

What controls attention in natural environments?

Abstract The highly task-specific fixation patterns revealed in performance of natural tasks demonstrate the fundamentally active nature of vision, and suggest that in many situations, top-down processes may be a major factor in the acquisition of visual information. Understanding how a top-down visual system could function requires understanding the mechanisms that control the initiation of the different task-specific computations at the appropriate time. This is particularly difficult in dynamic environments, like driving, where many aspects of the visual input may be unpredictable. We therefore examined drivers' ability to detect Stop signs in a virtual environment when the signs were visible for restricted periods of time. Detection performance is heavily modulated both by the instructions and the local visual context. This suggests that visibility of the signs requires active search, and that the frequency of this search is influenced by learnt knowledge of the probabilistic structure of the environment.

Key Words: attention, saccades

INTRODUCTION

In everyday life, visual operations are embedded in the context of ongoing behavior, and depend critically on this immediate context. We have little understanding, however, of how visual processes evolve in time in the context of extended behavioral sequences. Investigation of vision in the natural world has revealed that the pattern and duration of the fixations are highly specialized for each situation. In driving, Land has shown that drivers reliably fixate the tangent point of the curve to control steering around the curve (Land & Lee, 1994). In cricket, players exhibit very precise fixation patterns, fixating the bounce point of the ball just ahead of its impact (Land & McLeod, 2000). A similar pattern is seen in table tennis (Land et al, 1997). Ballard et al (1995) found stereotyped fixation patterns in a block-copying task. In tea making (Land et al, 1999) and sandwich making (Hayhoe et al 1999) observers' fixations are tightly linked, step-by-step, with task performance. The duration of the fixations is also different for different tasks. Epelboim et al (1995) showed different fixation durations for tapping and looking at a set of points on a table. In all these cases the visual stimulus is similar throughout the task, but the observer's goals are different. Observers actively select the specific information required for the momentary cognitive goal. What are the mechanisms that orchestrate the selection process?

Selection of just the task specific information from a scene is an efficient strategy. It is not possible to anticipate and compute all the information in a scene that might be needed ahead of time (Ullman, 1984). Some kind of selection is necessary to deal with the computational complexity of representing even simple properties of objects and scenes (Brooks, 1986; Bajcsy, 1986; Ballard, 1991). Task specific strategies not only circumscribe the information that needs to be acquired, but also allow the visual system to take advantage of the known context to simplify the computation (Ballard et al, 1997) However, this computational advantage comes at a cost. Any task driven, or top-down system must deal with the issue of how the particular computations are scheduled. There must be some mechanism for providing the observer with perceptual information that is not on the current agenda. The visual system must balance the

Hiroyuki Shinoda*, Mary M. Hayhoe, & Anurag Shrivastava
Center for Visual Science, University of Rochester, Rochester NY
14627, USA

*Present address: Department of Photonics, Ritsumeikan University,
Kusatsu Shiga 525-8577, Japan.

selectivity of ongoing task specific computations against the need to remain responsive to novel and unpredictable visual input that may change the task agenda. How do we select what we need to see without knowing what is there in the first place? This problem is particularly challenging in dynamic environments. In the context of driving, for example, it is usually not known ahead of time where the traffic signs are located, or when the car in front might turn. This issue has been described by Ullman (1984) as the “initial access” problem. We refer to it here and elsewhere (Hayhoe, 2000) as the “scheduling” problem.

How do observers switch from one behavior to another? What determines when an observer will attend to a traffic sign instead of the car in front? The traditional solution to this problem has been to assume that some ongoing “pre-attentive” analysis of the visual image takes place, and that the products of this analysis attract the observer’s attention to important aspects of the image for further processing (Neisser, 1967; Wolfe 1994, 1999; Treisman, 1993; Itti & Koch, 2000) Temporal transients, for example, are perceptually salient and typically attract attention (Yantis, 1998). In this view basic visual responses are driven “bottom-up” by the image and serve as a basis for more extensive processing. However, the extent and effectiveness of bottom-up vision is unclear. A number of studies have shown that the ability of even stimuli such as a unique color or abrupt onset to attract attention is modulated by the current attentional set (Yantis, 1998). For example, Folk et al (1992) showed that a color singleton attracted attention during visual search for a colored object, but not during search for an abrupt onset stimulus. They proposed that subjects adopt an “attentional control setting” that determines which features will control the deployment of attention in any given task. This may be a way to benefit from the advantages of attentional selection while solving the scheduling or initial access problem. There is still uncertainty, however, whether some stimuli such as onsets or novel objects invariably attract attention whatever the task (Yantis, 1998; Gibson & Jiang, 1998; Gibson & Kelsey, 1998; Turatto & Galfano, 2000).

Since the ability of particular stimuli to attract attention is very sensitive to the conditions of the experiment it is difficult to predict how feature-driven attention might generalize to the natural world. Retinal transients cannot be relied on to attract attention, since transients are continuously generated across the entire visual field as a result of the observer’s own movements. In ordinary life, the visual system needs to deal with the entire visual field, which presents a complex

unsegmented spatial array. In contrast, typical experimental displays subtend only a small portion of the visual field, and the stimuli are usually simple, easily segmented, geometric forms. Another important factor is time. The temporal evolution of behavior in natural environments occurs over seconds to minutes, and is hard to address in standard experimental paradigms. The issue of scheduling, or how behaviors are initiated, can only be addressed indirectly, since the observer’s task is defined by the experimenter, and remains constant within blocks of trials. Nor is it always clear how observers define the task. Thus the whole area of the display might be accorded some attention as potentially relevant in the experiment. Mack & Rock (1998) and Gibson & Kelsey (1998) show that observers change their attentional set in response to an unexpected event on a single trial.

Studies of the ability of pre-attentive mechanisms to control the agenda typically look for interference effects from unattended stimuli while subjects are engaged in another task. The present study takes a different approach, and asks to what extent can top-down mechanisms handle the demands of normal vision. This can be done by exploiting advances in rendering to create virtual environments that can be updated at a rate that creates a realistic impression of the natural world. In such environments, arbitrary changes can be introduced to manipulate the availability of task relevant information at critical points in the ongoing behavioral sequence, while maintaining the continuity of the input from the visual world. By manipulating the temporal location of task-specific information we can assess the observer’s strategies for obtaining it.

