New insights on ambient and focal visual fixations using an automatic classification algorithm

Brice Follet, Olivier Le Meur, Thierry Baccino

i-PERCEPTION

New insights on ambient and focal visual fixations using an automatic classification algorithm

Brice Follet(1)*[Present address: Technicolor,
1 avenue Belle Fontaine 35510 Cesson-Sévigné, FRANCE], email: Brice.Follet@technicolor.com;
Olivier Le Meur(2) email: olemeur@irisa.fr;
Thierry Baccino(3); email: thierry.baccino@univ-paris8.fr;

1. Technicolor
1 avenue Belle Fontaine
35510 Cesson-Sévigné, FRANCE;
2. Université de Rennes 1 - IRISA
Campus Universitaire de Beaulieu
35042 Rennes Cedex, FRANCE;
3. LUTIN
Cité des sciences et de l'industrie de la Villette
30 avenue Corentin Cariou
75930 Paris Cedex 19, FRANCE

Overt visual attention is the act of directing eyes towards a given area. These eye movements are characterized by saccades and fixations. A debate currently concerns the role of visual fixations. Do they all have the same role in free viewing of natural scenes? Recent studies suggest that there exist at least two kinds of visual fixations called focal and ambient ones. The former would be used to inspect accurately local areas whereas the latter would be used to grab the context of the scene. We investigate in this paper an automatic solution to cluster visual fixations in two groups using four types of natural scene images. We give new evidences supporting a focal-ambient dichotomy. Our clustering reveals that the determining factor is the amplitude of saccade. The dependence on the low-level visual features and the time course of these two kinds of visual fixations are examined. Results demonstrate that there is an interplay between both fixation populations and that focal fixations would be much more dependent on low-level visual features and centred than ambient ones.

Introduction

Set in natural conditions, observers carried out two or three visual fixations per second to perceive their visual environment (Findlay and Gilchrist 2003). Fixations can be characterized by their durations and their amplitude of saccades. While fixations and saccades are usually analyzed separately, several studies have demonstrated a close relationship between the two. One of the first was done by Antes (Antes 1974) who reported a variation in fixations and saccades: fixations duration increases whereas saccade size decreases over time. More recently, Over et al (Over et al 2007) have also reported a systematic decrease of saccadic amplitudes concomitantly with an increase of fixation duration over the time course of scene inspection. This behavior might be related to a matter of visual search strategy (such as the coarse-to-fine strategy) or a strategic adaptation to the demands of the task (Scinto et al 1986). To better understand this behavior, Velichkovsky and his colleagues (Unema et al 2005; Velichkovsky 2002) conjointly analyzed the fixation duration with the subsequent saccade amplitude. They found a non-linear distribution indicating that i) short fixations were associated with long saccades and conversely, ii) longer fixations were associated with shorter saccades (fig 6 of (Unema et al 2005)). This distribution suggested the existence of two kinds of visual processing: ambient processing involving shorter fixations and focal processing involving longer fixations. Focal and ambient fixations would occur in a sequential fashion: the first 4 or 5 fixations are the ambient fixations and they may have a role in the extraction of contextual information while the subsequent fixations, called focal fixations may be related to recognition and conscious understanding processes. As reported by (Pannasch et al 2008), the labeling of these fixations followed a neuropsychological dichotomy

used by (Trevarthen 1968) for disentangling between two visual processing in the brain: one ambient determining space processing and the other focal dedicated to object processing. This discrimination might rely on the two different neural pathways (the ventral pathway supposed to monitor focal fixations while the dorsal pathway would be dedicated to the ambient fixations) that would be related to the coarse-to-fine strategy (Oliva and Schyns 1994; Schyns and Oliva 1996). More recently, (Henderson and Pierce 2008; Pannasch et al 2008; Tatler and Vincent 2008) have given further evidence in favor of the existence of these two categories of fixations.

This paper aims at investigating the relationship between fixation duration and saccade amplitude and to test whether the two modes of processing (focal/ambient) are affected by the type of image content. Recent findings have shown that the manipulation of low-level image content (such as luminance) may affect scanpaths during scene recognition (Harding and Bloj 2010). This kind of study raises the question of the relative contribution of low and high level information in the eye movement guidance across images (Tatler 2007). Using four different scene categories (Coast, Mountain, Street and OpenCountry), we also investigate whether focal and ambient dichotomy is scene dependent. Rather than using distribution parameters to discriminate between focal and ambient fixations, the analyses of fixations/saccades rely on a k-means clustering algorithm that categorizes automatically fixations as focal and ambient ones. Furthermore, once the clusters fixations will be identified, the question will be to know whether they are affected by bottom-up features or higher levels factors. To test this, we will compare human fixation maps to different saliency maps stemming from recent computational models of visual attention (Follet et al 2010).

The conclusions of this paper are listed below:

- An automatic classification is used to label the visual fixations into two clusters. A first cluster is called focal and the second one ambient. The classification relies on the previous saccade amplitudes;
- These two modes are not sequential. There is an interplay between them;
- Ambient fixations are located near the screen's center after the stimulus onset but become sparser with the viewing time;
- Focal visual processing mode would be more bottom-up than the ambient one. Ambient fixations would be also bottom-up but to a lesser extent than the focal mode;
- The focal-ambient dichotomy is not scene-dependent.

Method

Participants

Forty voluntary participants (22 men, 18 woman, and mean age 36.7) of Technicolor Research and Innovation in Cesson-Sévigné (France) participated to this experiment (experiments were carried out in accordance with the relevant institutional and national regulations and legislation and with the World Medical Association Helsinki Declaration). All subjects were naïve to the purpose of the experiment and had normal or corrected-to-normal vision. Four out of 40 subjects were rejected due to an incomplete recording.

Stimuli

Each participant viewed 120 natural color images. These images were either personal images or were collected from the web. They had a resolution of 800 per 600 pixels. This set of images is composed of four categories, containing 30 images each. The four categories are Street, Coast, Mountain and OpenCountry similar to (Torralba and Oliva 2001). Figure 1 shows a sample of the 120 images used in this experiment. The use of these four categories relies on their structural differences as illustrated by Figure 1. Stimuli are visually selected to present an empty landscape without any salient or incongruent objects. The only objects existing in these visual scenes are congruent features as parked cars in street scenes or trees in pictures of the OpenCountry category. As a consequence, no human being or animals (for instance there is no pedestrian in pictures of the Street category) and no object standing out from the background (such as a ship in the Coast category), were present in the selected visual scenes. We expect to have a better discrimination between the two modes of visual processing by using these kinds of stimuli.

Eye movement recordings

Eye movements were recorded while observers viewed the images. Participants were given no specific task instruction, merely to watch images as natural as possible. However observers had to answer 4 questions randomly chosen in a predefined list. Answers were collected after every image. Observers' responses were not analyzed.

A SMI RED iViewX system (50Hz (Teltow - Germany)) was used to record the eye movements. All viewers sat at a distance of 60 cm from a computer screen (1256 x 1024 pixels) in a dark room. The images subtended 36 degrees horizontally and 29 degrees vertically of the observer's field of view. Images were presented for five seconds and were each followed by a uniform grey level image. Prior to the beginning of the experiment, a nine-point grid was used to calibrate the apparatus.

Fixations having their durations smaller than 80 ms and higher than 1s were removed from the analysis (they represent 0.02% of the total number of fixations). Saccades and fixations were detected from a fixation-dispersion algorithm provided in the SMI software (Begaze™). The first fixation in each trial was defined as the first fixation that occurs after the stimulus onset. Scanpaths with less than 4 fixations were removed (i.e. 14% of the total scanpaths). Table 1 gives the statistic of the collected fixations and subsequent saccades. One point concerning the duration of fixations has to be commented. The average duration is shorter than typically reported for scene viewing (200-300ms). This might be explained by the fact that our stimuli contained few objects or salient regions. As previously reported, the number of objects has a significant impact on the fixation duration (Irwin and Zelinsky 2002; Unema et al 2005).

