CHAPTER 18

How Much Do Novice Drivers See? The Effects of Demand on Visual Search Strategies in Novice and Experienced Drivers

David E. Crundall, Geoffrey Underwood and Peter R. Chapman
University of Nottingham

Abstract

Varying levels of visual and cognitive demand produce different visual search strategies. These effects differentiate between drivers on the basis of experience. Previous studies are reviewed with the aim of identifying a process which may account for the effects of changes in demand according to driver experience. One possible theory is that of perceptual narrowing, which suggests that the usable field of view shrinks with an increase in demands at the point of fixation. This theory is discussed in relation to novice and experienced drivers and a new methodology is put forward to test for such differences.
Introduction

This chapter discusses the effects of cognitive and visual demand upon visual search strategies during driving, and whether they can be used to differentiate between novices who have recently passed their test and more experienced drivers. If novices are placed under higher demands than experienced drivers due to the novelty of stimuli, lack of automated sub-routines, or the absence of relevant schemas, then this would compound any increases in the general level of demand. Assuming a limited capacity of attention, it may occur that novices' attentional limitations (and the resultant consequences such as a link with increased accident liability), only become apparent when the task becomes particularly demanding. If this is the case, then what process could underlie this effect, and what could be done, short of giving novice drivers the experience they need, to redress the matter?

The effects of increased demands on visual search strategies in drivers

Before addressing the question of what effect increased demand has on a driving task, it may be useful to draw analogies from areas of research more accessible to experimentation. In reading for example it is a widely reported effect that unfamiliar words require longer fixation durations than common words (Rayner and Polletsek, 1989). The explanation for this is that an unfamiliar word requires more processing, and within certain limitations, the measure of fixation duration is considered to reflect object identification time according to the eye-mind assumption (Henderson, Polletsek and Rayner 1987, Underwood and Everett, 1992). A similar effect can be noted in laboratory based driving tasks. Underwood, Crundall and Chapman (1997) asked novice and experienced drivers to watch a series of video clips taken from the drivers perspective and to respond to potential hazards by pressing a button. This hazard perception paradigm is discussed in more detail elsewhere in this book (see Chapter 17). Regardless of driving experience, it was found that fixation durations on the cause of any potential hazard (such as a car suddenly emerging from a side road) were greatly in excess of mean fixation durations for the rest of the scene. In essence, the emergent car in this example takes the role of the unfamiliar word in reading studies: it requires more processing than usual, and thus attracts increased fixation durations.

The transition from studying how people read to how people view moving scenes is not an easy one. Sentences have an agreed grammar which allows parsing. The hazard perception clips used in the above study have a structure which is just as complex as language though there is no current agreement on how to parse them meaningfully. We do not know what are the crucial elements which divide clips. With this in mind consider the problems involved with understanding vision during
real driving. How should we define demand and what constitutes an increase in the level of it.

A fundamental problem in defining demand is the confounding of increases in visual demand, such as an increase in visual clutter or complexity, and increases in cognitive demand, such as an increase in the processing demands of a particular stimulus perhaps due to an increase in its relevance to a current context (Williams, 1982). Visual demands are closely related with task demands, for as the dominant sensory modality in driving is vision, an increase in task difficulty will usually coincide with an increase in the complexity of the visual scene. In an ideal situation one could provide subjects with the same visual stimuli in two conditions, yet the information is only of relevance in one of the conditions. This would hold the visual demands constant, while varying the cognitive demands which would be reflected in the processing times of the stimuli. Alternatively adding visual clutter while holding the meaningfulness of the display constant would provide the opposite manipulation. Unfortunately, as stimuli become more realistic, the two become harder to separate as we shall see later. We will argue in the following section that though the two types of demand (visual/task demands and cognitive demands) are rarely fully separated, the majority of driving research has focused on task demands which are predominantly related to visual complexity, rather than cognitive demands.

**Previous manipulations of demand in driving research**

One problem in attempting to manipulate the level of demand in an experiment is the identification of a suitable independent variable. There is a lack of consistency in the relevant literature in the adoption of a demand manipulation, with the result that it is very hard to compare across studies. The one common feature that the majority of these studies share however is that their demand manipulation is concerned more with the task demands of factors such as road geometry or traffic density (both of which increase visual complexity), rather than assessing the cognitive demands placed on the subject. Despite the disparity between the factors chosen to represent demand on the road, the following discussion highlights the consistent results that an increase in the complexity of the driving task (and thus an increase in the visual complexity) tends to increase one’s active search of the scene, producing a wider spread of search and an increased sampling rate. This may initially seem to weaken our analogy with reading where infrequently used words tend to capture attention for longer than normal, but let us look at the results of these studies before drawing any conclusions.

The use of road geometry as a demand manipulation has focused mainly on how visual search strategies differ between driving along straight roads or when driving through curves. Shinar, McDowell and Rockwell (1977) were the first to note that
the increased processing demands associated with the negotiation of a curve were related to a more active visual search pattern, as compared to observations on a straight road. The increase in demand occurs due to a shift in the loci of important, visual information sources. Fry (1968) suggested that the focus of expansion is the most important point of information for driving as it maximises preview time for objects directly in the path of travel. Evidence confirms that experienced drivers tend to fixate close to the focus of expansion, while information concerning lane maintenance is obtained through peripheral vision from near the car (Land and Horwood, 1995). However, when driving through a curve the focus of expansion becomes less important for direction as the car's immediate heading is offset from the expansion point. Lane maintenance also becomes more difficult: a curve is rarely of constant arc and this necessitates constant monitoring of one's position in relation to the edge of the curve. The increased importance of road markings for lane maintenance, and the corresponding decrease in the importance of the expansion point create a more dynamic visual search pattern. Shinar et al. (1977) found the subjects tended to switch rapidly between fixating the road ahead for long-term directional information, and fixating the road edge or lane markings in order to stay within their lane. To accommodate the increased number of fixations on the road markers, fixation durations decrease during curve negotiation. Shinar et al. suggested that the visual processing of a high speed curve during driving suggests that the subjects were collapsing a two process system (directional information from foveating the focus of expansion, and lane maintenance information through peripheral vision) into one, where the fovea is attention switching between the two sources of information. Zwahlen (1993) also found curve negotiation to involve a more active search strategy than on straights. He noted that fixation durations were markedly shorter in the curve, and equated this finding with the American Automobile Association's "brief glance technique" where drivers are advised to keep fixations short in order to avoid "captured attention".

As mentioned earlier, within the laboratory the appearance of an unfamiliar word, or a potential hazard on a video clip of driving, tends to concentrate one's attention. Conversely, the increased processing demand involved in curve negotiation decreases fixation durations on the real road. A further problem is that not all the evidence points toward a more active search strategy on curves than straights. Liu, Veltri and Pentland (1997) have provided contradictory evidence. They utilised a first order Markov matrix to analyse drivers scan paths on straights and curves and discovered two distinct scan patterns on straights: a 'preview' search (from the middle preview distance down the road to either the near or far preview distances with equal probability, and then back to the middle), and a 'side-to-side' search (from the next road segment which is very far away, to either side with equal probability, and then back to the middle). During curve negotiation, however, though they identified the side-to-side pattern, they failed to find the preview
pattern. In this study, the search strategy became less dynamic under the increased demand of the curve, though it could still be feasible that the number of fixations increases (while the durations decrease) within the smaller confines of a reduced search space.

Another measure of cognitive demand that has been used is traffic density. As traffic increases, so does the danger of any driving situation up to the point of traffic congestion. Rahimi, Briggs and Thom (1990) looked at eye and head movements of a driver at two American intersections, one busy and one quiet. The subject performed 20 left turns (crossing the line of traffic) at each junction alternately, while head and eye movements were recorded via video. They found that the busy intersection produced more fixations than the quiet junction, which suggests a corresponding reduction in fixation durations as demand increased. There is also evidence that the proximity of other vehicles may affect the visual search patterns of drivers. The work of Hella, Laya and Neboit (1996) suggests that the closer one is to the car in front the shorter one’s fixation durations upon it become, though there is a corresponding increase in the total number of fixations upon it. They discovered this by comparing the eye movements of drivers on a three lane motorway. Interestingly, they did not discover any visual search changes due to the speed of the car (which was dictated in part by the lane they were in at the time). The decreased fixation durations support the suggestion that as drivers find the task demands and visual complexity increasing they respond by increasing the sampling rate of the scene.

Miura (1979) used four separate levels of task demand to investigate fixation durations. These were stable running, passing parked vehicles, entering into a narrower route, and overtaking. He found that entering the narrower route and the act of overtaking significantly reduced mean fixation durations. This mirrors the results of studies of curves and intersections.

Despite the lack of consistency in the manipulations of demand, it seems fairly well documented that general increases in task demands and visual complexity tend to reduce mean fixation durations and increase the sampling rate. Though this reveals the limits of our reading analogy it does not mean that the effects of increased fixation durations upon hazards viewed in a hazard perception test (Underwood, Crundall and Chapman, 1997) are not generalisable to the real road. The general increase in the task and visual demands of driving through a curve may necessitate an increase in the sampling rate, though the demand of a particular stimulus with high priority and a defined locus (i.e., a hazard) may still capture attention. This paradox has been noted in a study by Chapman and Underwood (this volume). They discovered that though subjects tended to increase fixation durations upon hazards relative to the overall scene, the mean fixation durations for each whole clip varied according to the visual complexity of the roads viewed. Cluttered urban and suburban roads produced shorter fixation durations than the empty, rural road clips. This difference between the two effects is probably due to the localisation
of the hazard stimuli compared to the diffused nature of increased visual complexity. It is the difference between a solitary car suddenly pulling out from a side road on to an otherwise empty road, and a busy road where any one of a number of vehicles is a potential hazard. The former captures attention whereas the latter requires an increased search in order to monitor all potential threats. The former is more akin to cognitive demands in that the saliency of a particular stimulus has increased and requires further processing, indicated by increased fixation durations, while the latter reflects an increase in visual complexity.

Are novices more susceptible to high demands than experienced drivers?

The hypothesised differences between novices and experienced drivers due to increases in cognitive demand concern the processing times involved in identifying particular stimuli such as hazards, and other measures of perceptual input such as the spread of search and sampling rate when the task is held constant, which relate to attentional capacity. We shall discuss these capacity related problems first, before addressing the issues of visual and task demands.

Novice drivers are likely to encounter capacity problems with attention more often and more severely than experienced drivers. Recently licensed drivers will have no doubt gained experience on actual roads though there will still remain much which will be novel. Faced with new stimuli a novice driver may take longer to process it in the same way that an infrequent word will tax a novice reader more than an experienced reader. In addition, depending of the amount of practice they have received, they still may have to automise certain sub-routines of the driving task. One such task which is widely believed to be automatic is that of changing gear. Novices have been noted as being slower gear changers than more experienced drivers (Duncan, Williams and Brown, 1991), which suggests a failure to completely automatisate the task. One of the benefits of automatising this is that the task will no longer need attention. The experienced driver can then allocate all attention to other matters, while the novice drivers may still have to apportion some to gear changing. This should not be a problem when cognitive demands on the driver are low, but as demand increases the novices may suffer a degradation of either the gear changing or the other tasks which are competing for attention.

The most cited studies of novice and experienced drivers are those of Mourant and Rockwell (1970, 1972). They found that novices have an increased frequency of pursuit tracking eye movements. These are fixations where either the stimulus or the viewer is moving. In order to maintain the stimulus in the same place on the retina, the eye must move to accommodate for other movements in the scene. A traffic sign may first be fixated at the focus of expansion but as one gets nearer to it, the sign will be displaced upwards and to the left of the visual field (on British roads). In order to
remain fixated on it for any length of time one must move the eyes with the optic flow rate of the sign. The increase in frequency of pursuit fixations found by Mourant and Rockwell (1972), suggests that the novices overall fixation durations were greater than those of the experienced drivers who did not linger on objects outside the focus of expansion long enough to significantly increase their amount of pursuit tracking. This higher level of fixation duration may reflect increased processing times, and attentional capacity limitations. They also discovered that novices tended to look in their mirrors less and at lane markings more than experienced drivers, and that they searched an area of the road ahead which was closer to the car. It was further noted that the spread of search along the horizontal axis was more compact than that produced by the experienced drivers. Similarly Renge (1980) identified a tendency for novices to search predominantly in the vertical plane. He asked subjects to verbalise what they were looking at while driving and noted the pattern of verbalisation was consistent with a vertically based search strategy. The reduction in horizontal scanning and high level of lane marker fixations has also been recorded after alcohol consumption (Mortimar and Jorgeson, 1972) and for drivers suffering from fatigue (Kaluger and Smith, 1970). Both alcohol and fatigue are considered to decrease attentional resources, which provides further support that the novice drivers may be suffering from a competition for resources.