Is feature-driven analysis needed to handle the unpredictability of the natural world? In order to address this we used a driving task, where the observer must select information necessary for task performance at the appropriate time, from a complex and constantly changing image. Driving has an obvious ongoing visual agenda. There are a number of tasks that require different information from the image: steering, avoiding obstacles, detecting traffic signals and braking, observing the layout of the surroundings etc. How is each of these tasks initiated? In a purely top-down system, mere presence of a stimulus such as a Stop sign will not be enough to ensure detection if it does not coincide with an episode where the observer is actively searching for signs. If the stimulus itself attracts attention, however, it should be detected whenever it is presented. Thus the logic of the experiment was to examine the visibility of Stop signs, which can be detected on the basis of a simple color

feature, when the signs are present only for brief periods during the driving sequence. One way that observers might handle environmental uncertainty in a top-down system is by using learnt knowledge of the probabilistic structure of the environment to initiate task specific computations at likely points. Chun & Yiang (1998, 1999) showed that visual search is facilitated by implicit learning of spatial structure in static and dynamic environments, and of the covariation between sets of objects. To examine whether this is a factor in normal viewing, we manipulated the a priori likelihood of the stimulus and observe the extent to which this modulates the probability of responding to the sign. In addition to this we manipulated the subject's overall goals. If observers rely on bottom-up scene analysis to initiate a particular visual computation, these manipulations should have little effect. However, subjects' behavior should be sensitive to these manipulations to the extent that it is controlled by top-down factors.

Another difficult issue is determining exactly when the Stop sign is visible to the observer. Conscious report is likely to be an incomplete measure of visual processing, since it necessarily reflects only what the observer remembers, and not the instantaneous processing (Wolfe, 1999). In addition, Hayhoe et al, (1998) showed that fixation durations revealed sensitivity to stimulus changes that could not be reported verbally. Since fixation locus and the focus of attention are tightly linked, eye movements are likely to be a more sensitive measure of the observers' attentional state. This too, is an incomplete measure, as subjects can easily distribute their attention covertly across the visual field in the absence of eye movements (Engel, 1971; Saarinen, J. & Julesz, 1991; Summala et al, 1996), and some kinds of information, like optic flow, require information from the entire visual field. However, a variety of psychophysical and imaging studies support the idea that the shifts in attention made by the observer are usually reflected in the fixations. Corbetta (1998) and Culham et al (1998) have demonstrated that the regions in posterior parietal and frontal cortex activated in fMRI studies of overt and covert attention are almost identical. Findlay & Gilchrist (1998) and Motter & Belky (1997) have also argued that fixations reflect attentional distribution in visual search experiments. The tight link between fixations and task performance in natural situations, as described above, also lends credence to the idea that fixations reflect the primary distribution of attention. However, because of the difficulty in equating fixations and attentional state, we also measured the drivers' braking behavior. Thus if a

driver detects a Stop sign in the peripheral retina without fixating it, this may be revealed by braking to stop at the sign. All three measures, eye movements, stopping, and verbal report, were used in the experiment, since they all provide evidence that the sign was visible to the subject in some way.

METHODS

Driving Simulator & Eye Tracker. The driving simulator was based on a go-kart frame in which a steering wheel, a gas pedal, and a brake pedal were instrumented with potentiometers. The graphic system, a Silicon Graphics Onyx processor equipped with four R10,000 processors and an Infinite Realty rendering engine, generated a pair of stereo images at up to 60 Hz. Stereo images were presented through a Virtual Research V8 head-mounted display (HMD). The V8 HMD had a pair of 1.3" LCD panels with a resolution of 640x480, which offered a visual field of 60x45 degrees. The actual field of view in the helmet was 54 deg. The position and orientation of the subject's head was measured by the Polhemus Fastrak 3-SPACE motion tracking system that updated the image at frame rates (typically about 30 Hz). The subject's left eye was tracked by an ASL Model 501 eyetracker mounted in the V8 HMD. Data was in the form of a video record of the image presented to the subjects, with gaze direction indicated by a white cursor. Car distance, speed of the subject's car, and their behavioral data such as steering, gas, and brake were store in the Graphics computer. For one subject, eye position data was collected in the data stream, in addition to the video record.

Virtual Environment. The virtual town, a Silicon Graphics 'Performer Town', consisted of irregularly sized city blocks, and extended over 600x400 m². The texture mapping and wide variety of buildings and objects gave the town a realistic appearance. There were no pedestrians and no vehicles, except the lead car during the experiment.

Procedure. Fourteen paid volunteers participated in the experiment. All the subjects had normal color vision, normal visual acuity without glasses, and abundant driving experience. They were all naive about the experiment. All subjects were given a practice trial at the beginning of the experiment to get used to the virtual environment. To control their path through the town and to provide an attentional load, we asked subjects to try to keep a constant distance behind a lead vehicle. In one

