HAL
open science

A Psychophysical Study of Fixation Behavior in a Computer Game

Veronica Sundstedt, Efstathios Stavrakis, Michael Wimmer, Erik Reinhard

To cite this version:

Veronica Sundstedt, Efstathios Stavrakis, Michael Wimmer, Erik Reinhard. A Psychophysical Study of Fixation Behavior in a Computer Game. Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization, Aug 2008, Los Angeles, United States. pp.43-50. inria-00606816

HAL Id: inria-00606816

https://inria.hal.science/inria-00606816v1

Submitted on 13 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Psychophysical Study of Fixation Behavior in a Computer Game

Veronica Sundstedt* University of Bristol
Efstathios Stavrakis† REVES-INRIA
Michael Wimmer ‡ Vienna University of Technology
Erik Reinhard§ University of Bristol

Abstract

Prediction of gaze behavior in gaming environments can be a tremendously useful asset to game designers, enabling them to improve gameplay, selectively increase visual fidelity, and optimize the distribution of computing resources. The use of saliency maps is currently being advocated as the method of choice for predicting visual attention, crucially under the assumption that no specific task is present. This is achieved by analyzing images for low-level features such as motion, contrast, luminance, etc. However, the majority of computer games are designed to be easily understood and pose a *task* readily apparent to most players. Our psychophysical experiment shows that in a task-oriented context such as gaming, the predictive power of saliency maps at design time can be weak. Thus, we argue that a more involved protocol utilizing eye tracking, as part of the computer game design cycle, can be sufficiently robust to succeed in predicting fixation behavior of players.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality;

Keywords: Psychophysics, Electronic Games, Eye Tracking, Visual Attention, Eye Movements, Saliency

1 Introduction

In the design of interactive applications, notably games, a recent trend is toward tapping the potential of psychophysics for the purpose of understanding player behavior. Predicting where players are likely to focus could be a very useful tool in the arsenal of game designers. The purpose of such predictive algorithms would for instance be to help game designers decide how and where to allocate rendering resources, leaving as many computing cycles as possible free to carry out a variety of tasks. Verifying game mechanics or improving game AI could be other uses.

A trend is toward the use of saliency maps [Itti et al. 1998], which analyze the imagery presented to the user for low-level features such as motion, contrast, and local luminance levels. Their success lies in the fact that these features are attention-grabbing, making it likely that users focus their gaze on them. Rendering such features at full fidelity and the remainder of the image at lower quality could therefore go unnoticed, while saving precious computing cycles. However, saliency maps work under the assumption that no specific task is present. This is an important limitation, as it is well known that the presence of a task may alter the way in which users focus their attention [Yarbus 1967; Hayhoe et al. 2003].

*e-mail:veronica@cs.bris.ac.uk
†e-mail:efstathios.stavrakis@sophia.inria.fr
‡e-mail:wimmer@cg.tuwien.ac.at
§e-mail:reinhard@cs.bris.ac.uk

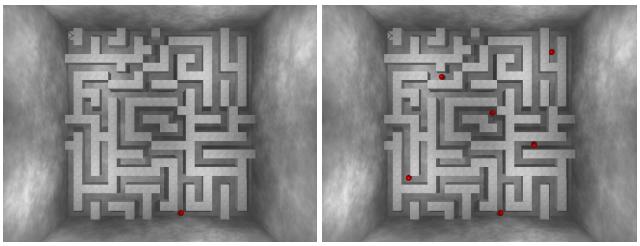


Figure 1: *The design of the maze using a single user-controlled object, a ball shown here in its starting position (left), as well as the maze with five additional distractor balls (right).*

In this paper, we perform a psychophysical eye tracking experiment, carried out on an easy-to-understand game (shown in Figure 1), which was purpose-designed for psychophysics. Our experiment demonstrates the extent to which gaze behavior may be altered by the presence of a task, even if participants are only instructed to observe a pre-recorded video, i.e. in the absence of a specific task.

Our main result shows that even passive user behavior is task dominated and cannot be statistically distinguished from active gameplay behavior. In our game design, saliency is therefore a very poor predictor for task relevance. Additionally, we experiment with further game variations where attention distracting or attention focusing elements are present. Neither case had a measurable effect on the observed fixation patterns, and therefore on where attention was focused.

On the basis of our experiment we argue that, at least in the context of electronic games, a task is nearly always implied. Rather than relying on saliency algorithms, we therefore advocate the use of eye tracking during the design cycle of electronic games, as we show that computational methods to predict fixation behavior are potentially unreliable.

2 Background

For the purpose of living in complex environments, humans rely strongly on vision, which consists of two broad components. The first is *perception*, which is pre-attentive. The second is *cognition*, which involves high-level processes such as thought, reasoning and memory [Palmer 1999]. The delineation between these two is not sharp, and much feedback exists between the two.