The majority of these findings can be explained in terms of the attentional allocation problems which novices may suffer from (Underwood, Crundall and Chapman, 1997). The smaller search area of novices may reflect an attempt to reduce perceptual input, as may the reduction in mirror checks. As previously mentioned, increased fixation durations suggest increased processing time which in turn may cause the driver to reduce the size of the visual search pattern in an attempt to avoid overloading. Underwood et al. (1997) also noted that novices tend to have longer fixation durations on hazards than experienced drivers. With all these studies the tasks were held constant between subjects (as much as is possible with on-road studies), and therefore can be considered more related to the differing levels of cognitive demand placed on novice and experienced drivers, than due to changes in visual complexity and task demand. The next section considers differences between novice and experienced drivers due to changes in visual demands.

An investigation into the effects of experience on different roadways

A further problem for novice drivers which is not directly related to attentional capacity problems, is the possible lack of schemas for certain road situations, or the use of inappropriate schemas. Earlier in this chapter, examples of task demand studies were reported, which demonstrated differences in drivers’ responses to differing road geometries, levels of traffic density, and types of manoeuvre. In an
attempt to identify hypothesised differences between novice and experienced drivers’ search strategies due to task and visual differences, Underwood et al. (1997) analysed the search strategies of 32 subjects while driving along roads of varying complexity in a real traffic situation. Half of these subjects had passed their driving test within three months of the study while the other half had at least five years' experience.

From a 20-minute drive three one-minute windows were selected to reflect differing levels of demand based upon the type of road they were on. Each window started from the same geographic spot. The roads used were a single lane, rural road with good visibility ahead, a single lane suburban road through a small village (which included shops, parked cars and pedestrians), and a dual carriageway which was joined by two slip roads, one from the left and one from the right. The main measures that were analysed from the windows were fixation duration, and variance of fixation locations along the vertical and horizontal axes. The latter was to provide information concerning the spread of search while the former was intended to gauge the sampling rate of the subject.

Fixation durations produced a significant interaction with novices producing their longest fixation durations on the dual carriageway while experienced drivers tended to display longer fixations on the rural road (see Fig. 1). It was only upon the suburban route that both groups of subjects agreed with the use of short fixations. If one accepts the experienced drivers’ strategy as the correct one, then their fixation durations do not increase with visual demand, for the rural road was considered to be the least demanding of the three road types. It had less traffic, less visual clutter, and only one lane. The speed restriction was 60 mph, though the evidence for the effect of speed on fixation durations is inconsistent (McDowell and Rockwell, 1978; Cohen, 1981) and there are some studies which suggest a very limited effect of speed on visual demand (Miura, 1985; Hella, Laya and Neboit, 1996). It seemed that the experienced drivers decreased their fixation durations on the busier routes, with the cluttered suburban route through a village producing the shortest durations (and therefore the greatest sampling rate). The number of fixations varied accordingly with the durations, which supported the theory of the sampling rate increasing as the complexity of the scene increased. The novices however produced the longest mean fixation durations on the dual carriageway. If one accepts that the increased traffic and danger of the additional lane makes the dual carriageway more demanding than the rural road, then the novices have responded inappropriately by reducing their sampling rate. The dual carriageway was most often the route where the subject had traffic ahead in the same lane. Mourant and Rockwell (1970) reported that novice drivers’ fixation durations tend to increase when following another car, as opposed to Hella et al. (1996) who found that experienced drivers fixation durations decreased but the total time spent on the car ahead increased. This suggests that the car in front is as important to experienced drivers as it is to novices, though the experienced
Fig. 1. Three graphs which show how the eye movements of novice (N) and experienced (E) drivers vary across the road types. The measures taken are (a) mean fixation durations, (b) variance of fixation locations in the horizontal plane, and (c) variance of fixation locations in the vertical plane.
drivers still manage to increase the sampling rate of other areas of the scene. One explanation could be that the novices suffered attentional capture by vehicles in front in the same way that subjects tend to fixate the hazards in hazard perception clips for longer (Underwood, Crundall and Chapman, 1997; and Chapter 17). The more experienced drivers however overcame this and maintained a high sampling rate. A further explanation stems from Mourant and Rockwell's (1972) finding that novices fixate lane markers more often when driving on a freeway. Such fixations made up 70% of the pursuit tracking that they noted in novices. As mentioned above, pursuit tracking fixations imply long fixation durations, which could account for the effect.

A subsequent category analysis performed on our data produced evidence to suggest that novice and experienced drivers spend similar amounts of time looking at certain items within the visual field. This analysis was conducted on a subset of the original sample (five novices and five experienced drivers with the most accurately calibrated data) and compared the total time spent fixating the car in front and lane markings. No significant differences were found between the subject groups in regard to total time on lane markers, or on a followed vehicle on the dual carriageway. This finding reconciles the results of Mourant and Rockwell (1972) and Heila et al. (1996) in that total time dedicated to fixating certain stimuli such as the car in front tends to be the same across experience though the experienced drivers may still have a higher sampling rate of the scene.

Comparison of the variance of fixation locations in the horizontal and vertical meridians also highlighted the dual carriageway as a main difference between the experienced and novice drivers' search strategies. Mean comparisons of the horizontal search interaction revealed that the experienced drivers increased their scanning in this meridian when on the dual carriageway. The other roads produced narrow, less dynamic search strategies. The novice measures of horizontal scanning were all similar to the experienced drivers' measures for the rural and suburban road. Analysis of fixation variance in the vertical axis did not produce a significant interaction though a main effect of road type was discovered. However, means comparisons of the levels of roadway found the spread of search for experienced drivers on the dual carriageway to be significantly different to the suburban and rural roads. Despite the lack of interaction there is a suggestion that experienced drivers increase their vertical search on the dual carriageway just as they increased their horizontal search. Whereas the novices tended to maintain a restricted horizontal search comparable to the experienced drivers' search on the rural and suburban roads, their vertical search is closer to the expanded scanning of the experienced drivers on the dual carriageway (see Fig. 1).

There are several points which should be drawn from this research. First, one should note that the differing levels of demand that each roadway places upon the driver do produce changes in the relevant search strategies. The experienced drivers
Demand and visual search

behave according to the prior published results. Their fixation durations decrease on more demanding roads, and their search strategy widens (Mourant and Rockwell, 1970; McDowell and Rockwell, 1978; Shinar et al., 1978; Rahimi et al., 1990; Zwahlen, 1993). The difference between the suburban route and the dual carriageway is of interest, the latter producing the greater scanning and the former producing the lower fixation durations. It may be that both roads are considerably demanding, but the responses to such demands are different. A second point to note is the lack of flexibility of novices' scanning strategies across the road types. The experienced drivers increased their scanning behaviour in both meridians according to the road type, while the novices maintained one level of scanning throughout. In their analysis of curve negotiation Shinar et al. (1978) reported that high levels of field dependence correspond with inflexible, narrow search strategies that are insensitive to increases in demand. A third important finding was the high level of vertical scanning and the low level of horizontal scanning produced by the novices. This fits with previous research which suggests that novice drivers require the experience which will sensitise them to the horizontal axis as the main source of information.

The result that is hardest to explain is the short fixations on the suburban route, yet a failure in both groups of subjects to increase the search space. The dual carriageway received both short fixations and increased scanning. Perhaps, as mentioned above, the different demands of the roadways require different responses. Unfortunately the myriad of visual and task oriented factors which correspond to a particular roadway prevent anything but a coarse grain view of the visual demands.

An alternative approach however is take subjective measurements of the levels of demand placed upon subjects. In a further follow up study 18 novices and 18 experienced drivers were asked to rate a set of video clips, taken from the roads where the eye movement data was recorded. The ratings were made on nine, seven-point Likert scales (e.g., How much risk would you have felt during that drive? How stressful would it be to be the driver in that drive? How hard would you need to concentrate to drive safely during that drive?) which loaded onto two constructs: Danger (How dangerous is this route at this particular time?) and Difficulty (How difficult would you find the route to drive at this particular time?). These two constructs were initially identified by Groeger and Chapman (1996) as useful in distinguishing between driver groups.

Analysis of the results revealed that novices rated all the road types as both more dangerous and difficult than the experienced drivers. On the danger ratings, both groups of subjects rated the dual carriageway as dangerous as the suburban route, while the rural route was considered relatively safe. With the difficulty ratings however, the suburban road was given the highest score. These subjective scores can be related to the eye movement data. One could postulate that a need to increase the search area and sampling rate of a scene may be related more to the danger of the situation than the difficulty of the drive. An increase in the number of potentially
hazardous stimuli would require a more active search strategy in order to monitor all the possible sources of danger. In support of this Beck and Emery (1985) believe that anxiety, or the unpredictability of events, produces in people a state of hypervigilence where search strategies become more active, and more of the environment is inspected in an attempt to locate any potentially dangerous stimuli. In times of difficulty however, evidence supports a concentration of attention in a few locations such as longer fixation durations on an unfamiliar word during reading. On the dual carriageway the experienced drivers had low fixation durations and a wide search pattern, perhaps due to the high level of danger. On the suburban route fixation durations were low also, but the search strategy was narrower. This may be due to difficulty of the suburban road. Though as dangerous as the dual carriageway, the suburb was considered more difficult and as such drivers may have constricted their search while maintaining short fixation durations due to the element of danger.

One cannot infer that perceptions of danger or difficulty lead to particular search strategies, especially as the ratings were made by subjects who viewed the roads on video. If one assumes that novice and experienced drivers would use the same search pattern while driving on a road or watching it on video, one could validly suggest that the search strategies themselves may lead to particular perceptual ratings. Differences in search strategies across these roads may however be linked to the description of the road according to the two constructs. Though the two groups of drivers do not differ in their ordinal ratings of the roads, they may differ in how they react to that information. If danger and difficulty impose different demands upon the driver then one may predict this to differentiate between novices and experienced drivers. For example, novice drivers may have developed strategies or schemas for coping with demands linked to difficulty rather than danger. The possible relationship between these two constructs and the nature of the demands they place on the driver in regard to visual search strategies may provide an interesting avenue of research for the future.

What is the underlying process whereby demand modulates visual search?

The evidence presented so far argues that the level of both cognitive and visual demand upon a driver will either constrain and direct visual search, or actively expand it. Whereas the effect of the visual and task demands on drivers are triggered externally, the cognitive demands depend on internally motivated factors in an individual’s ability to process stimuli in an efficient manner. The effect of varying roadways on novices could be plausibly explained in terms of inappropriate schemas, though as yet there has been no suggestion as to the nature of the process which underlies the effects of cognitive demand upon the search strategies of novice and experienced drivers. In search of a possible contender to explain the effects of both
forms of demand let us turn away from foveal vision and instead explore the periphery.

The term *usable* or *functional field of view* is used to describe the area of the visual field within which stimuli can be detected, and possibly processed to some extent. Engle (1971, 1974) proposed that though ultimately limited by the physiological boundaries of visual acuity, the area of peripheral vision available to analysis changes in size and shape according to circumstances. One such circumstance is cognitive load at the fovea. In a similar manner to the zoom lens model of attention (Erikson and Yeh, 1985; Erikson and St. James, 1986) the theory suggests that a high level of cognitive demand at the fovea should reduce the area of the usable field of view, as this allows limited resources to be concentrated upon the foveal region, increasing the resolving power. The reduction of the functional field of view due to increases in cognitive load is termed perceptual narrowing. If stimuli in the peripheral field are left outside the functional field of view as the attentional tide retreats, then preview benefits will be lost. Holmes, Cohen, Haith and Morrison (1977) discussed two models of perceptual narrowing: *general interference* and *tunnel vision*. The former is merely a general degrading of peripheral detection rates as cognitive load at the fovea increases, while the latter predicts an interaction with the eccentricity of the peripheral target. Tunnel vision suggests that as cognitive load increases, peripheral detection rates will suffer more at greater eccentricities. Evidence has been found for both models (Williams, 1982; Williams, 1985). Williams (1988) concluded that either model can be induced depending on the antecedent conditions. In order to invoke tunnel vision he suggested that three things were necessary: a high foveal load, an attentional strategy overtly focused on the central task, and speed stress. Regardless of which particular model one supports, the main prediction of this theory is that as a foveal load becomes more cognitively demanding, so less attention is given to peripheral items.