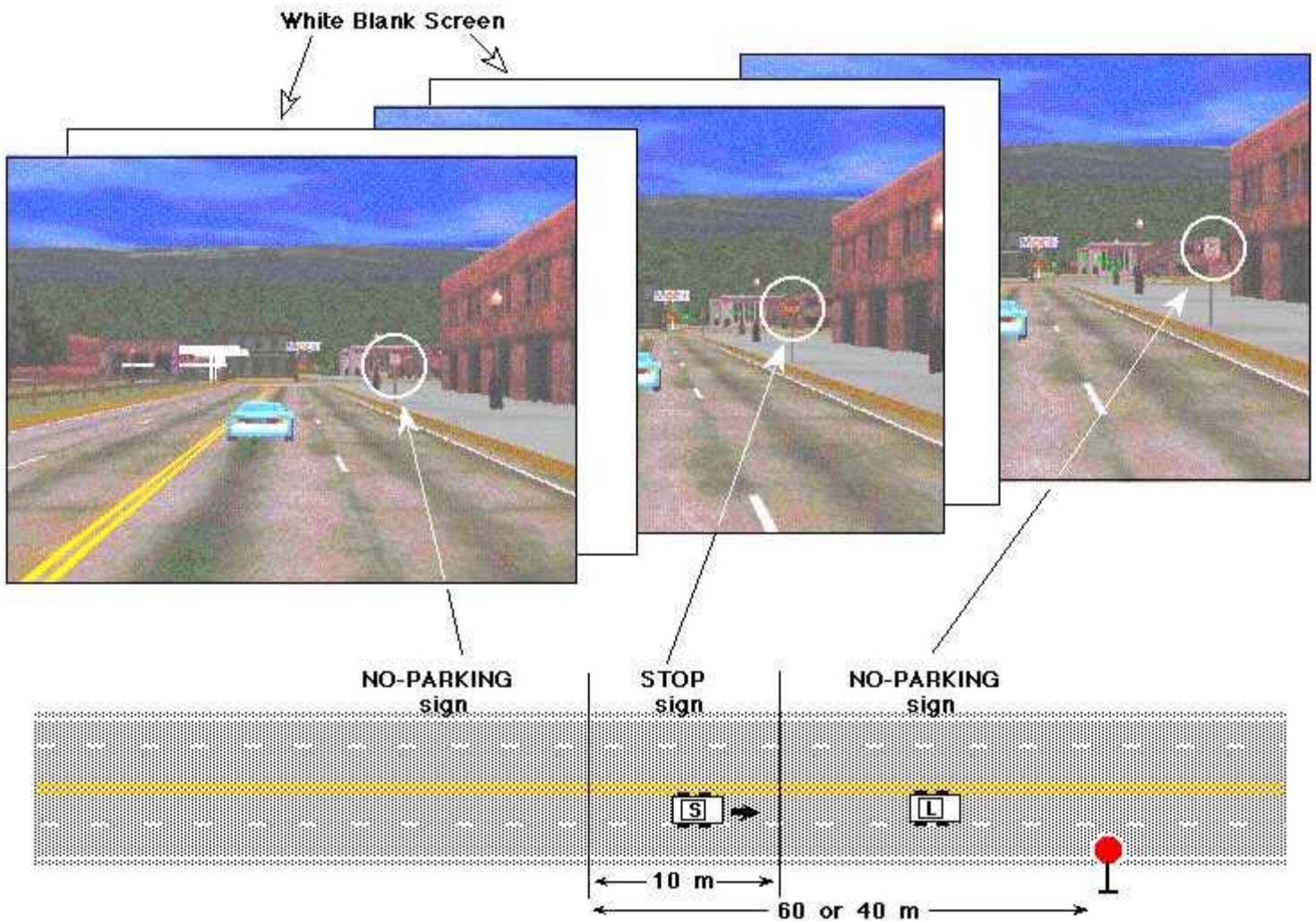


Figure 1. Sign alternation sequence. A No Parking sign was replaced by a Stop sign for 0.5 - 1.0 sec while the subject was in a predefined region, which was set at either 60-50m or 40-30m before the sign. The sign alternation occurred either at an intersection or in the middle of the block. An 80 msec white blank screen was interposed at the moment of sign change, to mask local retinal transients that might draw subjects' attention to the sign. This enabled us to analyze the effect of the presence of a Stop sign itself, rather than a retinal transient. The lead car (L) cruised along predefined path. The lead car's speed was sinusoidally modulated around 13m/s with amplitude of 2m/s at frequency of 0.25Hz so that it engaged the subjects' attention for a substantial fraction of the time.

particular block a No Parking sign was replaced by a Stop sign for a 0.5-1.0 sec period. The sequence of events is shown in Figure 1. A uniform white frame (130 msec) was inserted prior to the transition to mask retinal transient generated by the change. This technique has been used in a variety of change blindness experiments (Rensink et al, 1997; O'Regan et al, 2000). Similar white frames were inserted at random points during the

experiment, so that subjects would not associate the occurrence of the blank with a sign change.

To manipulate the behavioral relevance of the signs, Subjects were asked either to follow a lead car, keeping a constant distance behind it (F task), or else to observe normal traffic rules in addition to following (F+S task). In each of these tasks, no other instructions were given. In the F+S task, no mention was made of stop signs. Each subject drove through the town 4 times; the first 2 trials were for the F task and the last 2 trials were for F+S task. This order was used because it seemed likely that the F+S instruction might influence behavior in the F condition if it followed the F+S instruction. The temporary Stop sign appeared at a distance of 60-50m in the first and the third trial, and at closer distance (40-30m) in the second and the fourth trial. We did not know ahead of time the point at which subjects might look for signs preparatory to stopping. These two distances were chosen to span the range of likely distances. At these

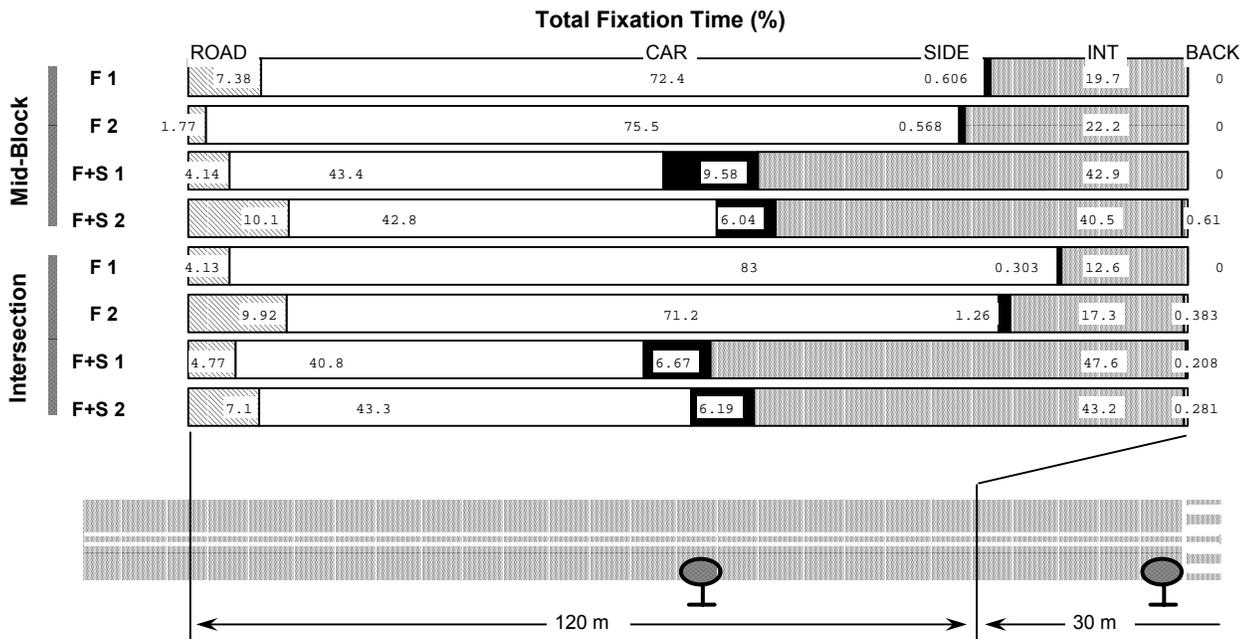


Figure 2. Proportion of time gaze location was either on the road (ROAD), the car in front (CAR), the side of the road (SIDE), the region of the intersection (INT) or the on the background (BACK). Data are shown for each of the four trials for Mid-Block and Intersection groups, each averaged over 7 subjects. Standard errors are given in the text. The data are for the block in which the sign change occurred up to the point when the driver was 30 m from the intersection. (After that point the details of the fixations are governed by the details of the intersection and by the observer’s decision to stop.)