When carrying out a task, human visual perception aggregates low-level features into higher level representations, thus informing cognitive processes while affecting gaze direction. In turn, cognitive processes can guide perception, for instance by actively focussing attention on a particular part of the scene [Yarbus 1967].

The focus of attention can thus be classified as a combination of perceptual (bottom-up) and cognitive (top-down) processes [Itti and Koch 2001]. Low-level features in the environment that trigger pre-

attentive focus are called salient. Features that attract attention as a result of performing a specific task are called task relevant.

The first computational models concentrated on modeling gaze behavior using low-level features such as color, intensity, and orientation [Triesman and Gelade 1980; Koch and Ullman 1985; Itti et al. 1998]. Such models compute for each pixel of a frame a measure of saliency, the result of which is called a saliency map. However, it was shown that task-related gaze behavior can dominate over saliency [Land et al. 1999]. Per-pixel measures of task relevance have more recently appeared, and these are called task maps [Cater et al. 2002; Navalpakkam and Itti 2005].

There are various application areas, including computer graphics, virtual reality, and games, where saliency and task models have been used with varying degrees of success. In graphics for example, these models have been used to inform global illumination algorithms [Yee et al. 2001; Haber et al. 2001; Cater et al. 2003; Sundstedt et al. 2007].

Experiments using virtual environments, with stimuli presented to the participant via a head-mounted display, have shown that saliency is not a good predictor when performing natural tasks [Rothkopf et al. 2007]. It has also been shown that when saliency is extended with top-down features, gaze behavior can be predicted more accurately [Canosa 2003; Lee et al. 2007].

Recent studies suggest that in adventure games, fixation behavior can follow both bottom-up and top-down processes [El-Nasr and Yan 2006]. Visual stimuli are reported to be more relevant when located near objects that fit players' top-down visual search goals. In first-person shooter games, gaze tends to be more focused on the center of the screen than in adventure games [Kenny et al. 2005; El-Nasr and Yan 2006]. In an experiment involving active video game play, nine low-level heuristics were compared to gaze behavior collected using eye tracking [Peters and Itti 2008]. This study showed that these heuristics performed above chance, and that motion alone was the best predictor. This was followed by flicker and full saliency (color, intensity, orientation, flicker, and motion). Nonetheless, these results can be improved further by incorporating a measure of task relevance, which could be obtained by training a neural network on eye tracking data matched to specific image features [Peters and Itti 2007].

Our hypothesis is that the success of low-level image features, collected into saliency maps through computational models, depends strongly on the type of task performed. In the context of games, saliency, as measured by contrast, luminance, edges, and the like, may correlate with task relevance if this is designed explicitly into the game. However, this need not be the case: saliency may correlate very poorly with fixation behavior, and therefore with attention. In this paper we present a study that demonstrates this lack of correlation. The consequence is that task relevance, while more difficult to model, should play a larger role in models of gaze behavior than currently assumed.

3 Psychophysical Experiment

We designed an experiment to analyze the validity of the concept of saliency in the context of games. Our first hypothesis is that each object type in the maze carries different significance with respect to solving the maze. The distribution we compute therefore relates directly to the importance that each object type carries for executing the task. Within the maze, for instance, we expect players to focus more on the one path that leads from the starting position to the target than they would focus on dead ends, and other task irrelevant areas and objects. Note that such a distribution cannot be predicted by current saliency maps.

The second hypothesis is that solving the maze actively does not lead to statistically different behavior from passively viewing a prerecorded game. This is on the basis that the task involved in solving a maze is directly evident, even if no specific instructions are given to solve the maze. Finally, we make a first attempt at assessing the ability of additional features to add to or distract from the task.

Our aim is therefore to design a game which is sufficiently complex to bear relevance to real games, while being controlled enough to allow rigorous analysis. We have found that a good trade-off between playability and visual complexity on the one hand, and the necessity to obtain robust and reproducible data, is afforded by a maze puzzle (Section 3.2). In conditions involving active game play, the task is to navigate a red ball from its starting position to a designated target position. To make the task more challenging, a physics engine simulates various attributes of the ball, including weight, acceleration, and how it bounces off other surfaces.

After a calibration step, performed for each participant individually, we employ a Tobii x50 eye tracking device to record gaze behavior. The raw eye tracking data are subsequently filtered to detect fixations (Section 3.3), which are mapped back to different object types in the scene (Section 3.4). For each participant, we compute the distribution of fixation points over the different object types that can be discerned in the scene (Section 3.5). This distribution is computed for every participant, over a set of different conditions (Section 3.6), before being analyzed using standard techniques (Section 4).

3.1 Experimental Setup

The Tobii x50 eye tracker, running at 50 Hz, was placed in front of the display, and angled toward the participant according to the manufacturer's instructions. We employ a chin rest, ensuring that the participants' eyes are 60 cm away from the display, and remain accurately positioned within the detection region of the eye tracker. The game itself was modified so that it can start and stop the eye tracker automatically at the start and end of each trial.

