Object and Gist Perception in a Dual Task Paradigm: Is Attention Important?

Maria Koushiou[1] (maria_koushiou@yahoo.com)

Elena Constantinou[1] (elena_constantinou@yahoo.gr)

[1]Department of Psychology, University of Cyprus,
P.O. Box 20537, 1678 Nicosia, Cyprus

Abstract

An experiment was conducted to investigate object and gist perception under conditions of inattention. Participants performed an attentionally demanding central task while responding to a secondary peripheral task involving object categorization or gist identification in natural scenes. Participants were unexpectedly more accurate on object than gist categorization, a finding attributed to a possible facilitatory effect of object figure saliency. This hypothesis was confirmed by a second experiment comparing the peripheral tasks under single-task conditions. A third dual-task experiment was conducted to compare object and gist perception when controlling for figure saliency by using the exact same stimuli for both peripheral tasks. No significant differences were found in the third experiment. Conflicting results and methodological issues are thus discussed.

Keywords: object categorization; gist perception; dual-task methodology.

Introduction

Modern computational models of visual attention propose as a necessary stage in the processing of visual stimuli a pre-attentive level during which early visual features are computed in a bottom-up fashion in order to direct attention towards stimuli of interest (Itti & Koch, 2001). However the extent of processing that takes place at this pre-attentive stage is not quite clear, with empirical evidence often offering contradicting results.

Phenomena like inattentional blindness and change blindness, provide evidence for the presence of significant pre-attentive perceptual limitations since humans fail to perceive highly visible objects (Mack, 2003) or to detect large changes in a visual scene (Simons & Ambinder, 2005) when their attention is allocated elsewhere. However, the use of priming methodology in inattentional blindness experiments, has shown that unconscious information prime subsequent behavior, indicating that visual information (e.g. the observer's name) may be processed perceptually prior the engagement of attention (Mack, 2003).

This empirical contradiction is also depicted in theory since early selection theories (Broadbent, 1958) support that only a rudimentary analysis of physical features occurs before attentional selection while late selection theories (Deutch & Deutch, 1963) claim that perception is achieved in an automatic and parallel fashion, with attentional selection intervening only after full perception of items is achieved. Lavie (1995) proposed a model in an attempt to resolve the early vs. late selection controversy. According to this model, the perceptual load of high-priority relevant stimuli determines whether selection is early or late.

Experimentally, one way of testing the applicability of the above theories is provided by the dual-task paradigm. In this paradigm attention is drawn to a perceptually demanding central task while irrelevant to the central task stimuli are presented on the periphery. In dual task experiments researchers assess whether secondary tasks involving peripheral stimuli interfere with performance in the concurrent central task.

Experiments using this procedure have shown that our perceptual system is subject to capacity limits when the stimulus load is increased (Pashler & Johnson, 1998). Efficient performance on each task requires perceptual analysis followed by central operations to produce a response. Since central processing of the two tasks cannot be done concurrently, the processing of one task presumably occurs in parallel with the perceptual analysis of the other task. Thus, in a dual task condition our brain seems to switch the central processing from one task to another, while buffering the essential information for both tasks concurrently (Pashler & Johnson, 1998).

Different levels of perceptual processing have been examined using dual task methodology to determine the type of information that can be processed at a pre-attentive level. The present study focuses on two kinds of processing natural scenes, namely the categorization of objects in scenes and the perception of the scene's gist. Our aim is to examine what types of information one can extract from a natural scene before focusing attention on certain aspects of it. An overview of previous research examining these two types of processes is presented next.

Perception of Objects in a Scene

According to the classical view, perception without attention is possible when detecting primitive, low-level features such as motion, orientation, texture and brightness. Objects, on the other hand, consist of superior stimuli that require focused attention and a higher level of processing (Braun, 2003 cited in Evans & Treisman, 2005).

Experimental evidence from studies using the dual task paradigm, Inattentional Blindness, and Rapid Serial Visual Presentation (RSVP), challenge the aforementioned view,

suggesting that high-level representations can be accessed under conditions of little or no attention. For example, Li, VanRullen, Koch and Perona (2002) examined performance on a natural scene categorization task under single-task and dual-task conditions and found no significant differences. In the single-task condition, participants had to detect an animal or a vehicle in natural images that were briefly flashed in front of them whereas in the dual-task condition participants were asked to perform the same task while simultaneously performing an attentionally demanding foveal letter discrimination task. Their results indicated that complex stimuli (e.g. animals) appearing in different environmental contexts can be detected even when attention is allocated on a demanding central task (Li et al, 2002).