Table 1, statistic of collected fixations and subsequent saccades (between brackets), after filtering.

	Number of fixations and subsequent saccade	Average of fixation duration (ms) (STD) and subsequent saccade amplitude	Maximum (last quartile)	Minimum (first quartile)
Street	15611	169 (79) [4.8 (4.7)]	198 [7.23]	119 [0.73]
Coast	15015	169 (78) [4.9 (4.6)]	198 [7.44]	119 [0.77]
Mountain	15089	167 (74) [4.9 (4.6)]	198 [7.41]	119 [0.87]
OpenCountry	15008	170 (77) [4.9 (4.9)]	198 [7.36]	119 [0.81]

A saliency map was computed by convolving a Gaussian kernel (the standard deviation is of one degree of visual angle) across the user's fixation locations. Figure 2 shows an example of fixation map (visual fixations are represented by a red circle) and heat map. The heat map is just a coloured representation of a saliency map. Red areas correspond to the most fixated parts of the image.

Figure 1 : A sample of the 120 images used in the experiment. Five images per category

are shown. First to fourth row: category Street, Coast, Mountain and OpenCountry.

a) b) c)

Figure 2: a) original image; b) fixation map representing all human fixation locations recorded during the eye tracking experiment. Note that the radius of the red circles doesn't correspond to one degree of visual angle; c) Heat map indicating the most fixated parts of the image.

Data analysis

All data analyses were carried out using our own software. The same notations introduced in (Tatler and Vincent 2008) were used (see Figure 3). As introduced in the first section, past studies (Unema et al 2005; Velichkovsky 2005, Velichkovsky 2002) found a non-linear distribution showing that i) short fixations were associated with long saccades and conversely, ii) longer fixations were associated with shorter saccades (fig 6 - (Unema et al 2005)). The non-linear relationship between subsequent saccade amplitude and fixation duration might be the result of two distinct modes of visual processing. As mentioned by (Unema et al 2005), a first mode is compared to a race to jump from one salient object to another one. The second would be *an inhibitory process that allows spatial selection and search selection to attenuate* the role of the saliency map. Although this interpretation is appealing, we raise the following issue: why did they use "only" the subsequent and would the use of the previous saccade amplitude instead give the same result? Recently, Tatler and Vincent (Tatler and Vincent 2008) found the same non-linear relationship between duration of fixation and the subsequent saccade amplitude (see fig. 6 (D) of their paper). They also analyzed the relationship between fixation duration and the amplitude of saccade that immediately preceded that fixation (see fig. 5 (D) of their paper). The two shapes of curves are dramatically different. Authors noted that in one case (with subsequent saccade amplitudes) *fixation duration can be used to describe the probable saccade amplitudes that follow* whereas in the other case (with previous saccade amplitude) *the duration of fixation can be used to characterize the saccade that brought the eye to this location*. Unfortunately, they did not go further in the analysis. In this study, both configurations are analyzed. Indeed it makes sense to carry out the analysis by taken into account either the subsequent or the previous saccades. For instance, the end points of saccades might be used to label a fixation as ambient or focal. If we assume that the ambient mode is used for large scale exploration or for space perception over the whole field, fixations preceded by a large saccade might be rather ambient than focal.

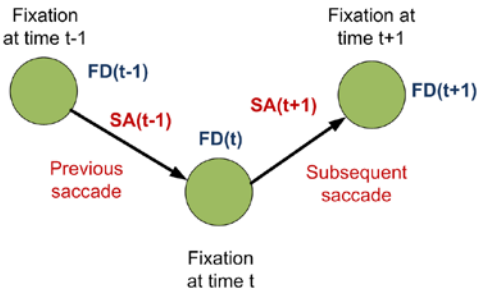

Figure 3: Notations used in this paper. FD and SA pertain for Fixation Duration and Saccade Amplitude, respectively. SA(t-1) and SA(t+1) correspond to the previous and next saccades of the fixation at time t.

Results

The purpose of this study is to investigate whether two different modes of visual processing are involved in free viewing inspection. Fixation durations and saccades amplitudes are first separately analyzed in function of the viewing time and secondly analyzed together. We will then propose an automatic solution to classify the visual fixations.

Fixations and saccades

Fixation duration according to time viewing

Figure 4a) gives the median values of fixation durations in function of viewing time. As in (Pannash et al 2008) early, middle and late visual processing periods are considered. Each processing period lasts 1.5 seconds: the early phase concerns the first 1.5 seconds of viewing, the middle phase is the period of 1.5 to 3 seconds whereas the late phase is the period of 3 to 4.5 seconds. We hypothesize that different contributions may sequentially occur in each phase. In the first one, it is reasonable to assume that the contribution is mostly bottom-up. Then, the bottom-up contribution might be progressively overridden by top-down influences (Parkhurst et al 2002). However, the extent to which these mechanisms contribute to the gaze deployment is still an open-issue.

Results indicate that the median value of fixation duration increases with the viewing time. A one-way ANOVA with the factor Time (Early, Middle, Late) shows that Time have a significant effect on the fixation duration ($F(2,53446)=11.2$; $p<.001$). Bonferroni-corrected t-test shows that the fixation durations of Early and Middle periods are not significantly different ($F(1,53446)=5.28$, $p<.06$)) whereas there is a significant difference between fixation durations of periods Middle and Late ($F(1,53446)=5.88$, $p<.05$).

This first result is similar to (Yarbus 1967; Antes 1974; Pannash et al 2008). Fixation durations are rather short after the stimulus onset compared to those occurring after three seconds of viewing. One explanation might rely on the contribution of bottom-up and top-down mechanisms. After the stimulus onset, our gaze might be mostly driven by low-level visual features. This bottom-up aspect is an unconscious and very fast mechanism. After several seconds of viewing, the top-down process becomes more influent on the way we look at the picture. Our gaze might be driven more by our own expectations and our own knowledge than by low-level visual features.

Saccade amplitude according to time viewing

Figure 4b) gives the median value of saccade amplitudes in function of viewing time. As previously, three time periods (Early, Medium and Late) are defined. Results are consistent with earlier studies (Yarbus 1967; Antes 1974; Pannash et al 2008). A one-way ANOVA shows a significant interaction between the saccade length and the factor Time ($F(2,49721)=57.68$, $p<.001$). Bonferroni-corrected t-test shows a significant difference between periods Early/Middle ($F(1,49721)=113.2$, $p<.001$) and Middle/Late ($F(1,49721)=14.02$, $p<.001$).

To sum up, long saccades are first performed probably to explore quickly the visual content. Then the amplitudes of saccade decrease over time. This second phase could be compared to a focal inspection used to explore in more details the scene.