All experiments were carried out on a Pentium 4 PC clocked at 3.46 GHz with 2 GB RAM, and an NVIDIA 8600 graphics card. This setup is sufficiently powerful to run the game (at 16 frames per second), record the game state, as well as drive the eye tracker. The display is a Viglen TS700 LCD display with dimensions of 34 $\times$ 27 cm, and a resolution of 1024 $\times$ 768 pixels. The setup was located in a blackened room with fluorescent lighting built into the ceiling, positioned to avoid undue reflections off the screen.

3.2 Game Design

The design of the maze is shown in Figure 1. Several features are included to help improve the robustness of the results. In many game designs, the camera follows an avatar representing the player, or, in the case of first-person shooter games, the player has full control over which part of the environment is explored. This would not affect our ability to collect data, but does not let us pool the data obtained from different participants within the same condition. Hence, we have fixed the view point, so that each participant has access to the same information during each trial.

The maze itself includes several design features which help us maintain control over the robustness of the results. The enclosing walls ensure that the playing field is located away from the edges of the display, where we have found the accuracy of the eye tracker to be unsatisfactory. The total lengths of the paths in each quadrant of the screen are approximately equal. In addition, we have inserted five separate regions of equal length in which we may place distracting objects. These regions are not connected to the paths

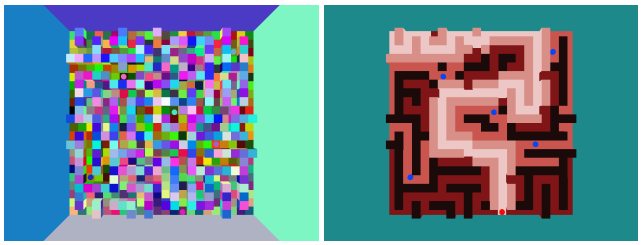


Figure 2: *Item buffer showing all color coded objects which enable us to relate fixation points to objects and object classes (left), as well as the different object classes in our experiment (right).*

accessible by the player to avoid polluting the task with the possibility of collisions. With the exception of the balls, all surfaces are textured with a random achromatic texture which exhibits $1/f$ image statistics [Deriugin 1956], bringing the imagery statistically closer to natural scenes. The contrast of the texture was reduced to ensure that paths and walls are easily distinguished, thereby not hindering the task. The balls are colored red to make them easily distinguishable from the remainder of the scene. The scene is illuminated approximately uniformly with four light sources, with one located above each quadrant of the playing field.

3.3 Fixation Detection

The eye tracker outputs a sequence of time-stamped (x,y) pixel coordinates, representing momentary gaze points, which can subsequently be analyzed. It is typically assumed that if gaze points remain within a small region during a time exceeding 200 ms, then we may speak of a fixation point [Salthouse et al. 1981]. The premise is that such fixation points correlate with the focusing of attention, and therefore constitute our features of interest [Duchowski 2003].

The eye tracker is rated to be accurate to within a solid angle of 0.7 degrees. Taking into account the distance between the observer and the display, and its size and resolution (see Section 3.1), this corresponds to an error of ± 11 pixels.

We have implemented a fixation filter which detects fixations of 200 ms or longer (i.e., > 10 consecutive sample points), whereby gaze points remain within a circle with a radius of 11 pixels of each other. The output of this filter is a set of time intervals with an associated center point of the fixation.

3.4 Item Buffer

We are interested in determining which objects in the scene are focused on, at any time during an experiment's session. The fixation points are therefore mapped to the objects making up the scene. To this end, each scene object is first assigned a unique number that represents a color value. Each frame of the session is then rendered out as an *item buffer* image, shown in Figure 2 (left), which color codes all scene objects. Each item buffer is also stamped with a start and end time over which it was displayed. Thus fixation points can be subsequently correlated, both temporally and spatially, with the scene objects using these item buffers.

3.5 Fixation Binning

The item buffer is used to map fixation points to objects. During a trial, we count how many times each object was fixated upon. In eye tracking experiments, it is common to use relatively sparsely populated environments, because the analysis of fixation points becomes more complicated if several objects are located in the vicinity of a fixation point [Lee et al. 2007]. A common technique is to simply select the object directly underneath the fixation point.

Name	Description
Closed paths	These are the parts of the maze that are separate from the main paths and may or may not contain distracting elements.
Correct path	The floor and walls of the path that must be traversed to go from the starting point to the designated destination.
Incorrect path	The floors and walls of dead ends.
User-contr. ball	The ball under the participant's control.
Distractor balls	Balls not under participant control.
Main walls	The four enclosing the playing field.
Top surface	The top surface of the walls that are not adjacent to the correct path.
Adj. to correct	The top surface of the walls that are adjacent to the correct path.

