Digital Creativity
2005, Vol. 16, No. 4, pp. 193–204

Gaze-guided viewing of interactive movies

Tore Vesterby, Jonas C. Voss, John Paulin Hansen, Arne John Glenstrup, Dan Witzner Hansen and Mark Rudolph

IT University of Copenhagen, Denmark

tore@itu.dk, jcv@itu.dk, paulin@itu.dk, panic@itu.dk, witzner@itu.dk, mrudolph@itu.dk

Abstract

The idea of gaze-interactive movies is illustrated by a simple example movie that unfolds non-deterministically via an analysis of the interest of the viewer measured by the interpreted input from an eye tracker. We demonstrate how the amount of relative attention paid to key subjects of narrative importance may guide the outcome of a narrative branching. An experiment was conducted to test the operation of gaze guided film. The experiment involved 11 subjects influencing a two-minute film clip by gaze in two scenarios. In the first case subjects were aware that gaze could be used to control the narrative, and in the second case the subjects were unaware of this control. The outcome was found to be quite uniform across subjects, and it was not influenced by repetitions or by knowledge about the control option. Comments from the aware users indicated that they were looking for confirmation of gaze selections from the system. Thus, non-intrusive feedback seems to be fundamental for a successful gaze-interactive media. We suggest a range of discrete audio and visual effects that may serve this purpose and present some narrative control principles.

Keywords: computer games, gaze tracking, interactive narrative, movies, multimodal interaction

1 Introduction

Non-intrusive gaze tracking holds potential for adding engaging game-like interaction to narrative media, while retaining the immersive experience associated with the viewing of motion pictures. When selections among action branches are seamlessly interwoven into the viewing process itself and wilful decisions are avoided, the viewer may interactively navigate the narrative structure while never interrupting the cinematic ‘suspension of disbelief’.

Interactive movies often have difficulties in maintaining an immersive experience throughout the story, because the engagement via keystrokes, joysticks or game pads prevents the viewers from being entranced by the dream-like immersion in the experience, and because the viewer may lose track of the storyline when required to make decisions about every new course of actions (Gerrig 1993, Murray 1997).

In contrast, Douglas and Hargadon (2000) noticed that when reading,

> *…our immersion in the fictions of a Dickens, Balzac, or Chekhov, stems from the relative lack of demands the act of reading places on us.*

We believe that watching an interactive movie should not be a stop-and-go affair. Rather, it should feel like being on a bus that doesn’t stop until the end of the tour. If interactivity is dependent on where the viewer looks on the screen, it could implicitly let the viewer decide which route to take. The changes should not be noticeable to the viewer. They should just happen as the story unfolds. With some experience, the interaction itself should

1462-6268/05/1604-0193$16.00 Routledge Taylor & Francis Group

Digital Creativity, Vol. 16, No. 4

possibly become unnoticed, in the same sense that good film camera work is invisible.

Gaze-based interactivity may add a range of new narrative options. Imagine, for instance, that the characters who initially received relatively more attention from the viewer became the most exposed during a comedy. The viewer could become actively involved in a gangster movie taking the role as the paterfamilias, deciding whom to kill simply by staring at them. In a murder mystery, clues could only affect subsequent action if the viewer had actually noticed them.

Unlike previous gaze tracking systems that were difficult to use, present gaze trackers are easy to use and quite robust. They achieve a precision that is comparable to the hand-mouse pointer. In this paper we suggest that gaze-based interactivity—if properly designed—may overcome some of the obstacles that have 'broken the spell' of most interactive narratives so far. We will present possible first answers to the following questions. How can we establish a set of narrative guides for gaze-controlled interactive media that do not interfere with the flow of the story? What parameters need to be considered when establishing this set? What directorial options are required when creating material for this new medium? How may a viewer interact seamlessly with the system?

2 Previous work

Richard A. Bolt (1984) first presented the idea of using eye movements for human-computer interaction. He suggested that the eye could be used as a pointing device: many simultaneous video images with sound would be presented to a viewer who was then able to select one of the sequences by looking at a certain image for a given period of time. A special video camera would record the eye position relative to the screens. The system zoomed in on the visual image with what Bolt termed "auditory zooming"—i.e. the sound associated with the image became the only sound playing—in order to indicate that the clip had been selected.

