respond to us in a human-like manner.

It must be noted that bottom-up attentional data (such as colour and intensity) were not considered as criteria because these data are not readily available in virtual reality systems without resulting in high computational expense as highlighted in Itti's work [8]. Furthermore, the model does not account for object displacement. An object moving sideways may not have the same impact as an object moving towards the agent.

We were keen to evaluate the character's behaviour in a commonly used application. Previous SL research used LSL script language to manipulate avatar functionality, but this has delayed updates because of slow responses from the central server and is not ideal for real-time evaluation. The use of the open SL viewer source code solved this problem with real-time updates.

Future work will concentrate on optimising the population of the object database such as including information from morph target animation. Furthermore, it would be interesting in the future to test the social behaviour using professional actors in the comparison videos. Extending the model to select plausible navigation targets in order to create an attention-driven explorative mode would also be of great interest. A valuable future improvement of the integration of this model in the SL viewer would be to send the model's updates via the SL web server, enabling other SL users to see our model-driven avatar. In addition, it would provide a great opportunity to run collaborative experiments.

ACKNOWLEDGEMENTS

We would like to thank the following sponsors: EPSRC Eyecatching, EU Presenccia and EU Beaming.

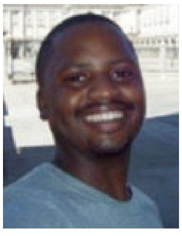
REFERENCES

1. Thalmann D. Challenges for the research in virtual humans. In *Workshop Achieving Human-like Behavior in Interactive Animated Agent*, 2000.
2. Oyekoya O, Steptoe W, Steed A. A saliency-based method of simulating visual attention in virtual scenes. In *Proceedings of the Virtual Reality Software and Technology*. ACM Press, New York, NY, 2009; 199–206.
3. Rymaszewski M, Au WJ, Wallace M, Winters C, Ondrejka C, Batstone-Cunningham B, Rosedale P. *Second life: the official guide*. SYBEX Inc., Alameda, 2006.
4. Pashler HE. *The psychology of attention*. The MIT Press, Cambridge, Massachusetts, 1998.
5. Khullar SC, Badler NI. Where to look? Automating attending behaviors of virtual human characters. *Autonomous Agents and Multi-Agent Systems,* 2001; **4**(1): 9–23.
6. Peters C, O'Sullivan C. Bottom-up visual attention for virtual human animation. In *Proceedings of the Computer Animation and Social Agents*, 2003; 111–117.
7. Gillies MFP, Dodgson NA. Eye movements and attention for behavioural animation. *Journal of Visualization and Computer Animation,* 2002; **13**(5): 287–300.
8. Itti L, Dhavale N, Pighin F. Realistic avatar eye and head animation using a neurobiological model of visual attention. In *Proceedings of SPIE*, 2003; 64–78.
9. Grillon H, Thalmann D. Simulating gaze attention behaviors for crowds. *Computer Animation and Virtual Worlds,* 2009; **20**(2–3): 111–119.
10. Xiaohan M, Zhigang D. Natural eye motion synthesis by modeling gaze-head coupling. In *Virtual Reality Conference, 2009. VR 2009. IEEE*, 2009; 143–150.
11. Lee SP, Badler JB, Badler NI. Eyes alive. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press, New York, NY, 2002; 637–644.
12. Rikert TD, Jones MJ. Gaze estimation using morphable models. In *Third IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE CS Press, Nara, Japan, 1998; 436–441.
13. Masuko S, Hoshino J. Head-eye animation corresponding to a conversation for CG characters. In *Computer Graphics Forum*, Vol. 26. John Wiley & Sons, Oxford, UK, 2007; 303–312.
14. Gu E, Badler N. Visual attention and eye gaze during multiparty conversations with distractions. In *Proceedings of Intelligent Virtual Agents*. Springer, Berlin/Heidelberg, 2006; 193–204.
15. Bilvi M, Pelachaud C. Communicative and statistical eye gaze predictions. In *International Conference on Autonomous Agents and Multi-Agent Systems*, 2003.
16. Poggi I, Pelachaud C, De Rosisc F. Eye communication in a conversational 3D synthetic agent, *AI communications,* 2000; **13**(3): 169–181.
17. Vinayagamoorthy V, Garau M, Steed A, Slater M. An eye gaze model for dyadic interaction in an immersive virtual environment: practice and experience. In *Computer Graphics Forum*, Vol. 23. John Wiley & Sons, Oxford, UK, 2004; 1–11.
18. Itti L, Dhavale N, Pighin F. Photorealistic attention-based gaze animation. In *2006 IEEE International Conference on Multimedia and Expo*, 2006; 521–524.
19. Istance H, Bates R, Hyrskykari A, Vickers S. Snap clutch, a moded approach to solving the Midas touch problem. In *Proceedings of the 2008 symposium on Eye Tracking Research & Applications*. ACM press, New York, NY, 2008; 221–228.
20. Friedman D, Steed A, Slater M. Spatial social behavior in second life. In *Proceedings Intelligent Virtual Agents*. Springer, 2007; 252–263.

21. Cranefield S, Li G. Monitoring Social Expectations in Second Life. *Lecture Notes in Computer Science* 2010; **6069**: 133–146.

AUTHORS' BIOGRAPHIES

Elena Kokkinara studied Computer Science at Athens University in Economics and Business. In 2010 she received her MSc in Computer Graphics Vision and Imaging from University College London. She currently works in the laboratory for Synthetic Perceptive, Emotive and Cognitive Systems at University of Pompeu Fabra. Her current research interests are virtual reality, mixed reality as well as psychological and behavioural principles underlying perception, emotion and cognition related to virtual humans.



Oyewole Oyekoya received his degree in Computer Science from Kingston University in 2001. He received his Ph.D. from University College London in 2007 for work on eye tracking for image search and retrieval. He currently works in the Virtual Environments and Computer Graphics group at University College London. His current research interests are virtual reality, mixed reality, attention models and computer vision.

Professor Anthony Steed is a head of the Virtual Environments and Computer Graphics group at University College London. He received his Ph.D. from University College London in 1996 for work on interactive virtual environments. His current research interests are in very large 3D model rendering, immersion and presence in virtual environment displays, and interaction and collaboration between users of virtual environments systems.