# Towards a Model of Information Seeking by Integrating Visual, Semantic and Memory Maps

Myriam Chanceaux, Anne Guérin-Dugué, Benoît Lemaire, Thierry Baccino

# Towards a model of information seeking by integrating visual, semantic and memory maps

Myriam Chanceaux[1], Anne Guérin-Dugué[1], Benoît Lemaire[1], Thierry Baccino[2]

[1] University of Grenoble, France
<first name>.<last name>@imag.fr

[2] University of Nice-Sophia-Antipolis, France
baccino@unice.fr

**Abstract.** This paper presents a threefold model of information seeking. A visual, a semantic and a memory map are dynamically computed in order to predict the location of the next fixation. This model is applied to a task in which the goal is to find among 40 words the one which best corresponds to a definition. Words have visual features and they are semantically organized. The model predicts scanpaths which are compared to human scanpaths on 3 high-level variables (number of fixations, average angle between saccades, rate of progression saccades). The best fit to human data is obtained when the memory map is given a strong weight and the semantic component a low weight.

**Keywords**: computational model, information seeking, visual saliency, semantic knowledge, memory

## 1 Introduction

Over the past decade, a large amount of writings or contents has become available to the web user. However in the same time, there has been relatively limited progress towards a scientific understanding of the psychology of human interaction with the web. Detailed integrated cognitive models are difficult to create, limited to very narrow experimental conditions of interaction and mostly unable to face with the semantics of web contents.

One of the most frequent tasks on the web consists in seeking information on pages. Searching for information requires defining a given goal that may be precise or vague, more or less variable along the navigation. When the goal is well-defined, top-down models like ACT-R [1] predict relatively well how the information is retrieved on the web [2]. When the goal is ill-defined, the user must rely on data displayed on the page and incrementally build/maintain into memory the goal to reach information. Such task requires substantial acquisition and integration of knowledge coming from external sources [3] in order to better define goals, available courses of action, heuristics and so on. Modeling in this case must take into account perceptual information carrying out mostly bottom-up processes that analyze data presentation along the visual exploration of pages

and guide the user attention. However, this visual processing is closely related to content processing and any computational model (psychologically valid) has to explain how this integration of information is generated on-line to verify whether the goal is reached or not.

While a number of studies have investigated this activity, very few cognitive models are satisfying. Some models are very general pointing out the activity of information seeking [4–6] but saying nothing or very few about the underlying cognitive processes. Others are too specific dedicated to web navigability as COLIDES [7] or some extensions of ACT-R model such as SNIF-ACT [8] or BSM [9]. The scope of the paper is to sketch an integrated cognitive model that account for both visual and semantic processing during information seeking.

## 2 Information Seeking: 3 components

Seeking information (i.e, a word) in a document requires from the user to process two sources of information: visual information (i.e, exogenous information) and semantic information (i.e, endogenous information). The former refers to low-level visual features involving bottom-up selective processes while the latter refers to word meaning represented in semantic memory and entails top-down processes. Both visual and semantic information have been shown to guide the visual scanpath, the gaze tends to move towards locations that are visually salient, but it is also attracted to regions that are semantically relevant with respect to the current search goal. Let's describe more precisely this respective influence.

### 2.1 Visual information

Computational models of selective visual attention have attracted growing levels of interest during this last decade. The purpose is to predict where humans look when they perform a visual detection task from a bottom-up perspective. Most of these models are mainly based on two original concepts: the Feature Integration Theory [10]. Among them, the most popular is proposed by Itti and Koch [11]. It is based on a feature decomposition of the visual stimuli (natural visual scenes), and a competition-fusion process between the parallel feature maps that extracts a visual saliency map. The highest salient regions are then segmented and sorted according to their saliency value. For the first eye fixations on a picture, these models fits well the eye movements data when the visual stimuli have little semantic information and when the task is free without explicit task driving the scene exploration [12]. In the case of more demanding visual search, the visual saliency is progressively modulated over time by semantic and cognitive controls, depending of the type of the scene (a priori knowledge of the scene) and the task [13]. See for example the discussions in [14]. Due to this complexity, few models integrate these two pathways. Among recent propositions, [15] extracts saliency regions through interactions with a working memory and a semantic and visual long term memory. In our context of information seeking, visual stimuli can be considered as more simple than natural scenes from the point of view of the

image features, but very complex from the point of view of the meaning and the semantic of the scene. Thus these bottom-up models must be highly simplified in the final model, but they must integrate specificities of reading tasks as it is proposed in [16].