By 1990 Starker and Bolt (1990) presented a working self-disclosing system, based on a children's book, *The little prince*. A gaze tracker would measure where on the display the user was looking and the narrator would then proceed to tell about the object the user had shown interest in. An object being looked at would "blush…momentarily lightening its colour on the display" in order to give feedback to the user. The system included an 'interest module', that computed the object looked at, and incremented a 'tally' bound to that object. In this way, the user's gaze-patterns produced aggregate data describing the user's cumulative interests over a time interval. Starker and Bolt tested three different ways of deciding what object the user was most interested in, and thus what the narrator should be talking about. They determined that all three algorithms would work, but two of them would "more gracefully" drop the interest level of objects if they had not been looked at. Starker and Bolt (1990) also described how the system would make distinctions between focusing on a single object and focusing on groups of the same type of objects. If the user looked at a certain staircase he or she would receive information on this staircase, and when looking at a group of staircases, the user received general information on staircases.

Hansen, Andersen and Roed (1995) suggested unfolding the narratives of interactive movies by counting fixations on objects or areas of narrative importance in specific scenes. For instance, an opening scene could introduce a man and a woman entering a hotel. If the woman had received most attention, the camera would then follow her to the elevator in the next scene. If the man had got the most attention it would follow him to the reception. Hansen, Andersen and Roed (1995) also suggested combining the indications of interest with measures of the viewer's affective reactions by monitoring blink frequencies and pupil dilation.

3 Gaze-controlled narratives

In this section we will present structures by which the gaze-interactive process may be organised. In the next section we will suggest ways to direct the viewer's attention toward interactive elements.

3.1 Branching

In a traditional branching structure, the user is presented with several choices or options upon arriving at pre-designated 'forks in the road' (Samsel and Wimberley 1998). This structure can be extended to 'branching with optional scenes' where alternate scenes appear at certain points in the story, but all of them return to the main spine of the story. An example may be given by expanding on the previous example of a hotel movie: even if the viewer had followed the woman into the elevator, the next scene would show how, later that evening, she bumped into the man she had met in the doorway. And those who had followed the man would see the exact same scene. This kind of 'bottlenecking' is crucial to not only keeping the narrative flow from spinning out of control, but also as a means of keeping production costs down, since each scene that has one or more branches requires that much more production time. We are aware that by depending on a branching narrative structure, the question also remains whether the user experience can actually justify the production costs. Attention-dependent branching will be investigated in section 7.

3.2 Exploratorium

In the exploratorium the user is free to explore objects on the screen in their own time. The main storyline is on hold until the user has finished exploring the different objects available. In our hotel movie, our main characters may discover a body in a hotel room. They can now inspect that room for some time, until the important pieces of information have been found. Then suddenly a police officer knocks on the door. Exploratoriums are well known from adventure games and children's games. An exploratorium may best be used when the user has to explore a restricted environment like that of Starker and Bolt's (1990) *The little prince*.

The use of an exploratorium in a gaze-interactive medium requires an active search process from the viewer. Otherwise the viewer will just be stuck in a limbo where nothing occurs. Active searching may best be reinforced by immediate feedback from all gazed interactive objects, but of the subtlest kind. Significant characters looked at could just turn their heads toward the camera; objects may blush or they may emit a discrete sound.

With an eye tracking system the viewer can pan a camera to inspect 360° interactive panoramic images. For instance, by use of a QTVR-viewer the viewer may navigate a 360° photograph of a hotel room. When the gaze passes quickly over a painting hanging on the wall, which then catches his interest, the camera movement will slow down and even pan back towards the painting at a slower pace if the camera has moved away from it.

Apart from a common experience of the gaze tracker being a bit too eager to spin out of control, subjects we have observed are able to give very detailed descriptions of panorama images inspected with a gaze controller. They often report feeling immersed in the images, wanting to go explore more. This suggests that being able to explore a virtual space in a gaze-interactive medium is a viable option, but the control logic should be designed to the way the eye and gaze tracker operates, e.g. by tuning the smoothening factor of the gaze tracker to minimise the stability problems that arise due to time delays.

3.3 Parallel streaming

Parallel streaming allows the writer to create a single linear story, but the user may switch between perspectives, paths or states. The user can then experience the same series of events from multiple points of view (POV). Switches between viewpoints can be achieved by measuring the viewer's interest level to determine the camera placement or perspective. In a dialogue, for instance, the character receiving most fixations is the one that speaks up, while the other character keeps quiet. When the viewer looks to the other character, he or she will then be encouraged to talk. As with branching, this type of interactivity proportionately increases the production costs of the movie.