In support of the previous findings, Li, VanRullen, Koch and Perona (2005) found that removing color information from a scene, lack of training, or training on a different type of scene categorization (other than animal), has no effect on the performance of participants on the natural scene categorization task under the same dual task conditions. Increasing distracting information by adding an extra copy of the natural scene in the periphery also failed to impair participants' performance thus indicating that natural scene categorization, as a secondary task, poses little or no attentional load on the central processing system.

Evans and Treisman (2005), on the other hand, examined the robustness of the human visual system in categorizing natural scenes with a series of RSVP experiments. The presentation rate of the images was high enough to prevent attention engagement in the processing of the natural scene. Even though participants were able to accurately detect the target (either animal or vehicle) in a string of 6 images that serially flashed in front of them they were not able to report the identity and the location of the target in the scene. It seems that participants are able to extract various features of a natural scene on which they rely for target detection but they are not able to bind these features together in order to form a complete representation and subsequently report the identity or the location of the object (Evans & Treisman, 2005).

Evans and Treisman (2005) reject the notion that natural scene categorization requires high-level processing including semantic processing but rather propose a feature-based detection that is in line with the two-stage perceptual model presented by Marois et al (2004). According to Marois et al (2004) detection and/or efficient categorization of items entail unconscious, attention-free perceptual analysis whereas identification of visually presented items requires processing that accesses awareness (e.g., binding and consolidation; Evans & Treisman, 2005).

Gist Perception

Another important aspect of high-level scene perception is the capture of the gist, i.e. the general meaning of a scene. Gist perception has been found to occur quickly and early in the visual processing of a scene (see Henderson & Hollingworth, 1999 for a review). It seems that only 100ms after the presentation of an image, observers can recognize the basic-level category of a scene, its spatial layout and other global structural and object-related information (Oliva, 2005). Thus, gist can be conceptual (related to the meaning of the picture) and/or perceptual (related to the perceptual properties of the image, like colour or texture; Oliva, 2005; Potter, Staub, Rado and O'Connor, 2004).

Empirical evidence suggests that a scene may be processed initially as a whole, extracting global information about the image. Segmentation of the scene in specific objects providing local information may occur later in a more elaborated gist formation process (i.e. with regards to intra-objects relations; Henderson & Hollingworth, 1999). Thus, both perceptual and conceptual representations of gist could be initiated without a priori object identification (Oliva, 2005).

On the other hand, Fei-Fei, Lyer, Koch and Perona (2007) found that after 107 ms of scene exposure people could name both object categories and scene environments (e.g. classroom). With shorter presentations, though, people recalled the shape and other low-level sensory features, with semantic information about objects or scenes appearing after 40-67ms of exposure (Fei-Fei et al., 2007). Thus, it seems that shape-related information of a scene exhibits a processing benefit over semantically meaningful recognition.

Fei-Fei et al. (2007) also point out what previous research suggests, i.e. a mutual facilitation of object and scene perception. Researchers hypothesize that scene processing stages occur in parallel and feed information back to each other, facilitating in this way the overall recognition of specific segments of a scene. Thus, as long as semantic information arises our brain takes advantage of it to further process objects and scenes (Fei-Fei et al., 2007).

As far as we know, object and gist perception have not been previously compared directly. In the first experiment we compare object and gist perception using a dual-task paradigm. We expect that gist (operationalized as environment identification) will be extracted more easily since it seems to occur first during scene perception. Object categorization is expected to be harder, since it seems to incorporate both low-level feature information and semantic information (gist) based on Oliva's (2005) theoretical framework and empirical findings.

Experiment 1

The first experiment aims at investigating the level of processing that occurs under conditions of inattention using a dual task paradigm. Participants had to complete an attentionally demanding central task, while performing a secondary task of detection involving a) object categorization and b) gist identification. Both object categorization and gist identification were tested using natural scenes.

Based on previous research we expected that: a) participants' performance on the secondary task will become inferior as the level of processing demands increase, i.e.

their performance will be better in terms of accuracy on gist identification than on object categorization and b) participants performing better on the central task (indicating attention deployment to the central task) will perform less accurately on object categorization than gist perception.

Methodology

Participants Thirty-seven undergraduate and graduate Psychology students from the University of Cyprus, with normal or corrected-to-normal vision, participated in the experiment.

Experimental Design A dual task experiment with two within-subject conditions (peripheral tasks: object categorization and gist identification in natural scenes) was conducted. Each participant performed 128 experimental trials, 64 for Object Categorization and 64 for Gist Identification.