Table 1: *Description of object classes. The color coding is consistent with Figure 2 (right).*

However, games environments can be cluttered and complex, making such a simple approach error-prone. We have therefore developed a novel approach, taking the distance between pixels and a fixation point into account. With the accuracy of the eye tracker at 0.7°, and the foveal region of human vision spanning approximately 2° of visual angle [Palmer 1999], the area over which a fixation point bears relevance is a circle with a radius of around 43 pixels (corresponding to a solid angle of 2.7°).

Considering frame number f, we assign a non-zero weight w_f to all pixels (x_p, y_p) in a window of 400×400 pixels (large enough to be accurate, and small enough to be computationally efficient) according to their distance to the fixation point (x_f, y_f):

$$w_f(x_p, y_p) = \exp\left(\frac{-\sqrt{(x_f - x_p)^2 + (y_f - y_p)^2}}{2\sigma^2}\right) \qquad (1)$$

where $\sigma = 43$ pixels. With the aid of the item buffer, the weights computed for frame f are then added to the weights of the different objects. This process is repeated for all frames for which a fixation point exists[1]. After normalization, necessary to account for the fact that each trial may last for a different amount of time, this produces a weight $\hat{w}^i$ associated with object i.

In the design of the maze, different objects may have the same relevance to the task. For instance, the walls are made up of different segments. Therefore, we group objects playing a similar role into object classes. The classes we distinguish are based upon their anticipated relevance to solving the maze (described in Table 1). A color-coded image showing the location of each of these elements is shown in Figure 2 (right). After classification, each participant produces a normalized distribution of fixations per object class. This set of distributions is then subjected to further analysis, as discussed in Section 4. We first describe the collection of conditions used in our experiment.

3.6 Conditions

The four conditions are developed to test the distribution of fixation points over object classes in active and passive tasks, as well as in

[1] Note that not all frames have an associated fixation point, as saccadic eye movements are of non-zero duration.

the presence of distracting and focus-enhancing elements. In the first condition, participants were asked to navigate the ball to the destination position, located in the top left corner of the playing field. No distracting objects or sound is present in this condition.

The second condition adds five static distracting spheres to the five separate areas. The task was otherwise identical to the first condition. In the third condition, we additionally add two types of sound, both connected to the sphere under participant control. A continuous sound change pitch in association with the current speed of the ball. A transient sound indicates collisions. It is anticipated that these sounds help with the control of the sphere, rather than distract.

For the remaining condition, the game was modified to enable its state to be recorded. This is achieved by writing all events (such as object movements) which affect the game state to a file. This is less data and compute intensive than encoding a video of the game. It can therefore proceed in parallel to gathering gaze data from the eye tracker, as well as driving the game itself without affecting the responsiveness of the game, or its frame-rate.

Further, the recorded game data can be played back in real time, with the same frame-rate as the original recording, yielding an exact replica of an earlier trial. Each participant in the second condition was recorded in this manner, producing data for the fourth condition. This last condition consists of participants observing pre-recorded game data, without having been set a task. As the same number of participants were used in each condition, each participant in the fourth condition was exposed to a different recording obtained in the second condition.

3.7 Participants

Forty participants (36 men and 4 women, age range $18-42$) with normal or corrected to normal vision were recruited for our experiments. As the eye tracker occasionally loses calibration during a trial, especially for participants wearing glasses, trials which have clearly produced unreliable data were replaced with new ones, for which new participants were recruited. No participant was exposed to more than one condition, so that learning effects between trials could be avoided. As a result, each condition was evaluated by exactly ten participants. The participants were asked to read instructions before the start of the experiment. The instructions for the active and passive conditions are shown in Appendix A.

4 Results

First, we wanted to analyze whether habituation plays an important role in the analysis. It may be anticipated that participants learn the path to the destination first, and then spend the remainder of the trial steering the ball. It is equally likely that participants frequently alternate between navigation and steering. To assess whether participants change their fixation behavior during the course of a trial, we have computed the number of fixations occurring per second for each participant and each condition. For each point t on the time line, we counted the number of fixations in a window spanning $t \pm 15$ seconds (except near the start and end of trials, where the window size was reduced to accommodate for the lack of data).

The results are plotted in Figure 3, and indicate that fixation behavior for each participant is relatively stable over time. As a result, we infer that participants do not change their solution strategy during the course of the trial. It is more likely that most participants use a strategy whereby navigation and steering are alternated. The average trial duration for the four conditions was 72 ± 12 seconds.

Given that the number of fixations per second is stable over the course of each trial, we conclude that habituation and learning effects do not play a major role, and therefore do not complicate our analysis. There is also no significant difference between the four conditions in the total durations of the trials ($F(3, 36) = 1.767$, $p = 0.171$), the number of fixations ($F(3, 36) = 0.291$, $p = 0.832$), or the total fixation durations ($F(3, 36) = 1.092$, $p = 0.365$).