### 2.2 Semantic information

Looking for a word entails also the automatic access to semantic memory. Only very few computational models of information seeking have taken into account this top-down process probably due to the difficulty to represent meaning for a computer. However, since the development of Latent Semantic Analysis [17], meaning representation can be computed and estimated. Basically, LSA takes a huge corpus as input and yields a high-dimensional vector representation for each word, usually about 300 dimensions. It is based on a singular value decomposition of a word $\times$ paragraph occurrence matrix, which implements the idea that words occurring in similar contexts are represented by close vectors. Such a vector representation is very convenient to give a representation to sentences that were not in the corpus: the meaning of a sentence is represented as a linear combination of its word vectors. Therefore, we can virtually take any sentence and give it a representation. Once this vector is computed, we can compute the semantic similarity between any word and this sentence, using the cosine function. The higher the cosine value, the more similar the words are.

One of the first models attempting to explain information seeking by using LSA was COLIDES (Comprehension-based Linked model of Deliberate Search) [7]. It describes how people attend to and comprehend information patches on individual webpages. It is a simulation model of navigation trying to extend a series of earlier models developed by Kitajima & Polson and the Kintsch's construction integration theory of text comprehension. In COLIDES, the description of a web page is made up of a large collection of objects competing for users' attention, which are meaningful units and/or targets for action. Users manage this complexity by a two-phased processes: 1) an Attention Phase where users segment the page into regions and focus on a region of the page; 2) an Action Selection Phase in which users first comprehend each of the objects (e.g., hypertext link, graphic, radio button, etc.) that can be acted on in the focused-on region, including the consequences of acting on the object. Then they select one of the actions, usually clicking on one of the available hyperlinks. In both phases, the user's behaviors are determined by the perceptions of semantic similarity between the user's goals and the descriptions of alternative regions or actions. This similarity is calculated by LSA. Despite this interesting semantic component, COLIDES does not describe precisely how low-level information coming from vision or attention processes can guide the user gaze and orient the selection.

### 2.3 Memory Mechanism

A model of selective visual attention and scanpaths would be incomplete without describing the process, by which the currently attended location is prevented

from being attended again, this mechanism is known as the Inhibition of Return (IOR). The IOR refers to an increased difficulty of orienting to a location to which attention has previously been directed. This possibly ensures that fixations are less likely to return to a previous point of high salience [18]. Pratt and Abrams [19] reported that the IOR of attention is found only for the most recently attended of two cued locations but other has shown that in more complex environments more possible locations may be involved [20].

## 3 Model

Our model is a spatio-temporal model based on the dynamic integration of visual and semantic information associated to a simple memory mechanism. It aims at describing in a cognitively plausible manner how visual, semantic and memory processes interact in order to predict eye movements in an information-seeking task. Each of these three components of our model is implemented by means of a conditional heat map of the current image, in which each of its elements is assigned a weight representing its relevance for the given component. Each of these maps is conditional on the location of the current fixation. The memory map is also conditional on the entire scanpath. These three maps are therefore continuously updated during the simulated visual search. Basically, they work in the following way: - the visual component integrates both the specific behavior of the human retina and the visual properties of the scene. Therefore, this component gives high weights to the fovea, but also to visually salient elements; - the semantic component gives high weights to the current zone if its semantic similarity with the goal is high, since the solution might be close; however, if this similarity is low, the current zone elements are assigned low values, meaning that is it probably not an interesting area; - the memory component strongly decreases the weight of the previous fixation zone, in order not to move back to it. However, since human memory is limited, these values tend to return to normal values over time. The three maps are integrated by a weighted sum and the simulated gaze is moved towards the best-weighted zone. Once the new fixation has been selected, maps are updated accordingly, then a new fixation is chosen, and so on. From an initial fixation point, our model thus produces a scanpath.

### 3.1 Task

Our final goal is to apply this model to complex web pages in which visual and semantic information are highly salient and generally not congruent. However, we first implemented and tested this general model on a simple task in order to control parameters as much as possible. Therefore, we largely simplified a general web page to only keep minimal visual and semantic data. We ended up with images containing 40 independent words (Fig. 1).

