THE EFFECTS OF GAZE CONTINGENT WINDOWS ON DETECTION OF PERIPHERAL TARGETS

by

Stephen Hiroyuki Ito

A thesis submitted in conformity with the requirements for the degree of Master of Arts
Graduate Department of Psychology
University of Toronto

© Copyright by Stephen Hiroyuki Ito 1997
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.
The Effects of Gaze Contingent Windows on Detection of Peripheral Targets

Degree of Master of Arts 1997

Stephen Hiroyuki Ito

Department of Psychology

University of Toronto

Abstract

Foveated displays achieve large computational savings by providing high resolution imagery only where the viewer is looking, and degraded imagery elsewhere. Applications of foveated displays are discussed. Area-of-interest (AOI) displays are used in flight simulators and consist of a high-resolution window within a low-resolution background. If the edges of the window are perceptible, they could cause impairments in detection of peripheral targets.

Experiment 1 examined the effects of two different types (high resolution and bright) of windows on detection of moving targets. Windows impaired target detection, and the impairment was partly due to the visible edges of the windows. Experiment 2, examining the effects of window size on detection of static targets, found no size effects. The results are discussed in terms of tunnel vision and general interference accounts of impairment of peripheral tasks.
Acknowledgements

I would like to thank my supervisor, Eyal Reingold, for providing me with the idea for this project and for his tireless help and support throughout the development of this thesis. I would also like to thank Dave Stampe for his considerable technical support and his suggestions. My sanity would not have survived the running of a seemingly endless pool of subjects without the assistance of Esther Meisels and Johnathan Nightingale, thanks for all the help and the silliness, Esther and Johnathan! Finally, thanks go to my father and my sister for their endless support.
# Table of Contents

Abstract  
ii
Acknowledgements  
iii
Table of Contents  
iv

## Introduction

The Bandwidth Problem

- Computationally-Intensive Graphics Applications  
  2
- Communications  
  3
- An Answer Inspired by the Human Visual System  
  5
- Real-Time Foveated Display Systems  
  7
- Predictive Foveated Display Systems  
  10

Applications
- Flight Simulation  
  13
- Indirect Vision for Pilots  
  21
- Remote Vehicle Control  
  22
- Teleoperation: Remote Manipulation  
  24
- Image Filing Systems  
  24
- Video Teleconferencing  
  25
- Medicine  
  26

## Human Factors of Foveated Displays

- Acceptability of Display  
  28
- Evaluation of AOI parameters  
  31
- AOIs May Interfere With Target Detection  
  37

## Experiment 1

- Method  
  39
  - Participants  
    39
  - Apparatus  
    40
  - Stimuli  
    41
  - Design  
    42
  - Procedure  
    42
    - Setup and Calibration  
      43
    - Trial Sequence  
      44
  - Measures and Analyses  
    44
    - Operational Definitions  
      44
    - Dropped Trials  
      45
- Results  
  45
  - Participants’ comments  
    47
  - Discussion  
    47

## Experiment 2  
52
Method
  Participants  52
  Apparatus  52
  Stimuli  52
  Design  52
  Procedure  53
    Setup and Calibration  53
    Trial Sequence  53
  Measures and Analyses  54
Results  54
Discussion  55
General Discussion  56
References  63
List of Figures  70
Appendix (Subject Instructions)  86
The Effects of Gaze Contingent Windows on Detection of Peripheral Targets

Advancements in computer imaging technology are revolutionizing defense, medicine, industry, communications, education, and entertainment. They have made possible very realistic and immersive flight simulators for Navy pilots, allowing them to train and practice difficult maneuvers without risk to people or equipment, allowed face-to-face communication of people thousands of kilometers apart, and even enabled a surgeon to perform surgical operations on a model of a pig a continent away. Much more is possible, but there is a problem to be hurdled: many of these applications require extremely computationally intensive imagery which is either beyond the reach of current technology, or too expensive.

The problem exists whenever the computational requirements of generating or transmitting images exceeds the available resources or bandwidth. Bandwidth refers to the amount of information transmitted in a certain period of time (usually expressed in bits per second, or bps). Bandwidth usually refers only to communication, but for the sake of simplicity, the term “bandwidth problem” will be used here to mean any time the requirements of generating or transmitting images exceeds available resources, in both communication and image generation. When the bandwidth problem occurs, it is necessary to reduce the processing requirements to generate the images or employ some form of compression to reduce the necessary bandwidth. Conventional compression techniques generally are not sufficient to perform the enormous reductions of bandwidth required for these applications. The problem is not new; nature has been faced with the problem of providing high-resolution, wide field of view vision. Nature solved it by varying the acuity of the retina, so that only a very small region of the retina, the fovea, possesses high acuity vision. Eye and head movements
combine to position the fovea on areas of interest in the visual scene. This property of the human visual system provides an avenue for reducing the processing demands of computer images.