The distribution of fixations over the different object classes for each participant and each condition is shown in Figure 4. We note that these distributions are markedly different from the distribution one would obtain by counting the number of pixels that are covered by each object type (shown in Figure 5). Thus, we conclude that none of the results presented next can be explained by random fixation behavior.

The first important observation is that the number of fixations attracted by the *correct path* and its neighboring *adjacent to correct* form the majority of all fixations points, regardless of condition. This can also be seen by plotting the fixation points on top of the maze stimulus, as shown for a representative participant of condition 2 in Figure 6 (a). The fixation points in this figure roughly map out the correct path between start and destination points, showing that the task of solving the maze has a strong impact on where fixation points are likely to occur. The fixation points for a representative participant of condition 4 are shown in Figure 6 (b).

The raw data of Figure 4 were subjected to a one-way analysis of variance (ANOVA), the null hypothesis being that no statistically significant differences between these distributions are found. All preconditions for ANOVA regarding independence and normality were fulfilled. By using the four conditions as groups (our independent variable), we found that the test of homogeneity of variances was met for all the object classes (our dependent variables), bar the *distractor balls* ($p = 0.001$).

Tukey's HSD post hoc test shows that there are statistical differences between condition 1 and the other conditions for this object class ($p = 0.000$). This is explained by the fact that no distractors exist in one of the conditions (their weight was set to 0 in condition 1). If we exclude condition 1, then the test of homogeneity of variances was met for all object classes. The ANOVA values for both tests are shown in Table 2.

We find that the null hypothesis is supported for all object classes over all four conditions, with the exception of the *distractors* class. Once again, if we exclude condition 1 from the analysis, then the null hypothesis is supported for all object classes in the three remaining conditions. Thus, we have found no statistical differences in fixation behavior of participants in each of the four conditions.

There are several implications of this result. First, comparing conditions 2 and 4, we see that there is no significant difference in the distribution of fixations, independent of whether the game was actively played or passively observed. This result is important, because it proves that the absence of a task (other than simple observation) does not guarantee that participants will not assume a task. For this particular game, free viewing of a pre-recorded game encourages the observers the solve the task "in their heads". As in this case the distribution of fixation patterns continues to follow that of the active condition, we conclude that any computational metric that assumes the absence of a task cannot be applied. Importantly, this includes saliency algorithms.

Second, comparing conditions 1 and 2, we see that the addition of five distracting objects did not have any effect on task performance: the distribution of fixations over all object classes remains the same, with the exception of the distracting objects themselves which were absent in condition 1. Once more, this result would not be predicted by saliency algorithms, which would indicate a difference between these conditions due to the contrasting color of the distracting balls.