It is important to note that Ryan (2001) has suggested that the structure of the interactive movie itself rather than the plot of the movie may be more interesting to the user. In other words, the novelty of playing with the system itself may overshadow the user's immersion in the plot.

4 Cues for gaze-interaction

4.1 Pictures

In its infancy, the film medium re-used most of its narrative elements from theatre, and just slowly evolved its own set of visual and auditory semantics. The inventions were based in large upon experiments with the new technical capabilities available to the directors (Gianetti 1990). We expect that gaze-interactive media will at first evolve using the semantics of traditional movies and computer games, but as time passes proceed to evolve narrative principles unique to the medium itself. Therefore, we will also present some of the cues from the history of movies and games that gaze-interactive media could effectively re-use.

Velichkovsky and Hansen (1995) suggested four generic formats a gaze selection opportunity could take (see Figure 1):

A: presented as a traditional GUI button, to be activated by a fixation;

B: indicated by a framing of the selectable objects with e.g. a halo;

Figure 1. Four suggested formats of gaze selections in the hotel movie.
A: Traditional GUI button. B: Framing of the selectable objects.
C: Blurring the non-selectable areas. D: No visible selector.
(Adapted from Velichkovsky and Hansen (1995)).

C: visually emphasised by blurring the non-selectable areas;

D: not visible at all.

Only B, C and D should be considered for a truly non-intrusive media.

4.1.1 Lighting, perspective and level of detail

In order to direct attention to vital information, it may be put in the front of the image, highlighted, or given greater level of detail as occurs naturally when the object is in focus relative to a background. This is probably the most fundamental principle of picture composition, and has been used for centuries by painters and for decades by film directors. Recently, control of gaze by graphical means has been re-emphasized by Baudisch et al. (2003), who demonstrated how in a classical painting, by controlling luminance, colour contrast and depth cues, a painter is guiding the viewer's gaze. We envision that these effects could be used in a gaze-controlled interactive media, for instance when an object vital to the story is introduced. The halo—or inner glow—has been used in computer games in various ways to indicate selections or objects that can be manipulated. It is a faint aura of light either emitted by an object or character, or a faint glow above or below the object. The halo can also be used as a discreet feedback on a selection, similar to the blushing effect suggested for *The little prince* by Starker and Bolt (1990).

4.1.2 Blurring

By having two different texture versions for each important scene object, one where the objects are blurred and one focused, dynamic changes in blur levels could be used to direct the viewer's attention. An interesting possibility would be to initially display the interactive areas of a new scene in high detail and then gradually update the rest of the scene picture. After a few seconds the whole image would then appear in equal resolution. These 'fade-away-buttons' are a most discreet way of visualising the interaction possibilities, we imagine. However, they are likely to work best on more stationary scenes and may not be very useful for short, fast shifting shots of, for example, action movies.

4.2 Sound

In a gaze-interactive media, music or off-screen sounds could be used in the classic way, to set the mood of a scene. Off-screen sounds are normally used to establish a connection with something happening off-screen, or to emphasise the mood of a scene. The invoking of these sounds could be controlled by continuous measures of pupil dilation, indicating the emotional involvement of the viewer. Partala et al. (2000) found statistically significant pupil size variations as a function of emotionally arousing audio stimuli. Subjects' reaction peaked 2–3 seconds after stimulus onset. Pupil sizes were larger during negative stimuli (e.g. a woman screaming) than positive stimuli (e.g. people laughing). However several external factors need to be considered if this is to be implemented. External ambient light also effects pupil dilation. Therefore pupil dilation should be normalised according to the light intensity of the images and the surroundings in order for this feedback measure to work properly.

Object sound is a theme or sound that is connected to the object's presence, form or history and is played, either explicit or as a variant, over the current theme to accentuate its presence in the scene. In a gaze-interactive movie, the object sound could be used as a way to discreetly verify that a user has seen an object vital to the story, to tell what kind of object it is, or to imply interactive possibilities with an object. For instance, looking at a suitcase in our hotel movie for the first time could release a "TIC – TOC" sound, that would hereafter blend into the ambient music of the vestibule, to indicate that the suitcase contains a time bomb.

Changing the sound level of the soundtrack is a classical means of intensifying a scene in a film, to increase the emotional effect of the scene, best illustrated by the shower scene in Hitchcock's *Psycho* from 1960. Bolt (1984) suggested changing the level of the soundtrack to complete the cocktail mélange imagery of the 'world of windows'. All windows have a soundtrack that is played at equal level, until the user fixates on a window for several seconds. The systems then zooms the window being fixated, and plays only the soundtrack related to that window.