Such an effect has been discovered in a number of areas of vision research. In reading it has been noted that fixation durations on words can be lengthened by placing increasingly unfamiliar words before the target. The suggestion is that the unfamiliar word reduces perceptual span and removes preview benefits for the subsequently fixated words (Rayner, 1986; Henderson and Ferreira, 1990). Identification of objects is also susceptible to perceptual narrowing. Reynolds (1993) found that errors identifying a peripheral target 4° from the fovea increased when a complex picture was displayed at the point of fixation rather than when a letter or geometric shape was presented instead. Williams' (1982) complaint applies to this study however. He noted "those few studies that have examined dual task performance within a single glance have intentionally manipulated the visual complexity of the foveal task or have confounded the visual and cognitive aspects of the foveal task" (p. 684).
In a recent study the authors attempted to circumvent this problem by using the same stimuli in both the high and low demand conditions. The study was designed to test the hypothesis that peripheral preview benefits could be removed by increasing the cognitive demands of a central stimulus while holding visual demands constant. Thirty subjects were given two blocks of trials on a visual discrimination task. Each of the presentations consisted of a central, red-bordered, triangular warning sign (which subtended 1°) with either a vowel or a consonant in it, and a peripheral red-bordered triangle (4.6° to either the right or left of centre) with either a staggered junction sign, or a right bend junction. The subject’s task was to distinguish between the peripheral targets after making a saccade to it from the central sign. To ensure each subject was looking at the centre at the start of each presentation the trials were only presented when the computer was satisfied that the subjects were focused on a central cross. Two counterbalanced blocks were given to subjects consisting of 24 presentations, with the only difference between the two blocks being the instructions that subjects received. In the high demand block subjects were told to respond to the peripheral target only if the central letter was a vowel. In the low demand block subjects were told to ignore the central letter, and to decide on the peripheral target as soon as possible.

The first measure that was analysed was saccade latency. This is the time taken to disengage from the central stimulus and to saccade to the peripheral target. In the low demand task the central stimulus did not hold any relevant information, though in the high demand task the same central stimulus had to be processed before saccading to the target. The increase in cognitive complexity between the tasks was reflected in a significant main effect of cognitive demand with saccade latencies for the high demand task greater by 327 ms on average ($F(1,28) = 165.9, p < 0.01$).

Comparison of subjects' first fixation durations on the peripheral target produced a main effect of task demand ($F(1,28) = 31.9, p < 0.01$), with the high demand task attracting nearly 100 ms of attention more than the low demand task on average. This difference increases to an average of 180 ms when re-fixations are included in a measure of total gaze duration on target ($F(1,28) = 18.0, p < 0.01$). These differences can be viewed in Fig. 2. Other measures included analysis of saccade inaccuracy (distance from the target after the first saccade) and the duration of any fixations which fell short of the target, but these failed to reveal any significant differences. Reaction times for the discrimination task also showed a main effect of demand ($F(1,28) = 90.61, p < 0.01$), with targets in the high demand condition taking an extra 664 ms to respond to on average, more than double the average increase in saccade latencies, suggesting that the differences noted in the first fixation durations and gaze durations on target do actually represent processing differences that influence the response. Error rates tended to vary between 2 and 4%.
These results are consistent with the theory of perceptual narrowing. The peripheral preview that was afforded subjects in the low demand condition was removed when they had to process the central stimulus producing nearly 100 ms benefit in the duration of first fixations at an eccentricity of 4.6°.

**Do novice drivers see less of the world?**

The evidence suggests that a reduction in the peripheral field may well occur due to an increase in the cognitive demands of a foveal stimulus, but what evidence is there to suggest that novice drivers may be more prone to perceptual narrowing than experienced drivers?

Unfortunately no research addresses this question directly. Two sub-questions can be answered however. First, one should ask whether experience in any task can influence one's usable field of view, and secondly, whether perceptual narrowing occurs at all in driving. If such narrowing does occur in drivers, and experience has been shown to be a factor in other task domains, then it is a short step to predict that driving experience may influence visual search through a cognitive demand-based reduction of the usable field of view.

With regard to the effects of experience, Holmes et al. (1977) suggested that the adaptation of the functional field of view is a skill which is learned, rather than a natural response to the changing environment. One example of this is the preview windows of Israeli subjects reading English and Hebrew (Polletsek et al., 1981). While reading English the subjects had a preview window of up to 15 letters to the
right of fixation, yet only three or four letters to the left. When reading Hebrew however, which reads from right to left, this visual asymmetry was reversed. In this instance the functional field (where specifically defined in terms of preview benefits for reading) was adapted to the particular language. Experience in reading produced the two opposing attentional strategies that Polletsek et al. discovered. Other studies of picture or shape identification in the functional field of view have noted a training effect (Engel, 1971; Ikeda and Takeuchi, 1975). When subjects have experience in peripheral detection experiments they become more resilient to perceptual narrowing.

In a comparison of aviators and non-aviators Williams (1995) found that the aviators had better accuracy than non-aviators in identifying peripheral targets under conditions of high cognitive load at the fovea. The experiment consisted of a foveal memory task involving letters presented in the centre of a tachistoscope field, and the identification of digits at various eccentricities in the peripheral field. This has little immediate relevance to the task of flying which suggests that the perceptual strategies of the aviators did generalise to tasks other than piloting a plane to some extent. If experience in areas such as aviation can improve peripheral detection rates, then driving experience may also have an effect.

The second question concerns whether perceptual narrowing has ever been recorded in the driving domain. An early series of in-car studies of peripheral detection rates was conducted by Lee and Triggs (1976). Their experiments consisted of up to 12 subjects driving along various roadways such as a freeway, a suburban road and a shopping centre route, or along a private road attempting to keep the vehicle following a thin line on the road surface, while verbally responding to peripherally presented lights. Four target lights were mounted on the dashboard and body of the car, the furthest two at 70° from a fixation straight ahead, and the nearest two 30° from fixation. Though they questioned the appropriateness of the term "perceptual narrowing" they noted that as the processing demands increased, such as when driving through the shopping centre or when the margin of error for line following was reduced, peripheral detection rates fell with a pronounced decrement occurring in the two targets furthest from centre.

Miura (1990) reported an experiment involving two subjects and 120 hours of driving. The subjects drove along a number of roads selected on the basis of traffic density and task demands. During the drive subjects had to verbally respond to peripherally presented target lights in a similar manner to the studies of Lee and Triggs (1976). Miura noted that as the demands of the roadway increased there was a corresponding increase in reaction times. From this he concluded that perceptual narrowing was occurring. He also identified a negative correlation between response eccentricity (distance of the target from the fixation point at the time of response) and the demands of the roadway. As the roadway becomes more complex the subjects saccaded closer to the target before responding, and used a greater number of fixations to do so. Miura's explanation is that as the usable field of view shrinks, drivers tend to
search toward the extremes of this field to increase their active search space. This can be described as a compensatory strategy developed to overcome the limits of peripheral vision under conditions of high demand, and it corresponds with the on-road data reported earlier from Underwood et al. (1997) which focused on the varying task demands of different roadways, and with Beck and Emery's (1985) suggestion of hypervigilance under anxiety provoking circumstances. Though the dual carriageway and the suburban route are viewed by novices and experienced drivers alike in regard to danger, the lack of difficulty on the former road may allow a compensatory strategy to be employed to overcome any reduction in the peripheral field.

From the work of Lee and Triggs (1976) and Miura (1990) perceptual narrowing ostensibly transfers to the driving task. Evidence also suggests that task experience can influence the shape and size of the usable field of view (e.g. Williams, 1995). The proposition that experience may play a role in the effective size of the peripheral field of drivers is supported by evidence from culmination of these two research areas. It is also possible that perceptual narrowing could be partially responsible for the increased accident liability of inexperienced drivers. Land and Horwood (1995) have demonstrated in a rudimentary simulator that experienced drivers take in information about lane position through peripheral vision, rarely fixating the lane markers close to the vehicle. Mourant and Rockwell (1972) found that novice drivers tended to fixate road markers more often than experienced drivers. If novices do suffer greater perceptual narrowing, then lane maintenance information will not be available through peripheral vision, necessitating foveating the markers. This reduces the amount of time spent fixating the focus of expansion, thus decreasing the preview time for potential hazards, which may in turn increase the likelihood of an accident.

**A laboratory methodology for assessing the usable field of view in novice drivers**

The current aim of the authors is to devise a laboratory test which will distinguish between experienced and novice drivers on the basis of peripheral vision performance. An initial study on 10 novice drivers was completed to assess the validity of the methodology. This chapter will conclude with a brief look at the method that has been adopted, and whether initial testing supports the current theory of perceptual narrowing. Though other studies have measured peripheral detections in a real driving task, the subjects that participated in this research were experienced drivers. The safety implications of conducting such a study on a group of novice drivers are considerable. For this reason a laboratory approach has been adopted.

One issue in the development of any test is the choice of measures that should be recorded. Miura (1990) said that the two most important indices of peripheral performance are response time and response eccentricity.
However, the use of reaction time as a valid measure is dependant on the presentation of the peripheral targets. If the targets are only presented for a few hundred milliseconds then a response time can add little information to our knowledge of when the light was seen and will mainly consist of post-detection response bias, unless the difference in reaction times between groups is shorter than the presentation time of the target. If the light remains on until a response is made, then the time of response is more informative about when the light was noticed. During the time between target onset and response however, one cannot identify the motivations underlying the search strategy. The subject may note the stimulus and saccade toward it for verification, or they may simply “stumble” across it in their inspection of the visual field. For this reason it was decided to use simple detection rates of short duration targets as the main indicator of perceptual narrowing.

Similarly the measure of response eccentricity can be misleading. Miura’s findings suggest that response eccentricity is inversely correlated with demands and the size of the usable field of view. This means that the smaller one’s visual field, the nearer one must be to the target before responding. However, if one saccades toward a target, then this presupposes that the stimulus has captured exogenous attention and has produced a reflexive saccade (Serano, 1992). If this is the case, the usable field of view must be at least as wide as the furthest eccentricity from which a peripheral target elicits a saccade. Instead of using response eccentricity, this initial study has focused on onset eccentricity — the distance from fixation to target at target onset. Coupled with the detection rate of peripheral targets which are presented for extremely short durations, these measures reflect the size of the subjects’ usable field of view.

**Method**

**Subjects**

Ten novice drivers were paid to take part (5 male, with a mean age of 19 years, and a mean experience since passing the driving test of 1.5 months). All the subjects had normal vision and were recruited via questionnaires distributed through the Driving Standards Agency (DSA) in Great Britain to newly qualified drivers.

**Materials and apparatus**

Thirty-nine MPEG video clips taken from a driver’s perspective were presented to the subjects via a P90 PC. Each clip contained at least one potentially hazardous event such as a car emerging from a side road or a pedestrian stepping in front of the vehicle. Overlaid on the video clips were four red place holders, each positioned half way along one of the sides of the video display. The place holders subtended $0.7^\circ$. 
The left and right place holders were 6.8° from the centre of the screen, while the top and bottom place holders were 4.4° from the centre. The peripheral targets were 200 ms lights which appeared in a random order in the centre of the four place holders. The lights subtended 0.3°. On average one light was presented every 5 seconds, with the added stipulation that two lights could not appear within one second of each other. In total 297 peripheral lights were presented to the subjects over 45 minutes of video clips. Subjects responded to the peripheral lights by pressing a button. While they watched the clips their eye movements were monitored using a SRI Dual Purkinje Generation 5.5 eye-tracker produced by Fourward Technologies.