Foveated display systems distribute display resolution according to the spatially variant acuity of the retina, in order to obtain substantial savings in resources without perceptibly degrading the image. Foveated display systems must be able to monitor or predict either head position or gaze position (gaze position refers to the spot where the observer is looking, and is a combination of head and eye position), and generate a variable resolution image accordingly. Systems that respond to head and eye movements are known as head- or eye-slaved systems, or gaze-contingent displays, and have already been implemented in flight simulators, and are also being developed in image and video display systems and volume rendering. The first section will discuss in more detail the bandwidth problem, the human visual system, and foveated display schemes. The second section will delve into application-specific requirements of computer imaging, and implementations of this technology in each domain. The third section will discuss human factors-type work on these systems, and the fourth and final section will discuss a potential problem and describe the current experiment.

**The Bandwidth Problem**

*Computationally-intensive graphics applications.* Flight simulators require high resolution or detail and extremely wide fields of view. Many flight simulators use polygons to generate imagery, and are thus limited in terms of the number of polygons they are capable of drawing, rather than pixels, but the problem is the same. Flight simulators require highly detailed imagery, which requires large numbers of polygons. Even the most powerful modern high-speed computers are not capable of generating the huge number of pixels or polygons required by uniformly high resolution, wide field of view displays in real-time. Usually there is an upper limit to the number of pixels or polygons
which can be displayed, which necessarily entails a trade-off between resolution and field of view; increasing the field of view necessitates a cost of decreased resolution, and vice versa.

*Communications.* The bandwidth problem arises in communication because computer imagery is an extremely high-bandwidth application, and communications links are limited in terms of the amount of information that can be carried, and tend to be very expensive. Following are the bandwidth requirements for the following applications: slow-motion video (10 frames per second (fps), 176x120 pixels, 8 bits/pixel) 5.07 Mbps, video file transfer (15 fps, 352x240 pixels, 8 bits/pixel) 30.41 Mbps, broadcast video (30 fps, 720x480 pixels, 8 bits/pixel), 248.83 Mbps, and HDTV (59.94 fps, 1280x720 pixels, 8 bits/pixel), 1.33 Gbps (Bhaskaran & Konstantinides, 1995).

Compare the bandwidth requirements given above with transmission capacities of several types of commercially available communication links: basic-rate ISDN (Integrated Services Digital Network, or dial-up telephone lines), can carry either 0.064 or 0.128 megabits per second (Mbps), fractional T1 can carry 0.384 Mbps, T1 can carry 1.536 Mbps, and T3 can carry 44.7 Mbps. T1 is required for good-quality teleconferencing, and is extremely expensive; the estimated annual intercity cost of T1 service for telemedicine is $65,000, compared to $18,000 for basic-rate ISDN (Hiatt, Shabot, Phillips, Haines, & Grant, 1996). Higher bandwidth lines include TAXI (100 Mbps) and SONET (155 Mbps and higher; Cabral, Deforge, & Kim, 1996). Other types of transmission are more specialized and include fiber optic links, which can support high transmission rates but have limited range and are prone to breakage, microwave links which can also support high transmission rates but require line of sight between transmitter and receiver, radio links, which typically have low transmission rates; state-of-the-art tactical communications links for example support data rates between 16 and 64 Kbps, and hydrophone links, which have even lower transmission rates (Weiman, 1994). Clearly bandwidth requirements must be drastically reduced to make image communication
applications feasible using existing communication networks. For example, DePiero, Noell, and Gee (1992) estimated that video data rates for remote teleoperation must be reduced by a factor of at least 1000, based on video rates of 60 Mbps over a 16-64 Kbps radio link.

The standard answer to limited bandwidth requirements is compression of the data, or reduction of the number of bits required to specify an image by exploiting redundancies and irrelevancies in the data. Many different compression schemes have been developed, including lossless and lossy schemes. Lossless schemes allow the full recovery of the original image when it is decompressed, but have low compression ratios, between 2:1 and 4:1. Lossy schemes involve some irreversible loss of data when the image is decompressed. However, at compression rates between 10:1 and 20:1, it is still possible to recover images which are "visually lossless" or perceptually identical (Cabral & Kim, 1996). The most commonly used compression standard for still images was developed by the Joint Photographic Experts Group (JPEG). One problem with this standard is that it uses discrete cosine transform (DCT) functions which can yield blocky "quantization artifacts". Other compression schemes such as wavelet transforms do not generate these quantization artifacts.

Video compression standards include motion JPEG, MPEG-1, which is designed to compress VHS quality video into a 1.2 Mbps bitstream, and MPEG-2, which can compress many different types of video, ranging from VHS to HDTV, and which can compress 720 x 480 video at 30 fps into a 5-15 Mbps bitstream (Cabral & Kim, 1996). These compression rates generally fall far short of those required for teleoperation, telemedicine, teleconferencing, and other applications. For these applications to become feasible, these conventional compression schemes must be combined with some other method of data compression.

An answer inspired by the human visual system. One answer to both the problems of finite computer resources inherent in computationally intensive graphics applications and limited bandwidth

"
communication applications is to emulate or exploit certain low-level features of the human visual system, especially the spatially variant acuity of the retina. A brief overview of relevant aspects of the human visual system follows.