distances, the sign was between 7° and 14° eccentricity, and the sign subtended from 0.8° to 1.8°. For each of the distances, a different route was used, one for trials 1 and 3, the other for trials 2 and 4.

To test whether subjects make use of learned probabilities about the environment to control visual sensitivity, we manipulated the location of the sign. For half of subjects, the sign was at an intersection. For the other half, it was in a less likely location in the middle of a block. (Thus the location factor involved independent groups, whereas the instruction manipulation involved repeated measures on the same subjects.) Image parameters of Stop sign, such as size, eccentricity, and duration, were identical regardless of its location (intersection or mid-block). On each trial there was only one critical block where the sign alternation occurred. As well as monitoring gaze and braking performance, the subjects were asked at the end of the experiment if they noticed anything unusual. If they did not report the sign change, they were then asked about it explicitly. As discussed above, each of these measures was taken as evidence of visibility, that is, reporting the sign alternation (perceptual criterion), stopping at the sign

(behavioral criterion,) or making a saccade to the sign (eye-movement criterion).

RESULTS

The video records of gaze, superimposed on the image sequence, were used to classify gaze location into 5 categories, as shown in Figure 2. In the critical block where the sign change occurred, gaze was classified as either on the car in front, on the road, on the side of the road to the right where signs might be present, in the region of the intersection, or on the background (buildings, sky, etc). The proportion of the time that gaze is in each of these categories is shown for each trial for both groups. This figure shows the effect of the instructions on the distribution of gaze. When subjects are asked to follow normal traffic rules, they spend much more time looking in the neighborhood of the intersection than when asked only to follow. This is true for both trials for Intersection and Mid-Block groups. Averaging over trials and groups gives a value of 45% of the time in the F+S condition (SEM between subjects = 5%) as opposed to 15 % (SEM between subjects = 3.8%) in the F condition. Similarly, time spent fixating the side of the road went from less than 1% (SEM = 0.5%) to 6% (SEM = 1.6%). All subjects except one showed this pattern, and the between subject variability is quite small. This pattern is unaffected by the condition (Mid-block versus Intersection) and does not change between the first and second trials for the same instruction. This shows that subjects deploy systematically different gaze patterns depending on their overall goal.

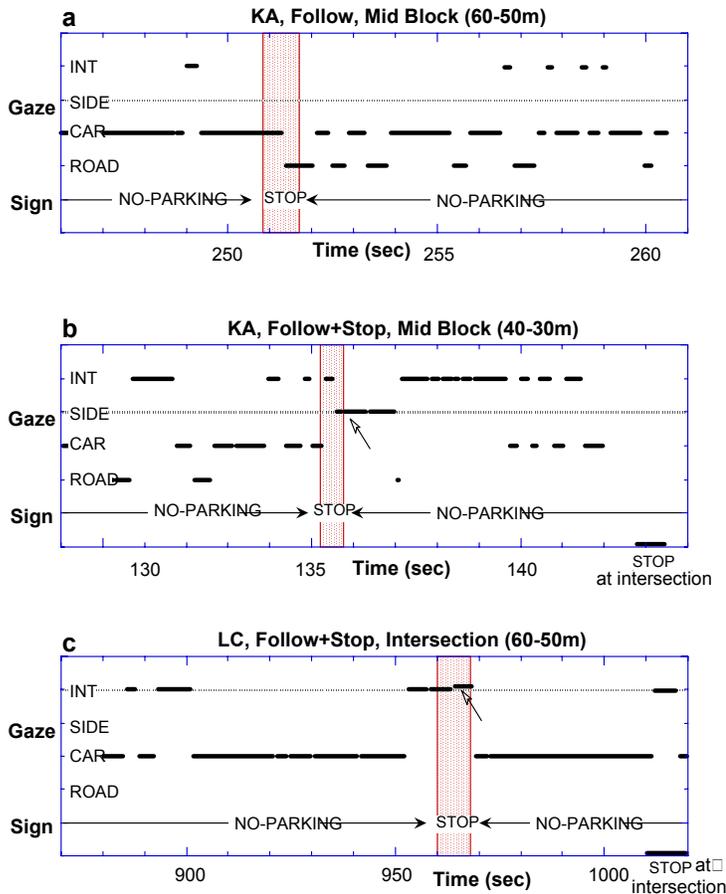


Figure 3. Gaze sequences in the block where the sign alternation occurred. All fixations were categorized according to their locations (road, car, roadside, intersection, or background) using the video record, and shown by thick horizontal lines. Note that each line does not necessarily correspond to a single fixation, but indicates only that gaze fell in the categorized area. Shaded areas show the period when the Stop sign was being presented. Horizontal dotted lines indicate the location of the Stop sign. The thick line on the bottom indicates the period when the subject Stopped. (a) Subject KA, F task, mid-block (60-50m). Gaze was directed onto CAR or ROAD for most of the time and no fixation was made on the sign on the roadside. (b) The same subject for the F+S task, mid-block (40-30m). The subject made saccade to the Stop sign at 135.5 sec. A thick line on the bottom shows the subject Stopped at the intersection. (c) Subject LC, F+S task, intersection (60-50m). The subject saw the Stop sign during the presentation period, and Stopped at the intersection. She did not make re-fixation at the sign (No Parking sign) and did not report seeing the sign alternation.