**User goal** We formalized the goal the user is pursuing by considering that this user is seeking a particular piece of information. This item is defined by *the class*

*it belongs to* and its specific features within this class. For instance, the user may look for *a scientific publication* given its title, *a tennis result* for a specific player, a *restaurant* that is open on Sunday, etc.

Our experimental users are thus instructed to find a specific word in the image which belongs to a given *category* and is the best at satisfying a given *feature*. This question is linguistically expressed by such a sentence: *find the most [*`feature`*] [*`category`*]*; for example, *find the most alcoholized beverage* or *find the roughest sport*. In the remainder of this paper, this user goal will be called *instruction* to avoid any confusion with the word the user should find.

**Visual and semantic features of words** Each of the 40 words of each image has a visual feature and a semantic feature.

For the moment, the only visual feature of words is their font size, from 13 to 19pt. In a future experiment, words will be also characterized by colors.

We also organized the 40 words in order to reproduce a very common property of our world which is that objects that are similar to each other tend to be near each other. This semantic-spatial congruency helps us a lot when we search information: in newspapers, football and tennis results are close to each other; in supermarket, all vegetables are in the same place and they are close to fruits.

In each of our images, 7 words belong to the same category, including the target word. For instance, there are 7 alcoholized beverages in the first image mentioned previoulsy, 7 names of sport in the second one, etc. All 33 other words are of decreasing semantic similarity with the instruction.

Figure 1 presents such an image, the instruction being *Find the sweetest food*. The target is *confiture* (jam). The six words that belongs to the same category are *citron* (lemon), *crème* (cream), *salade* (salad), *viande* (meat), *soupe* (soup) and *chocolat* (chocolate). Close to the target are also words related to food, like *saveur* (flavour), *marmite* (cooking-pot), *litre* (liter), but the more distant words are from the target, the less similar they are.

We tried to be as much objective as possible in the design of these images. To that end, we created a semantic space by applying Latent Semantic Analysis (LSA) [17] to a 13 million-word French corpus composed of novels, newspaper articles and encyclopedia chapters. Many articles in the literature have shown high correlations between LSA cosines and human judgements of similarity [21]. We thus relied on this LSA semantic space to define our 18 images. The procedure was the following:

- 18 [`feature`]-[`category`] instructions were defined;
- for each instruction, semantic similarities between the instruction and all words whose length was between 5 and 9 letter long and LSA weight was between .3 and .7 (medium frequency) were computed;
- the 7 best words which belonged to the category were selected;
- 33 other words were randomly selected at regular intervals between a 0 semantic similarity and the semantic similarity of the $7^{th}$ previous word. For instance, the list corresponding to the instruction *sport brutal (rough sport)* is the following (similarity values are in parentheses):

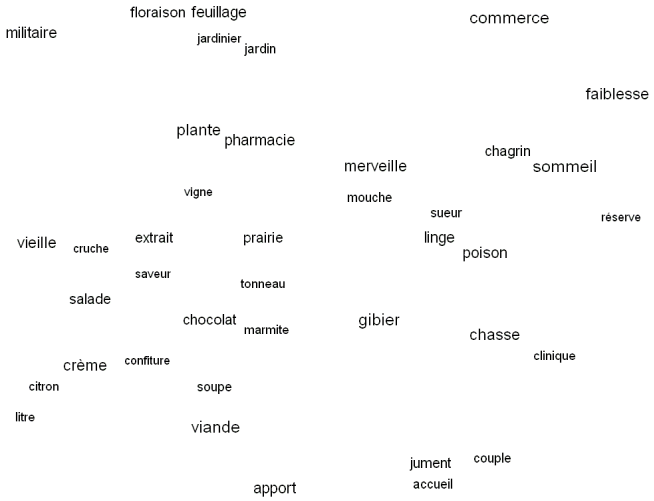


**Fig. 1.** Example of image. Instruction is: *quel est l'aliment le plus sucré ? (what is the sweetest food?)*. Correct answer is *confiture (jam)*.