Visible light enters the eye through the pupil, is focused by the lens, and forms an image on the retina, a sheet of neurons in the back of the eye. The light is converted or transduced into neural impulses by a class of neurons called photoreceptors. There are two types of photoreceptors, rods, which are highly sensitive and are responsible for vision in low light conditions (scotopic vision), and cones, which are responsible for high acuity, colour vision under daylight conditions (photopic vision). It is photopic, or cone-driven vision, that is important for computer displays. Cones are distributed unevenly across the retina (see Figure 1), and are most densely packed in a 1° circular region of the retina called the fovea, where cones are spaced at about 0.5 arc min apart (Wandell, 1995). From the fovea cone density falls off smoothly but rapidly into the parafovea, which extends out 5° from the centre of the fovea, and then into the periphery (see Figure 1). Note that the dropoff of cone density is smooth, and the fovea/parafovea/periphery distinctions do not correspond to distinct changes in cone density, and are thus somewhat arbitrary and definitions differ from investigator to investigator. As one might expect, the acuity of the retina closely corresponds to the density of cone photoreceptors in that region (see Figure 1). Acuity is highest in the centre of the fovea, where maximum resolution varies from 0.5 arc min to 1.0 arc min, depending on the measure of acuity used, and drops off sharply and smoothly toward the periphery.

Eye (and head) movements are used to position the high acuity foveal region at regions of interest in the visual scene. Combinations of eye and head movements are called gaze shifts, and the term gaze position refers to where the viewer is looking and includes both head and eye position. Saccades are rapid, ballistic movements of the eyes to bring the fovea to the target area quickly.
Ballistic means that once initiated, saccades cannot normally be stopped or altered. Saccades can range in amplitude (size) from about 3 arc min to about 90° (Avanzini & Villani, 1994), although normally saccades do not exceed 40°. Usually gaze shifts greater than 20° are accompanied by head movements. Saccade size and velocity are closely related: the larger the saccade, the higher the peak velocity, up to about 700°/sec. During saccades, and for up to 70 msec (Ishida & Ikeda, 1979) following saccades, sensitivity to visual stimulation is reduced; this phenomenon is known as saccadic suppression (Matin, 1974).

Fixations refer to the periods of stable gaze between saccades. Visual integration across fixations is perceptually seamless and complete, and thus perception of the world is not of a series of discontinuous “snapshots” but rather a seamless, integrated whole. Smooth pursuit eye movements are designed to keep a moving target on the fovea. When tracking is perfect, eye velocity matches target velocity, and the gain, defined as eye velocity/target velocity, equals 1.0, but usually pursuit gains are somewhat lower than one. Smooth pursuit can be maintained for target speeds of up to about 90°/sec (Avanzini & Villani, 1994). Optokinetic eye movements are similar to pursuit eye movements but are caused by large visual field motion. Vestibulo-ocular reflex, also known as compensatory eye movements, keep the fovea on the target by compensating for movements of the body and head. Vergence eye movements are paired and opposite inward and outward rotations of the eyes which align both eyes’ foveas on the target, depending on its distance from the observer.

The spatially variant acuity of the visual system can be exploited to minimize transmission bandwidth and computer processing demands, and is the focus of this experiment. In normal vision, low resolution peripheral vision is used to identify regions of interest in the visual field, and head and eye movements are used to position the fovea at those regions. This system works so well that human perception of the world is not of a high resolution central region and a severely degraded
peripheral region, but rather a compelling illusion of a continuous, high resolution, wide field of view image.

If gaze position is known, large bandwidth savings can be realized by spatially varying the resolution of images according to the acuity function of the eye, so that high acuity foveal and parafoveal regions receive high resolution imagery, and low acuity peripheral regions receive compressed, low resolution imagery. The savings over conventional displays, which display high resolution everywhere, occur in the peripheral regions of the image. Greater compression ratios can be achieved with larger field-of-view displays: for example standard NTSC video will benefit less from foveated displays than HDTV, because owing to the greater field of view of the HDTV display, more of the display is located in peripheral vision, and thus can be compressed without perceptible loss. Foveated display systems can be divided into two types: real-time systems that monitor the viewer's gaze position and update the display accordingly, and predictive systems that predict the viewer's gaze position(s), and preprocess the image accordingly.

*Real-time foveated display systems.* Foveated display systems which dynamically change the position of the high resolution portion of the image in response to head and eye movements *during viewing* are known as “real-time” systems. The US Department of Defense has long been using displays of this type, called “area-of-interest” (AOI) displays in flight simulators. Typically they consist of a relatively large (18° to 40°) rectangular head-, eye-, or target-slaved high-resolution inset, or AOI, within a low resolution background, usually with a blending region in between (Thomas & Geltmacher, 1993). Specific systems will be described in more detail later.

Abdel-Malek and Bloomer (1990), using a gaze contingent foveated display system with four levels of resolution, demonstrated that extremely large savings in the number of pixels displayed can be obtained, with larger savings for larger displays. On a workstation with a 40° x 40° field of view
displaying approximately 1 million pixels, number of rendered pixels can be reduced by 91% (11:1 compression ratio), and on a large dome display used for flight simulation displaying approximately 8 million pixels, a 98% savings can be realized (50:1 compression ratio).