Material and apparatus Testing took place in a laboratory at the Department of Psychology of the University of Cyprus. Participants were seated in front of a computer screen, which presented the stimuli for each task.

The peripheral task stimuli were taken from various sources including databases like LabelMe for gist images (Oliva & Torralba, 2001), SIMPLIcity (Li & Wang, 2003) and Caltech 101 (Fei-Fei, Fergus, & Perona, 2004) as well as by using the Google image search tool for some images. The object categorization images included pictures of animals, buildings, vehicles and pictures of food. The gist category images included pictures of each of the following natural scenes: coasts, cities, mountains and forests.

Central Letter Discrimination Task Each trial started with a fixation cross presented 300 ms before the onset of the first stimulus. The fixation cross was then substituted by the central task's stimuli (a combination of 8 letters). The 8 letters were either all identical, (i.e. all L's presented at random orientations) or included one letter that was different from the other 7 (i.e. one T among L's). The letters remained onscreen for 100 ms before the peripheral stimulus appeared. Stimuli for both tasks remained on the screen for 180 ms. Then, the peripheral task was masked while the central stimulus remained onscreen either until the participant entered a response or after 2000 ms. Participants had to decide whether all letters were the same or not by pressing the appropriate button. For half experimental trials (64 out of 128) the letters were the same while in the remaining half one letter was different. Reaction time (RT) and accuracy scores were recorded for each trial.

Peripheral Task In half of the trials the peripheral task was an Object Categorization task (64 trials) and in the other half it was a Gist Identification task. The order of the presentation of the two blocks of trials was counterbalanced across participants. In each trial, the stimulus was presented 100 ms after the central stimulus onset. The peripheral stimulus remained onscreen for 180ms and it was followed by a mask.

Peripheral stimuli consisted of coloured images depicting animals, food, buildings or vehicles for the object categorization trials (Figure 1a), and images of coasts, mountains, forests and cities for gist identification trials (Figure 1b). In all cases, an image was flashed randomly for 180ms at one of four possible locations in the periphery and it was followed by a blank image (mask). Images were of the same size and appeared at a constant distance from the central stimulus.

a)  b) 

Figure 1: Examples of pictures used in Experiment 1.

After the termination of the central task, participants were asked to report what they saw in the periphery by means of a multiple choice question. The possible answers referred to the four main image categories for each task appearing in a different random order in each trial. Accuracy scores were recorded for each trial.

Procedure Each participant initially completed 16 practice trials of the dual task, comprised of the central task (letter discrimination) and a peripheral task (8 trials with object categorization and 8 trials with gist identification). The two experimental blocks (total of 128 trials) followed the practice block. The timeline of each trial is illustrated in Figure 2. Participants were instructed to focus attention on the central task.

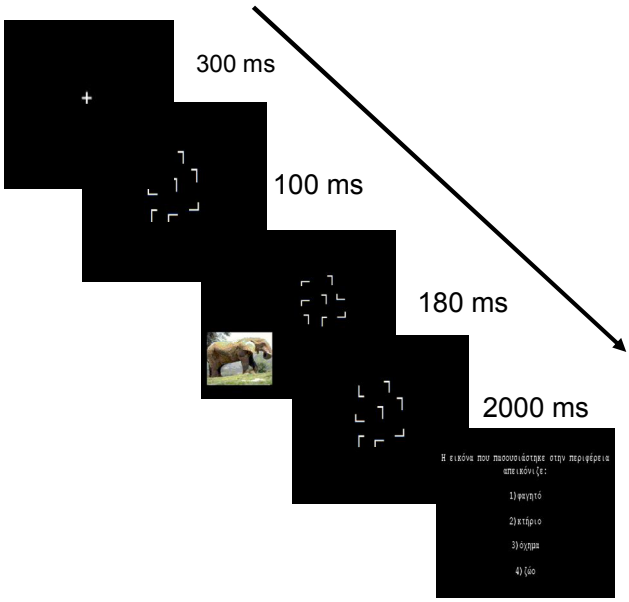

Figure 2: Dual task experiment sequence as appeared in Study 1 trials.