The different types of cue can be combined in multiple ways. Imagine, for instance, groups of characters having a conversation around two separate tables in a restaurant. The party that gets the least attention is blurred out and their conversation fades out of the soundscape. The camera zooms in on the other group and increases the volume of their conversation at the same time.

5 Execution of a movie

Figure 2 delineates the principal components of a system to control the execution of a gaze-interactive movie (not yet implemented). The gaze tracker sends the viewer's current gaze position and pupil diameter to an interpreter module, which maps eye coordinates to the pre-defined areas of narrative importance (ANIs), that the director has chosen to build the story upon. This information is sent to the producer module, which can determine what to change in the given scene—or what to follow—accordingly.

The producer module handles its tasks with the aid of sub-modules shown in the diagram below. When being passed information from the interpreter module, the producer consults the scene script sub-module to see if the viewer has been paying attention to the objects needed to understand the scene, and to understand the history to follow. If not, the producer calls the cue handler telling it to insert, for instance, a short cut with the briefcase in an advantageous angle, and maybe even increase the "TIC-TOC" sound from the bomb inside it.

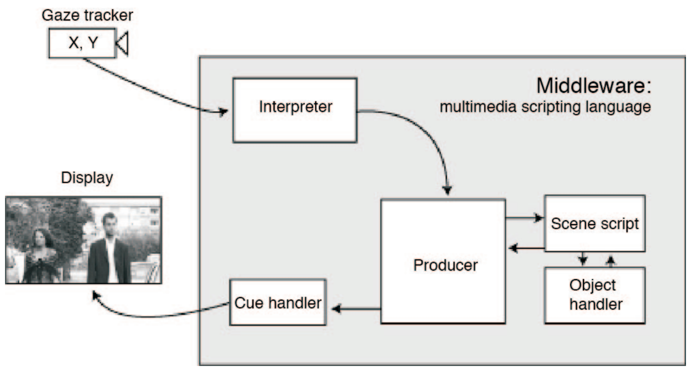

Figure 2. The gaze tracker sends the viewer's current gaze position and pupil diameter to an interpreter module, which maps eye coordinates to the pre-defined areas of narrative importance. This information is sent to the producer module, which can determine what to change in the given scene—or what to follow—referring to the scene script, the object handler and the cue handler. These changes determine how the plot of the hotel movie develops.

The scene script is the narrative core of the system. This is where the producer selects which story elements should be played and in what order. It is similar to a stage manager who makes sure the actors are ready to come on stage, or that the props for the next sketch are ready. The scene script holds a list of which narratively important objects—which can be both characters and physical objects—the viewer needs to have noticed before the scene may progress. The scene script keeps track of which objects vital to the story the viewer has noticed by checking if the object has had a macro-fixation set on it. The check is made by consulting the object handler. Basically this could be set up as a boolean value on each interactive object: either the object has been noticed or it has not. Additionally, objects can be set with a boolean value determining whether they are vital to the story or not.

The object handler keeps track of which objects in the scene have been noticed (N) and which objects are vital to the continuation of the story (VO). Objects in this case can both be actual objects such as the suitcase or a telephone, but also the actors appearing in the scene e.g. Tony, Veronica, or the Mysterious Man. The scene script also handles which scene in the system is playing, and has information about what objects vital to the story that must have been noticed by the viewer in order for the scene to be played. Hence it is a form of safety net that ensures that the narrative does not spin out of control, as discussed in section 3. If certain objects vital to the story have not been noticed by the viewer the scene script passes this information to the producer who can then enforce a flashback scene or open an exploratorium –which (re-) introduces the objects to the viewer.

6 Determining interests

In order to determine the viewer's interest, the interpreter module needs to be able to register which ANIs in a scene received the most fixations or evoked the most significant response from the viewer. Stelmach, Tam and Hearty (1991) found that gaze positions of different people cluster into a few (1 to 3) groups on most video material. On the basis of their find-

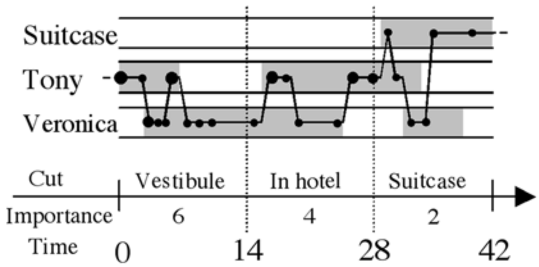

Figure 3. A fictitious set of eye movement data from a scene that shows which object or actor the viewer is fixating on, the gray background colour indicates when an object or person is visible and the sizes of the points are meant to indicate the pupil dilation at various intervals. These can be compared by different processes at the end of a scene to determine where viewer's interest lies.

ing it seems reasonable to limit the number of ANIs to a small select set.