Kortum & Geisler (1996), developed a gaze contingent foveated display system with a graceful degradation of resolution from centre to periphery. Their formula for screen resolution closely matches available perceptual and anatomical data for visual resolution, so that if the foveal “SuperPixel” (SuperPixels are collections of screen pixels that have been assigned the same gray level value) is less than the foveal resolution limit, then the increasingly larger SuperPixels which occur toward the periphery should remain less than the resolution limit of the corresponding retina, and the foveated image should be perceptually indistinguishable from the original, unprocessed image. However if the foveal SuperPixel is larger than the foveal resolution limit, then the visual system should be capable of resolving all of the SuperPixels in the image, and the foveated image will be discriminable from the original. They displayed three 256 x 256 8-bit gray scale images with frame rates of 60 Hz and eye movement update rates of 30 Hz. A letter chart, selected to evaluate visual clarity in a reading task, a face, for evaluation of teleconferencing systems, and a natural environment scene, for evaluation of cluttered, high detail images, were used as stimuli. Subjects were generally aware of the reduced resolution of the peripheral visual field only in the letter chart condition, and even then they noted that the effect had a very small effect on the perceived quality of the image. Note that the foveal SuperPixels in such low-resolution images were resolvable by foveal vision; the authors contend that had higher resolutions been used, such as 1024 x 768, the image would be perceptually identical to the original image. Kortum and Geisler achieved compression rates of 94.7% (a factor of 18.8) using relatively small images. Increasing the field of view will result in greater bandwidth savings. Furthermore, this foveation scheme can be used in conjunction with other
compression schemes. The authors estimate that compression rates of greater than 100:1 are possible with the addition of lossless compression schemes.

Foveated imaging scheme have also been developed for three dimensional images as well. As in flight simulation, more efficient use of computer resources, rather than data compression per se, is the goal of foveated display schemes in volume rendering. Levoy and Whitaker (1990) developed a gaze contingent ray tracer for rendering volumetric data that incorporates gaze direction directly into rendering algorithms. This is similar to Kortum and Geisler's approach and in contrast to the AOI approach of most flight simulators, in which the inset and background images are generated separately and combined later. The advantage of this method over the conventional AOI method is that it can support graceful degradation of resolution from fovea to periphery, thus more closely approximating the degradation of retinal acuity, and the image can be displayed on a conventional television or computer monitor, obviating the need for specialized (and expensive!) displays. Levoy and Whitaker employ a spatially adaptive ray tracer using 3D mip maps (Levoy & Whitaker, 1990). The target resolution falls off smoothly from one ray per pixel within a circular area 6° from the gaze position to one ray per 16 pixels outside a circular area 10° from the gaze position. Their system, implemented on a DEC 3100 workstation rendered a full-resolution 256 x 256 x 109 voxel three dimensional MRI scan in 59.6 sec, using 44,100 rays and 316,500 voxels. The same image rendered using the variable resolution method was generated in 13 sec, using 9,657 rays and 55,438 voxels, for a time savings of almost 80%. Unfortunately, although they had planned to implement their system on a Pixel-Planes 5 machine (anticipating a speedup of 32:1), the authors never implemented their system on a more powerful machine (Levoy, 1996, personal communication), so it is unknown what level of performance is attainable using current high-performance computers.
Ohshima, Yamamoto, and Tamura (1996) added a stereoscopic twist to 3-D gaze-directed volume rendering by adding binocular disparity cues to the display. Thus, objects closer or farther away in depth than the object being focused on are blurred and appear in double images, to simulate binocular disparity in normal stereoscopic vision. Their foveated volume rendering scheme reduced the rendering time per frame from 230 msec per frame to 50 msec per frame.

*Predictive foveated display systems.* A disadvantage with real-time foveated display systems is that they require sophisticated and bulky equipment for measuring head or eye position and dynamically updating the display. Furthermore, these systems can only support one viewer at a time. However, foveated display techniques can still be useful for some applications even without special equipment, if the image is preprocessed, and only regions of the image where most people are expected to look are displayed at high resolution. The feasibility of this method critically depends on whether different people tend to look at the same spots in the displayed images. Regions of an image which contain most of the gazing positions have been called the gazing areas of the image (Tsumura, Endo, Haneishi, and Miyake, 1996). Stelmach, Tam, & Hearty (1991) demonstrated a high degree of viewer agreement in terms of gaze positions in video sequences; for example, in a hockey game videotape, 78% of viewer gaze positions fell within the dominant gazing area of the image. This data suggests that at least for some video applications, people tend to look at the same spots in the image, which affords the possibility for foveated displays.

Stelmach and Tam (1994), using a predictive foveated coding scheme, determined the predicted gazing area for a video sequence. They then degraded the sequence outside of the predicted gazing area, outside of the central area of the screen, or throughout the whole image using one of two methods: blurring or quantization. Quantization was used to simulate the effects of DCT quantization-based compression schemes, such as the popular JPEG format. They then asked
viewers to rate the acceptability of the degraded and unprocessed sequences. The benefits of gaze-contingent coding were assessed by the gazing area vs. central area comparison. Viewers changed their viewing patterns while watching the video sequence, so the authors reported only modest benefits of providing high resolution in a gaze contingent manner rather than simply in the centre of the image. They concluded that predictive gaze contingent processing is not suitable for general purpose image coding.