Results

Overall participants' performance on the central task was assessed based on mean RT. The overall RT was to 813.18 ms (S.D. = 265.84) and the overall percentage of accurate responses was 79.3%. Participants' accuracy in each condition of the peripheral task was also assessed by the percentage of accurate responses. Overall accuracy was 79.1% for object categorization and 66.9% for gist perception. Paired-samples t-tests were conducted revealing no significant differences between the object categorization and gist perception conditions for either RT or accuracy on the central task. A statistically significant difference, though, was found on the participants accuracy on the peripheral task with better performance on the object categorization than gist perception, $t(36) = 5.35$, $p < .001$. (Figure 3).

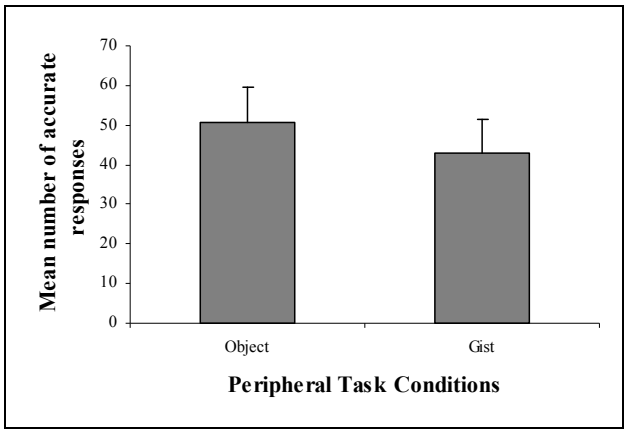

Figure 3: Mean accuracy for the two conditions of the peripheral task[1].

In order to examine whether participants shifted their attention to the peripheral task and to assess the possible effects of this shift on performance on the peripheral task, we divided our sample into groups according to their performance on the central task. Participants were assigned into two groups based on their mean RT (fast/slow) and two groups according to the mean number of accurate responses (high/low accuracy). A Multivariate Analysis of Variance (MANOVA) revealed that the high accuracy group performed significantly more accurately than the low accuracy group in the object categorization condition but not in the gist perception condition, $F(1, 33) = 4.22$, $p < .05$. (Figure 4). Central task RT (fast/slow groups) did not differentiate significantly the performance in the peripheral tasks.

[1] Error bars represent standard deviations.

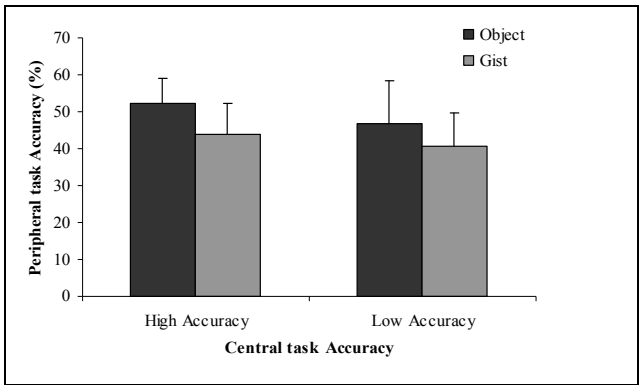

Figure 4: Participants Central Task Accuracy (high/low) and accuracy on the two peripheral tasks[1].

A one-way ANOVA showed a significant effect of the type of stimuli on the accuracy for both object, $F(3, 144) = 5.06$, $p < .01$, and gist perception, $F(3, 144) = 5.24$, $p < .01$. Post hoc comparisons using the LSD test indicated that participants were better on building recognition ($M. = 14.11$, $SD = 2.61$) than on animal ($M. = 11.81$, $SD = 2.56$) or food ($M. = 11.84$, $SD = 3.44$). In the gist identification task, participants could recognize more accurately city scenes ($M. = 12.05$, $SD = 3.14$) than mountain ($M. = 9.13$, $SD = 3.67$) or forest scenes ($M. = 10.71$, $SD = 3.47$).

Discussion

Present results indicate that participants could perform above chance level on both the central and the peripheral task, showing that object and gist perception can occur under dual task conditions. However, participants performed better on the object categorization than on the gist identification task, a finding that contradicts our initial hypothesis.

One possible explanation for this finding could be that objects in natural scenes are more salient and their features (e.g. contour, texture, and surfaces) facilitate the object recognition procedure without a significant effect on the central task performance. In support of this, participants were more accurate at building images, which include large figures, distinct from their background and consist of straight lines and surfaces, as compared to images with food and animals. Even more, participants are more accurate in recognizing the gist of city scenes than any other category possibly due to the fact these scenes include salient features (e.g. buildings).