Eye trackers made by consumer video cameras are expected to become available within a foreseeable future (e.g. Hansen and Pece 2003). They should be expected to have a lower resolution and possibly more frequent loss of eye data than the dedicated but expensive eye tracking systems of today. If this kind of standard hardware is to be used in order to make gaze-interactive media widely accessible, the system should merely count the number of gaze samples that fall inside each ANI presented within a specific period (i.e. scene). This method resembles what Salvucci and Goldberg (2000) have termed "macro-fixations". We decided to ignore that some of the hits may stem from data recorded during saccades—rapid movements of the eyes between fixation points—or during tracking failures. We assume that these sources of errors are equally distributed among the ANIs. This is, of course, a very rough assumption as areas of different sizes will get different amounts of, for example, false positives from saccade data. Salvucci and Goldberg (2000) also point out that the method may fail to identify a fixation should it fall just outside a pre-defined area. Again, we assume that this will happen randomly. Another source of error for movies could be false positives that stem from gaze locations in the previous cut. Therefore we expect a short time-out for data recording (e.g. 100 milliseconds) to be used after each cut as well as to filter out all macro-fixations that do not span a minimum duration threshold of, for example, 200 milliseconds.

When determining the interest the user has shown to the ANIs of a scene there are several aspects one must take into account. To illustrate this, we present a fictitious set of eye movement data from a scene (Figure 3). The curve shows which object the viewer is fixating, and the sizes of the points are meant to indicate the pupil dilation at various intervals. The director of the movie has defined the importance of cuts or objects.

We note some effects from Figure 3:

1 when Veronica enters, she gets fixations because she is a novel object;
2 when Veronica is the only object during the vestibule cut, she receives all the fixations;
3 when the opening scene is over, Veronica has received 4 macro-fixations, and Tony 5.

Two possible methods can be applied at the end of scene for determining the user's interest in the areas of narrative importance: 'winner-take-all' and 'weighted-decaying-interests'. Under the 'winner-take-all' principle, the area with the highest number of macro-fixations relative to its visible time is the ANI of most interest to the viewer, and should thus decide the outcome. The principle of 'winner-take-all' can be made sensitive to the relative narrative importance of the scene—as estimated by the director. The number of fixations gained by each ANI should be multiplied with the relative importance of the cut (c.f. 'Importance' in Figure 3). If the number of other important elements (people and objects) in the cut is also taken into account, the total number of objects in the cut should be, for example, multiplied with the number of times an object

was seen in that cut. Finally, if we multiply the importance of the scene with the number of competitors and divide it with their overall time being present, we get:
The suitcase: (2(2 x 3))/12 = 1.0
Tony: (6(2 x 1) + 4(2 x 1) + 2(1 x 3))/24 = 1.08
Veronica: (6(2 x 1) + 4(2 x 1) + 2(1 x 3))/29 = 0.90
—and Tony is the winner. This weight is particularly sensitive to the relative attention an ANI gets among its competitors in the most important cuts. Obviously, the settings of, for example, the cut importance by the director can be quite arbitrary and done without much consideration. An alternative approach would be to give the viewer an opportunity to select initially among different versions of the story in which the viewer would mainly follow Tony, Veronica or the suitcase. To handle this pre-selection, all weights should be adjusted as the importance factor in favour of the chosen one.

Under the strategy of 'weighted-decaying-interest', the interest registered at each fixation is weighted according to the cut importance and also how many objects are visible at fixation time. Furthermore, the interest of each object decays over time with some preset intensity *a*. This decay intensity factor is introduced to account for the fact that a viewer's interest gained for a given object will decay over time, but be renewed when the object is fixated anew (c.f. Hochberg and Brooks 1978). Thus the interest calculated by weighted-decaying-interest reflects the viewer's interest at the end of the scene, where the most recently seen objects are given priority over objects seen at the start of the scene. Weighted-decaying-interest can be used on individual fixations as well as macro-fixations. Figure 4 shows how the interest in the three objects of the opening scene change over time, based on individual fixations. Note how the interest stays constant when the object is not present. The final interests are 2.4 for the suitcase, 14.9 for Veronica and 23.7 for Tony.