**Design**

Two factors were of importance: demand and eccentricity of the target from fixation. Task demand was decided on the basis of a median split of a prior hazard perception study in which an average of 16 experienced drivers and 16 novices watched each video clip and pressed a button when they spotted a potential hazard. The number of button presses per subject was calculated for each 5 second segment of film, and a median split of 0.1842 defined half the 5 second windows as high demand and the other half as low. Eccentricity of the target varied according to where the subject was looking. The target was considered “near” if it fell within 6° of the current fixation. This roughly equates to an onset eccentricity within the same hemifield as the target.

**Procedure**

Subjects were instructed to search the scene as if they were the driver, attempting to spot any potential hazards. At the end of each clip they were instructed to judge it along two ratings: danger and difficulty (Groeger and Chapman, 1996). This was done on the computer with a cursor on a seven-point scale controlled by the PC mouse. They were also instructed to press a button whenever they saw a peripheral light, though the experimenter emphasised the importance of maintaining a relatively normal search pattern and searching for the hazards rather than waiting for the lights to appear. The video clips were viewed in four counterbalanced blocks.

**Results and Discussion**

An analysis of variance was conducted on those peripheral targets which had been assigned onset eccentricities by the computer. This removed targets where calibration problems occurred, or where the possibility of a blink or a saccade would have meant the target would have been missed regardless of its onset eccentricity. A
significant main effect of both demand level ($F(1,9) = 14.54, p < 0.01$), and onset eccentricity ($F(1,9) = 20.27, p < 0.01$) were discovered, but an interaction was not found. The means can be viewed in Fig. 3.

The results are consistent with the theory of perceptual narrowing and somewhat validate the experimental methodology. The consistent effects of demand on peripheral detection rates support the method of apportioning demand according to button presses in a prior hazard perception study. In a constantly changing perceptual scene, one cannot define a priori cognitive levels of demand without the confounding effects of visual complexity. This method however used a direct, self-report measure of demand which has provided a moment to moment index of the demands of all the video clips. While this does not totally solve the problem it is a vast improvement over labelling whole roads as demanding or otherwise, due to traffic density or visual clutter, and it also allows novice drivers to be tested in a safe environment.

The lack of an interaction between onset eccentricity and level of demand suggests that the applicable model is general interference rather than tunnel vision. As mentioned earlier, Williams (1982, 1985, 1988) concluded that three things were necessary to induce tunnel vision: a high level of cognitive demand at the fovea, an attentional strategy overtly focused on the central task, and speed stress on the central task. The first two were present in this initial study but the latter was absent, which may account for the results.

Regardless of which model of perceptual narrowing the results are in accord with, the initial success of this methodology is encouraging. The natural progression is to compare drivers of different levels of experience in order to identify differences in the reaction of their usable fields of view according to cognitive, foveal demand. This methodology could provide an understanding of the process which underlies
the effects of cognitive demand on visual search strategies. The added possibility of compensation strategies (Miura, 1990) may also provide an insight into the expanded search strategies which occur with increases in task demands and visual complexity. If such differences can be identified and linked with accident liability, the possibility of training interventions could become a feasible proposition.

References


Shinar, D., McDowell, E.D. and Rockwell, T.H. (1977). Eye movements in curve negotia-
ing driving and driving-related tasks. In: G.B. Grayson (Ed.), Behavioural Research in
Underwood, G. and Everatt, J. (1992). The role of eye movements in reading: some limita-
tions of the eye-mind assumption. In: E. Chekaluk and K.R. Llewellyn (Eds.), The Role
111–169.
Williams, L.J. (1982). Cognitive load and the functional field of view. Human Factors, 24,
683–692.
Williams, L.J. (1985). Tunnel vision induced by a foveal load manipulation. Human Factors,
27, 221–227.
Zwahlen, H.T. (1993). Eye scanning rules for drivers: how do they compare with actual ob-
served eye scanning behaviour. Transportation Research Record, 1403, 14–22.
CHAPTER 19

The Development of the Eye Movement Strategies of Learner Drivers

Damion C. Dishart and Michael F. Land
The University of Sussex

Abstract

Land and Horwood (1995) showed that experienced drivers obtain visual information from two sections of their view of the road ahead, in order to maintain a correct position in lane whilst steering their vehicle around a curve. The more distant of these two sections is used to predict the road’s future curvature. This section is optimally 0.75–1.00 s ahead of the driver and contains the tangent point. (That point where the inside edge of a curve reverses its apparent direction, and the driver’s line of sight forms a tangent to the road edge.) This section of road is used by a feedforward (anticipatory) mechanism which allows the driver to match the curvature of the road ahead. The other, nearer, section is about 0.5 s ahead of the driver and is used by a feedback (reactive) mechanism to ‘fine tune’ the driver’s position in lane. As either lane edge approaches the vehicle, the driver steers away from it, correcting his/her road position.

This combination of these two mechanisms enables the trained driver to steer an accurate course on roads of varying curvature. Experiments using video based eye-head tracking equipment have shown that the feedback mechanism is present in most people regardless of their experience of driving (although its accuracy is higher in those with experience), but that the feedforward mechanism is learned through experience of steering tasks (that can include riding a bicycle, computer driving games, etc.). Eye-head tracking experiments on learner drivers during their tuition have indicated that use of the section of the road containing the tangent point increases with experience, then decreases as drivers learn to optimise their visual search patterns, allowing them to spend more of their visual resources on other visual tasks both related and unrelated to driving.
Introduction.

One of the fundamental skills involved in driving is the ability to steer a vehicle accurately around a course of varying curvature. In order to do this a driver must be able to estimate the curvature of the road and convert that estimate into appropriate movements of the vehicle’s controls (steering wheel, accelerator, brake, etc.).

Land and Lee (1994) demonstrated that three experienced drivers fixated a particular feature of their view of the road ahead while driving around bends in the road. This feature was the tangent point, that point where the driver’s line of sight forms a tangent on the inside edge of a road bend. The subjects of the study would begin to fixate this feature approximately 3 s prior to entering the bend and continue to fixate it until between 3 and 6 s after entering the bend. The highest percentage of fixations to the tangent point was at approximately 0.5 s into the bend where the drivers' fixations rested on the tangent point 80% of the time. Land and Lee hypothesised that the simple relationship between the position of the tangent point and the curvature of the road, as shown in eq. (1), combined with the stability of the tangent point on a bend of constant curvature, made the tangent point so useful to the visuo-motor processes involved in driving.

\[
C = \frac{1}{d \cos \theta} - \frac{1}{d}
\]  

where \( C \) is the curvature of the road, \( d \) the lateral distance of the driver from the road edge and \( \theta \) the angle between the tangent point and the driver’s current course as subtended on the retina.

Land and Horwood (1995) demonstrated that during simulated driving, three experienced drivers relied on two sections of their view of the road ahead for visual information. These sections were at approximately 4° (far-road) and 7° (near-road) below the vanishing point of the road at a speed of 16.9 ms\(^{-1}\). Land and Horwood also found that the near-road system was effective when used by itself at low speeds (described as similar to driving in fog), whereas the far-road system was never sufficiently accurate when used on its own.

Land and Horwood hypothesised that two independent mechanisms using visual information were at work during driving, an anticipatory or feedforward mechanism that estimated the future curvature of the road, possibly from the offset of the tangent point from the heading direction, and a reactive or feedback mechanism that maintained the vehicle’s position in lane (the far- and near-road mechanisms, respectively). This model of driver behaviour agreed with a similar model put forward by Donges in 1978.

A preliminary study of inexperienced drivers during simulated driving indicated that the two mechanisms suggested by Land and Horwood are not innate but are gained with experience of driving. The study also suggested that these mechanisms
do not require experience of actually driving a vehicle but may be gained in part if not in total through simulated driving (i.e., computer/video games, etc.). This study is intended to explore the development of these mechanisms in learner drivers. To that end, the equipment used by Land and Lee (1994) has been used to measure the eye movements of drivers undergoing tuition in the United Kingdom.

**Methods**

This study has two main subject groups. Those undergoing testing in the laboratory using a simulated car drive and those undergoing testing whilst driving an actual vehicle. Each group consisted of experienced and inexperienced subjects. For simplicity these groups will be known in the remainder of this document by the mnemonics RE, RI, SE and SI, where R and S stand for Road and Simulator (experiment type) and E and I stand for Experienced and Inexperienced (subject type). The groups RE, SE and SI were selected from volunteers from the University of Sussex, England (faculty, staff and students), while the RI group was selected from volunteers provided by a Driving school in Surrey, England. Subjects were excluded from the groups RI and SI if they had any experience of driving a vehicle in the past, while subjects were excluded from the RE and SE groups if they had not had one years driving experience *after* passing their British driving test.

The subjects were asked to drive either a car or the simulator whilst wearing eye tracking headgear. This equipment makes a simultaneous video recording of the subject’s eye and the subject’s view ahead. Before and after each recording is made, the subject performs a calibration routine by looking directly at objects named on the audio recording by the experimenter. Further details can be found in Land (1993).

The recordings made from this equipment are then processed to produce a second-generation recording bearing a spot representing the position of the subject’s gaze during the recording. The position of the subject’s eye in its socket on the recording are transformed into the position of the gaze spot using the information gathered during the calibration routines. This second-generation recording can then be processed by a number of different methods. The relative movements of eye and head that move the gaze spot can be extracted by reprocessing the second-generation recording. The position of an anchor point in the subject’s field of view is tracked, on computer, to record the direction of the subject’s head. This anchor point must, therefore, be static with respect to the subject’s body (and therefore the vehicle). To this end pieces of marked masking tape are placed on the windscreen of the vehicle during the primary recording. The tape can then be used as anchor points, as can any other static point in the field of view (e.g., the rear-view mirror). The anchor point can be changed during the processing without upsetting the data collection if, for example, it moves out of the field of view. The data collected via this method is
combined with the position of the gaze spot recorded during the first process. This combined data can then be printed out in graphical form and analysed. The second-generation recording can also be analysed with respect to the distribution of gaze points in time and space relative to a particular feature of the road.

The speed of the vehicle during on road recordings can be calculated. The vanishing point (horizon) is found by extending the sides of the road until they meet. The vertical offset of a target point, static with respect to the road, from the vanishing point is then recorded for a defined period in time. Equation (2) gives the speed of the vehicle in $\text{ms}^{-1}$, where $\theta_n$ is the angular vertical offset of the target point at time $n$ and $\Delta t$ is the time interval in seconds between time 1 and time 2. The value 1.1 represents the height of the driver's eyes from the road, assumed to be 1.1 m throughout this study. The apparent speed of the simulator is set by the experimenter prior to the recording.

$$\frac{(1.1/\tan \theta_1) - (1.1/\tan \theta_2)}{\Delta t}$$  (2)

The subjects in groups RE and RI were recorded while driving on public roads in the Crawley area of Sussex, England. The subjects in group RI were recorded three times during their tuition. The first recording was made as soon as possible after the subjects had begun driving tuition. In practice this was usually the subject's second two-hour driving lesson, the subject being approached to participate in the study during their first two-hour lesson. The subject's first lesson generally consists of one hour of induction and preparation, where the subject does no driving, and one hour of driving practice. The second recording is made when the instructor feels that the subject has made sufficient progress to start being taught the more complex parts of driving (manoeuvres, etc.). The third recording is made prior to the subject's first driving test. The content of the lesson received by the student during the recording was typical of a lesson for a student of that level of ability.

The subject groups SE and SI were recorded while driving a simulator program displayed on a 55 cm wide television at approximately 80 cm from the subject's head. The subjects were free to move their heads as they required. The simulator program displays a simple road scene consisting of a horizon and left and right road edges in white on a black background. The scene also contained an arc at the bottom of the screen representing the vehicle's bonnet (that was stationary with respect to the driver). The simulator program gives both graphical and numerical outputs of the subject's performance at the end of a drive. This output gives the standard error of mean of the vehicle's position with respect to the centre of the roadway and graphical outputs of the steering wheel angle with respect to the road angle and the vehicle's position with respect to the centre of the road for the duration of the drive. All experiments were performed at a simulated speed of $12.5 \text{ ms}^{-1}$. This speed,
lower than that used in Land and Horwood (1995), was chosen as some of the more inexperienced subjects were prone to 'losing the road' at the higher speed. (Where the simulated roadway is lost from the display and is almost usually unrecoverable.) The task given to the subjects in all cases was that they should attempt to drive the vehicle along the road, keeping as near to the centre of the road as possible which simulated maintaining a correct position in a road lane. This task would be the same for drivers on most, if not, types of road and would also eliminate any differences between drivers who have greater experience of driving on the right side of the road (although, in the end, none were included in the experiments).