Duchowski and McCormick (1995) note that peripheral vision is used to identify potential future gaze targets. They call these “visual attractors”, and contend that foveated coding schemes should not diminish these peripheral visual attractors. They furthermore argued that uniform degradation of the periphery was not acceptable because the acuity of the human visual system is not uniformly degraded in the periphery. Thus, in their predictive foveated coding method, they used multiple regions of interest and spatially-varying Gaussian smoothing. They did not evaluate the subjective image quality of their coding scheme, nor did they report any compression ratios.

Tsumura et al. (1996) developed a gaze contingent progressive image transmission system. Progressive image transmission involves transmitting an image in stages, with each successive stage containing a better approximation of the final image. Thus an overall vague impression of the image is available very quickly, and the details of the image are gradually fleshed out as the transmission proceeds. The benefit is that the user can quickly recognize and abort transmission of unwanted images. Tsumura et al’s innovation was in making this progressive image transmission gaze contingent, so that the gazing area of an image is transmitted first, and then other regions are transmitted later. Not only does their method decrease the total size of the compressed image (76.3 KB using their gaze contingent method compared with 93.0 KB using the conventional method), their method results in quicker mismatch identification and abortion (1.97 stages were transmitted
before abortion using their gaze contingent method, compared with 2.74 stages using the conventional method).

In summary, both real-time and predictive foveated displays can deliver drastic savings in the number of display pixels, and thus offer large compression ratios. These compression ratios can be used in conjunction with other compression schemes to obtain extremely high rates of compression, which increases the feasibility of a number of different graphics- and video-intensive applications. However, as illustrated by the single and multiviewer problem, different applications have different requirements for foveated display systems. In the next section, application-specific requirements and specific implementations of foveated display systems are discussed for a number of application domains.

Applications

The bandwidth problem is ubiquitous, cropping up in flight simulation, teleoperation, image archiving systems, teleconferencing, telemedicine, and even television. However, the foveated image scheme has only been extensively used in flight simulation. Undoubtedly part of the reason is the traditional expense and complexity of implementing gaze-tracking systems and sophisticated imaging systems; until recently these were out of reach for most people outside of the defense industry.

Before an imaging system is implemented for any application, it is important to define the imaging requirements for that application. Factors to consider include field of view size for each resolution region, level of resolution for each region, colour, available bandwidth/computing resources, and acceptable delay. Delay for real-time systems refers to the latency from an eye or head movement to the update of the display.
The bandwidth problem and the potential for the use of a foveated display system occurs in the following domains: flight simulation, indirect vision for pilots, remotely piloted vehicles, teleoperation, image archiving systems, teleconferencing, television, and telemedicine. A description of the specific requirements of each application and implementations of foveated image systems within these domains follows.

**Flight simulation.** Flight simulators were developed in order to help reduce training costs and to eliminate the risk of injury during the training of dangerous maneuvers. Their cost-effectiveness for training has been demonstrated for aerial refueling, as contracted simulator training of aerial refueling resulted in a savings of one inflight aerial refueling sortie per pilot (Lee & Lidderdale, 1983). Researchers investigating transfer of training from the Advanced Simulator for Pilot Training (ASPT), a head-slaved AOI system, found that simulator training under high density ground threat conditions can improve the survivability of aircrews (Hughes, Brooks, Graham, Sheen, & Dickens, 1982). This is particularly important given that the first few missions a pilot flies are critical for survival, and that after those few missions, the pilot's survivability significantly increases (Hanson, 1983). Side benefits include reduced airport congestion, noise, and pollution because of less training flights.

The problem of conflicting requirements of field of view size and resolution was recognized early on by developers of flight simulators in the US Department of Defense. This was the first implementation of foveated imaging schemes, and still represents the bulk of research performed in this area. The first requirement, field of view size, varies widely with the types of tasks being trained. For example, Collyer, Ricard, Anderson, Westra, and Perry (1980), found that safe and acceptable landings and take-offs could be performed when field of view was restricted to $10^\circ \times 10^\circ$. However, a $180^\circ$ (horizontal) x $90^\circ$ (vertical) field of view is not sufficient for maintaining a normal tactical
two-ship formation (Thomas & Gelmacher, 1993). A study by the USAF demonstrated a wide range of field of view requirements for different tasks; the most demanding was a barrel roll, which required a 299° (horizontal) x 142° (vertical) field of view (Leavy and Fortin, 1983). To be able to train a large range of tasks in the flight simulator, a wide field of view is required. Thomas and Gelmacher (1993) point out the ability to support tactical two-ship operation as a minimum requirement for flight simulators. Flight simulators require high resolution for long range detection, classification, and orientation of targets (Thomas & Gelmacher, 1993). Detection of a MIG-21 at a range of 5 nautical miles requires a resolution of about 2.86 arc min/line pair, and recognition of a tank or armoured vehicle at a range of 2 nautical miles requires resolution of about 3.43 arc min/line pair (Turner, 1984). Kennedy, Berbaum, Collyer, May, and Dunlap (1988) noted that although current Navy and Air Force simulators permit detection of targets at great ranges, they are not able to provide enough spatial information for estimating the orientation of aircraft at real-world distances (>1.61 km), which is dangerous because it may prevent the learning of proper maneuvering tactics. Finally, Turner (1984) described a study conducted in the Advanced Simulator for Pilot Training that indicated that the background scene and formation aircraft could be displayed at a resolution of 9 arc min/line pair for effective training.