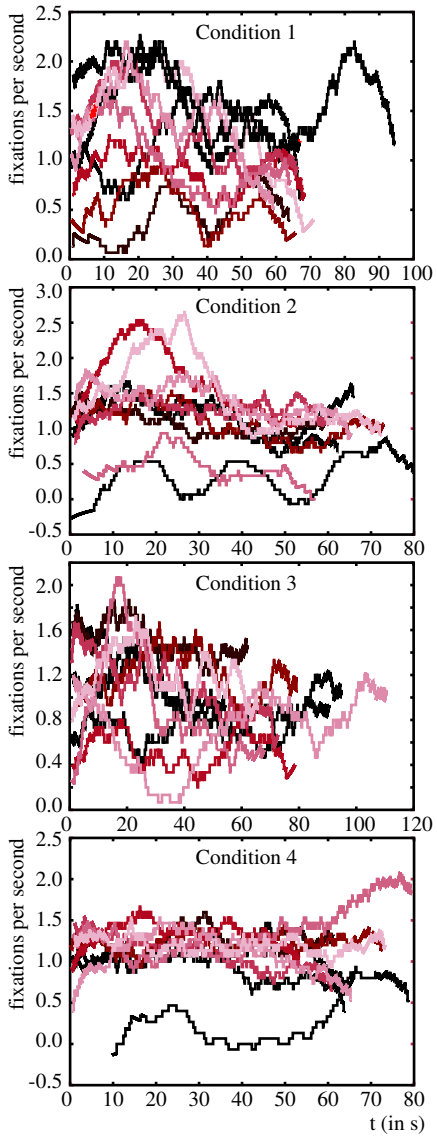


Figure 3: *The number of fixations per second for each participant.*

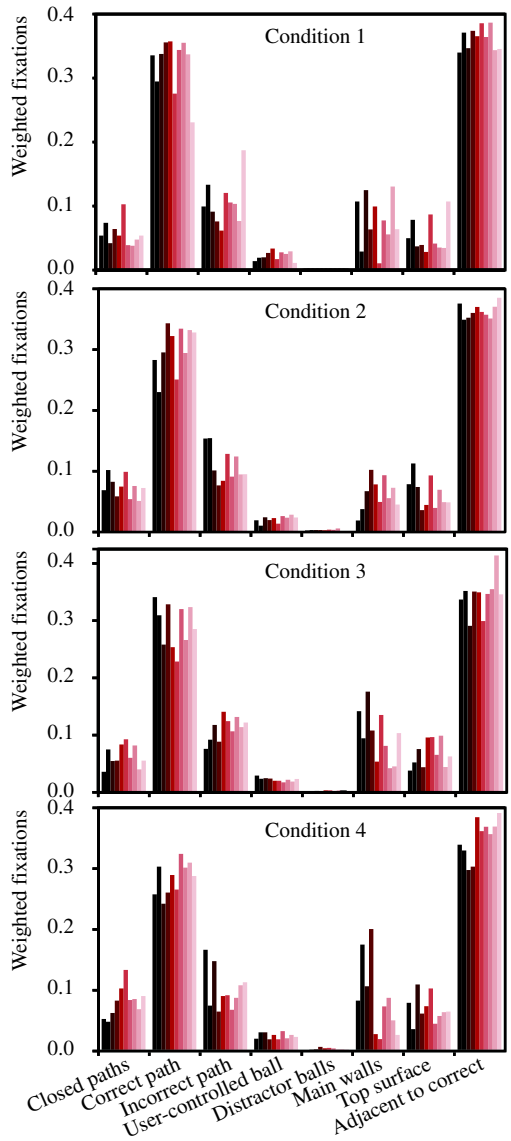


Figure 4: *The output of our fixation classification algorithm.*

Third, conditions 2 and 4 enable us to assess the influence of sound, which was present in condition 4. Here, rather than distract, the sounds were designed to support the steering component of the task. We would anticipate this to have an influence on the distribution of fixations over object classes. However, we did not find such an influence, showing that the task was not facilitated by including an auditory indication of speed and the occurrence of collisions.

4.1 Comparison with Saliency Maps

While saliency maps are currently advocated as a useful tool to understand fixation behavior in games [Peters and Itti 2008], on the basis of our experiments, we are now able to argue that such algorithms can only be used with the utmost care, and possibly in the context of games should be avoided.

In a sense, the design of the maze forms a worst-case scenario for saliency maps. These algorithms determine saliency on the basis of low-level image features. In our design, these image features are more or less evenly distributed across the playing field. A saliency map will therefore have a substantially uniform distribution, as shown in Figure 6 (c), which plots the saliency map computed using Itti and Koch's [1998] algorithm. Comparing this with the actual distribution of the fixation points (Figure 6 (a and b)), we see that the actual fixation pattern is not predicted by the saliency algorithm, even if no task is specified.

The number of times that the fixation point was in the vicinity of the ball that was being controlled was unexpectedly low (see Figure 4). As this was the only object that moves, we cannot attribute the differences between the outcome of our experiment and the saliency maps shown in Figure 6 (c) to the fact that we did not take motion into account for the computation of the saliency map.

Figure 6 (d) shows the relative importance assigned to the object classes based on the eye tracking results from condition 2. Note that the unevenness of the importance distribution suggests that our initial division of objects into object classes was highly correlated with the task performed. Such correlation does not exist between

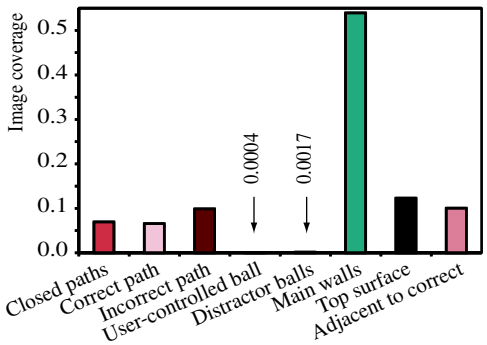


Figure 5: *Pixel coverage for the object classes. If fixations were randomly distributed over the image, this distribution would be obtained. Compare with Figure 4, showing the actual distribution of fixations. The color coding is consistent with Figure 2 (right).*

Object classes	Four Conditions	Conditions 2, 3, and 4
Closed paths	(F(3, 36) = 2.748, p = 0.057)	(F(2, 27) = 1.843, p = 0.178)
Correct path	(F(3, 36) = 2.052, p = 0.124)	(F(2, 27) = 0.619, p = 0.546)
Incorrect path	(F(3, 36) = 0.239, p = 0.869)	(F(2, 27) = 0.397, p = 0.676)
User-contr. ball	(F(3, 36) = 0.885, p = 0.458)	(F(2, 27) = 1.738, p = 0.195)
Distractor balls	(F(3, 36) = 21.99, p = 0.000)	(F(2, 27) = 3.002, p = 0.066)
Main walls	(F(3, 36) = 1.149, p = 0.343)	(F(2, 27) = 1.554, p = 0.230)
Top surface	(F(3, 36) = 0.775, p = 0.516)	(F(2, 27) = 0.105, p = 0.901)
Adj. to correct	(F(3,36) = 1.376, p = 0.266)	(F(2, 27) = 1.292, p = 0.291)