1. football (soccer) (.71)
2. rugby (rugby) (.69)
3. tennis (tennis) (.57)
4. basket (basketball) (.48)
5. cyclisme (cycling) (.43)
6. course (run) (.42)
7. voile (sailing) (.38)
8. gardien (goalkeeper) (.37)
9. victoire (win (.36)
10. vainqueur (winner) (.35)

...

37. charbon (coal) (.02)
38. chêne (oak) (.01)
39. domicile (home) (.01)

– semantic similarities between all pairs of words were computed and a Multi-Dimensional Scaling procedure was run to assign all words 2D coordinates;
– all coordinates were scaled in order to fill an entire 1024x768 screen;
– in order to avoid word overlapping, 80 non-overlapping positions (NOP) were randomly defined and each word was moved to a close NOP such that the sum of these moves was minimum.

In order to select targets and be more precise about the 7 first words which play an important role in the task, we controlled their semantic similarity with the instruction by asking 28 participants to assess their similarities with the instruction on a 5-point scale. The target was defined as the word which was best-rated. We also computed a Student test between the first and second best-rated words to make sure there was no ambiguity on the target. In 4 cases out of 18, this difference was not significant. We then removed the second best-rated and replaced it by a word which was not highly similar to the instruction.

**Experimental conditions** Last but not least, in order to investigate the relative contribution of semantic versus visual factors, we defined three visual conditions and two semantic conditions. Semantic conditions are (1) semantic organization of words as defined previously; (2) no semantic organization : words are randomly assigned to the locations of the previous condition.

Visual conditions are (1) random assignment of visual features to words from font size 13 to 19; (2) no visual features at all; (3) visual features are congruent to spatial locations: words that are close to the target have higher font size.

We therefore ended up with 108 images: 18 instructions x 2 semantic conditions x 3 visual conditions. We now present how we implemented our visual, semantic and memory maps to predict scanpaths on these images.

### 3.2 Maps

The unit of our maps should normally be the pixel, but for the sake of psychological validity, we are currently using the word since our images contain nothing but words. Actually, users are not looking for a region of pixels but for an object, here a word. Let us take an example. Suppose our model has already made 7 fixations on an image from which the instruction was: *find the most dangerous fish*. lt is now fixating the word *baleine (whale)*.

Figure 2 displays the 3 maps corresponding to this scanpath as well as the integrated map. Word colors represent weights: the darker, the higher. When summing up the 3 maps, words are given new weights. The next fixation in made on the word that obtained the highest weight (*requin (shark)* in our example).

Let us now detail how we implemented each map. Each one dynamically assign a weight to each *word*, conditional on the current scanpath. These weights are in-between 0 and 1: the higher the more attractive.

**Visual map** Two kinds of information could affect bottom-up gaze movements: the physiological features of the retina which tend to promote local areas over further ones, and the visual saliency of the current image. Our visual map is therefore a classical visual saliency map multiplied by a filter corresponding to physiological parameters, namely the visual acuity per degree of eccentricity.

$$weight_v(word) = visualSaliency(word) * visualAcuity(word, currentFixation)$$

In our experiment, the saliency of each word only depends on the printed surface of the word, roughly computed as the square of its font size multiplied by its number of characters and normalized to be in the [0,1] range.

$$visualSaliency(word) = nbChar(word)/9 * (fontSize(word)/19)^2$$

The visualAcuity function depends on the distance between the current fixation and the given word, following the classical curve of visual acuity as a function of eccentricity.

**Semantic map** As we mentioned previously, the underlying assumption of our semantic component is that spatial proximity reflects semantic similarity: things