Finally, slaving system delays have been linked to impaired performance on flight tasks, (e.g., Frank, Casali, & Wierwille, 1988; Grunwald & Kohn, 1994), impaired training (Ricard, 1995), and simulator sickness (e.g., Frank, Casali, & Wierwille, 1988; Grunwald & Kohn, 1994). Furthermore, large head-slaving delays tend to result in pilots minimizing head movements, so that the full wide field of view is not fully utilized (Grunwald & Kohn, 1994). Thus it is desirable to minimize slaving system delays. These field of view, resolution, and delay requirements are rather stiff, and indeed they have prevented the use of flight simulators by the USAF for combat training (Thomas &
Geltmacher, 1993). Still, they have been useful for weapon-system familiarization, procedure training, and emergency training, and development continues in the hopes of achieving a useful general purpose combat simulator.

Simulators currently in use by the USAF at the Aircrew Training Research Division at the Air Force Armstrong Laboratory fall into two general categories, dome displays and helmet mounted displays. With the exception of low-cost helmet displays, all of the systems which follow require the use of two or more channels of a computer image generator called the Advanced Visual Technology System (AVTS), which updates 8000 faces, distributed among a maximum of 10 active channels at 60 Hz (Thomas & Geltmacher, 1993).

Before continuing, a distinction should be made between “resolution” and “level of detail”. Resolution refers to the size of individual pixels or line pairs in the display. Level of detail is defined as the density of computer image generator edges, or edges per unit visual angle. Computer image generator databases for flight simulators are modelled with a number of levels of detail. For example, a house at the lowest level of detail is just a block, which is sufficiently detailed when it is very distant. The same house at a higher level of detail has doors and windows, which are details required to identify it as a house at close range. For AOI systems, it makes sense to display objects at high levels of detail in the inset, but it is uneconomic to display objects at high levels of detail in the background. Thus the AOI inset can have a higher resolution than the background, a higher level of detail than the background, or both. The preferred system will have both.

Dome displays are considered the most successful design and the standard against which other designs must be matched. Domes that use target projectors in addition to a few channels of background imagery are quite effective for training air-to-air tasks because high target resolutions are possible with just a few projectors and video channels.
The limited-field-of-view dome (LFOVD) was the first of the domes built. It consists of two channels of projectors that paint imagery on the inside surface of a 24-ft diameter dome. A central high-resolution AOI is optically combined and blended with a low-resolution background. The instantaneous field of view, which as its name suggests, refers to the field of view available at any moment in time, is a 140° (horizontal) x 60° (vertical) oval, and the field of regard, which includes total field of view with head movements, is 180° (horizontal) and 90° (vertical). Although it was originally designed to support eye-slaved operation, early attempts at eye-tracking were unsuccessful, so currently the system is head-slaved. The AOI inset is 40° (horizontal) x 30° (vertical), and optical blending filters smooth the transition between the high resolution and low resolution regions. The size of the blending region is 5° (Warner, Serfoss, & Hubbard, 1993). The inset resolution is approximately 2 arc minutes/pixel with a peak luminance of 4 fL and a contrast ratio of 20:1. The LFOVD can also display a smaller AOI, about 26° (horizontal) x 21° (vertical), with a resolution of about 1.5 arc minutes/pixel, and a 2.5° blending ring (Warner, Serfoss, & Hubbard, 1993). Despite considerable efforts by the designers to give the LFOVD an acceptable field of view for most training maneuvers (Leavy & Fortin, 1983), it has been found that this simulator's field of view is too restrictive to allow tactical two-ship operation (Thomas & Geltmacher, 1993).

The full-field-of-view dome (FFOVD) display uses more projectors (6) and video channels than the LFOVD to provide a full field of regard and instantaneous field of view for a tactical fighter cockpit. A seventh projector displays the high resolution AOI, which can be head-, eye-, or target-slaved. The AOI resolution can equal that of the LFOVD. The peak luminance is about 3 fL, and the contrast ratio is around 20:1.

The display for advanced research and training (DART) was designed to provide bright, high-contrast wrap-around imagery at a much lower cost than the other dome displays. Nine CRT rear-
screen projectors wrap around the cockpit. A head tracker determines which CRTs are invisible to
the pilot, so six AVTS IG channels can be switched to cover all nine projectors. The real image is
3.5 ft from the pilot's eye, has a peak luminance of 25 fl and a contrast ratio of 50:1. The field of
regard is identical to that available in an actual F-16C cockpit, and the entire display is medium
resolution (4.75 arc minutes/pixel; Thomas & Geltmacher, 1993). Kelly, Shenker, and Weissman
(1992a, 1992b) have proposed adding a head-slaved, helmet-mounted AOI to the DART, to address
its most glaring weakness, its relatively poor resolution.

The mini-display for advanced research and training (Mini-DART) uses a high-resolution
front screen and seven wrap-around screens for lower resolution peripheral imagery. Currently there
is no head-slaved or eye-slaved AOI system built into this display, so this system will not be discussed
further.