Table 2: *Statistics for each object class. The second column shows the results if all four conditions are included. Due to the absence of distractor balls in condition 1, we have repeated the analysis for conditions 2, 3, and 4 in the last columns.*

the saliency map (see Figure 6 (c)) and the eye tracking results. Comparing conditions 1 and 2, the distracting objects do not alter the distribution of fixations over object classes. However, their contrasting color does produce a pattern of saliency, which is not supported by our experiments.

We assume that for many games, designers frequently adjust the amount of detail, motion, contrast, luminance, etc., to direct players toward their goal: objects more relevant to the task to be accomplished may exhibit more motion, contrast, etc., than objects with lower task relevance. This may explain the fact that in some games, saliency maps are able to predict fixation behavior better than chance [Peters and Itti 2008]. However, our experiment shows that this correlation does not necessarily generalize to other game scenarios. Finally, there is a possibility that the predictive power of saliency maps is in essence a measure of how much game designers make their task-relevant objects stand out from the background.

5 Conclusions

The aim of our work is to assess the distribution of fixation points over a set of pre-determined object classes. The object classes were chosen to correspond with the expected relevance to the task of solving a maze and steering a ball through it. We find that the distribution of fixation points deviates significantly from those which would be predicted by current saliency algorithms. Moreover, no difference was found between active game play and passive observation.

We attribute this result to the fact that tasks and objectives in games are typically designed to be easily apparent or understandable. This means that even if no task is specified, observers will assume a task. Related work have shown that when an observer is watching a block stacking task, the gaze prediction is highly similar to when they perform the task themselves [Flanagan and Johansson 2003].

Whereas existing studies explored average-case behavior by using existing games [Kenny et al. 2005; El-Nasr and Yan 2006; Peters and Itti 2008], i.e. without control of salient features in the game design, our work has for the first time created a worst-case scenario for saliency algorithms. As our maze design shows, actual fixation behavior can deviate by an arbitrary amount from the predicted saliency. Further, objects which a saliency algorithm would flag as distracting do not necessary impede task efficiency.

We hypothesize that the correspondence between fixation behavior and saliency algorithms depends strongly on the efforts of the designer to impose salient characteristics to objects that happen to be task-relevant. This may explain the above chance performance of saliency algorithms in existing games [Peters and Itti 2008]. As a result, in the context of game design we would argue for caution when using saliency algorithms.

Instead, the direct use of eye tracking during the design of games can be extremely valuable. We have shown that a relatively straightforward modification to the game engine (namely tagging each object with a color-coded identifier), is sufficient to map fixation points back to objects. While eye tracking sessions during actual game design by necessity will be less controlled than our experiment, simply accumulating fixations over different object classes may prove to be a fruitful approach in understanding where game players focus their attention. Such information cannot currently be extracted from an analysis of low-level features alone.

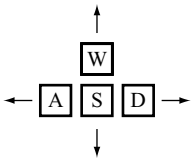
Acknowledgments

This research was sponsored through the European Union within the CROSSMOD project (EU IST-014891-2). We would also like to thank Daniel Scherzer, Heinrich Fink, Frühstück Anton, Matthias Bernhard, Isabelle Viaud-Delmon and Clara Suied.

A Appendix

For participants in conditions 1, 2, and 3 (all actively playing the game), the instructions are as follows:

"Please read this information carefully: The task in this experiment is to play a game. The goal of the game is to guide a ball through a maze. The finish point of the maze is indicated with a cross, as shown below. When you reach the finish point the game will stop and the experiment is over. The ball can be moved in four directions by using the keys shown below. Please ask now if you have any questions."



An example maze was also shown to the participants. For participants in the passive viewing condition, the instructions are:

"Please read this information carefully: In this experiment you will be shown an animation. Please watch it. If you have any questions please ask now."