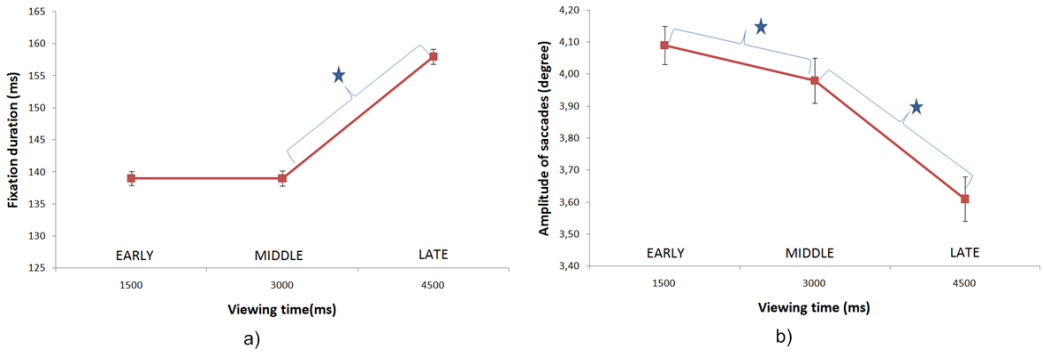

Figure 4: Median fixation duration a) and saccade amplitudes b) as a function of the

viewing time. The error bars represent 95% confidence interval. A star indicates a significant difference (Bonferroni-corrected t-test, p<.05).

Relationship between fixation duration and saccade amplitude

The two previous sections investigated the time course of fixation duration and saccade amplitude *separately*. Figure 5 gives the relationship between fixation durations and median saccade amplitudes (both previous and subsequent). It is difficult from this curve to define a classification of the visual fixations or to identify two kinds of visual processing. However, it is interesting to note that the saccade amplitude could be used to perform an estimation of the fixation durations (Velichkovsky et al 2005; Unema et al 2005). For instance, saccade amplitude of 4 degrees of visual angle might be used to split the visual fixations into two parts.

To obtain an objective classification, an automatic method to segment visual fixations into two groups was used. Note that the fixation durations and the saccade amplitudes were analyzed conjointly. The method is described in the next section.

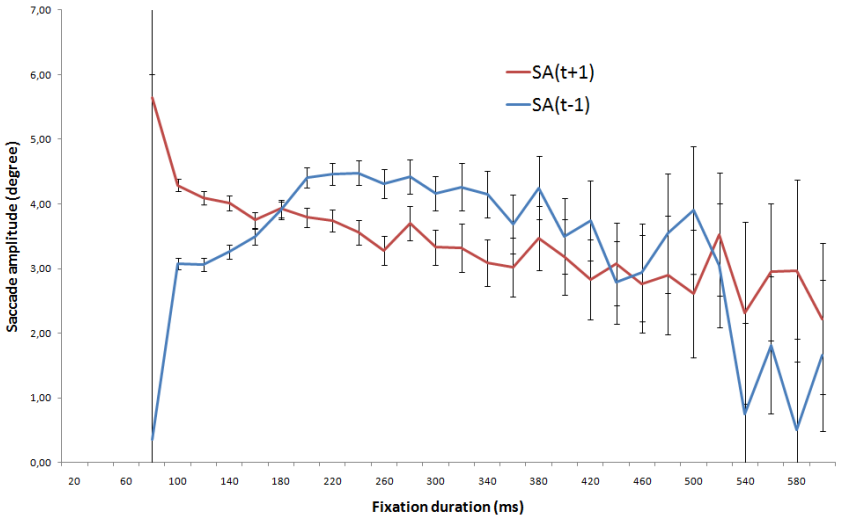

Figure 5: Median saccade amplitude (with 95% confidence intervals) as a function of current fixation duration (FD(t)). The saccade amplitudes stem from either the previous (SA(t-1)) or the subsequent one (SA(t+1)).

Classification by using a k-means algorithm

In order to verify the existence of two populations of fixations, a k-means clustering was used. It is a method for finding clusters which aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean. The algorithm iteratively moves the centers of each cluster so as to minimize the within-cluster sum of squares. It should be noted that each variable (fixation duration and saccade amplitude) has been firstly standardized (z-scores) to face with the problem of homoscedasticity usually reported (Tatler, Baddeley & Vincent, 2006; figure 1). However, even with this standardization, violations to normality may appear (i.e, skewed distribution). Consequently, we transformed the variables to get a distribution as close to normality as possible by the Box-Cox method. The Box Cox method works out a power transform of the data (Osborne, 2010). The outcomes of that transformation may be easily demonstrated by calculating the skewness. Box-Cox transformations reduced greatly the skewness of our variables rendering them close to normal distribution: before transformation (skew (FD) = 1.49; skew (SA) = 1.21), after transformation (skew (FD) = 0.09; skew (SA) = -0.13). The k-means clustering was carried out on both z-scores and Box-Cox Transformed data. The outcomes were very similar (i.e, same clusters).

Two clusters were used to identify two modes of visual processing. We follow the standard hypothesis of the two modes of visual processing, one used to accurately inspect an area and the other to perform large exploration. Table 2 to Table 5 give details regarding the clusters for each visual scene category. Results of classification indicate that the relevant dimension to segment data into two clusters is the amplitude of saccade. Except for the category Mountain, the two

clusters are significantly different in term of saccade amplitudes. For instance, there is a significant difference between clusters for the category OpenCountry ($t(14077)=168$, $p<.001$). There is however no significant difference between clusters if the fixation durations is considered ($t(14077)=1.08$, ns for the category OpenCountry). Results are interesting for different reasons.

Table 2, Parameters of the clusters provided by the k-mean algorithm for the category OpenCountry. (***) means that there is a significant difference (t-test, $p<.001$) between the cluster in term of FD or SA. (ns) stands for non significant.

	FD (ms)	SA(t-1)	Number of cases	%
Cluster 1	168.6	10.86	4057	28.81
Cluster 2	170.19(ns)	2.52(***)	10022	71.18
	FD (ms)	SA(t+1)	Number of cases	%
Cluster 1	169.07	10.89	4330	28.85
Cluster 2	170.54(ns)	2.51(***)	10678	71.14

Table 3, Same as Table 2 but for the category Coast.

	FD (ms)	SA(t-1)	Number of cases	%
Cluster 1	169.36	10.84	4185	29.69
Cluster 2	168.41(ns)	2.47(***)	9907	70.3
	FD (ms)	SA(t+1)	Number of cases	%
Cluster 1	167.9	10.92	4428	29.49
Cluster 2	170.03(ns)	2.47(***)	10587	70.50

Table 4, Same as Table 2 but for for the category Street.

	FD (ms)	SA(t-1)	Number of cases	%
Cluster 1	167.93	11.18	3954	26.96
Cluster 2	169.41(ns)	2.49(***)	10708	73.03
	FD (ms)	SA(t+1)	Number of cases	%
Cluster 1	169.41	11.2	4223	27.05
Cluster 2	169.64(ns)	2.47(***)	11388	72.94

Table 5, Same as Table 2 but for the category Mountain

	FD (ms)	SA(t-1)	Number of cases	%
Cluster 1	172.84	11.01	4072	28.73
Cluster 2	164.77(ns)	2.53(***)	10099	71.26
	FD (ms)	SA(t+1)	Number of cases	%
Cluster 1	164.85	11.05	4299	28.49
Cluster 2	168.99(ns)	2.55(***)	10790	71.5

First, the two clusters might be interpreted as being a focal and ambient mode. Indeed the first cluster might represent here a focal processing mode since the saccade amplitudes are relatively small (Mean: 2.5°) compared to those of the second cluster (Mean: 10.5°) which might concern the ambient mode. Second, the population of the two clusters is dramatically different. There is in average 70% and 30% of focal and ambient visual fixations in each cluster respectively. This difference in term of population is coherent with the assumed role of ambient and focal visual mode. The former might be used to extract the gist and the layout of the scene. It would act as a sampling of the scene to extract some sparse local patches. From these dispersed patches, we

might be able to infer fundamental information about the visual scene, such as its type. Contrary to the ambient fixations, the focal mode might be used to perform an accurate inspection of a small area. Several fixations would be required to inspect it. A logical consequence is a decrease of the saccade amplitudes indicating periods of "local" fixations. Third, the clustering is almost the same whatever the scene category. This might suggest that this dichotomy of the visual fixation would be independent of the visual scene type. This systematic tendency might underline an automatic viewing process that could be linked to the motor aspects of visual attention (Rizzolatti et al 1987). Fourth the automatic clustering shows the centroids are the same when the subsequent and the previous saccade amplitudes are considered. However, the meaning of these two configurations is different. Indeed, if we want to label fixations as being focal or ambient, it makes more sense to consider the previous saccade amplitude than the subsequent one. A fixation preceded by a small saccade would be labeled as focal whereas an ambient fixation is characterized by a longer previous saccade.