Another unexpected finding of this study was that better performance on the central task was associated with higher accuracy in object recognition. Higher accuracy on the central task seems to be an index of successfully divided attention as opposed to full attention towards the central task, as it was related to higher accuracy in the peripheral object recognition task. On the other hand, participants with low performance on the central task did not manage to

divide their attention successfully in order to perform more accurately either the central or the peripheral task. This effect was observed solely in the object recognition task and was not present for the gist identification task, perhaps because extracting the figures in object recognition images was easier than extracting the gist.

A significant limitation of the study is the lack of a single-trial condition that would indicate whether the object and gist pictures we used were comparable in terms of perceptual saliency. This was deemed necessary since all images underwent editing before being used in the experiment and especially those used in the gist identification task could have been more indiscernible than the object pictures.

Experiment 2

We conducted a second experiment in order to investigate whether the advantage of object over gist perception found in Experiment 1, can be attributed to image related factors like figure saliency, regardless of attentional focus.

Methodology

Fourteen students from the University of Cyprus performed the peripheral tasks of Experiment 1 under single-task conditions. The procedure was the same as in Experiment 1 (with the same images appearing at the periphery) but without the central letter discrimination task. Both object categorization and gist identification were performed by all participants.

Results

An independent samples t-test revealed that participants were more accurate on object categorization than on gist identification, $t(26) = 3.65$, $p < .001$. However, RTs were not statistically different between conditions.

Discussion

These results show that object categorization task is easier than gist identification even in the case of full attention. This can be attributed to the possible facilitatory elements of the object pictures (e.g. edges and surfaces). The advantage of object categorization over colour discrimination found in the Li et al. (2002) study was confirmed herein, even when compared to other also semantic stimuli (gist). A third study is however mandated in order to eliminate the effect of different and probably non equivalent in terms of features images used in each category (gist and object images).

Experiment 3

The aim of Experiment 3 was to investigate object and gist perception under dual task conditions, by using the same images for each peripheral task. The images selected for this study, are real-life photos of various contexts (city, forest and beach), all containing objects, animals or people usually found in these sceneries.

Methodology

Twenty students from the University of Cyprus participated in the experiment. The procedure was the same as in Experiment 1, although some modifications were made for the peripheral task. The peripheral task was used as a between-subject variable, with half of the participants performing object categorization on the periphery and the others performing gist identification.

The experiment consisted of 135 experimental trials for all participants and 18 practice trials. There were three types of objects (people, animal, and vehicle) that appeared equally often in three different contexts (city, beach, forest). Fifteen coloured images for each combination (beach-animal, beach-people, beach-vehicle, city-animal, city-vehicle, city-people etc.) were used for both tasks (Figure 5). Therefore, all participants viewed the same 135 peripheral images, but half of them had to identify the object in the scene and the other half the gist of the scene. Both accuracy and reaction time were recorded.

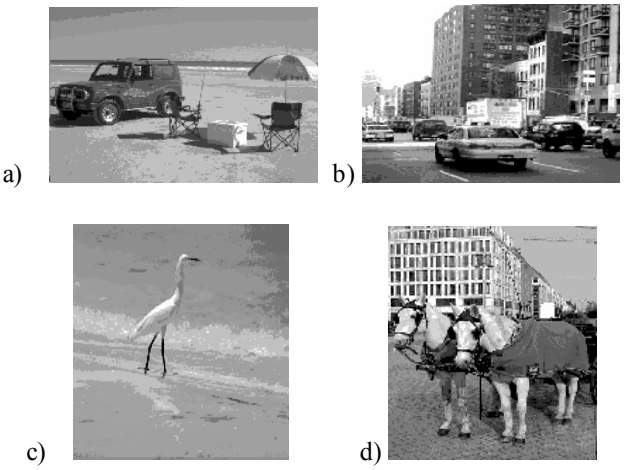

Figure 5: Examples of peripheral images from Experiment 3.

Results

An Independent Sample T-test indicated no significant differences among the two groups of participants on their performance on the central task, neither on the accuracy nor on their reaction time means. The percentage of accurate responses on the central task was 77.77% for the object categorization condition and 78.51% for the gist identification condition, which indicates comparable attention allocation to the central task. Importantly, no statistically significant differences were found on participants' accuracy on the two peripheral tasks, with object categorization accuracy reaching 57.63% and gist identification accuracy 58.30% respectively.

Discussion

Results from Experiment 3 contradict the findings from both Experiments 1 and Experiments 2 suggesting that under

dual-task conditions people have comparable ability to perceive objects and the gist of a scene. We consider the findings of Experiment 3 more valid, since the exact same real-life images were used for both conditions.