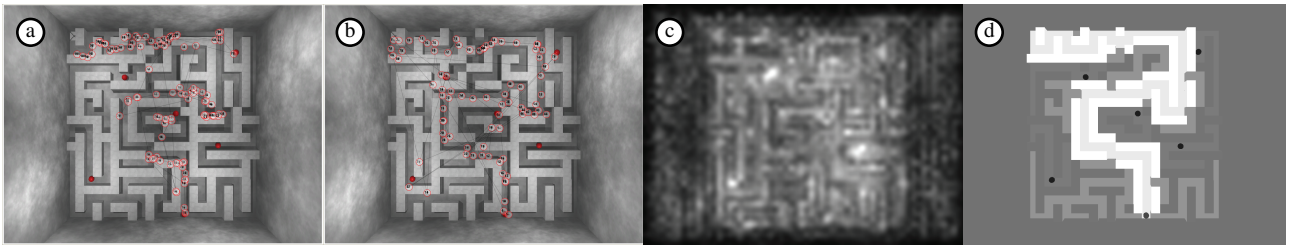


Figure 6: *The location of fixation points for a representative participant of condition 2 (a) and the passive condition 4 (b). Image (c) shows a saliency map of the maze, computed using a saliency algorithm [Itti et al. 1998] and (d) shows our distribution map over object classes generated from the ten participants in condition 2.*

References

CANOSA, R. 2003. *Seeing, sensing, and selection: modeling visual perception in complex environments*. PhD thesis, Rochester Institute of Technology, Rochester College of Science.

CATER, K., CHALMERS, A., AND LEDDA, P. 2002. Selective quality rendering by exploiting human inattentional blindness: looking but not seeing. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, 17–24.

CATER, K., CHALMERS, A., AND WARD, G. 2003. Detail to attention: Exploiting visual tasks for selective rendering. In *Proc. of the 14th Eurographics Workshop on Rendering*, 270–280.

DERIUGIN, N. G. 1956. The power spectrum and the correlation function of the television signal. *Telecommunications 1*, 7, 1–12.

DUCHOWSKI, A. T. 2003. *Eye tracking methodology: Theory and practice*. Springer, New York.

EL-NASR, M. S., AND YAN, S. 2006. Visual attention in 3d video games. In *Proc. of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, 22.

FLANAGAN, J. R., AND JOHANSSON, R. S. 2003. Action plans used in action observation. *Nature 424*, 6950, 769–771.

HABER, J., MYSZKOWSKI, K., YAMAUCHI, H., AND SEIDEL, H.-P. 2001. Perceptually guided corrective splatting. *Computer Graphics Forum 20*, 3, 142–152.

HAYHOE, M. M., SHRIVASTAVA, A., MRUCZEK, R., AND PELZ, J. B. 2003. Visual memory and motor planning in a natural task. *J. Vis. 3*, 1 (2), 49–63.

ITTI, L., AND KOCH, C. 2001. Computational modeling of visual attention. *Nature Reviews Neuroscience 2*, 3, 194–203.

ITTI, L., KOCH, C., AND NIEBUR, E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence 20*, 11, 1254–1259.

KENNY, A., KOESLING, H., DELANEY, D., MCLOONE, S., AND WARD, T. 2005. A preliminary investigation into eye gaze data in a first person shooter game. In *19th European Conference on Modelling and Simulation*.

KOCH, C., AND ULLMAN, S. 1985. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology 4*, 4, 219–227.

LAND, M., MENNIE, N., AND RUSTED, J. 1999. The roles of vision and eye movements in the control of activities of daily living. *Perception 28*, 11, 1311–1328.

LEE, S., KIM, G. J., AND CHOI, S. 2007. Real-time tracking of visually attended objects in interactive virtual environments. In *ACM Symp. on Virtual Reality Software and Technology*, 29–38.

NAVALPAKKAM, V., AND ITTI, L. 2005. Modeling the influence of task on attention. *Vision Research 45*, 2, 205–231.

PALMER, S. E. 1999. *Vision science: Photons to phenomenology*. MIT Press, Boston.

PETERS, R. J., AND ITTI, L. 2007. Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–8.

PETERS, R. J., AND ITTI, L. 2008. Applying computational tools to predict gaze direction in interactive visual environments. *ACM Transactions on Applied Perception 5*, 2, 22.

ROTHKOPF, C. A., BALLARD, D. H., AND HAYHOE, M. M. 2007. Task and context determine where you look. *Journal of Vision 7*, 14 (12), 1–20.

SALTHOUSE, T. A., ELLIS, C. L., DIENER, D. C., AND SOMBERG, B. L. 1981. Stimulus processing during eye fixations. *Journal of Experimental Psychology: Human Perception and Performance 7*, 3, 611–623.

SUNDSTEDT, V., GUTIERREZ, D., ANSON, O., BANTERLE, F., AND CHALMERS, A. G. 2007. Perceptual rendering of participating media. *ACM Trans. on Applied Perception 4*, 3, 15.

TRIESMAN, A. M., AND GELADE, G. 1980. A feature-integration theory of attention. *Cognitive Psychology 12*, 1, 97–136.

YARBUS, A. L. 1967. Eye movements during perception of complex objects. In *Eye Movements and Vision*, Plenum Press, New York, 171–196.

YEE, H., PATTANAIK, S., AND GREENBERG, D. P. 2001. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Trans. on Graphics 20*, 1, 39–65.