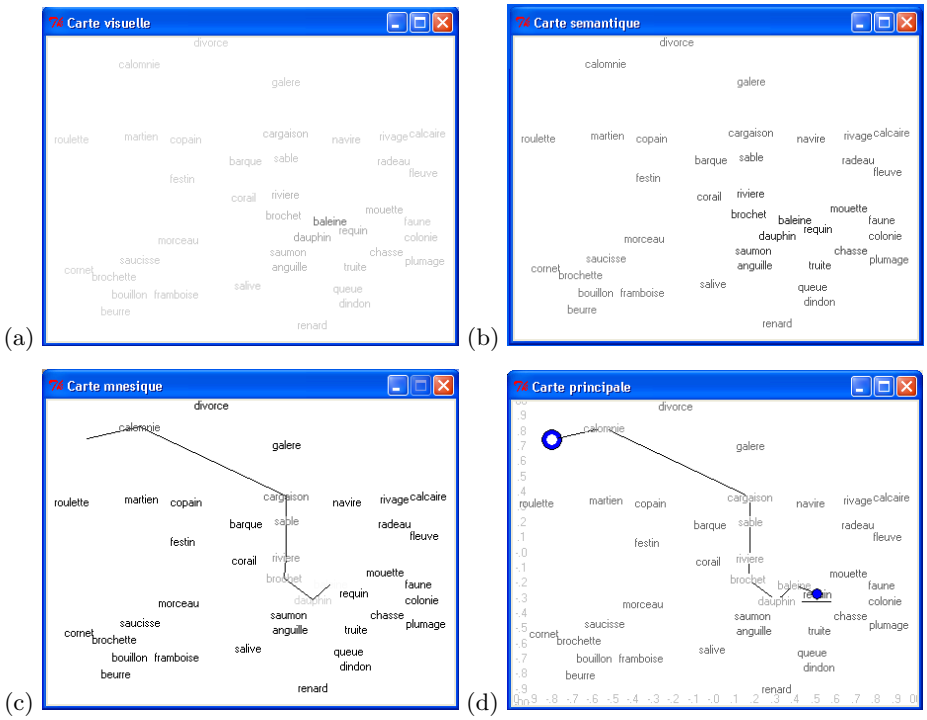


**Fig. 2.** Examples of (a) visual, (b) semantic, (c) memory and (d) integrated maps.

that belong to the same category are usually near each other in our world. If you are seeking cauliflowers in a brand new supermarket and you are in front of laundry soaps, you would better avoid the current area and search elsewhere. However, if you are in front of carrots, you are almost there! We implemented this idea in the following way: first we computed the semantic similarity between the fixated word and the instruction, using LSA. If this similarity is under 0.2[3], the weights of the current zone are given low values, following a Gaussian around the current fixation. If the similarity is above 0.2, the weights are given higher values still following a Gaussian around the current fixation. The height of the Gaussians depends on the similarities: weights are maximum for a word close to the current fixation and a high similarity between the current word and the instruction. Figure 3 shows the height of the Gaussian as a function of the semantic similarity and examples of Gaussians as a function of the distance from the current fixation.

[3] This value of 0.2 is usually considered in the LSA literature as a threshold under which items are unrelated.

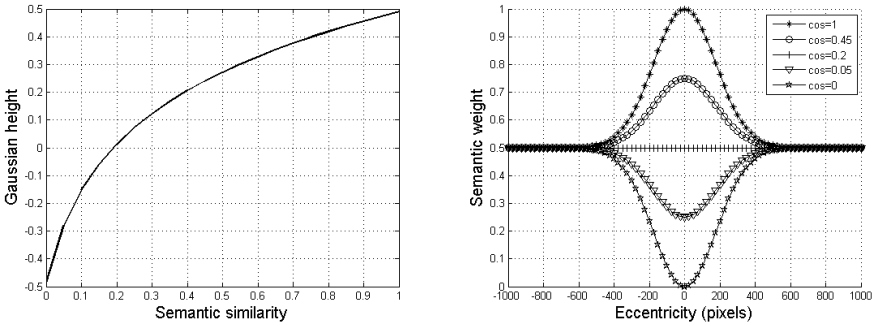


**Fig. 3.** (a) Semantic Gaussian height as a function of semantic similarity; (b) Examples of Gaussians as a function of the distance from the current fixation.

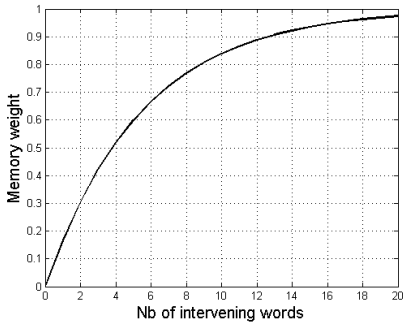


**Fig. 4.** Memory weights as a function of the number of intervening words.

**Memory map** Humans generally do not move back to locations previously visited, although this can be sometimes observed. In the literature, this functionality is usually implemented by means of an inhibition of return mechanism which prevent models from moving back to the previous fixations. In order to have a more flexible mechanism, our memory map strongly decreases the weight corresponding to the last fixations, but slightly increases all previously visited fixations. The latter aims at modeling a forgetting mechanism, so that the model could still go back to words that were visited several saccades before. Basically, the current fixation is given a weight of 0, meaning that we do not want to go back to it, and weights increase as the number of intervening words in the scanpath increases (Fig. 4).