Results

This study is currently in progress, and so the results in this section are from those subjects tested so far, or who are currently in the process of being tested. Thus the experimental process is incomplete and that, as yet, few definite statements can be made. This, combined with the low numbers of volunteer subjects, means that the data are generally not yet statistically significant.

The initial study of groups SE and SI indicated that there was a difference in performance between those inexperienced subjects who had experience of performing a similar task to driving (e.g., use of computer driving simulation games) and those who had not. The subjects with relevant experience were able to better perform the tasks given to them on the simulator.

A qualified driver (SE-4), a non-driver with relevant experience (SI-2) and a complete novice (SI-3) were recorded using the eye-tracker while driving the simulator. After processing, the position of the gaze spot was recorded, once per second, for the entirety of the test. These points were then sorted into bands depending on their vertical position relative to the horizon, as shown in Table 1 along with the horizontal and vertical spread of fixation points and the mean number of saccades made by the subject per second. This last value was calculated by the observing number of saccades made by the subject during a particular 20-s section of the test. This value could not be calculated from the gaze point data in Table 1 as it represents a once per second sampling of gaze points for tests of approximately the same duration.

The pattern of gaze spots for the three subjects appears to increase in horizontal spread with increasing experience, while the frequency of saccades decreases correspondingly. This suggests that experience of driving teaches the subject to make fewer, longer, fixations and saccades. The increasing horizontal spread of gaze points occurs in a band between 1 and 3° below the horizon, the area where tangent points occur on the simulation. The vertical spread of gaze points appears unrelated to experience.
Table 1

Summary of gaze point positions relative to horizon for three subjects of differing driving skill

<table>
<thead>
<tr>
<th>Degrees below horizon</th>
<th>Subject</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SE-4</td>
<td>SI-2</td>
<td>SI-3</td>
<td></td>
</tr>
<tr>
<td>-1.0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>-0.5</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>2</td>
<td>7</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>2</td>
<td>11</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>3</td>
<td>12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>18</td>
<td>16</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>16</td>
<td>8</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>4.5</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6.0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6.5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>82</td>
<td>77</td>
<td>83</td>
<td></td>
</tr>
</tbody>
</table>

Horizontal spread 17° 14.5° 7.5
Vertical spread 6° 10° 7°
Saccades per second 1.45 1.80 3.40

The SE and SI groups were also asked to drive the simulator with only certain vertical sections of the roadway visible. These 2° sections were at 0, 6 and 9° below the horizon, respectively. These sections were chosen as they were at the 'top, middle and bottom' of the area of the display between horizon and vehicle. The standard deviation (s.d.) of the vehicle's distance from the centre line throughout the tests are shown in Table 2. A value of less than 0.3 for s.d. indicates that the subject did not allow the vehicle to leave the road during the test. All tests were performed at a simulated speed of 12.5 ms⁻¹ (see above).

Subjects SI-3 and SI-4 were the subjects with no experience of driving-like situations and performed the worst at all of the tests. However, there were no obvious differences between experienced drivers and those used to 'games driving'. All of the subjects were then given a test where the steering wheel on the simulator
Table 2
The standard deviation for subjects from groups SE and SI driving with only parts of the road visible

<table>
<thead>
<tr>
<th>Subject</th>
<th>Distance of 2° section of road below the horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0°</td>
</tr>
<tr>
<td>SE-1</td>
<td>0.29</td>
</tr>
<tr>
<td>SE-2</td>
<td>0.34</td>
</tr>
<tr>
<td>SE-3</td>
<td>0.35</td>
</tr>
<tr>
<td>SE-4</td>
<td>0.32</td>
</tr>
<tr>
<td>SI-1</td>
<td>0.24</td>
</tr>
<tr>
<td>SI-2</td>
<td>0.28</td>
</tr>
<tr>
<td>SI-3</td>
<td>2.44</td>
</tr>
<tr>
<td>SI-4</td>
<td>5.01</td>
</tr>
</tbody>
</table>

did not control the vehicle (although the subjects were unaware of this) and the vehicle maintained a central course for the duration of the test. This test was designed to observe the subjects steering wheel movements in the absence of visual feedback from the road. All of the subjects made vastly exaggerated steering wheel movements on this test, except for SI-3 and SI-4 who made very small, ineffectual steering wheel movements, indicating that they had not yet learned what visual feedback to expect. None of the subjects, when subsequently asked, realised that the steering wheel was no longer controlling the vehicle.

To investigate subjects' behaviour during real driving, the locations and durations of fixations were obtained twice from a group RI subject on the same stretch of road (Tables 3 and 4). Where a target of fixation is noted, the fixation point was within 1° of that object for the duration of the fixation. The saccades and fixation associated with a mirror check are not included in the averages at the bottom of these tables as they are to a within-car object. Table 1 was taken after four hours of tuition and Table 2 after 12 hours.

The road curve from which these figures were taken was a right hand bend approaching a roundabout. The data comes from 8 s of recording, starting 1 s prior to entering the bend. As the data comes from a right hand bend, the distance of the fixation point from the tangent point formed on the road edge opposite the subject has been obscured in some cases by oncoming vehicles. The mnemonics in the target of fixation column represent tangent point (T.P.) and oncoming vehicle (O.V.).

Tables 3 and 4 show similar, but not repeatedly identical results for the subject driving around a given right hand curve.
Table 3
Fixation durations and positions for a subject with four hours driving experience

<table>
<thead>
<tr>
<th>Duration (s)</th>
<th>Distance from tangent point (°)</th>
<th>Far road edge</th>
<th>Centre line</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.16</td>
<td>6</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>5</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.36</td>
<td>6</td>
<td>&gt;10</td>
<td>O.V.</td>
<td></td>
</tr>
<tr>
<td>0.16</td>
<td>6</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.54</td>
<td>4</td>
<td>7</td>
<td>O.V.</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>3</td>
<td>10</td>
<td>O.V.</td>
<td></td>
</tr>
<tr>
<td>0.48</td>
<td>–</td>
<td>&gt;10</td>
<td>O.V.</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1.16</strong></td>
<td>–</td>
<td>–</td>
<td><strong>MIRROR</strong></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>–</td>
<td>6</td>
<td>O.V.</td>
<td></td>
</tr>
<tr>
<td>0.28</td>
<td>4</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.66</td>
<td>5</td>
<td>&gt;10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.24</td>
<td>3</td>
<td>5</td>
<td>O.V.</td>
<td></td>
</tr>
</tbody>
</table>

Mean fixation duration 0.36 s
Mean distance from nearest tangent point 4.1°

Discussion

Godthelp (1986) describes the necessary steering-wheel angle for a curve of constant curvature by eq. (3), where \( c_r \) is the road curvature, \( G \) the steering ratio, \( K \) the stability factor, \( l \) the wheel base, \( u \) the vehicle’s speed and \( \delta_s \) the steering-wheel angle. No matter how variable the curvature of a road, it can be considered as a series of curves of constant curvature.

\[
\delta_s = \frac{[Gl(1 + Ku^2)]c_r}{1000} \tag{3}
\]

Godthelp (1986) proposed a model of driver steering behaviour which was derived from Donges’ (1978) model. This model had two mechanisms that allowed the above equation to be solved in a vehicle, which he dubbed the anticipatory and the compensatory mechanisms. Both mechanisms received an estimate of the road
Table 4

Fixation durations and positions for a subject with 12 hours driving experience

<table>
<thead>
<tr>
<th>Duration (s)</th>
<th>Distance from tangent point (°)</th>
<th>Far road edge</th>
<th>Centre line</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>6</td>
<td>&gt;10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.34</td>
<td>3</td>
<td>&gt;10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.24</td>
<td>8</td>
<td>1</td>
<td>T.P.</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>–</td>
<td>–</td>
<td>MIRROR</td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>9</td>
<td>&gt;10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.78</td>
<td>5</td>
<td>&gt;10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.26</td>
<td>8</td>
<td>1</td>
<td>T.P.</td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.52</td>
<td>1</td>
<td>8</td>
<td>T.P./O.V.</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.72</td>
<td>7</td>
<td>1</td>
<td>T.P.</td>
<td></td>
</tr>
<tr>
<td>0.38</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.38</td>
<td>–</td>
<td>–</td>
<td>MIRROR</td>
<td></td>
</tr>
<tr>
<td>0.16</td>
<td>6</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.36</td>
<td>–</td>
<td>–</td>
<td>MIRROR</td>
<td></td>
</tr>
<tr>
<td>0.32</td>
<td>–</td>
<td>5</td>
<td>O.V.</td>
<td></td>
</tr>
<tr>
<td>1.16</td>
<td>4</td>
<td>8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean fixation duration: 0.50 s
Mean distance from nearest tangent point: 3.8°

curvature \((c_o)\) from the visual system and produced a component of the desired steering-wheel angle. These components were combined and output to the vehicle via the steering-wheel. The movement produced by the vehicle (which is dependent on the other variables in the equation) is then fed back into the compensatory mechanism.

Land and Horwood (1995) suggested a similar mechanism, where the anticipatory mechanism received its input from the visual system’s estimate of future road curvature (from the tangent point) while the compensatory system received its input from the visual system’s estimate of the current distance of the vehicle from the edge of the lane it is in.
The first study detailed above indicated that the development of the anticipatory and compensatory mechanisms can occur with any form of driving-like experience. This is shown by the increase in the ability of subjects to perform the experimental tasks. The ability appears to reach a ‘steady state’ (i.e. little or no further improvement occurs in the ability to perform this task) with relatively little experience, as both of the subjects in group SI with relevant experience (SI-1 and SI-2) achieved s.d. scores similar to, and in some cases better than, the qualified subjects. Subject SI-2 displayed a fixation point spread more like the qualified subject and was, like the qualified subject, making fewer, longer saccades. Subject SI-3, who had no relevant experience, was making more, shorter saccades.

The subject whose eye movements were recorded on the road showed the development of a logical efficient pattern of eye movements while driving. On a right hand curve the subject alternates between fixating around the far road edge tangent point, where future potential hazards (e.g. oncoming vehicles, pedestrians, etc.) will first become visible, and the tangent point on the centre line of the road. As the Land and Horwood (1995) model of driver behaviour relies on the driver being able to estimate the distance to the lane edge, it is more logical for a driver to use the lane edge closest to the vehicle, as that edge is less likely to be obscured by other traffic, as is the tangent point.

The recording taken after four hours driving tuition shows that the subject has already begun to develop this pattern of fixations. After 12 hours the subject is able to use this pattern efficiently, for the most part using the pattern except when other visual stimuli require attention (e.g., oncoming traffic or the rear-view mirror). The subject also has a higher mean fixation duration, suggesting that the subject is processing more useful information from the fixations made, lowering the need to change fixation position.

Conclusions

From the experiments completed so far it would appear that the two-system model of driver behaviour, in the form proposed by Land and Horwood (1995), is developed by the brain in response to visual stimuli from a task that requires a relation to be established between the curvature of a road (or road-like object) and the muscle action involved in steering. The compensatory mechanism, that uses the vehicle’s lateral position relative to the road or lane edge, is obtained first, followed by the anticipatory mechanism using the angular position of the tangent point relative to the vehicle’s current heading.

Godthelp’s (1986) equation of steering dynamics (eq. 3) includes speed as a variable. A third stage of development of the behaviour would seem to be, from personal observations, the ability to adjust the behaviour for different speeds and to
be able to estimate an optimum speed for a given corner. Equation three also includes vehicle wheel base and stability factor as variables. This would suggest that a period of re-calibration would be necessary when driving different vehicles, which also seems likely from personal observation.

This study of driver-steering behaviour will continue with subjects recorded during their driving tuition as well as those recorded while driving the simulator. A new series of experiments, suggested by the results obtained thus far would be to observe the entire development of the behaviour in subjects learning the behaviour solely on the simulator. These experiments would allow the observation of the relative development of the observed and suggested behaviours under controlled conditions.