Time course of ambient and focal visual fixations

Two populations of visual fixations have been identified. In this section, we are interested in the time course of ambient and focal visual fixations. Previous studies (Unema et al 2005; Irwin and Zelinsky, 2002; Tatler et al 2003; Tatler and Vincent 2008) found that the ambient mode is mostly met at the beginning of the viewing whereas the focal mode would appear after several milliseconds of viewing. To address this point, the probability of occurrence of focal and ambient fixations is computed in function of time. A histogram is then built for each population stemming from the k-means clustering. The bins of the histogram represent the viewing time. For each bin of size 100ms, we count the number of fixations falling into the bins. Two probability density functions (one for the focal cluster and another for the cluster ambient) are then obtained by dividing the population of each bin by the total number of fixations (either focal or ambient). Figure 6 gives the probability density functions (pdf) for the ambient and focal fixations according to the time viewing. In others words, it gives us information about the probability that at a given time an ambient (or focal) fixation occurs.

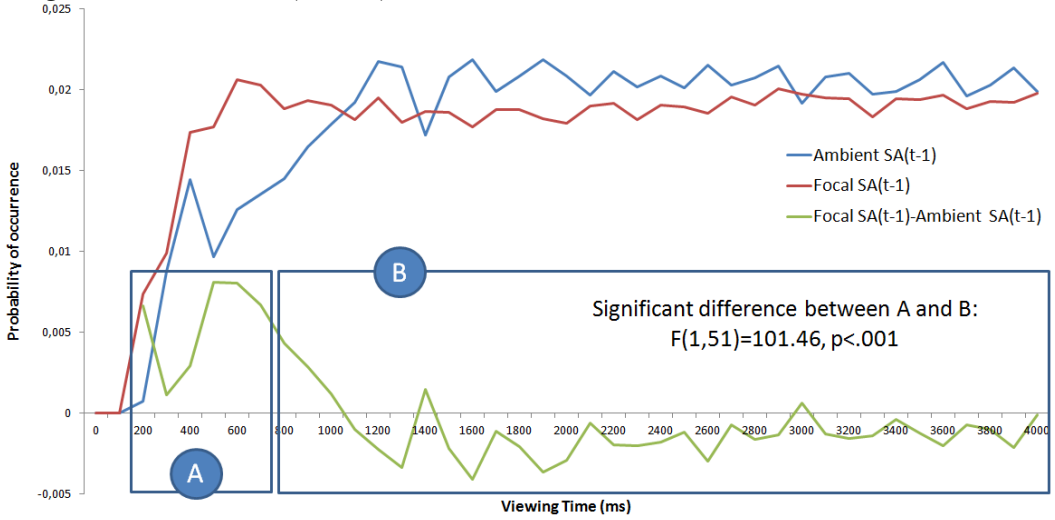

Figure 6: probability of occurrence of focal and ambient fixations in function of viewing time. The pdfs are computed by taken into account the number of focal and ambient visual fixations. Two clusters were identified: cluster A with more focal fixations than ambient ones (before 700ms); cluster B showing no difference between focal and ambient fixations (after 700ms).

Results indicate that there is a dominance of focal fixations just after the stimuli onset. By analyzing the difference between the probability of occurrences of such fixations (green curve on Figure 6), a significant difference is observed between focal and ambient probability. By using a k-means, two time intervals were identified as illustrated by Figure 6: a first one consists of fixations occurring before 700 ms of viewing time (A) whereas the second (B) is composed of fixations occurring after 700 ms. The difference between the two intervals was significant,

(F(1,51)=101.46, p<.001). The pdf of focal fixations increases up to 600 ms and stays almost constant over time ([average, standard deviation]=[0.019+/-0.00076]).
Figure 6 gives the probability to meet a focal or ambient fixation over time. To compare directly the contribution over time of each population, two probability density functions are again computed by dividing each bin of the focal and ambient fixations histogram by the total number of visual fixations (i.e the sum of focal and ambient fixation, see Figure 7). Results indicate that the ambient mode contribution increases up to 1000 ms to reach an asymptote.

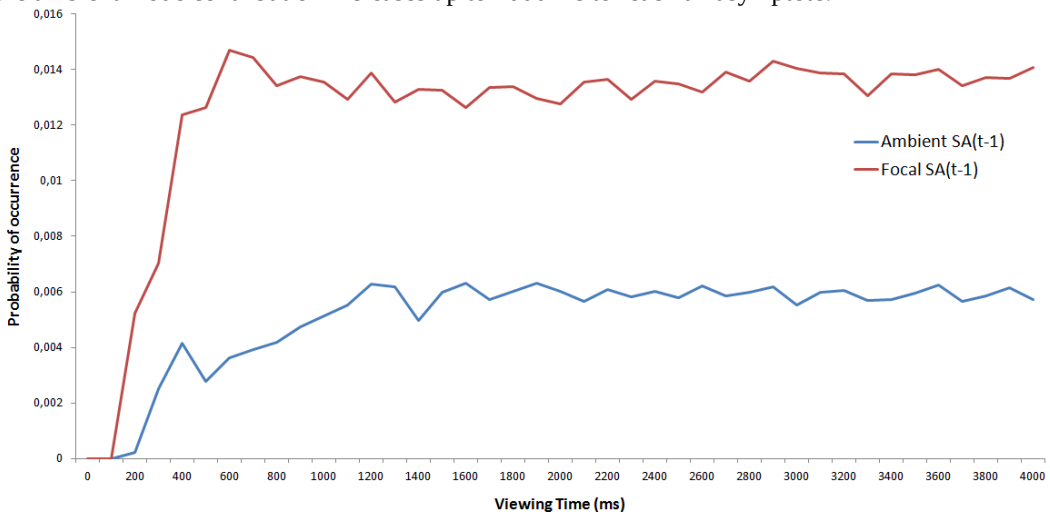

Figure 7: probability of occurrence of focal and ambient fixations in function of viewing time. The pdfs are computed by taken into account the total number of visual fixations.

It is not consistent with previous findings (Irwin and Zelinsky 2002; Tatler et al 2003; Unema et al 2005; Tatler and Vincent, 2008) in which ambient mode was the most important mode at the beginning of the viewing. Results also indicate that the ambient mode is still present after several seconds of viewing and that the two populations of visual fixations are mixed together, suggesting an interplay between both modes (Pannasch et al 2008).

Summary

By using an automatic classification, we found that the previous saccade amplitude can be used to classify the visual fixations into two clusters that are statistically different (the fixation duration is not required to cluster the data). Two centroids have been found: a first one is centred on 2.5 degrees and the second is about 11 degrees. The former might be related to a focal mode whereas the latter would represent the ambient mode. This finding is consistent over the four visual scene categories. As these categories represent various contents, this kind of classification might be considered as a systematic and fundamental phenomenon. It is however difficult to interpret the role of each visual mode. A parallel might be draw between the focal-ambient dichotomy and the bottom-up vs top-down visual attention. In order to investigate the extent to which the focal mode is bottom-up, the degree of similarity between saliency maps stemming from computational models of the bottom-up visual attention and focal (and ambient) saliency maps is assessed.