Figure 3 shows three examples from the video record to show how subjects distribute their gaze over a 15 second time interval. In Figure 3a, the sign was presented mid-block, and the instruction was to follow (F task). In this case, the subject made no fixations on the roadside area in the neighborhood of the sign and did not brake at any point on the block. Gaze was largely confined to the car ahead or to the roadway. In the

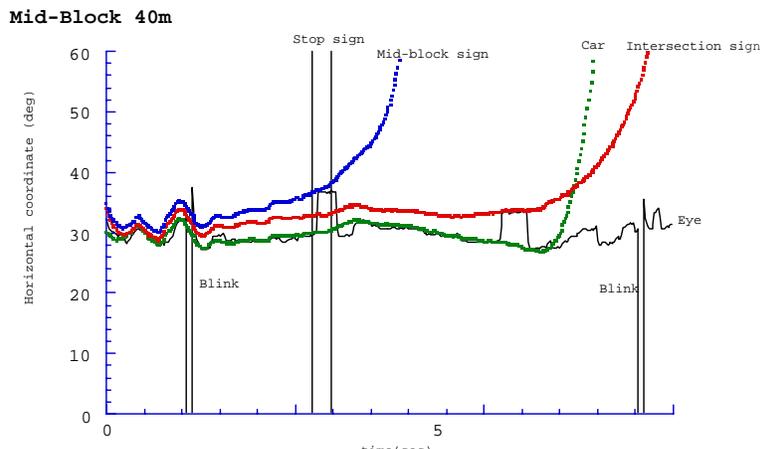


Figure 4. Horizontal gaze location for one subject whose gaze coordinates were recorded in addition to the video record. The figure shows the trajectory of the No Parking sign across the visual field in the mid-block condition. The vertical lines indicate the period when it was a Stop sign. The trajectories of the(permanent) sign at the intersection and the lead car are also shown. Gaze is shown by the solid line.

subsequent questioning, the subject did not report seeing the sign alternation. Thus we categorize this as a failure to detect the sign. In contrast, in Figure 3b, the subject fixated the Stop sign during the presentation period. The sign alternation was also reported subsequently. Thus the sign was detected on this trial. In this condition, the sign was presented mid-block in the F+S task. The influence of the instructions on gaze distribution is evident in these two trials. Figure 3c shows the gaze sequence for another subject in the F+S task. Here the Stop sign was at the intersection (shown by the dotted line). The subject fixated the sign and stopped a few seconds later (indicated by the thick line at the bottom of the figure). At this point the subject's gaze was directed at the Stop line on the road. Interestingly, after seeing the sign, the subject did not re-fixate it even though it had turned back to a No Parking sign. Nor was she aware of the alternation. This suggests that once the decision to stop was made, that region of the visual field was not analyzed further.

Figure 4 shows an example trial in one subject for whom gaze location was available in the data record. These records show the metrical gaze position information in addition to gaze category. (The data from this subject is not included in Figures 2 or 5 as it was obtained after the main body of the experiment. It is included here to give an indication of the relative locations of the relevant stimuli in the visual field as the subject approaches an intersection.) Horizontal gaze is plotted as a function of time, together with the position of

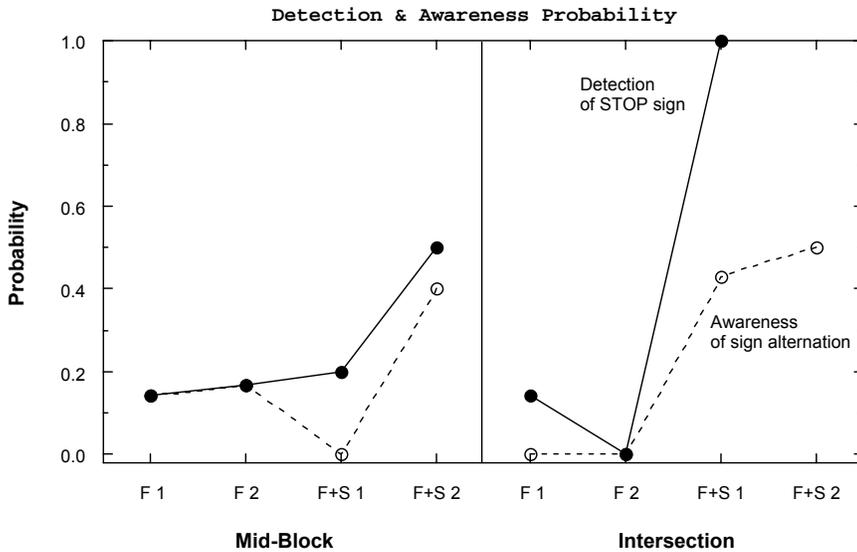


Figure 5. Detection probability for both sign locations (mid-block, intersection), and both tasks (F, F+S). The two trials for the F and F+S conditions, corresponding to the different distances, are also shown. Detection includes either noticing the sign alternation (perceptual criterion), Stopping at the sign (behavioral criterion) or making a saccadic eye movement to the sign (eye-movement criterion). Probability of awareness of the alternation was obtained for the perceptual criterion only.

the lead car and two Stop signs, one mid-block and one at the intersection. The mid-block sign was a No Parking sign except for the period indicated on the figure, where it became a Stop sign. For most of the period the subject's gaze is fixed of the lead car. In this trial the subject fixates the mid-block sign during the period when it is a Stop sign, and then fixates the permanent Stop sign at the intersection several seconds later. In both cases the sign is at a retinal eccentricity of about 10° , and the sign itself subtends about 1.3° . At the end of the trace, gaze is transferred to the road in the region of the intersection, in preparation for stopping. Note that in this instance the subject fixates the real sign at about the same distance as the temporary sign, suggesting that the locations chosen in the experiment (40-50m and 50-60m) are probably reasonable. Analysis of the point at which subjects first fixated close to the permanent stop signs (in the region of the intersection) varied widely across subjects and across trials for the same subject. This suggests that the choice of distances in the experiment is not critical.

Figure 5 shows the probability of detection over all the subjects in the four trials, for both Mid-Block and Intersection conditions. To calculate the probability of detection shown in the figure, we used only those trials up to, and including, the first detection, since detection on one trial may influence the probability of detection on

subsequent trials. This partially removes contamination from possible order effects. Probability of detection is shown both for the perceptual awareness criterion alone, and for the combination of all the measures. The combined measure is dominated by the eye movements. When subjects look at the sign, they are frequently unaware of the sign alternation. Subjects reported the sign alternation without fixating the sign on only two occasions (of 56 possible occasions, and 16 instances when the alternation was reported). This suggests that an attentional shift to the peripheral retina in the region of the sign is almost always accompanied by an overt gaze change. On one of these two occasions, the subject stopped at the intersection. This was the only time a subject stopped without a prior fixation on the sign. Thus fixation is the most sensitive measure of detection in this context. Fixation was usually accompanied by stopping when the sign was at the intersection, but subjects never stopped for the Mid-block sign.