**Map integration** Our model was designed in order to limit as much as possible the number of free parameters. We tried to put a cognitively-based rationale behind each component parameters. However, we did not set the weight of each

component in the general performance of the model. Which component is playing the larger role? Is the semantic process important in such a task? Is memory so crucial? Is visual information necessary at all? There is no reason to assume that visual, semantic and memory processes play identical roles. Therefore, they are integrated to form the general map by means of a weighted sum:

$$M_{general} = \alpha_V.M_{visual} + \alpha_S.M_{semantic} + \alpha_M.M_{memory}, \qquad \alpha_V + \alpha_S + \alpha_M = 1$$

The comparison to experimental data will tell us which are the best values for these weights. Before explaining this comparison, we now present the way experimental data were collected.

## 4 Comparison to experimental data

### 4.1 Experiment

**Participants** Forty-three students (average age 20.9 years) participated in the experiment. Each received 5€ for their participation in this experiment, and all had normal or corrected-to-normal vision.

**Stimuli** Images were displayed on a computer monitor (1024*768 pixels) at a distance of 50 cm from the seated participant, generating a image subtending 42 horizontal deg. of visual angle.

**Procedure** Each subject was in the semantic condition or not and there were 6 images with visual feature facilitating (closer the target, bigger the font size), 6 neutral, and 6 with randomly visual feature (18 trials). Each trial begin with an instruction, followed by a cross fixation. After the subject gazed their fixation on a corner of the display, an image appeared containing the 40 words and observers were instructed to find the best answer as soon as possible. An image was displayed until the subject responded without maximum delay. None of the images used in the experiments were used during the training.

**Apparatus** The right eye was tracked using EyeLink II, which is a head mounted eye tracking device (500 Hz Sampling Rate). This was used to record the position and duration of fixations during the search.

### 4.2 Comparison

Remind that our goal is to find the $\alpha_V$, $\alpha_S$ and $\alpha_M$ parameters. Our method is to compare human and model scanpaths on various combinations of $\alpha_V$, $\alpha_S$ and $\alpha_M$ in order to highlight relevant values. Directly comparing human scanpaths and model scanpaths would not be informative though, because there is too much variability in the way a target is found. Therefore, we had to select high-level variables which would be able to characterize such scanpaths.

**Variables** We defined the following variables:

- **number of fixations until target is found**. This variable is an indication of the difficulty participants had in finding the target;
- **average angle between saccades**. Humans tend to rationalize their scanpaths in order to minimize their effort and this variable is able to characterize the general shape of the scanpath. For instance, humans seldom show 180° angle (half-turn) between saccades.
- **rate of progression saccades**. Saccades could either be progression saccades if the new fixation is closer to the target than was the previous fixation, or regression saccades. This variable aims at characterizing the general behavior of our participants which is to go more or less quickly to the target.

**Discriminating power** One manner to investigate the power of these variables is to look at their ability to discriminate humans scanpaths with respect to different conditions. We compared the average angle between saccades in the semantic and non-semantic conditions. We found a statistical significant difference between these conditions using a Student test ($p=.036$). In the same way, we compared the rate of progression saccades in the semantic condition only, for the first 6 images the participants saw, for the next 6, and for the last 6. There is indubitably a learning process occuring in the semantic condition of our task because participants should realize after a while that words are semantically organized. We actually found that the rate of progression saccades is able to capture the difference between the first images and the last ones ($p<.001$). The number of fixations until the target is found is also different in the first 6 images seen and the last 6 ($p<.02$) Therefore, all variables seem good candidate to compare human and model behaviors.