Focal and ambient saliency maps

Fixation and saliency map of the focal and ambient processing

Visual fixations are labeled as being focal or ambient. This classification relies on the amplitude of the previous saccade. The label of each fixation is determined by comparing the amplitude of the current saccade to the average saccade amplitude of the two clusters. If the distance between the center of the focal cluster and the current amplitude of saccade is smaller than the distance between the center of the ambient cluster and the current amplitude of saccade, the current fixation is labeled as a focal fixation, otherwise as an ambient fixation. Note that the fixation duration is not used since this dimension is not significant in the clustering.

Comparison between focal-ambient maps and computational saliency maps

As there are considerable differences between focal and ambient fixation maps (illustrated by Figure 8), a comparison between these maps and computational maps is carried out. To test the relevance of our assumption that the ambient processing is less bottom-up than the focal one, the highest similarity degree would be observed by comparing focal maps with computational maps. The robustness of the comparison is an important factor which should not be undermined. To be as independent as possible from model's architectures, four different computational models were used to compute saliency map. The three first models, Itti (Itti et al., 1998), Le Meur (Le Meur et al., 2006) and Bruce (Bruce & Tsotsos, 2009), rely on two seminal works: the biologically plausible architecture for controlling bottom-up attention proposed by (Koch & Ullman, 1985) and the Feature Integration Theory (Treisman & Gelade, 1980) positing that the visual processing is able to encode in a parallel manner visual features such as color, form, orientation, and others. The major difference between Bruce's model and the others is that a probabilistic framework is used to derive the saliency. The last model is Judd's model (Judd et al 2009). This model is the result of learning on a large database of eye tracking data. Compared to the previous ones, this model includes higher-level information such as the position of the horizon line, human face, a detector of cars and pedestrians and a feature indicating the distance to the center for each pixel. As previous studies (such as (Le Meur et al 2006; Tatler 2007; Bidemann 2010)) noted that there is a bias towards the center of the screen, a centered model is also used. The maximum value 1 is located at the picture's centre. The values decreased with the eccentricity. A value of 0.5 is obtained at 3.5 degrees of visual angle.

Figure 8: ambient a) and focal b) fixation maps.

Finally, a random model was designed. The input of the model is the saliency map computed by the best model in average (Judd's model). The random model randomizes the input map into non-overlapping blocks of size 32x32 pixels. Figure 9 shows the predicted saliency maps computed by the different models. Bright areas correspond to the most salient locations.

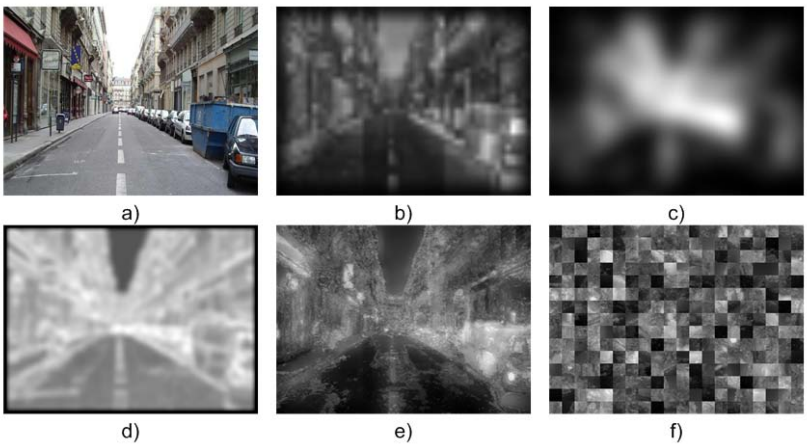

Figure 9 : Predicted saliency maps for different models. a) Original picture, b) Itti's model, c) Le Meur's model; d) Bruce's model; e) Judd's model and f) random model.

To quantify the degree of similarity between predicted saliency maps and experimental maps (focal or ambient), a ROC analysis is conducted (Fawcett 2006; Le Meur et al 2010). Pixels of aforementioned maps are then labeled as being fixated or not. The ROC analysis provides a curve that plots the false alarm rate (labeling a non-fixated location as fixated) as a function of the hit rate (labeling fixated locations as fixated). The reference or the ground truth is here the binarized focal/ambient maps. A fixed threshold was chosen to keep in average the top 20% salient locations (a set of thresholds (5, 10 and 15%) has been tested leading to similar results). Regarding the predicted maps, thresholds that are uniformly distributed between the minimum and maximum values of the predicted maps were used to label the pixels. A perfect similarity between two maps gives an Area Under the Curve (AUC) equal to 1. An AUC of 0.5 suggests that the similarity is at the chance level. Figure 10 shows for a given picture the ambient and focal maps after the threshold operation.

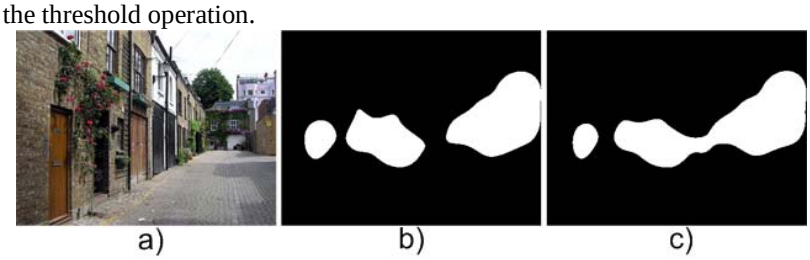

Figure 10: a) Original picture; b) Ambient map is thresholded to keep the most fixated areas of the images (around 20%); c) Same as b) for the focal map.

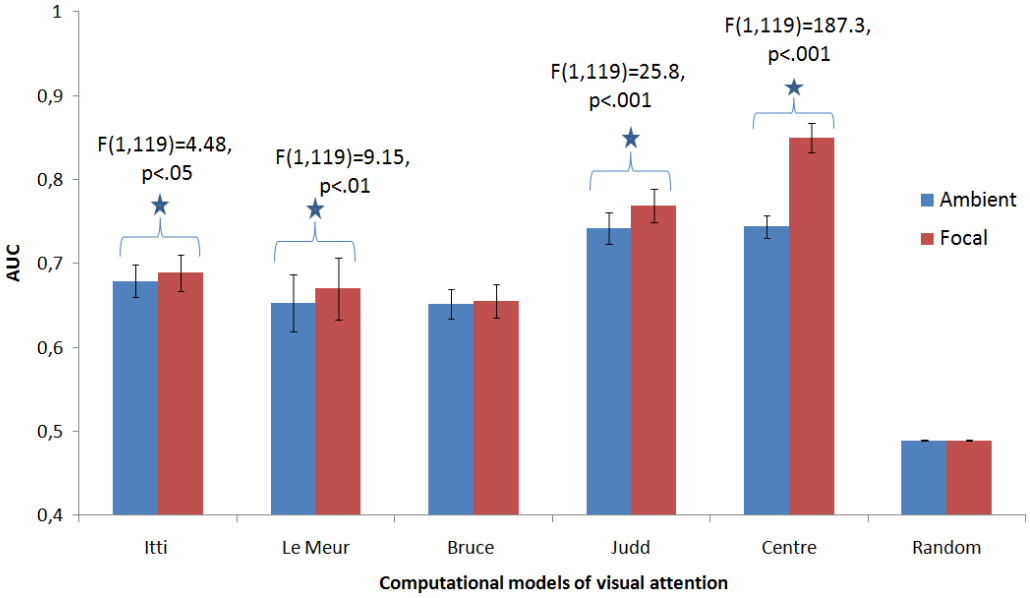

Figure 11: AUC values indicating the difference between computational saliency maps (models of Itti, Le Meur, Bruce, Judd) and focal (or ambient) maps. A value of 0.50 indicates random performance whereas 1.00 denoting perfect performance. Error bars are the standard error of the mean.