Detection probability is heavily modulated both by task relevance and by the environmental context of the stimulus. The sign was rarely detected in the F task, regardless of its location (0.15 in mid-block, 0.08 at the intersection, averaged over the two trials). In the F+S task, subjects are much more likely to detect the sign. It was detected with probability 0.33 at mid-block (averaged over trials 1 and 2), and in all trials at the intersection (probability 1.0)¹. Because the four trials were done in a fixed sequence, there it is possible that the effect of instructions is contaminated by order effects. However, there is little difference between the first and second trials, except for the Mid-Block, F+S condition, suggesting that order effects are modest. In addition, only the first detection was used, as described above. No order effects between trials one and two were present in the gaze distributions shown in Figure 2.

¹ In order to evaluate the reliability of these differences, it is necessary to assume binomial variability, which ranges between 0.1 and 0.17 for this small number of subjects. On this assumption, the effect of instructions in the Intersection condition, and location in the F+S condition are reliable, but the effect of instructions in the mid-block condition is too small to be statistically reliable. A larger sample is needed to get a reliable estimate of the Mid-Block, F+S probability.

There is also an effect of sign location in the F+S condition. When the Stop sign is in the Mid-Block it will not be detected on about 2/3 of the occasions. This suggests that subjects are more likely to actively search the peripheral visual field for a sign in a context where it has high probability. This is consistent with the greater time spent fixating in the neighborhood of intersections than along the side of the road, shown in Figure 2. The number of fixations does not reveal the frequency of covert search, however, as some searches may not result in an eye movement. The low probability of detection in the F task also suggests that detecting the sign requires an active search, and this is unlikely to be performed when it is not needed. The smaller number of overt fixations on the intersection shown in Figure 2 is also consistent with this. Subjects in this condition were not told to ignore the traffic signs, but this seemed to be the subjects' default assumption. It might be argued that the low detection probability results from the relatively small size of the Stop sign, or relatively low contrast with respect to its surroundings. However, the fact that it was invariably detected when at an intersection, even for a brief period, argues against this. Thus perception appears to depend heavily on active search initiated by the observer, based on learnt probabilities.

DISCUSSION

These results suggest that a top-down system can handle many of the scheduling demands of detecting signs when subjects are attempting to follow normal traffic rules, and the environment is predictable. Subjects reliably fixated the sign at the intersection even though it was presented for a restricted period. Although the sign was detected in mid-block only a third of the time, the likelihood of detection would no doubt be greater if the sign were present continuously. The reduced likelihood of fixating the sign in mid-block shows that observers rely on learnt regularities in the environment to control the scheduling of their searches. Consistent with this, overt fixations in the intersection were much more frequent than fixations by the side of the road (45% versus 6%) when subjects were obeying traffic rules. This reliance on learnt scene regularities is consistent with the results of Chun & Jiang (1998, 1999), that familiar contexts facilitate visual search. It is hard to evaluate the magnitude of the bottom up component, as subjects might detect the sign covertly in the F condition, but give it less behavioral significance, and fail to respond either by fixating it, stopping, or by being consciously aware of it.

However, the greater sensitivity of the gaze measure than perceptual awareness supports the idea that overt gaze location specifies the primary distribution of attention, as discussed above. Perhaps more importantly, subjects spend nearly half of the time deploying gaze in the region of the intersection, when required to drive normally. This suggests that active search is necessary, and that bottom up mechanisms cannot be relied on to attract attention, at least in the case of Stop signs. Stimuli that are larger or higher contrast, such as other cars, would no doubt have greater intrinsic salience.

The current experiment is consistent with the idea that the complex task of driving might be decomposed into sequential application of simple routines for the various sub-tasks, such as route following, obstacle avoidance, speed control, and detection of traffic signs. Salgian & Ballard, (1998) designed an automated vehicle based on this principle that ran on the same complex images viewed by subjects in this experiment. The vehicle had a hierarchical organization, with specialized behaviors for stopping at traffic lights and Stop signs, and car following. These behaviors used sub-set of specialized routines, which in turn used sets of basic operations on the image such as spatial frequency and chromatic analysis. The routine for Stop sign detection involved examining a restricted region of the visual image (on the right) looking for red blobs, and, if one is found, examining the spatial frequency content for a match to Stop sign image. The vehicle was able to analyze complex images in real time and detect traffic signs and follow another vehicle. Observers' behavior in the current experiment can be well described by such a model. In Salgian & Ballard's model, however, the issue of scheduling was not addressed. Their vehicle simply cycled through the behaviors at a fixed rate. It appears, however, that human observers take advantage of prior experience. The finding that subjects are more likely to search in the region of intersections, suggests that the probability of deployment of a particular routine at a given point in the task must depend on learnt schemas. Similar evidence for control of information acquisition by learnt schemas in several everyday tasks is demonstrated by Land and colleagues (Land & Furneaux, 1997; Land & McLeod, 2000). In cricket, for example, observers fixate the ball for about 100 msec as it leaves the batsman's hands, saccade to the anticipated bounce point of the ball, and use the trajectory information to modulate the swing of the bat. Chapman & Underwood (1998) also demonstrated different eye movement patterns in expert and novice drivers. While such stereotyped patterns are

not surprising for skilled behavior, it may be the case that similar principles pertain to general settings.

How does the context of the intersection facilitate detection of the sign? There are three ways that this might occur. One is that drivers actively search for intersections, and then schedule searches for Stop signs at a higher rate when one is found. Drivers in this experiment frequently look to the end of the street upon turning a corner, consistent with this strategy. The greater frequency of overt fixations in the intersection than on the side of the road is also evidence for more frequent searches. Another possibility is that when the observer fixates the intersection, the eccentricity of the sign is reduced and this facilitates detection. On 3 of 14 trials the observer was fixating the No Parking sign at the point when it changed to a Stop sign, so this factor must play some role. A third possibility is that the familiar visual configuration of the intersection facilitates detection or recognition of the sign. There is considerable evidence for this kind of facilitation (eg Biderman et al, 1982; de Graef, 1992), although it is not clear whether semantic content or basic featural properties of a scene guide attention to peripheral locations (Henderson & Hollingworth, 1998; de Graef, 1998). The present data do not allow us to discriminate between these possibilities.