Overall, participants' performance on the peripheral task under dual-task conditions was better when they were presented with images containing a single object that appeared centrally in the image (79.1% percentage of accurate responses on object categorization task in Experiment 1). This does not seem to hold when more complex images are used. The images used in Experiment 3 are considered more complex and rich since they contained either more than one object from a category (people, animal or vehicle) or other surrounding context-related objects (i.e. umbrellas on the beach), which probably explains the lower performance (57.6%) compared to Experiment 1. These results are in line with the finding of Walker, Stafford and Davis (2008), who also reported diminished accuracy on object categorization in scenes with multiple foreground objects under dual-task conditions.

Based on the results of our last experiment, we concluded that there is no clear difference in the processing demands required by object and gist perception as these were examined with complex real-life scenes. Lack of differences may be attributed to image properties, since even with the exact same images (as in our third experiment) it is very difficult to control for factors such as saliency of an object within the same context e.g. a coloured parrot in a forest, and a gorilla in the same forest. Furthermore, the exposure time we used was much longer than 107 ms, at which object and gist perception occur mutually facilitating each other according to Fei-Fei et al. (2007). It is not, though, clarified if shorter exposure time can differentiate the two processes.

What seems to emerge from all three experiments is the need for creating different image databases to be used in these sorts of studies, i.e. studies that compare object and gist processing, since the different images typically used for each process (landscape images for gist, single central object images for object processing) may confound conclusions about attentional demands.

Conclusively, future research is needed to delineate the processes required in object and gist perception using larger samples and should compare these processes using images with shorter exposure time that are equal in properties (e.g. number of objects, colour and size of the objects) and processing load.

Acknowledgments

Special thanks to Dr. Marios Avraamides for his valuable contribution to the methodological design and realization of this study and foremost for his encouragement and trust. We are also grateful to all fellow students who participated in the experiments and our reviewers for the useful comments.

References

Evans, K. K. & Treisman, A. (2005). Perception of objects in Natural Scenes: Is it Really Attention Free? *Journal of Experimental Psychology: Human Perception and Performance, 31(6),* pp. 1476-1492.

Fei-Fei, L, Fergus R., & Perona P. (2004). Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. *IEEE. CVPR Workshop on Generative-Model Based Vision.*

Fei-Fei, L., Lyer, A., Koch, C. & Perona, P. (2007). What do we perceive in a glance of a real-world scene?, *Journal of Vision*, 7(1):10, pp.1-29.

Henderson, J.M. & Hollingworth, A. (1999). High-level scene perception, *Annu. Rev. Psychol.*, 50: 243-271

Itti, L. & Koch, C. (2001). Computational modelling of visual attention, *Nature Reviews/Neuroscience*, 2, pp. 1-9.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention, *Journal of Experimental Psychology: Human Perception and Performance*, 21(3), pp.451-468.

Li, F. F., VanRullen, R. Koch, C. & Perona, P. (2002). Rapid Natural Scene Categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of America, 99(14)*, pp. 9596-9601.

Li, F. F., VanRullen, R. Koch, C. & Perona, P. (2005). Why does Natural Scene Categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition, 12(6),* pp. 893-924.

Li, J., & Wang, J. Z. (2003). Automatic linguistic indexing of pictures by a statistical modeling approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9), pp. 1075-1088.

Mack, A. (2003). Inattentional blindness: looking without seeing, *Current Directions in Psychological Science*, 12 (5), pp.180-184.

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, and J.K. Tsotsos (Eds.) *The Encyclopedia of Neurobiology of Attention*, Elsevier, San Diego, CA (pages 251-256).

Oliva, A.& Torralba, A. (2001). Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope**,** International Journal of Computer Vision 42(3), pp. 145–175.

Pashler, H., & Johnson, J.C. (1998). Attentional limitations in dual-task performance. In H. Pashler (Ed.), *Attention* (pp. 155–189). Hove, England: Psychology Press.

Potter, M.C., Staub, A. & O'Connor, D.H. (2004). Pictorial and conceptual representation of glimpsed pictures, *Journal of Experimental Psychology: Human Perception and Performance*, 30(3), pp. 478-489.

Simons, D.J. & Ambinder, M.S. (2005). Change blindness: Theory and Consequences, *Current Directions in Psychological Science*, 14 (1), pp. 44-48.

Walker, S., Stafford, P. & Davis, G. (2008) Ultra-rapid categorization requires visual attention: Scenes with multiple foreground objects, *Journal of Vision,* 8(4):21, 1–12.