Helmet mounted displays have some inherent advantages over dome displays. They require
fewer video channels to provide a full field of regard, they are smaller, and can provide very high
resolution. Furthermore, they can support two viewpoints for two-place cockpits by using two
helmet displays in one cockpit. However, they cannot provide a full instantaneous field of view and
they require placing a lot of heavy equipment on the head (e.g., 6 lbs plus the helmet weight,
Burbidge & Murray, 1989), which can induce fatigue and affect head and eye movements. Gauthier,
Martin, & Stark (1986), demonstrated that the added inertia resulting from helmet-mounted weights
can cause slower head velocities and increased eye movement velocities and amplitudes; in other
words helmet mounted displays may cause unnatural head and eye movements, which can result in
decreased perceptions of visual stability. Thus weight becomes a critical consideration for helmet-
mounted displays, and increasing field of view, resolution, or luminance almost always means
increasing weight (Welch & Kruk, 1986).
The fibre-optic helmet-mounted display (FOHMD) consists of two components, the helmet-mounted optics and the off-helmet support optics. Two channels of the AVTS image generator provide the imagery, one channel per eye, and transmits them to the helmet via flexible fibre optic cable consisting of approximately four million fibres, with an effective capacity of about ten million full-colour pixels per bundle. Two versions of the display are in operation at the Aircrew Training Division of the Air Force Armstrong Laboratory, one with an instantaneous field of view of 126° (horizontal) x 67° (vertical), and the other with a field of view of 160° (horizontal) and 80° (vertical). These instantaneous fields of view, along with a 25° (horizontal) x 19° (vertical) high resolution inset (1.5 arc min/pixel), can be head- or eye- slaved and both systems have an unlimited field of regard. The luminance of the display has been measured at over 40 fL, with a contrast ratio of 30:1 (Thomas & Geltmacher, 1993). The wide horizontal field of view was achieved at the cost of losing a large degree of binocular overlap. Only 38° of the 82.5° monocular fields of view overlap (Robinson, Thomas, & Wetzel, 1989). A similar solution to the horizontal field of view problem has been proposed for the helmet mounted displays of helicopter pilots. Normally, binocular overlap for humans is 114° (Howard & Rogers, 1995), so binocular vision with the FOHMD system represents a significant departure from the normal condition. The objections raised to partial binocular overlap helmet mounted displays for helicopter pilots can apply to the FOHMD as well. One consequence of this type of partial binocular overlap is a perceptual effect known as luning, which is a subjective darkening of the monocular regions of the field of view. This can manifest in behaviour as an impairment in target identification (Klymenko, Verona, Beasley, & Martin, 1994).

Low-cost helmet displays have also been developed with instantaneous fields of view of 40° (horizontal) and 30° (vertical), with an unlimited field of regard. Currently there is no provision for a high resolution inset (Thomas & Geltmacher, 1993). Although not specifically designed for flight
simulation, a system has been proposed for low-cost head-mounted AOI displays with the following specifications: an instantaneous field of view of 50° (horizontal) x 40° (vertical) and a resolution of 4.69 arc min, and an inset size of 12.5° x 10° and a resolution of 1.17 arc min (Yoshida, Rolland, & Reif, 1995a, 1995b)

Simulators have also been in use in the Navy. The Visual Display Research Tool (VDRT), located at the Visual Technology Research Simulator facility at Orlando Florida, uses a helmet-mounted laser projector to paint imagery on the inside of a 20 foot diameter dome. Two channels of imagery are transmitted to the helmet projectors via fibre optics links by a digital image generator (GE Compuscene 1) and a photo-based image generator (LTV). The photo-based image generator can provide photorealistic imagery at 30 Hz (Dalton & Deering, 1989). One video channel is used for the instantaneous field of view, which is 140° (horizontal) x 100° (vertical). The other video channel is used for a high resolution AOI inset, which is 27° (horizontal) x 24° (vertical). A 5° blending region is used to soften the transition between the two regions. The resolution of the AOI is approximately 3.5 arc min, and the resolution of the instantaneous field of view is approximately 18 arc minutes. Note that the AOI resolution is not enough to detect or recognize objects at real-world distances, nor is the peripheral resolution adequate for portraying the background and formation aircraft. Measured luminance is 10 fl. Contrast ratio is not given. Both the AOI and the instantaneous field of view can be either eye- or head-slaved (Browder, 1989).

Turner (1984) described an intriguing application for gaze contingent flight simulators. As pilots perform high-G maneuvers, their visual fields shrink progressively (the “tunnel vision” phenomenon), eventually to zero, if the G-levels are sustained beyond the pilot’s tolerance. Flight simulators with gaze contingent displays can simulate this situation. In fact, by measuring the size of the visual field loss for various G-levels for each pilot prior to entering the simulator, it may be
possible with some systems, such as the NASA/TI Programmable Remapper (Juday, Smith, & Loshin, 1992) to tailor the visual field loss to the pilot operating the simulator.