**Fitting** We ran the model for 21 values of $\alpha_V$ and 21 values of $\alpha_M$ both from 0 to 1. $\alpha_S$ is directly obtained since $\alpha_V + \alpha_S + \alpha_M = 1$. To generate several scanpaths per condition, we introduced a bit of noise in the model by not choosing just the best-weighted word in the general map, but randomly selecting between the best one and the second best one. We ran the model 20 times for each combination.

Before going into details in the results, let us discuss some extreme cases.

- If the visual weight $\alpha_V$ is set to a very low value, the model is not dependent on visual information and, more important, is not bound to stay in the local area. The model therefore shows long saccades, sometimes going from one side of the screen to the other.
- If the semantic weight $\alpha_S$ is set to a very low value, the model does not take into account semantic information: it seems to wander from word to word. If the memory is set to a high value, it looks like an exploration behavior.
- If the memory weight $\alpha_M$ is set to a very low value, the model tends to go back to words it has just seen, or even continuously refixates the same word.

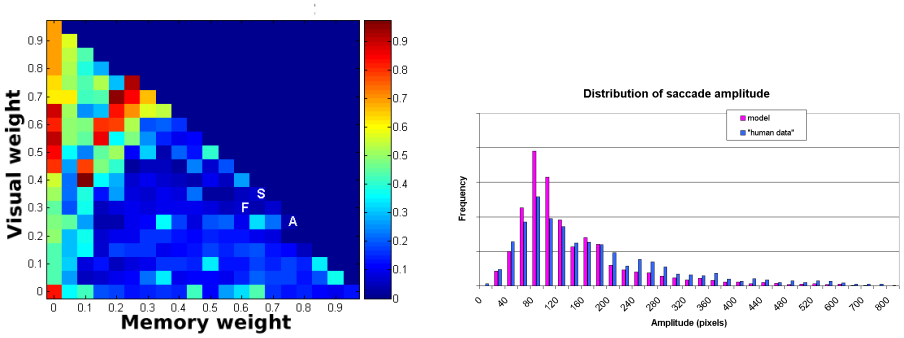


**Fig. 5.** (a) Average relative errors for all values of $\alpha_V$ and $\alpha_M$ (b) Saccades distribution of human data and best model

The only human data we kept for comparison with the model concerned the semantic condition and, more, the only last 6 images each participant see in this condition, to make sure the semantic organisation was understood by participants. We also removed some bad quality scanpaths. We ended up with a total of 4085 fixations.

We computed the relative error $\delta$ of the model data with respect to the human data for each variable and for each combination of $\alpha_V$, $\alpha_S$ and $\alpha_M$. Figure 5a displays the relative errors averaged over the three variables for all values of the visual and memory weights. Values corresponding to minimum relative errors for specific variables are marked F (number of fixations), A (angle between saccades) and S (rate of progression saccades). Minimum errors and corresponding weights are the following:

- **number of fixations until target is found**: $\delta = 0.0045$ corresponding to $\alpha_V = 0.3, \alpha_S = 0.1, \alpha_M = 0.6$;
- **average angle between saccades**[4]: $\delta = 0.0031$ corresponding to $\alpha_V = 0.25, \alpha_S = 0, \alpha_M = 0.75$;
- **rate of progression saccades**: $\delta = 7.10^{-5}$ corresponding to $\alpha_V = 0.35, \alpha_S = 0, \alpha_M = 0.65$;

All three variables give very similar results. Basically, memory plays the highest role. It requires a weight of 60 to 70% to best simulate humans. The visual component needs about 30% and the semantic component plays a minor role in our task.

To perform one more validation, we compared the human distribution of saccades with the ($\alpha_V = .30, \alpha_S = .10, \alpha_M = .60$) model. Actually, a good cognitive model of eye movements should exhibit a distribution of saccades close

[4] We took the second best relative error for this variable since the best one correspond to a non-relevant model with almost no memory and a extremely high level of refixations from which the angle cannot be calculated most of the time.

to the human one. We found a pretty good fit (Fig. 5b) which is another evidence in favor of the respective contributions we found.