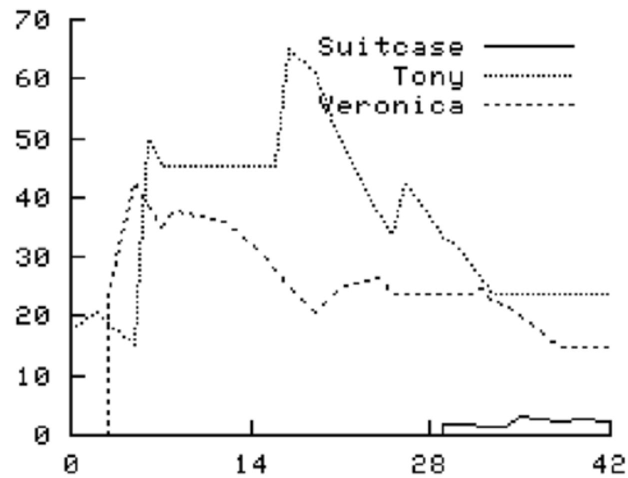

Figure 4. Weighted-decaying-interest in the three vital objects of the opening scene of the hotel movie over time, based on individual fixations.

A piece of code that calculates weighted-decaying-interest may be found on: http://www.itu.dk/~panic/Eyegaze/

7 User tests

We decided to build a small interactive movie and observe its use to evaluate our intuitions of how to solve some of the problems outlined previously in the paper. We wanted to explore the following questions:

1 If the viewers aren't told that the film is interactive, do they notice it anyway? And if the test persons are told that they can control the movie, how do they experience this control option?
2 Would the viewers find out, if they unknowingly had decided to view another outcome of the movie, or would they even notice a variation in the movie if it were shown repeatedly?

7.1 Procedure

Six male and five female subjects participated in the observations, aged between twenty-four and thirty. The participants were split into two groups, one of three males and three females, and one of three males and two females. The first group was only told that the purpose of the test was to identify what people looked at when watching thrillers.

The second group was told that they could in fact control the outcome of the movie depending on where they looked during playback. We did this with the intention to see if test subjects alerted to their influence via gaze would be overly aware about what and where they were looking at on screen, and if they would actively try to manipulate the outcome of the movie.

A QuickGlance system from EyeTech Digital Systems was used for eye tracking with an update rate of 15 frames per second and a smoothing factor of 7 samples. The prototype was played on a 1024x768 resolution 17-inch monitor. The movie took up an area of 720x324 pixels. The ANIs took up 360x600 pixels each and were separated by an empty space of 4 pixels, which meant that they exceeded the height of the movie in order to compensate for slight inaccuracies in the calibration of the gaze tracker.

The prototype of a gaze-interactive movie consisted of a short snippet of our hotel movie outlined previously. It contained one plot presented to the viewer with one simple ‘choice’ that would branch the main narrative in a bottleneck structure. The middleware, written in *Lingo*, Macromedia’s *Director* scripting language, would increment a variable each time the eye dwelled on one of two areas of the screen. By the end of that scene a simple winner-take-all calculation would decide where to go next. These active areas were made as non-visible rectangular layers on top of a 6.12 second clip showing the man and the woman walking side-by-side along the street.

After the first viewing we asked the subjects what they thought had happened. We then asked the subject to view the clip again a second time, and again asked them if they had noticed something different—even if they had seen the same sequence. Then we asked the subject to view it a third time, repeating the questions above, but also asking if they felt that they could control the film if we had told them to, or asking them why they thought they had seen different clips if they were in the group that was unaware that they could control the film.

7.2 Results

For each viewing of the hotel movie, the system counted the number of registered ‘mouse-within’ events (hits) on each of the tracking areas in the tracking sequence. All 11 subjects saw the movie 3 times. Data from 4 of the trails were excluded due to loss of calibration.

A three-way ANOVA with 2 (aware or unaware) x 2 (characters) x 3 (number of exposures) was performed with subjects nested in level of consciousness. Attention paid to the two characters was treated as a dependent measure. A significant effect of character was found $F(1,57) = 26.88$; $p<0.0001$. In average the Tony-character got 83.8 fixation counts (63 %) while Veronica got 49.1 (37 %). Only in 6 out of the 29 trails did the subject look most at Veronica and only one subject looked more at Veronica during the very first trail. There were no significant effects from the repeated exposures or the level of consciousness on the attention paid to the characters.