The current experiment is similar to a number of experiments on the phenomenon of “change blindness”, where observers are insensitive to many changes made in scenes either during a saccadic eye movement or some other masking stimulus. (See Simons, 2000 for a comprehensive review.) Perhaps the most similar experiments are those of Simons and Levin (1998), who demonstrated such change blindness when observers watched film clips or were involved in a conversation in the real world. The current experiment provides a unifying principle for understanding what changes are noticed by observers. Sensitivity to changes has been related to areas rated as having ‘central interest’ in pictures of scenes (Rensink et al, 1996; O’Regan et al, 1999). However, attentional state in the change blindness experiments is typically not controlled so it is hard to make a definitive link. The current experiments show that sensitivity is directly related to the particular information that observers extract from a scene, and to the precise time when it is extracted. This is supported by the findings of Wallis & Bulthoff (2000). In a similar driving simulator, they found less sensitivity to changes for drivers than passengers exposed to the identical image sequence. It is also consistent with the findings of Henderson & Hollingworth (1999), who showed that

subjects are more likely to notice a change in a target made during a saccade to the object than during a saccade away from the object. It also explains O’Regan et al.’s (2000) finding that even objects near fixation can undergo a change without the subjects’ awareness. This might occur if the particular information being extracted differed from that which was changed. O’Regan et al offer a similar explanation. Simons (2000) points out that there is evidence to support a variety of different accounts of change blindness: subjects might represent information in the initial view before the change, or in the final view, or represent both but fail to compare them. These hypotheses are not exclusive, however, since each of these possibilities can be seen as a manifestation of the active and highly specialized nature of visual computations. Thus in Henderson & Hollingworth’s (1999) experiment, for example, attention is focussed on the target region before a saccade (Kowler et al, 1995; Gottleib et al, 1998; McPeck et al, 1999), making it likely that a change will be noticed during the subsequent fixation.

The experiments also discriminate between the possibility that the studies of change blindness really test only what is remembered from prior views, not what might be processed while the image is on the retina (Wolfe, 1999). Rapid decay of information in memory must certainly lead us to underestimate the information represented within a fixation. In the present experiment, subjects were not asked to detect changes, an arguably rather unnatural task, but only asked this subsequent to the experiment. By using fixation patterns, the experiment tests more directly what information is accessible while the image is present. Since the Stop sign is relevant to the task, one would expect it to attract attention when it is present in the image, if there is some bottom-up analysis of that information. Instead, Subjects rarely fixate the sign or apply the brakes unless it is at an intersection, and they are explicitly following traffic rules. Similarly, the change from Stop sign to No Parking sign at the intersection was unlikely to be noticed after a subject’s gaze returned to the road, even though that location had recently been attended. Instead, once the decision to Stop had been made, information in that location was no longer relevant to the task. This is consistent with the prediction of O’Regan et al (2000), that changes in parts of a scene already encoded would not be noticed. Thus our results strengthen the claim that representations within a fixation are restricted (Irwin, 1991; Rensink et al, 1997; O’Regan et al, 1999; Mack & Rock, 1998; Simons, 2000), and are not simply consequence of limited

memory. Further examination of other information in the scene is necessary before this can be concluded with any confidence, however.

In summary, the present study suggests that observers manage some of the demands of scheduling behaviors in a normal driving environment on the basis of top down mechanisms. Bottom-up responses appear to be limited in this context. Instead, visibility of traffic signs depends on active search according to an internally generated schedule, and this schedule depends both on the observer's goals and on learnt probabilities about the environment. We propose that vision involves the dynamic application of specialized routines initiated by the observer, whose current goals specify the information that is extracted from a scene, and the time it is extracted. Thus a fuller understanding of the mechanisms of attention, and how attention is distributed in a scene, needs to be situated in the observer's natural behavior.

Acknowledgements

This work was supported by NIH Grants EY 05729 and RR06853. Correspondence and requests for materials should be addressed to H. Shinoda (e-mail: hshinoda@se.ritsumei.ac.jp).

REFERENCES

- Bajcsy, R. (1988) Active perception. *Proceedings of the Institute of Electrical and Electronic Engineers*, **76**, 996-1005.
- Ballard, D.H. (1991) Animate vision: An evolutionary step in computational vision. *J. of the Inst. of Electronics, Information, and Communication Engineers*, **74**, 343-348.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K. & Rao, R. P. N. (1997) Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, **20**, 723-767
- Biderman, I., Mezzanotte, R., & Rabinowitz, J. (1982) Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, **14**, 143-177.
- Brooks, R.A. (1986) A robust layered control system for a mobile robot. *IEEE J. Robotics and Automation*, **2**, 14- 22.
- Chapman, P. & Underwood, G. (1998) Visual search of dynamic scenes: event types and the role of experience in viewing driving situations. in *Eye guidance in reading and scene perception* (ed Underwood, G.) 369-394 Elsevier, Oxford,
- Chun, M & Nakayama, K. On the functional role of implicit visual memory for the adaptive deployment of attention across scenes. *Visual Cognition*, **7**, 65-82 (2000)
- Chun, M. & Yiang, Y. (1998) Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, **36**, 28-71.
- Chun, M. & Yiang, Y. (1999) Top-down attentional guidance based on implicit learning of visual covariation *Psychological Science*, **10**, 360-365.
- Corbetta, M. Frontoparietal cortical networks for directing attention and the eye to visual locations: Identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences*, **95**, 831-838 (1998).
- Culham, J., Brandt, S., Cavanagh, P., Kanwisher, N., Dale, A., & Tootell, R. (1998) Cortical fMRI activation produced by attentive tracking of moving targets. *J. Neurophysiol.*, **80**, 2657-2670.
- De Graef, P. (1992) Scene-context effects and models of real-world perception. In Rayner, K (ed) *Eye movements and visual cognition* 243-259 Springer, Berlin,
- De Graef, P. (1998) Prefixational object perception in scenes: Objects popping out of schemas. in *Eye guidance in reading and scene perception* (ed Underwood, G.) 313-336 Elsevier, Oxford,
- Engel, F. L. (1971) Visual conspicuity, directed attention and retinal locus. *Vision Research*, **11**, 563-576.
- Epelboim, J., Steinman, R., Kowler, E., Edwards, M., Pizlo, Z., Erkelens, C., & Collewyn, H. (1995) The function of visual search and memory in sequential looking tasks. *Vision Res.* **35**, 3401-22.
- Findlay, J. & Gilchrist, I. 1998) Eye guidance and visual search. in *Eye guidance in reading and scene perception* (ed Underwood, G.) 295-312 Elsevier, Oxford,
- Folk, C., Remington, R., & Johnston, J. (1992) Involuntary covert orienting is contingent on attentional control settings. *J. Exp. Psychol. HP&P*, **18**, 1030-1044.
- Gibson, B. & Jiang, Y. (1998) Surprise! An unexpected color singleton does not capture attention in visual search. *Psych. Science*, **9**, 176-182.