## 5 Conclusion

This paper presents a threefold model of information seeking which was implemented and tested on a word seeking task. We had to make several choices in the design of the 3 components of the model but we tried to be as objective as possible: the visual component is based on the decrease of visual acuity as a function of eccentricity ; the semantic component is based on similarities produced by a cognitive model of semantic associations and the memory component attempts to account for both inhibition of return and forgetting mechanisms. We compared the model output with experimental data and found that the best fit is obtained when the memory is given a strong weight and the semantic component a low weight. This is obviously dependent on our material but we believe that, after being tested on various tasks, this model could be used to predict the respective contributions of human visual, semantic and memory processes in a given task. Much remains to be done to apply it to web pages and especially towards more sophisticated visual maps which could be more related to saliency maps. Our next work is to apply this model to images with slightly different visual features but also to images with text paragraphs instead of just words.

## References


1. Anderson, J., Matessa, M., Lebiere, C.: ACT-R: A theory of higher level cognition and its relation to visual attention. Human-Comp. Interact. **12**(4) (1997) 439–462
2. Pirolli, P., Fu, W.: SNIF-ACT: A model of information foraging on the world wide web. In Corbett, A., de Rosis, F., eds.: 9th Int. Conference on User Modeling. Volume 2702 of Lecture Notes in Artif. Intelligence. Springer Verlag (2003) 45–53
3. Simon, H.: The structure of ill structured problems. Artificial Intelligence **4**(3) (1973) 181–201
4. Marchionini, G.: Information seeking in electronic environments. Cambridge University Press (1995)
5. Kuhlthau, C.: Learning in digital libraries : an information search process approach. Library Trends **45**(4) (1997) 708–725
6. Guthrie, J.: Locating information in documents : examination of a cognitive model. Reading Research Quarterly **23** (1988) 178–199
7. Kitajima, M., Blackmon, M., & Polson, P.: A comprehension-based model of web navigation and its application to web usability. In Waern, Y., Cockton, G., eds.: Proceedings of HCI 2000, People and Computers XIV, Springer (2000) 357–373
8. Pirolli, P., Card, S.: Information foraging. Psychological Review **106**(4) (1999) 643–675
9. Fu, W., Gray, W.: Suboptimal tradeoffs in information seeking. Cognitive Psychology **52** (2006) 195–242
10. Treisman, A., Gelade, G.: A feature-integration theory of attention. Cognitive Psychology **12**(1) (1980) 97–136

11. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence **20**(11) (Nov 1998) 1254–1259
12. Mannan, S.K., Ruddock, K.H., Wooding, D.S.: The relationship between the locations of spatial features and those of fixation made during visual examination of briefly presented images. Spatial Vision **10** (1996) 165–188
13. Yarbus, A.: Eye movements during perception of complex objects. In Riggs, L.A., ed.: Eye Movements and Vision. Plenum Press: New York (1967) 71–196
14. Henderson, J., Brockmole, J., Castelhano, M., Mack, M.: Visual saliency does not account for eye movements during visual search in real-world scenes. In van Gompel, R., Fischer, M., Murray, W., Hill, R., eds.: Eye movements: A window on mind and brain. Oxford: Elsevier (2007) 537–562
15. Navalpakkam, V., Itti, L.: Bottom-up and top-down influences on visual scanpaths. In Rogowitz, B., Pappas, T.N., Daly, S., eds.: Proc. SPIE Human Vision and Electronic Imaging XI (HVEI06), San Jose, CA. Volume 6057., Bellingham, WA, SPIE Press (Jan 2006)
16. Feng, G.: Eye movements as time-series random variables: A stochastic model of eye movement control in reading. Cognitive Systems Research **1**(1) (2006) 70–95
17. Landauer, T., Foltz, P., Laham, D.: An introduction to latent semantic analysis. Discourse Processes **25** (1998) 259–284
18. Klein, R.: Inhibitory tagging system facilitates visual search. Nature **334**(6181) (1988) 430–431
19. Pratt, J., Abrams, A.: Inhibition of return to successively cued spatial locations. Journal of Experimental Psychology: Human Perception and Performance **21**(6) (1995) 1343–1353
20. Tipper, S.P., Weaver, B., Watson, F.L.: Inhibition of return to successively cued spatial locations: A commentary on Pratt and Abrams (1995). Journal of experimental psychology : human, perception and performance **22**(5) (1996) 1289–1293
21. Landauer, T., McNamara, D., Dennis, S., Kintsch, W., eds.: Handbook of Latent Semantic Analysis. Lawrence Erlbaum Associates (2007)