8 Discussion

This result indicates a general pattern of viewing across subjects, which is robust to changes from repeated exposures and which is not significantly influenced by knowledge of interaction possibilities.

In our specific case, there may be several possible causes for Tony being the most viewed. First of all, the sequence in which the dominant fixation was calculated has more movement in the right hand side of the screen. Pedestrians walk by in the background, and

Digital Creativity, Vol. 16, No. 4

traffic passes by on a road in the distance. Secondly, Tony and Veronica are displaced a bit to the right of the screen when the scene begins—albeit the tracking areas are placed over the man and woman throughout the sequence. Tony moves slightly towards the left hand of the screen as the scene progresses. Thirdly, Tony is also taking a puff on a cigarette, while Veronica is just walking throughout the scene. Fourthly, Tony was played by one of the men conducting the test, and the subjects might be looking more at him, as they have already have met and seen him. This effect probably resembles the attention advantage a well-known actor will have among less known actors. All combined, these factors tip the visual balance towards Tony.

Even though no quantitative differences could be measured, the aware and unaware groups had quite different comments about their experiences. Through qualitative interviews with the subjects we learned that the consciously-tracking subjects claimed that they rarely noticed details of the story in the first viewing because they were concentrating on the ability to control the outcome of the movie. However, when asked more in depth about details from the first viewing, they were still able to recall details such as the shifty man, his bandaged hand, the lock on the suitcase, and the facial expressions of the actors. The subjects unaware of their gaze-tracking interaction were not at all aware that they were controlling the narrative branching of the movie. Of the gaze-aware subjects, only one suggested that we were measuring the attention in specific scenes. The others thought we were measuring throughout the movie, while two subjects thought we were measuring in the lobby scene, where three actors are in the same scene. In other words, none of them could say for sure at which point they were being tracked, whether it was in a certain scene, in several scenes, or throughout the movie.

One of the gaze-aware subjects expressed that he felt like being a passenger on a bus, i.e. he just went along for the ride, but did not necessarily know where the bus was going. Another such subject said she completely forgot that she could control the film, and instead focused on what was happening. This is an indication that at least some viewers would be able to ignore or even forget about gaze-guided control rather easily.

Neither the gaze-aware or the gaze-unknowing subjects were forced to stop at any point in the narrative of the story because of interface issues. However the subjects who were conscious of the ability to control the narrative said that they constantly looked for some sort of confirmation that they were indeed controlling the narrative choices. In other words, they were looking for feedback from the system. The issue of how to provide non-intrusive feedback seems to be fundamental for a successful development of a gaze-interactive media. We have suggested some of the possibilities in the previous sections but only experiences gained from real productions of a substantial length will be able to show if they actually work.

9 Conclusions and perspectives

The key principle guiding the choice of techniques for gaze-interactive media is to preserve the viewer's immersion in the story. We have demonstrated that gaze-guided selection of story branches may occur without the viewer's awareness. We have suggested means by which the system could attract user attention, or give feedback on user actions, in ways that do not interfere with the immersive viewing experience of movies. We consider it to be a successful pilot project for further research on how gaze-interactive media can be effectively developed. The shortness of the test sequence, however, necessitates further investigations using more complex and lengthy productions before definitive conclusions are made.

Widely available eye tracking equipment based on consumer hardware will have a fairly limited resolution in the foreseeable future. As we have argued, there is no need to define tiny or fine resolution areas of narrative importance. Effective gaze interactivity can be implemented using as few as two broad areas of narrative importance key to determining narrative branching. Given these observations from the research conditions we believe that gaze-interactivity is very applicable to the wide consumer market.

Acknowledgement

The research was partly supported by the Danish Ministry of Science, Technology and Innovation and by the EU Network of Excellence "COGAIN", Communication by Gaze Interaction, funded under the FP6/IST programme of the European Commission.

References

Baudisch, P., DeCarlo, D., Duchowski, A. T. and Geisler, W. S. (2003) Focusing on the essential: considering attention in display design. *Communications of the ACM* **46**(3) 60–66. Available: http://doi.acm.org/10.1145/636772.636799

Bolt, R. A. (1984) *The human interface – where people and computers meet*. Lifetime Learning Publications, Belmont, California.

Douglas, Y. and Hargadon, H. (2000) The pleasure principle: immersion, engagement, flow. *Proceedings of the 11th ACM on Hypertext and Hypermedia*, San Antonio, Texas, United States. ACM Press, pp. 153–160.