- Gibson, B. & Kelsey, E.. (1998) Stimulus-driven attentional capture is contingent on display-wide visual features. *J. Exp. Psych: HPP*, **22**, 1496-1504.
- Gottlieb, J., Kusunoki, M. and Goldberg, M.E.: (1998) The representation of visual salience in monkey posterior parietal cortex. *Nature* **391**: 481-484.
- Hayhoe, M. M. (2000) Vision using routines: a functional account of vision. *Visual Cognition*, **7**, 43-64.
- Hayhoe, M., Land, M., & Shrivastava, A. (1999) Coordination of eye and hand movements in a normal environment. *Invest. Ophthalmol & Vis. Sci.* **40**, S380.
- Henderson, J. M. (1992) Visual attention and eye movement control during reading and picture viewing. in *Eye movements and visual cognition* (ed Rayner, K.) 261-283 Springer, Berlin,
- Henderson, J. & Hollingworth, A (1998) Eye Movements during scene viewing: an overview. in *Eye guidance in reading and scene perception* (ed Underwood, G.) 295-312 Elsevier, Oxford,
- Henderson, J. & Hollingworth, A. (1999) The role of fixation position in detecting scene changes across saccades. *Psychological Science*, **10**, 438-443.
- Irwin. D. E. (1991) Information integration across saccadic eye movements. *Cognitive Psychology*, **23**, 420-456.
- Itti, L. & Koch, C. (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, **40**, 1489-1506
- Kowler, E., Anderson, E., Doshier, B. & Blaser, E. The role of attention in the programming of saccades. *Vision Research*, **35**(13), 1897-1916 (1995)
- Land MF, & Lee DN (1994) Where we look when we steer. *Nature* (Lond.) **369**: 742-744
- Land, M. & Furneaux, S. (1997) The knowledge base of the oculomotor system. *Phil. Trans. R. Lond. Soc. B*, **352**, 1231-1239.
- Land MF, Mennie N, & Rusted J (1999) Eye movements and the roles of vision in activities of daily living: making a cup of tea. *Perception* **28**, 1311-1328
- Land, M.F. & McCleod, P. (2000) The eye movement strategies of batsmen in cricket. *Invest. Ophthalmol. & Vis Sci.*, (Suppl.) **40**, S815 (abstract).
- Levin, D. & Simons, D. (1997) Failure to detect changes in attended objects in motion pictures. *Psychonomic Bulletin & Review*, **4**, 501-506.
- Mack, A. & Rock, I. (1998) *Inattention blindness*. (MIT Press, Cambridge, MA,).
- McPeck, R., Maljkovic, V. & Nakayama, K. (1999) Saccades require focal attention and are facilitated by a short-term memory system. *Vision Research*, **39**, 1555-1565.
- Motter, B., & Belky, E. (1998). The guidance of eye movements during active visual search. *Vision Research*, **38**, 1805-1815.
- Neisser, U. (1967) *Cognitive Psychology*. New York, Appleton, Century, Crofts.
- O'Regan, J. K., Deubel, H., Clark, J. & Rensink, R. A. (2000) Picture changes during blinks: looking without seeing and seeing without looking. *Visual Cognition*, **7**, 191-211.
- O'Regan, J. K., Rensink, R. A. & Clark, J. J. (1999) Change-blindness as a result of 'mudsplashes'. *Nature*, **398**, 34.
- Rensink, R. A., O'Regan, J. K. & Clark, J. J. (1997) To see or not to see: the need for attention to perceive changes in scenes. *J. Psychol. Sc.*, **8**, 368-373.
- Saarinen, J. & Julesz, B. (1991) The speed of attentional shifts in the visual field. *Proceedings of the National Academy of Sciences*, **88**, 1812-1814.
- Salgjan, G. & Ballard, D. (1998) Visual routines for autonomous driving. *Proceedings of the 6th ICCV*, 876-882.
- D. Simons (1996) In sight, out of mind: When object representations fail, *Psychological Science*, **7**, 301-305.
- Simons, D. J. (2000) Change blindness and visual memory. *A special issues of Visual Cognition*. (Psychology Press, Hove, UK,).
- Simons, D., & Levin, D. (1997) Change Blindness. *Trends in Cognitive Science*. **1**, 261-267.
- Summala, H., Nieminen, T. & Punto, M. (1996) Maintaining lane position with peripheral vision during in-vehicle tasks. *Human Factors*, **38**(3), 442-451
- Treisman, A. (1993) The perception of features and objects. in Baddeley, A. & Weiskrantz, L (eds.) *Attention: selection, awareness, and control. A tribute to Donald Broadbent*. 5-35 Clarendon Press, Oxford, UK,

Turatto, M. & Galfano, G. (2000) Color, form, and luminance capture attention in visual search. *Vision Res.* 40, 1639-1644.

Ullman, S. (1984) Visual Routines. *Cognition*, **18**, 97-157.

Wallis, G. & Bulthoff, H. (2000) What's scene and not seen: Influences of movement and task upon what we see. *Visual Cognition*, **7**, 175-190.

Wolfe, J. (1994) Guided search 2.0. A revised model of visual search. *Psych Bulletin & Review*, **1**, 202-238.

Wolfe, J. M. Inattentional amnesia. in *Fleeting Memories* (ed Coltheart, V.) 71-94 (MIT Press, Cambridge, MA, 1999).

Yantis, S. (1998) Control of visual attention. in *Attention* (ed Pashler, H.) 13-74 Psychology Press, Hove, UK.