References


CHAPTER 20

What the Driver’s Eye Tells the Car’s Brain

Andrew Liu
Nissan Cambridge Basic Research

Abstract

The analysis of drivers’ eye movement may provide useful information for an intelligent vehicle system that can recognize or predict the driver’s intention to perform a given action. Such a system could improve the interaction between the driver and future vehicle systems and possibly reduce accident risk. It has been experimentally demonstrated that the pattern of eye fixations reflects, to some degree, the cognitive state of the observer. Also, a Markovian analysis has been used to quantitatively characterize the eye movement patterns associated with specific mental states. This approach is quite similar to that used to model human driver behaviour for the aforementioned intelligent vehicle system. This strongly suggests that eye movement analysis could be readily incorporated into these systems.
Introduction

In 1993, over 6 million vehicular accidents were reported in the United States, including over 40,000 fatalities and over 3 million injuries (Gross and Feldman, 1995). A large number of these accidents are attributed to driver error stemming from driver inattention, misallocation of attention, or misperceptions which lead to inappropriate decisions. Without considering the intentions or state of the driver, the concurrent use of in-car devices such as cellular telephones, navigation systems, automated safety systems and perhaps even personal computers could draw the driver's attention from the traffic situation at inappropriate times and lead to an accident. An example is the use of cellular telephones in vehicles which quadruples the risk of a collision (Redelmeier and Tibshirani, 1997). Consider the situation where a driver is following a vehicle in the same lane and a second vehicle is in the adjacent lane slightly behind the driver's vehicle. An alarm or warning to the driver is appropriate if the driver's intention is to pass the car but it could be annoying and potentially distracting if the driver was only following the car ahead. Such false alarms could result in the driver ignoring future warnings. Thus a "smart car" would be able to predict or recognize the driver's intentions and then take the appropriate course of action based on that prediction.

The major hurdles in developing such a system are the model of the driver's behaviour and a non-invasive and unobtrusive technique for checking the state variables of that model. One promising approach for modelling the behaviour of human drivers is the hidden Markov dynamic model (HMDM) (Pentland and Liu, 1995; Liu and Pentland, 1997). In this approach, observations of the driver's control actions are used to infer what action the driver is intending to perform. In this chapter, I propose that by analyzing the pattern of a driver's eye movements, it may be possible to improve the performance of a system recognizing driver intentions and possibly even allow the prediction of intentions. Consider the following thought experiment: If you were sitting in the passenger's seat of a car, would you be able to determine what actions the driver was performing or was going to perform by merely observing the driver's head and eye movements? It certainly seems straightforward to determine whether the driver is looking at the mirrors, or instruments, or to one side of the road, and so one could make fairly accurate predictions of the driver's intentions from the eye movement patterns.

The actual utilization of eye movements for practical applications is problematic due to the unperceived dynamic nature of eye movements. Consider the efforts to design computer-human interfaces based on eye movements. Early efforts linking cursor control to the gaze position of the user proved to be quite unnatural as eye movements generally represent the motor output of a number of cognitive processes, not only foveal information acquisition. More successful approaches have made inferences about the user's intentions from the pattern of eye movements. In this
The driver’s eye paradigm, interaction is enabled from the user’s natural eye movements rather than from certain prescribed and ad hoc movements (Starker and Bolt, 1990; Jacob, 1991). Jacob (1991, p. 168) commented that “... when the system is working well, it can give the powerful impression of responding to its user’s intentions rather than his explicit inputs.” Analyzing patterns of eye movement is more critical when attempting to interpret “non-spatial” intentions such as zooming into or out of a picture (Goldberg and Schryver, 1995) where there is no explicit feature or object in the visual scene that indicates the intention to zoom. This latter application is much closer to the problem of inferring the underlying state of drivers (e.g., what manoeuvre they will execute). However, the driving application is also much more difficult given the dynamic nature of the visual scene.

The remainder of the chapter provides more detailed evidence why driver eye movement patterns hold much promise as an indicator of the driver’s intentions and then how the analysis of eye movement patterns might actually be implemented into a smart car. I will briefly review some examples from the literature that illustrate patterns of eye movements that can be associated with some mental state. I will also mention the issues concerning explicitly modelling this relationship. One statistical approach, the Markov model, has been successfully used to model the patterns of eye movements associated with specific mental states in visual tasks and in driving. This approach is quite similar to the HMDMs mentioned previously, which suggests that the information in eye movement patterns might be easily incorporated into such a model of the driver. At this point of the chapter, I will describe the hidden Markov dynamic models in further detail and outline how the analysis of eye movement patterns might be incorporated into the system.

Modelling the relationship between eye movements and cognitive processes

Numerous studies have tried to establish the level at which the relationship between eye movements and higher cognitive processes can be modelled. The least controversial conclusions simply state that the pattern of eye movements generally reflect the observer’s thought processes, indicating to some degree the goals of the observer and perhaps even the main areas of interest. The strongest conclusions assert that the eye movements are directly observable indicators of underlying cognitive processes, revealing the nature of the acquired information as well as the computation processes.

There are numerous examples illustrating the general relationship between eye movements and cognitive processes. The fixation patterns of observers can change rather substantially depending on numerous factors, such as the information they are trying to discover from a scene (e.g., Yarbus, 1967), or whether they recognize a hidden or embedded figure (Stark and Ellis, 1981) or according to the current mental interpretation of an ambiguous figure such as the Necker cube (Ellis and Stark, 1978). However, to postulate a more detailed link between eye movements and
cognitive processes, three important questions should be considered (Viviani, 1990). First, how do the various cognitive processes map onto the sequence of eye fixations? Clearly, the movement of the eyes is a strictly serial process yet, there is abundant evidence suggesting that certain cognitive processes can also work in parallel. In this case, the eye movements may represent the motor output from a multiplexed command signal which includes eye movement commands from a number of cognitive processes and it becomes virtually impossible to associate specific eye movements with specific processes without a theory of how concurrent processes combine eye movement signals.

Second, how does attention move with respect to the gaze direction? It has been shown that attention may shifted spatially from the fixation point (e.g., Posner, 1980; Eriksen and James, 1986), or even in depth within the visual scene (Nakayama and Shimojo, 1992), but it remains an open question whether these are totally independent processes (e.g., Klein, Kingstone and Pontefract, 1992) or closely linked (e.g., Rizzolatti, Riggio and Shelig, 1994). Further discussion of this question can be found in Chapter 13.

The third question arises from the second, that is, how can one determine the nature of the information being acquired? The concept of “information” in vision research has numerous definitions ranging from the physical features of a visual scene, such as edges or texture changes (Mackworth and Morandi, 1967) to the features that aid the interpretation of the meaning of a scene (Yarbus, 1967; Antes, 1974). The latter form of information is highly context dependent. When the observer is in one state, then fixations may cluster on areas of the picture relevant to that state. When the state changes, then hitherto meaningless regions acquire “informativeness” and may be fixated. The concept of information is not a low-level aspect of the scene but rather must be a higher level construct dependent on the mental models being tested. More in depth discussion of scene perception and information acquisition can be found in Chapters 12 and 14.

A well-known example of a theory modelling the connection between eye movements and cognitive processing is the Scanpath Theory of Recognition (Noton and Stark, 1971a, 1971b). The theory postulated that repetitive fixation sequences indicated the creation and storage in memory of visual-motor traces, or “feature-rings”. These traces were pattern-specific and idiosyncratic and moreover, remained unchanged. In order for recognition to occur, the eyes had to move through the same scanpaths. However, numerous studies (e.g., Biederman et al., 1974; Potter, 1975; Chapter 13) have shown that picture recognition can occur very quickly and without eye movements which indicates some parallel processing of the image and no need for visual-motor traces in recognition. Furthermore, the presence of scanpaths does not influence recognition presence (Locher and Nodine, 1974). This evidence suggests that the scanpaths are merely typical of, but probably not necessary for some level of recognition.
Markovian analysis of eye movements

While the details of the link between eye movements and cognitive processes remain in question, a model of eye movement behaviour is necessary to at least identify characteristic patterns of behaviour that can be associated with a cognitive state. Stark and Ellis (1981) described a probabilistic approach where the eye movements are modelled as a Markov process. Harris (1993) re-analyzed data from Buswell (1935) and found that the data were well modelled by a first-order Markov model. In addition, Stark and Ellis (1981) showed that the transition matrices for eye movements preceding recognition of a fragmented figure were statistically different from the matrices modelling fixation behaviour after recognition.

In this approach, the location of the current fixation is dependent on the location of the nth previous fixation. The pattern of eye fixations can be captured in the set of conditional probability matrices which can be empirically measured. The zero-order Markov matrix $M_0$ simply gives the probability that a fixation will be in a given region of the scene. Assuming that the fixations tend to cluster in the set of regions $R_m$ in the scene, then it is possible to generate the $m \times m$ first-order Markov matrix $M_1$ whose elements indicate the probability $p_{ij}$ that a fixation in region $R_i$ is immediately followed by a fixation in region $R_j$. If only a few entries in each row have high transition probabilities, then the eye movement sequences have a high probability of passing through a cyclic sequence (i.e., a scan path). First and higher-order Markov matrices characterize the dependency of transitions on previous history (i.e., the probability of reaching region $R_j$ from $R_i$ through $n$ intermediate steps). However, as $n$ increases, longer fixation sequences are required to obtain accurate estimates of transition probability, so that second-order models and higher have generally not been considered. In a similar approach, individual scan patterns have been discriminated by the relative frequency of fixation triplets (Groner, Walder and Groner, 1984).

It should be noted that the existence of scanpaths as shown by Markov matrix analysis does not necessarily imply that eye movements are controlled by a higher level process. The distribution of highly salient features across the scene may lead to repeated patterns of eye movements which resemble first-order behaviour (Viviani, 1990; Ellis, 1986). However, these cases can be distinguished by statistically comparing the empirically obtained transition probability matrix to a transition probability matrix derived from the zero-order probabilities. Ellis and Stark (1986) used this type of statistical test to show that the eye movement behaviour of airline pilots viewing a flight display was not simply determined by the area of the regions of fixation which would be zero-order behaviour but exhibited first-order properties. An additional limitation of this Markovian analysis is the exclusion of temporal aspects of the fixation sequences such as fixation duration (Viviani, 1990). The temporal properties of the pattern of fixations might provide another parameter...
to differentiate the cognitive processes. Fixation duration seems to provide interesting insights into road type and possibly a driver’s assessment of the danger of the current situation (see Chapter 17).

**Driver eye movements**

Returning to the subject of driving, I will now review some of the research literature on driver eye movements. The goal is to identify patterns of eye movements that can be associated with a particular mental state of driving. These states form the building blocks for a model of a particular driving intention such as changing lanes and ultimately, the patterns of eye movements will be used to distinguish these states. Most studies in the literature report only zero-order statistics such as fixation location and duration. While these statistics are quite informative, a moment by moment analysis of behaviour would be equally desirable (see Chapter 17).

**Eye movements for vehicle control**

The most basic task in driving is probably keeping the vehicle on the road. Therefore, the majority of studies have examined eye movement behaviour in the context of vehicular control. Consider first the case of driving on a straight section of roadway with little or no traffic. Generally drivers spend most of their time looking somewhere ahead of the car but also make fixations that are closer to or farther from the car or to either side of the road. One suggestion for the distant fixations is that drivers use the focus of expansion to determine their heading for lane-keeping purposes (Gordon, 1966b; Olson, Battle and Aoki, 1989). But more likely, this is simply the most advantageous position to maximize anticipation time (Wohl, 1961) and detect possible hazards. Fixating on the road ahead provides a uniform distribution of the visual field to either side of the road in which potential hazards can be detected and allows the driver time to react appropriately. Mourant and Rockwell (1972) suggested that experienced drivers tend to look further down the road, while novice driver fixate on the road closer to the car but this result has not been found in more recent studies (see Chapters 17 and 18). The fixations closer to the car and towards the edge of the road may be used to acquire information for lateral control (Mourant and Rockwell, 1972). Thus for driving on straight road segments, a characteristic fixation “pattern” would be to alternately fixate on or near the focus of expansion and any other point in the scene.