Figure 11 gives the median value of AUC for the different models. AUC values are higher for the focal maps than for the ambient maps. Except the random and Bruce's model, they are statistically significant (paired t-test between AUC values coming from focal and ambient processing. Statistics are reported on the figure). Focal maps are better predicted by saliency models than ambient ones. These results are consistent with the assumption that the ambient processing is less bottom-up than the focal one and support that the ambient process is more concerned with scene layout. Three bottom-up models (Itti, Le Meur and Judd's model) give better results when the focal map is used in the comparison. Ambient mode is also bottom-up but to a lesser extent. This finding is consistent with studies (Tatler et al 2006; Rajashekar et al 2007) reporting that large saccades are less dependent on the low-level visual features than short saccades. Ambient mode is supposed to be used to sample the scene in order to quickly explore it

and to select in a second phase the most interesting scene's parts. Therefore, ambient mode might behave more or less as a random mode and then would be less dependent on the low-level visual features. The comparison with a random model allows us to rule out that the ambient visual fixations are purely random. Results indicate that the overall prediction is significantly better than chance ($p<<.001$). It would suggest that both visual processing modes are more or less driven by the low-level visual features and that ambient mode is not based on a random sampling.

It is also important to consider the role of the central bias. The magnitude of the AUC value obtained by a focal map when compared to a centre model is statistically higher than the one obtained for the ambient processing (paired t-test $p<<.001$). We already knew that the screen's centre plays an important role in the visual attention deployment and that a centered Gaussian model often provides better quantitative performance than computational saliency maps (Le Meur et al 2006). A number of factors can explain this central tendency. The centre might reflect an advantageous viewing position for extracting visual information (see (Tatler 2007; Renninger et al 2007)). Tatler (Tatler 2007) found that this central tendency was not significantly affected by the distribution of visual features in a scene. More recently, Bindemann (Bindemann 2010) showed that this central bias is not removed by *offsetting scenes from the screen's center*, by *varying the location of a fixation marker* preceding the stimulus onset or even by *manipulating the relative salience of the screen*. Bindemann concluded that *the screen-based central fixation bias might be an inescapable feature of scene viewing under laboratory conditions*. These issues are consistent with our hypothesis. Ambient maps are less predicted by the centred model suggesting that the screen's centre is more neglected in the ambient mode than in the focal one. Farthest positions from the centre would be more favoured in ambient mode. From these observations, two preliminary conclusions can be drawn:

1 – the focal map are more bottom-up than ambient map;

2 – the degree of similarity between focal and centred maps is significantly higher than the ambient map compared to the centred map. This second conclusion is consistent with previous studies (Tatler et al 2006; Rajashekar et al 2007).

As the focal and ambient visual processing modes depend on the viewing time (see Figure 6), the degree of similarity between computational and experimental maps is computed on the first two and on the last two seconds. Figure 13 and Figure 14 give the results. All statistics are reported on the figure for the sake of readability. Results indicate that the centered model has a strong impact in early (0 to 2s) phase. Its contribution dramatically drops down for the late ambient fixations (2s to 4s) whereas it increases for the late focal fixations. It confirms the second conclusions given previously: the focal fixations are more centered than the ambient ones. However, it is important to underline that the ambient fixations in the early phase are also well predicted by the centre model. This is more or less consistent with our first hypothesis. Indeed, we assume that ambient mode is used to make a large scale exploration of the scene, especially after the stimulus onset. Our results suggest that the exploration scale during the first two seconds is not as large as we would expect. The exploration seems to be restricted to an area located around the center of screen. However, for the late phase, the degree of similarity between late ambient map and centered map dramatically falls down. It is more consistent with our preliminary hypothesis. Figure 12 shows the amplitude of previous saccades for two temporal phases: an early phase corresponding to the interval 0-2s and a late phase for 2-4s. It is interesting to note that the median amplitude of saccades significantly increases with the time viewing for the ambient mode ($F(1,11114)=138$, $p<.001$). It is not consistent with what we are used to observing: a decrease of amplitudes of saccade with the viewing time as shown by Figure 4. It indicates that just after the stimulus onset the scene exploration might start locally (small saccades) and might become more globally (longer saccades). This local to global behavior is consistent with results obtained by the centre model on Figure 13. This trend might reflect the efforts used to explore the scene. After the stimuli onset, the visual inspection would concern the periphery of the screen's center whereas, after several seconds of viewing, it would be required to go farther.

Finally, the first aforementioned conclusion, namely the focal is more bottom-up than the ambient mode, is consistent with results of Figure 13 and Figure 14. Whatever the phase, the focal maps are indeed more bottom-up than ambient maps.

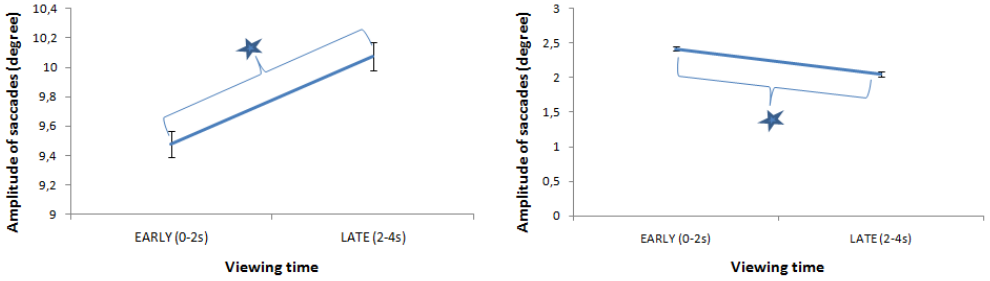

Figure 12: median amplitude of previous saccades in function of the viewing time and for the two modes of visual processing. On the left-hand side: ambient fixations; on the right-hand side: the focal fixations.

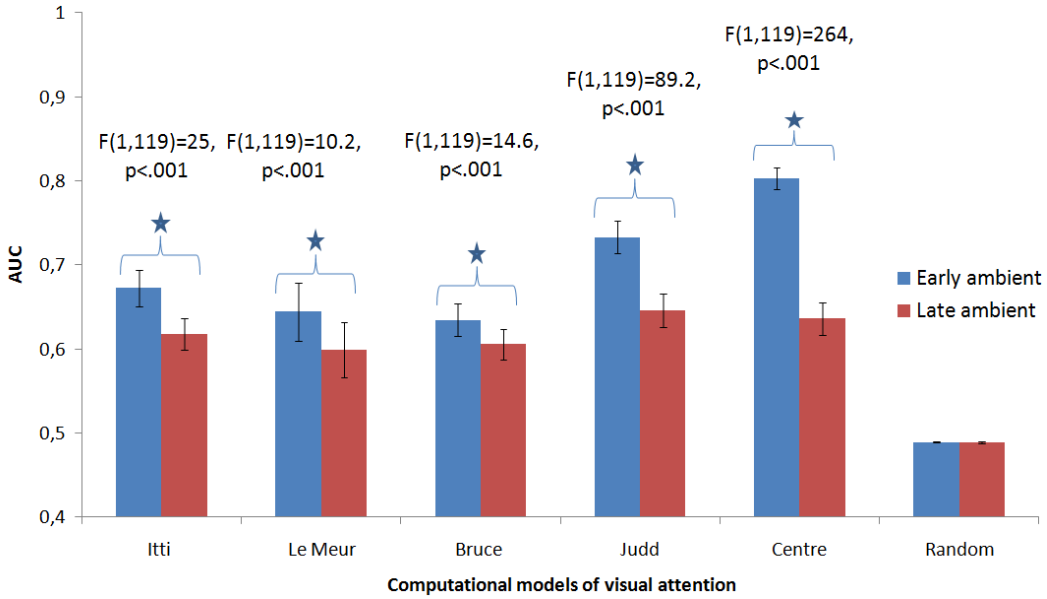

Figure 13: AUC between early-late ambient saliency maps and computational saliency maps.