Gerrig, R. J. (1993) *Experiencing narrative worlds. On the psychological activities of reading*. Westview Press, New Haven and London.

Gianetti, L. (1990) *Understanding movies*. 5th ed. Prentice Hall, Englewood Cliffs, NJ.

Hansen, D. W. and Pece, A. E. C. (2003) Iris rracking with feature free contours. *Proc 2003 IEEE International Workshop on Analysis and Modelling of Faces and Gestures*, Nice, France. pp. 208–214.

Hansen, J. P., Andersen, A. W. and Roed, P. (1995) Eye-gaze control of multimedia systems. In Anzai, Y., Ogawa,K. and Mori, H. (eds) *Advances in human factors/ergonomics: symbiosis of human and artifact*. Elsevier Science B.V., pp. 37–42.

Hochberg, J. and Brooks, V. (1978) Film cutting and visual momentum. In Senders, J. W, Fisher, D. F. and Monty, R. A. (eds) *Eye movements and the higher psychological functions*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 293–313.

Murray, J. (1997) *Hamlet on the holodeck: the future of narrative in cyberspace*. The Free Press, New York.

Partala, T., Jokiniemi, M. and Surakka, V. (2000) Pupillary responses to emotionally provocative stimuli. *Proceedings of the Symposium on Eye tracking research and applications 2000*. ACM Press, Palm Beach Gardens, Florida, pp. 123-129.

Ryan, M.-L. (2001) *Narrative as virtual reality; immersion and interactivity in literature and interactive media*. The Johns Hopkins University Press, Baltimore, Maryland.

Salvucci, D. D. and Goldberg, J. H. (2000) Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the symposium on Eye tracking research and applications 2000*. ACM Press, pp. 71–78.

Samsel, J. and Wimberley D. (1998) *Writing for interactive media – the complete guide*. Allwoth Press, New York.

Starker, I. and Bolt, R. A. (1990) A gaze-responsive self-disclosing display. *Proc of the SIGCHI conference on Human factors in computing systems*. ACM Press, Seattle, Washington, , pp. 3–10. Available: http://doi.acm.org/10.1145/97243.97245

Stelmach, L. B., Tam, W. J. and Hearty, P. J. (1991) Static and dynamic spatial resolution in image coding: an investigation of eye movements. *SPIE* **1453** 147–152.

Velichkovsky, B. M. and Hansen, J. P. (1995) New technological windows into mind: there is more in eyes and brains for human-computer interaction. *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, Vancouver, Canada, pp. 496—503.

Tore Vesterby holds an MSc Degree in Information Technology from the IT University of Copenhagen and a BA in Danish from Roskilde University. He has previously worked for private companies in CD-ROM production and web design. His research interests lie in the use of digital media as entertainment with a strong focus on gender issues.

Jonas C. Voss holds a BSc in Human Geography. He is currently undertaking a Masters Thesis on eye-tracking and computer games at the IT University of Copenhagen (ITU). His research interests lies within the field of eye tracking and attentive user interfaces.

John Paulin Hansen is Associate Professor at the IT University of Copenhagen. He received an MSc in 1984 and a PhD in 1992 both in Psychology from the Institute of Psychology, University of Aarhus. John Paulin Hansen is currently project leader of the Eye Gaze Research Team, designing and managing the development of gaze-based communication systems for disabled people.

Dan Witzner Hansen is Assistant Professor at the IT University of Copenhagen. He received an Msc in 1999 from the University of Aarhus and a PhD degree in 2003 from the IT University of Copenhagen. His main research topics are within real-time computer vision with a special emphasis on obtaining information about human activity, recognition and tracking in time sequences.

Arne John Glenstrup received his BSc in 1995 from the University of Copenhagen in Computer Science and Information Psychology with special interest in eye gaze tracking and its connection to user interest detection and multimodal user interaction. He received his MSc in 1999 from the University of Copenhagen and his PhD from the Technical University of Denmark in 2002. Since 2003, he has been Assistant Professor at the IT University of Copenhagen, where his research includes programming languages for mobile and context-dependent applications, as well as ubiquitous computing and location models.

Mark Rudolph is an interactive 3D artist and director, Java designer and musician. He worked at Silicon Graphics in California as a VRML director and Java designer, and at AT&T Bell Labs undertaking multimedia research. He also has a PhD in Mathematics in Transfinite Set Theory. Currently he is an Associate Professor at the IT University in Copenhagen.