Fixation behaviour in curves shows a more interesting pattern of eye movement. Numerous studies have shown a pattern of looking between the road ahead and nearer the car (e.g., Gordon, 1966a; Shinar, McDowell and Rockwell, 1977; Jurgensohn, Neculau and Willumeit, 1991). Again the general conclusion is that this
pattern indicates a switching between a preview of the road ahead and near-car fixations for controlling lateral position. Jurgensohn et al. showed a distinctive saw-tooth pattern of fixations suggesting that drivers fixate on a single point and track it as it nears the vehicle. When it reaches a preview distance of about one second, the eye makes a saccade to a new point and repeats the cycle. Other studies indicate that the driver makes fixations to the near and far areas of the road without the tracking (Shinar et al., 1977; Land and Horwood, 1995). Anticipation of the curve is also evident in numerous studies where eye movements to the upcoming curves preceded actual entrance into the curves by 1–4 s (Cohen and Studach, 1977; Shinar et al., 1977; McDowell and Rockwell, 1978; Land and Horwood, 1995). Different road conditions and vehicle speeds account for the range of preview times. Other studies (Land and Lee, 1994; Veltri, 1995) indicate that a large proportion of fixations on the curve ahead are concentrated near the tangent point of the curve which provides information about the curvature of the upcoming curve (Land and Lee, 1994). The curvature of the road has an effect on the pattern of eye movements in that drivers tend to make more fixations on higher curvature roads (Serafin, 1994). As the curvature of the road increases the separation between the tangent point and the area of road furthest from the driver also increases. Thus drivers may not be able to acquire information from both locations simultaneously and must make eye movements between them. This behaviour may be a useful indicator that the driver will be negotiating sharp curves.

Naturally, when eye movements of drivers are studied under normal traffic conditions, there is a significant change in fixation patterns. A vehicle in front of the driver and in the same lane tends to attract most of the eye fixations (Mourant and Rockwell, 1970; Olson et al., 1989; Veltri, 1995) Therefore the car following situation might be characterized by a pattern of eye movements centered on the lead vehicle. The mirrors and instruments inside the car attract only around 10–15% of the driver's fixations but may also be an excellent indicator of the current state or intentions of the driver. Glances at mirrors certainly indicate retrieval of information that can be related to specific driving tasks. For lane changes and merge maneuvers, one interesting trend in the pattern of looking is evident. American drivers tend to use the left-side mirror much more than the inside mirror for left side maneuvers, while the opposite is true for right-side maneuvers (Mourant and Donohue, 1977). Tasks involving the use of instruments inside the car also change the general fixation behaviour of the driver. Antin et al. (1990) compared fixation patterns when navigating with a moving-map system versus the control case of navigating a memorized route. Using a link probability measure (Wierwille, 1981), they showed that the dominant eye fixation pattern was between the road and either the mirrors or instrument panel when driving along a memorized route. However, when using the moving map system, the dominant pattern was between the center of the road and the map display. Thus the use of a navigation system could be signaled
by such a change in the driver's eye movement pattern. Interestingly, the link probability is another first-order statistical measure of the probability of a transition between two regions without regard to the direction of the transition. The link probabilities $P_{Lij}$ can be derived from the joint probabilities $p_{ij}$ of the Markov transition matrices, although the converse is not true. The two measures are related by the following equation:

$$P_{Lij} = \frac{p_{ij} + p_{ji}}{1 - \sum_{i=1}^{n} p_{ui}} \quad \text{where } i < j.$$  

Markovian analysis of driver eye movements

Liu, Veltri and Pentland (1996) used a Markovian analysis to examine whether characteristic fixation patterns for drivers could be identified (i.e., is there a unique Markov transition matrix which describes the eye fixation sequences of a driver for a given driving situation or task?). If driving is considered as a combination of a number of "basis" situations (i.e., one-task situations), then the Markov transition matrix for the general driving situation should be predicted by a linear combination of the Markov transition matrices characterizing the "basis" situations. Mathematically, the new first-order Markov matrix $M_i$ could be computed as the sum of $n$ elemental Markov matrices $M_i^i$ as follows:

$$M_i = \sum_{i=1}^{n} a_i M_i^i \quad \text{and} \quad \sum_{i=1}^{n} a_i = 1$$

where $a_i$ represents a weighting parameter for the $i$th basis driving situations. If this relationship holds, it suggests that drivers do not create a new fixation strategy for each new situation, but rather, combine existing strategies according to the perceptual or cognitive resources available.

An experiment to investigate this hypothesis was carried out on the Nissan CBR simulator to control the visual stimuli as much as possible. Eye movement data was collected at 120 Hz using a head-mounted ISCAN (Burlington, MA, USA) head/eye tracking system. Four subjects with more than five years experience drove on a simulated single-lane road composed of alternating straight and curved sections with no road markings under two conditions. In one condition, drivers were instructed to simply stay in the middle of the road. Their speed was kept constant throughout the trial. In a second condition, drivers were instructed to follow a lead car at a comfortable distance of their own choosing but also to keep the lead car always in view. The speed of the lead car was fixed but the subjects controlled their own speed.
Table 1

Description of regions used in Markovian analysis

<table>
<thead>
<tr>
<th>Region</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Near preview: current view point to 1 s ahead or up to tangent point</td>
</tr>
<tr>
<td>2</td>
<td>Middle preview: 1–2 s ahead or 1° around tangent point</td>
</tr>
<tr>
<td>3</td>
<td>Far preview: 2 s or tangent point to end of segment</td>
</tr>
<tr>
<td>4</td>
<td>Next segment ahead</td>
</tr>
<tr>
<td>5</td>
<td>Left side of road</td>
</tr>
<tr>
<td>6</td>
<td>Right side of road</td>
</tr>
<tr>
<td>7</td>
<td>Car ahead</td>
</tr>
<tr>
<td>8</td>
<td>All other samples/fixations</td>
</tr>
</tbody>
</table>

**Fixation definition**

The visual scene was subdivided into eight non-overlapping regions whose exact boundaries changed with car speed and the road geometry: four preview distances along the road, either side of the road, the tangent point, and the lead car (Table 1). Any consecutive string of raw gaze samples from a single region was labelled as a “fixation” if the duration of the consecutive string was greater than 50 ms (i.e., the number of consecutive samples from one region was greater than six). If two sample strings in the same region were separated by one or two samples (<16.6 ms) in a neighbouring region, the samples from the neighbouring region were relabelled and the samples were combined into one long string. Using this classification scheme, it was not possible to detect fixation transitions within the same region.

**Zero-order analysis**

To ensure that the drivers were making realistic eye movements in the simulator, we compared our results with those from a previous on-road study (Olson et al., 1989). In this study, six male subjects of ages between 20 and 34 years drove on a one mile rural road consisting of a straight segment followed by three curves (approximately 90°). Speed was controlled by the driver and was approximately 30 mph. The route was driven in both directions, under both day and night conditions and with and without a lead car. The drivers were not given specific instruction to follow the lead car, but the lead car adjusted its speed to stay 200–300 ft ahead.

Direct comparisons between studies are difficult because of different road and traffic conditions, as well as the different regions specified for fixation analysis. The regions from both Liu et al. (1996) and Olson et al. (1989) were combined into the
Table 2

A comparison of M₀ for a simulator (Liu et al.) and on-road study (Olson et al.)

<table>
<thead>
<tr>
<th>Region</th>
<th>No lead car</th>
<th>With lead car</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road ahead</td>
<td>0.31</td>
<td>0.24</td>
</tr>
<tr>
<td>Far field</td>
<td>0.37</td>
<td>0.25</td>
</tr>
<tr>
<td>Left of road</td>
<td>0.18</td>
<td>0.15</td>
</tr>
<tr>
<td>Right of road</td>
<td>0.12</td>
<td>0.16</td>
</tr>
<tr>
<td>Car interior</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>Lead car</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: The regions listed are combinations of smaller regions from the individual studies to make comparison possible.

six regions listed in Table 2 to facilitate quantitative comparison. The “Road ahead” region is the road surface 100–300 ft ahead of the car or <2 s preview. The “Far field” covers the road >300 ft ahead or >2 s preview. The other four categories are self-explanatory. Data represents all subjects and all speeds combined for straight segments. The fixation probabilities for two driving conditions (driving on an open road, and following behind a lead car) were comparable. In both cases, the presence of a lead car resulted in a significant shift of fixations to the lead car. In the simulator, the drivers shifted their fixations from the sides and just ahead to the lead car. In Olson et al., there were fewer fixations to the road ahead and the far field. This difference might be explained by the relative position of the lead car, which tended to be further ahead in the simulated driving.

**First-order analysis**

To study the sequences, we separated the data into four different situations: (1) Driving on straight segments, (2) driving on curved segments, (3) following a lead car on straight segments, and (4) following a lead car on curved segments. The first two situations represent the most basic driving task of lane keeping. From the literature, the first-order transition matrices should indicate significant transitions between near and far regions to acquire preview information and perhaps transitions from side to side for lateral control. The latter two tasks represent a more complex scenario where both lane keeping and car following are being executed. In these cases, significant transitions should occur to and from the lead vehicle. If the eye movement behaviour of the tasks is indeed additive, then the transition matrices of
Table 3
First-order Markov matrices $M_i$ for driving on a straight segment with and without a lead car

<table>
<thead>
<tr>
<th>Previous region</th>
<th>Current region</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No lead car</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.89</td>
<td>0.00</td>
<td>0.05</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.27</td>
<td>0.52</td>
<td>0.09</td>
<td>0.09</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.00</td>
<td>0.50</td>
<td>0.07</td>
<td>0.30</td>
<td>0.13</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.02</td>
<td>0.06</td>
<td>0.08</td>
<td>0.44</td>
<td>0.40</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.05</td>
<td>0.11</td>
<td>0.34</td>
<td>0.37</td>
<td>0.13</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.03</td>
<td>0.05</td>
<td>0.17</td>
<td>0.50</td>
<td>0.25</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>With lead car</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.32</td>
<td>0.00</td>
<td>0.13</td>
<td>0.00</td>
<td>0.00</td>
<td>0.55</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.19</td>
<td>0.18</td>
<td>0.09</td>
<td>0.01</td>
<td>0.01</td>
<td>0.52</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.00</td>
<td>0.12</td>
<td>0.05</td>
<td>0.11</td>
<td>0.02</td>
<td>0.70</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.05</td>
<td>0.02</td>
<td>0.04</td>
<td>0.38</td>
<td>0.09</td>
<td>0.42</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.00</td>
<td>0.02</td>
<td>0.20</td>
<td>0.45</td>
<td>0.06</td>
<td>0.27</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.00</td>
<td>0.04</td>
<td>0.29</td>
<td>0.39</td>
<td>0.14</td>
<td>0.14</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.10</td>
<td>0.20</td>
<td>0.39</td>
<td>0.24</td>
<td>0.06</td>
<td>0.01</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note: The italicized cells are transitions in which the frequency of occurrence is much greater than predicted from $M_0$. The regions are described in Table 1.

the car following situations should also contain the significant transitions associated with lane keeping.

By following the highest transition probabilities in each row of $M_i$, the most likely sequences of fixation were determined. In straight segments (Table 3), the expected "preview" and "side-to-side" patterns were found. The preview pattern (seen in regions 1–3) moved entirely up and down the roadway ahead of the car. The side-to-side pattern was partially a preview pattern in that highly likely transitions went from a region on the roadway ahead of the car to either side with almost equal
probability and then back to the preview region (regions 4–6). For curved segments, a similar side-to-side pattern was present, but not a preview pattern. The most likely transition in curves is to a distant preview region followed by a transition to either side and from there back to the preview region.

When a lead car is present, most of the transitions are to the lead car. On straight segments however, the two basic patterns are still evident. The preview pattern is approximately one-third as likely to occur while the side-to-side pattern is almost undiminished. On curves, the side-to-side pattern is slightly less likely, whereas transitions between the lead car and the far preview region are most probable.