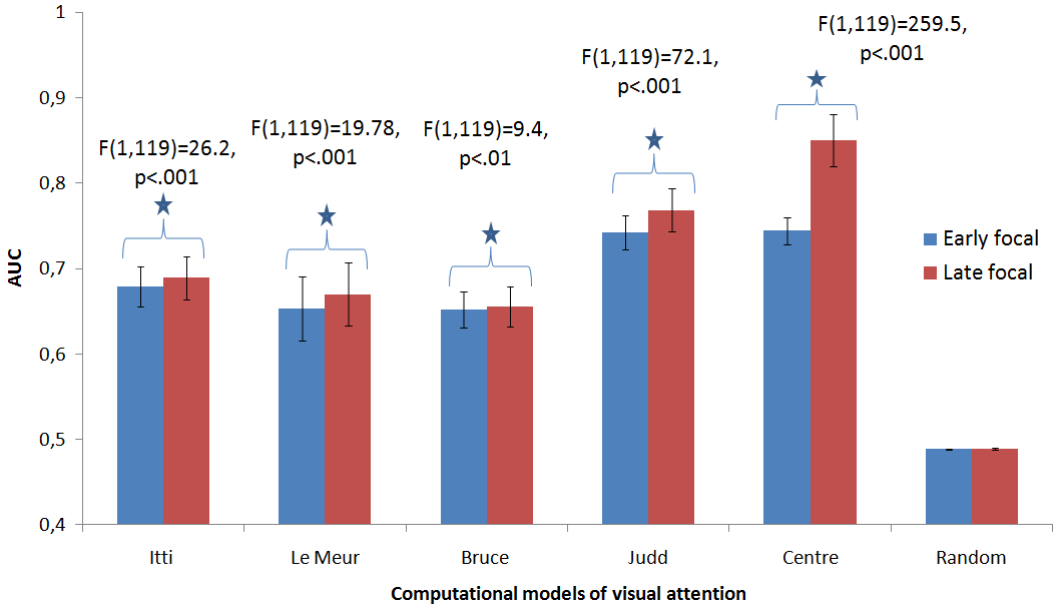

Figure 14: AUC between early-late focal saliency maps and computational saliency maps.

Discussion and conclusion

The present research investigates the existence of two different populations of visual fixations. We examined the relationship between fixation durations and saccade amplitudes, as previously done in the work of (Unema et al 2005). This relationship is non-linear and time-dependent, suggesting the existence of two kinds of visual fixations. From this observation, these visual fixations were classified and their relation with computational saliency investigated. The list of our findings is dressed below.

Two visual fixation populations

Visual Fixations are first classified using a k-means clustering algorithm. The clustering shows the amplitude of the saccade is the determining feature for the clustering. A first cluster groups together visual fixations characterized by small saccade amplitude (in average 2.5°) whereas the second group gathers together fixations with larger amplitude of saccade (in average 11°). Our results confirm the existence of two visual processing (Trevarthen 1968; Unema et al 2005; Pannasch et al 2008). The first cluster would represent the focal visual processing used to accurately inspect areas. The second one would concern the ambient fixations used to explore our visual field. As in previous studies (Unema et al 2005), we observed a larger proportion of focal fixations:70% of the visual fixations are indeed labeled as being focal whereas the remainer ambient fixations. There are two or three more focal fixations than ambient ones which is coherent and consistent with their presumed role. This finding is confirmed across scene category.

A different sensibility to the central tendency

The second finding concerns the central fixation bias. In this study, the central fixation bias is observed just after the stimuli onset for both populations of visual fixations. Even ambient fixations, which are deemed to be used to explore the scene, are located near the centre. However, the ambient fixations become less dependent on the central tendency when the viewing time increases. This temporal behaviour might reflect a local to global scene inspection. The scene exploration might start from locations near the center to locations far from the scene's center. In other words, the screen's center might be a good place to begin further exploration of the scene. Regarding the focal fixations, an opposite effect is observed over time. The central fixation bias is more pronounced after 2 seconds of viewing. These observations confirm the importance of the scene's center. The increase of the central bias contribution might be related to two important features shown by previous studies. First fixation locations tend correlate with low-level visual features (Reinagel and Zador 1999; Parkhurst et al 2002; Tatler et al 2005) and the second could concern the fact that interesting objects are often located in the centre of natural scenes (photographers tend to place objects of interest at the center of the picture).

Time course of focal and ambient fixations

The third finding concerns the time course of focal and ambient fixations. There are few studies concerning this matter. They suggest that the ambient processing is mainly present just after the stimulus onset and that the contribution of the focal mode is rather late compared to the ambient contribution (Pannasch et al 2008; Unema et al 2005; Norman 2002). Our results are not consistent with previous findings. We observed that the focal mode is the most important over time and appears just after the stimulus onset. This unexpected result may simply be explained by the central bias. Indeed, just after the stimulus onset, the screen's center attracts our attention for different reasons (some of them are given in the previous paragraph). This phase might be considered in our classification as a focal one. Concomitantly with this phase, the contribution of the ambient mode also increases over time but not as rapidely as the focal mode. It reaches its maximal influence after 1 second and stays almost constant over time.

Focal and ambient visual fixations are bottom-up or not?

The role of focal and ambient fixations has also been investigated. The focal mode is more dependent on the low-level visual features than the ambient mode. This conclusion stems from the comparison between focal and ambient maps and predicted saliency maps. Special consideration has to be given to this interpretation. For instance, the focal mode can be considered, in our study, to be more related to a bottom-up mode than the ambient one, simply because of the central bias

and (or) high-level object understanding. This last reason is supported by recent studies (Elazary and Itti 2008; Le Meur and Chevet 2010; Masciocchi et al 2009) suggesting that bottom-up models are good predictors of interesting hand-label regions of a scene. Although purely based on bottom-up features, bottom-up models succeeds in predicting areas of interest chosen consciously by observers. It might indicate that computational saliency models predicts fixations based on bottom-up features but also those (to some extent) based on higher-level information (less driven by bottom-up features). To go further on the role of ambient and focal fixations, it would be interesting to combine several sets of behavioural data. One possibility would be to use the EFRP technique (Baccino 2011) that combines the EEG technique with the eye-tracking technique. EFRPs are extracted from the EEGs by averaging the brainwaves occurring from the onset and offset of eye-fixation. The analysis of these EFRPs components may reveal the time course of attention or semantic processing. Separating these components with statistical procedures but also the localization of the activation (on which electrode) may be highly informative for labelling fixations. These findings would contribute greatly for interpreting scanpaths and fixations on some region of interests in real life activities. For example, EFRPs have shown to be a useful technique to investigate early lexical processes and for establishing a timeline of these processes during reading (Baccino & Manunta 2005) or during object identification (Rama & Baccino 2010)

The duration of fixations is not discriminant

Finally we would like to underline a point concerning the fixation duration. It is generally believed that the fixation duration reflects the depth of processing (Velichkovsky 2002) and the ease or difficulty of information processing. This behaviour has been shown when observers look at a picture (Mannan et al 1995) or read a text (Daneman and Carpenter 1980). Here, our findings show that fixation durations are not useful for classifying visual fixations into clusters. However, if the focal mode relies on a top-down visual processing, it probably involves cognitive mechanisms and the duration of focal fixations should be higher than ambient ones. How could we explain that the duration of fixation does not play an important role? A plausible reason might be related to the lack of task to perform or goal to achieve. Indeed, a free-task viewing requires the examination of the spatial environment and the casual observation of the picture. The “relevance” of fixation durations would also depend on the material used.

References

Antes J R, 1974 "The time course of picture viewing" Journal of Experimental Psychology 103(1), 62‑70

Baccino T 2011"Eye Movements and concurrent ERP's: EFRPs investigations in reading" In S. Liversedge, Ian D. Gilchrist & S. Everling (Eds.), *Handbook on Eye Movements* (pp. 857-870). Oxford University Press.

Baccino T and Manunta Y 2005 "Eye-Fixation-Related Potentials: Insight into Parafoveal Processing" Journal of Psychophysiology, 19(3), 204-215

Bindemann M 2010 "Scene and screen center bias early eye movements in scene viewing" Vision Research

Bruce N D B and Tsotsos J K 2009 "Saliency, attention and visual search: an information theoretic approach” Journal of Vision, vol. 9, pp. 1-24

Daneman M and Carpenter P A 1980 "Individual differences in working memory and reading" Journal of Verbal Learning and Verbal Behavior, 19, 450-466

Fawcett T 2006 "An introduction to ROC analysis", Pattern Recognition Letters 27 pp. 861–874

Findlay J M and Gilchrist I D 2003 "Active vision: the psychology of looking and seeing" Oxford: Oxford University Press.

Follet B, Le Meur O and Baccino T 2010 "Modeling visual attention on scenes" Studia Informatica Universalis, 8(4), 150-167.

Harding G and Bloj M 2010 "Real and predicted influence of image manipulations on eye movements during scene recognition" Journal of Vision, 10(2)

Henderson J M and Pierce G L 2008 "Eye movements during scene viewing: Evidence for mixed control of fixation durations" Psychonomic Bulletin & Review, 15(3), pp. 566-573

Irwin D E and Zelinsky G J, 2002 "Eye movements and scene perception: memory for things observed" Perception & Psychophysics 64(6), pp. 882-895

Itti L, Koch C and Niebur E, 1998 "A model for saliency-based visual attention for rapid scene analysis" IEEE Trans. on PAMI, vol. 20, pp. 1254-1259

Judd T, Ehinger K, Durand F and Torralba A 2009 "Learning to Predict Where Humans Look" ICCV

Koch C and Ullman S 1985 "Shifts in selective visual attention: towards the underlying neural circuitry" Human Neurobiology, vol. 4, no. 4, pp. 219-227

Norman, J 2002 "Two visual systems and two theories of perception: an attempt to reconcile the constructivist and ecological approaches" Behavioral and Brain Sciences, vol. 25, no. 1, pp. 73-144

Le Meur O, Le Callet P, Barba D and Thoreau D 2006 "A coherent computational approach to model the bottom-up visual attention" IEEE Trans. on PAMI, vol. 28, pp. 802-817

Le Meur O and Chevet J C 2010 "Relevance of a feed-forward model of visual attention for goal-oriented and free-viewing tasks" IEEE Trans. On Image Processing, vol. 19, no. 11, pp. 2801-2813.

Le Meur O, Ninassi A, Le Callet P and Barba D 2010 "Overt visual attention for free-viewing and quality assessment tasks. Impact of the regions of interest on a video quality metric", Elsevier, Signal Processing: Image Communication, vol. 25, Issue 7, pp. 547-558.

Mannan S, Ruddock K H and Wooding D S 1995 "Automatic control of saccadic eye movements made in visual inspection of briefly presented 2D images" Spatial Vision, 9, pp. 363-386

Masciocchi C M, Mihalas S, Parkhurst D and Niebur E 2009 "Everyone knows what is interesting: salient locations which should be fixated" Journal of Vision, 9(11)

Oliva A and Schyns P G 2000 "Colored diagnostic blobs mediate scene recognition" Cognitive Psychology, 2000.

Osborne J W 2010 "Improving your data transformations: Applying the Box-Cox transformation. Practical Assessment", Research & Evaluation, 15(12), 1-7.

Over E A B, Hooge I T C, Vlaskamp B N S and Erkelens C J 2007 "Coarse-to-fine eye movement strategy in visual search" Vision Research, 47(17), pp. 2272-2280

Pannasch S, Helmert J R, Roth K, Herbold, A K and Walter H 2008 "Visual Fixation Durations and Saccade Amplitudes: Shifting Relationship in a Variety of Conditions" Journal of Eye Movement Research, 2(2):4, pp. 1-19

Pannasch S and Velichkovsky B M 2009 "Distractor Effect and Saccade Amplitudes: Further Evidence on Different Modes of Processing in Free Exploration of Visual Images" Visual Cognition 17(6/7), 1109-1131

Parkhurst D, Law K and Niebur E 2002 "Modelling the role of salience in the allocation of overt visual attention" Vision Research, vol. 42, pp. 107–123

Rajashekar U, Van der Linde I and Bovik A C, 2007 "Foveated analysis of image features at fixations" Vision Research, 47(25), pp. 3160-3172

Rama P and Baccino T, 2010 "Eye-fixation related potentials (EFRPs) during object identification" *Visual Neuroscience*, 27, 1-6.

Reinagel P and Zador A M 1999 "Natural scene statistics at the centre of gaze" Network, 10, 341–350

Renninger L W, Vergheese P and Coughlan J, 2007 "Where to look next? Eye movements reduce local uncertainty" Journal of Vision 7(3):6, 1–17

Rizzolatti G, Riggio L, Dascola I, Umilta C 1987 "Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention" Neuropsychologia 25: 31–40.

Scinto L F, Pillalamarri R and Karsh R 1986 "Cognitive strategies for visual search" Acta psychologica, 62(3), pp. 263-292

Schyns P G and Oliva A 1994 "From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition" Psychological Science, 5, 195-200

Tatler B W, Baddeley R J and Gilchrist I D 2005 "Visual correlates of fixation selection: Effects of scale and time" Vision Research, 45(5), pp. 643-659

Tatler B W, Baddeley R J and Vincent B T 2006 "The long and the short of it: spatial statistics at fixation vary with saccade amplitude and task" Vision Research, 46(12), 1857-1862

Tatler, B W 2007 "The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions" Journal of Vision 7(14):4, 1–17

Tatler B W, Vincent B 2008 "Systematic tendencies in scene viewing" Journal of Eye Movement Research 2(2):5, pp. 1-18

Torralba A, Oliva A 2001 "Modeling the shape of the scene: a holistic representation of the spatial envelope" International Journal of Computer Vision 42 145-175

Trevarthen, C B 1968 "Two mechanisms of vision in primates" Psychologische Forschung, 31(4), pp. 299-337

Unema P J A, Pannasch S, Joos M and Velichkovsky B M, 2005 "Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration" Visual Cognition, 12 (3), pp. 473-494

Velichkovsky B M 2002 "Heterarchy of cognition: The depths and the highs of a framework for memory research" Memory, 10(5), pp. 405-419

Velichkovsky B M, Joos M, Helmert J R and Pannasch S 2005 "Two visual systems and their eye movements: evidence from static and dynamic scene perception" Proceedings of the XXVII conference of the cognitive science society, pp. 2283-2288

Yarbus A L 1967 "Eye movements and vision" (L.A. Riggs, Trans.). New York: Plenum Press