To distinguish whether the patterns are different from a pattern predicted by the zero-order matrix \( M_0 \), we compared the observed first-order transition frequency matrix \( F \) with the transition frequency matrix \( F^R \) which is predicted by \( M_0 \). To obtain \( F^R \) we first calculate the expected transition probability matrix \( M^R \) by first taking the outer product of \( M_0 \) and \( M_0^T \)

\[
M_1^R = M_0 \times M_0^T
\]

Since separate fixations within a single region are not identified, the transition probabilities \( a_{ij} \) of \( M_1^R \) are renormalized as follows

\[
p_{ij}^R = \begin{cases} 
0 & \text{if } i = j \\
\frac{a_{ij}}{\sum_{j=1}^{n} a_{ij}} & \text{if } i \neq j
\end{cases}
\]

Finally, the expected frequency for the random transitions \( f_{ij}^R \) of \( F^R \) is given by

\[
f_{ij}^R = p_{ij}^R \sum_{j=1}^{n} f_{ij}
\]

A transition was flagged as significant according to the following criterion

\[
|f_{ij} - f_{ij}^R| > 2\sqrt{f_{ij}^R}
\]

If one assumes that the number of transitions has a Poisson distribution, then this is roughly equivalent to \( f_{ij} \) being more than two standard deviations from the mean \( f_{ij}^R \). The results of this “significance” test when driving on straight segments is shown in Table 4. With no lead car present, the frequency of transitions is significant among regions 1–3 (preview pattern) and regions 4–6 (side-to-side pattern). The
### Table 4

Transitions with greater than expected occurrences in two driving conditions

<table>
<thead>
<tr>
<th>Previous region</th>
<th>Current region</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2019 Region</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2020 Region</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2021 Region</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2022 Region</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2023 Region</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2024 Region</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2025 Region</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>2026 Region</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. The + symbols indicate significant transitions when no lead car is present. The ⊕ symbol indicates significant transitions when following another car.

The addition of the car following task adds a pattern of significant transitions to the lead car emanating from region 7 but the other two patterns are mostly preserved. This suggests that the additive model may be a reasonable approximation for the driver’s fixation behaviour.
Extending the Markov model for recognition of driver cognitive state

In the preceding section the results show that given a specific driver state, it is possible to statistically analyze the eye fixation behaviour and determine characteristic scan patterns that are associated with the task. However, to implement a “smart car” as described at the beginning of the chapter, the inverse problem needs to be solved. That is, given an observed pattern of eye movements, is it possible to determine the driver state that produced it?

In fact, this is one of the key problems which can be solved using hidden Markov models (HMMs) (Rabiner and Juang, 1986; Rabiner, 1989) They have been applied with great success in speech recognition and more recently, for recognition of other human actions, such as gestures (Starner, 1995) or telerobotic manipulation (Hannaford and Lee, 1991; Yang, Xu and Chen, 1994; Yang, Xu and Chen, 1997). Unlike the simple Markov model assumed in the previous analysis, HMMs are a doubly stochastic process in which an underlying stochastic process can only be observed through another set of stochastic processes that produce the observed behaviour. For the simple Markov models, each state is associated with a single output (i.e., fixation in a particular region). With HMMs, the second stochastic process “hides” the state from direct observation. Each of these states generates some observable behaviour (changes in vehicle heading or acceleration, or possibly eye movements) that can be characterized by a set of probability distributions, typically Gaussian. Therefore, a fixation in a single region does not imply the current state of the observer since many states might produce such a fixation. Instead, the whole sequence of fixations must be examined to determine the underlying state.

Hidden Markov dynamic models

A conceptual variant of this approach, hidden Markov dynamic models (HMDMs), has been applied to the problem of recognizing driver intentions (Pentland and Liu, 1995; Liu and Pentland, 1997). In this approach, the driver is considered to be a Markov device with a (possibly large) number of internal mental states. Each of these states has its own characteristic control behaviour and a set of interstate transition probabilities. For driving, the states might include the centering of the car in the current lane, checking whether the adjacent lane is clear, steering to initiate a change in heading, and centering the car in the new lane. Together, the sequence of these internal states comprises a lane change. Figure 1 shows a three-state model of a driving manoeuvre with both recurrent transitions and transitions to the following state. Of course, this is not to suggest that human drivers operate in a stochastic manner (at least not everyone!), but the HMDM framework provides a rich set of mathematical tools with which to perform this recognition task.
The driver's eye

Training Set for Task A

Example 1 of Task A

Example 2 of Task A

Example 3 of Task A

Estimate Parameters of HMDM

HMDM for Task A

Fig. 1. The parameters of the HMDM are estimated from a training set using the Baum–Welch re-estimation formulae. Each example in the training set is a time series of observed variables (e.g., gaze location or vehicle heading) that was collected as the particular task was executed. A separate HMDM must be constructed for each driving manoeuvre.

**HMDM parameter estimation**

Using HMDMs for real-time recognition of driving maneuvers is a two-stage process. First, the parameters of the HMDM for a single manoeuvre are recursively estimated from a training set of examples of the manoeuvre (Fig. 1). Each example is an evenly sampled time series of the parameters which characterize the driver (e.g., vehicle heading or gaze direction) during a single instance of the manoeuvre. The training set will have examples of varying duration since humans drivers cannot perform each instance of a manoeuvre exactly the same. Initial estimates of the mean and variance of each state output distribution are made by segmenting the training examples evenly among the states. Next the maximum likelihood state sequence of this HMDM is generated using a procedure such as the Viterbi algorithm (Rabiner and Juang, 1986; Rabiner, 1989) which determines a new segmentation of the observation vector. The parameters of the HMDM are re-estimated and this is repeated until the changes in the parameter estimates are less than some threshold.
Fig. 2. The conditional probability Pr(O|M_i) that model M_i generated the observation vector O is computed simultaneously for all models using the Viterbi algorithm. The driving manoeuvre modelled by the M_i which reaches an acceptable likelihood first is assumed to be the task that the driver intends to perform.

**Recognition of actions**

In the recognition phase, the likelihood Pr(O|M_i) of each HMDM having generated an observed pattern of behaviour O is performed by the Viterbi algorithm (Fig. 2). The observation vector O is also a time series of sampled parameters like the training examples described previously, although it does not necessarily represent a complete manoeuvre as in the case of the training examples. As O increases over time, the likelihood of the HMDM modelling the manoeuvre being executed will continue to increase while the other models increase less or even decrease. The driving manoeuvre modelled by the HMDM reaching the acceptable likelihood threshold first is considered to be the task intended to be executed by the driver. Once an observed pattern of behaviour is recognized, it can be used to re-estimate the parameters of the respective driving manoeuvre HMDM, which essentially "tunes" the models to that particular driver.

**Recent results**

Recent efforts using this approach to recognize driving maneuvers from steering actions and acceleration have produced encouraging results (Liu and Pentland, 1997). In recognition tests on the training data, the HMDMs successfully recognized their respective driving tasks with greater than 90% accuracy using only the first half second of data from the manoeuvre. Accuracy improved slightly with
longer segments of data. In a test of real-time recognition of a driver’s actions in the simulator, accuracy was up to 60% (three times chance) within three seconds of the initiation of the manoeuvre.

Eye movement behaviour and HMDMs

Incorporating eye movement behaviour into the HMDMs of the maneuvers seems to be a natural addition. Recall that drivers typically make eye movements to a curve roughly 2–3 s before entering the curve. Also, eye movements to the side-mirrors typically precede lane changes and passing maneuvers. Using eye movement information might allow the car to predict the upcoming manoeuvre, rather than performing the recognition after the task has been initiated as in the studies described above.

The primary consideration for including driver eye movements in HMDMs is how the eye movement behaviour will be coded (e.g., by gaze locations or by gaze or saccade direction). Gaze location seems to be the natural choice given that characteristic patterns of driver eye movements have been identified using a Markovian analysis of gaze location (Liu et al., in press). Unlike the Markovian analysis based on fixation location, it is possible to incorporate information similar to fixation duration into the HMDMs based on gaze location. Fixations will be characterized by high transition probabilities in the diagonal of $M_t$, indicating recurrent transitions to the same location. Thus states that drive gaze to similar regions but with different durations in those regions due, perhaps, to context dependent information content (see Chapter 12) will have very different transition probability matrices.

This approach also requires the system to perform real-time segmentation of the driving environment in order to classify the gaze location. This appears to be technically feasible given the current state of vision-based autonomous vehicles. For various projects, the forward scene has been segmented into the roadway, roadside, lanes, tangent point regions, and even other vehicles (e.g., Raviv and Herman, 1994; Weber et al., 1995). Information from global positioning systems (GPS) and digital maps could further aid the segmentation of the scene. Furthermore, some level of calibration will be needed to register the eye movements to the visual scene acquired by the smart car. A brief procedure where the driver looks at two or three known points in the vehicle (e.g., the rear view mirror or speedometer) might suffice.

Another possibility might be to use gaze or saccade direction, which lessens the needs for a complicated calibration procedure. Establishing a single fixation direction as a frame of reference may be sufficient, as long as the eye tracker’s operating characteristics remain constant between uses. The exact spatial location may not be as informative as the distribution of gaze over time in terms of determining the
underlying strategy or intention of the driver. If the distribution is spatially fixed for a long time, then the driver may be fixating on an object moving at similar speed, such as a lead car during car following. A single distribution to one side or the other may indicate an upcoming curve. A bi-modal distribution may indicate normal lane keeping as the driver moves his/her gaze from a nearby location on the roadside to a preview position ahead of the car. Transition matrix analysis of saccade direction has proven to be a useful measure of visual search strategy in video display tasks (Ponsoda, Scott and Findlay, 1995). However, the road environment seen by the driver is constantly changing as she is driving and the surrounding traffic is constantly changing as well. Only the parts of the car which might be viewed by the driver (i.e., the mirror or instruments) remain reasonably fixed with respect to the driver's eye position. Thus it might be quite difficult to deduce the nature of the visual space from such low-level information, even with additional information provided from GPS or digital maps.

Other considerations

One interesting problem is the case of driving at night. Under these conditions, driver eye fixations tend to be more concentrated in the area ahead of the car (Olson et al., 1989) which could result in significantly different patterns of eye movement. HMDMs trained on examples of daytime behaviour may have difficulty recognizing behaviour at night. The easiest solution would be to train night-time HMMs and switch to using these models when the driver turns on the headlights. But it is also possible that some behaviour, such as looking at the mirrors before passing, will not change significantly such that the models will work equally well in daytime or nighttime. It could even be argued that since there is less to see at night (i.e., fewer stimulus-driven eye movements) that the gaze patterns might be more homogeneous across drivers and thus, improve recognition of driving maneuvers. There are a number of other combinations of conditions which could alter eye movement behaviour, including rain, glare on the road, fog, etc. More studies are needed to understand how eye movement behaviour changes under these conditions. Hopefully, the studies will indicate that separate models for every condition will not be required and that the parameters of a basic model might simply be adjusted by a "gain" factor to fit the current condition. Also, it must be hoped that basic eye movement and control behaviour is fairly general so that a single set of HMDMs will work reasonably well across drivers. The evidence from the literature seems to support this claim but individual differences maybe large enough that a self-learning system will be needed to tune recognition to individual drivers.
Summary

In this chapter, I have explored the possibility of using the link between eye movements and the underlying mental processes for improving driving comfort and safety. The main question is whether the mental state or intentions of the driver can be inferred from eye movements before an action or manoeuvre takes place. The numerous examples of characteristic patterns of eye movements associated with a particular mental state such as pattern recognition or with a particular driving situation strongly suggest that this is a reasonable possibility. In particular, the Markovian approach to characterizing eye movement behaviour seems able to model the eye movement behaviour that can be associated with a particular task. A similar statistical approach used to model human drivers, the HMDM, provides a set of mathematical tools with which to actually identify the mental state most likely to be associated with an observed pattern of behaviour. The results from preliminary experiments using HMDMs without utilizing eye movements are promising. The addition of eye movement analysis should enhance the system performance by enabling recognition of driver intentions rather than just recognition of manoeuvres as they begin. Although the wide variety of potential driving circumstances may impede the implementation of a generally smart car (e.g., what happens to eye movements while driving at night?), the general outlook for success seems optimistic.

References


matic vehicle guidance. Proc. of the IEEE Intelligent Vehicles Symposium, Detroit, MI, 530–535.