*Indirect vision for pilots.* Indirect vision systems are useful in situations where direct vision is poor or impossible. These include vision in low-visibility conditions (e.g., night operations) and in future aircraft designs with “windowless cockpits.” Helmet-mounted displays and infrared cameras enable Apache and Cobra helicopter pilots to perform operations in low-visibility conditions, such as at night. The requirements for these displays are similar to those in flight simulation; pilots need high resolution for target detection, identification, and orientation, and wide fields of view for maneuvering, combat, and tactical formations with other aircraft. However, there is an additional requirement for see-through displays, so that pilots can also “see with their own eyes”. The solution has been to provide pilots with imagery from a head-slaved forward looking infrared radiation (FLIR) camera mounted at the nose of the aircraft in helmet-mounted displays. Pilots report having considerable difficulties using these devices, and these difficulties can be traced in part to the small instantaneous field of view of the display (40° horizontal x 30° vertical) and head slaving delays and errors (Grunwald & Kohn, 1994). Kooi (1993) suggested a binocular configuration of this head-mounted display, in which one eye receives low resolution peripheral information and the other eye receives high resolution foveal information. A distinct frame separates the two regions. Kooi reports good fusion between the two images, and relatively high comfort ratings from observers testing a version of this display. However, viewing times were limited.

The trend in high-performance aircraft design is toward highly swept, low drag forebody designs inconsistent with forward-looking windows. A helmet-mounted display system has been developed and tested in a Boeing 737 test plane with head-slaved cameras mounted in the nose (Rolwes, 1990). Pilots were able to land the aircraft with reasonable accuracy with indirect vision
alone. However low resolution was cited as a shortcoming of the design which may have decreased landing accuracy and increased pilot apprehension. A foveated display scheme could remedy this shortcoming.

**Remote vehicle control.** Remote piloting of vehicles allows for operation in hostile environments, such as in combat, underwater, or in space. The nature of these applications tend to restrict remote vehicles to low-bandwidth communication links. Two possible high-bandwidth links for remote teleoperation of vehicles are line of sight microwave and fibre optic links. Line of sight links are typically not feasible because they put the vehicle in exposed conditions (especially bad for combat applications), and because the link can be lost in rugged terrains. Fibre optic links on the other hand restrict range and are prone to breakage. Thus most remote vehicle applications require a low bandwidth radio link (Weiman, 1994).

These low-bandwidth radio links must carry a lot of information; the video requirements of remote vehicles are heavy. Remote driving requires both a wide and deep field of view, and enough resolution to be able to discern textures and identify objects. Operators are not comfortable operating a Jeep Cherokee with a 40° field of view system, especially turning corners. Confidence increases with an expansion of the field of view to 120° (McGovern, 1993). Resolution is important when there are many sizes and types of obstacles to be avoided and for off-road operation for route-finding. Furthermore, remote driving requires colour for contextual discrimination, for example road vs. sky. Also, the difference between dirt and asphalt is important for driving, but is not apparent in a black and white display (McGovern, 1993). Relatively fast frame rates are required to maintain good optic flow depth perception. Frame rates of at least 10 fps are required for perception of smooth movement (Weiman, 1994).
Juday and Fisher (1989) proposed using a foveated display scheme using the NASA/TV Programmable Remapper. High display resolutions are transmitted in the centre of vision and low display resolutions are transmitted elsewhere. Their proposal called for joystick control of the high resolution region, but they suggested the use of an eye tracker for future implementations. Their display algorithm produced very blocky images in the periphery. The advantage of their system is that the programmable remapper can create mappings of retinal acuity which are tailored to the individual user. In fact, they have even used their system to restore vision to people with scotomas, or holes in their visual field (e.g., Ho, Loshin, Barton, & Juday, 1995; Juday, Barton, Johnson, & Loshin, 1994).

Weiman (1990, 1994) used a combination of log polar mapping, which is a form of foveation, so that detail in the centre is preserved and peripheral details are compressed, edge coding, and a conventional compression technique to achieve very large compression ratios. In the example Weiman gives in his 1994 paper, he starts with a 5.9 Mbit image (512 x 480 pixels x 8 bits x 3 colours) and ends up with a 3.5 Kb image, a 1700:1 compression ratio. With 3.5 Kb per frame, a 30 fps video sequence requires 105 Kb per second of digital data channel. A 10 fps video sequence will require 35 Kb per second, which is well within the range of radio capacity links.

DePiero, Noell, and Gee (1992) used a Laplacian image pyramid method of compression in their foveated display scheme. Their system has been implemented in a high-mobility multipurpose vehicle, and supports teleoperation up to speeds of 15 mph on a moguled dirt track. Compression rates per frame ranged from 105:1 to 145:1, depending on the content of the scene. The frame rate varied with compression ratios, ranging from 3 to 6 Hz., which is slower than is required for perception of smooth motion.

*Teleoperation: Remote Manipulation.* Remote manipulation enables the performance of dextrous tasks in hazardous or inaccessible environments. For example, in fire fighting, defusing
mines or bombs, underwater or space maintenance and repair, and nuclear reactor inspection human skills are required because machine technology cannot deal with such unstructured environments.

Requirements for remote manipulation displays differ for specific tasks but in general they do not require wide fields of view; often they require magnification (narrow field of view), high resolution, and shallow depth of field (Weiman, 1994). Furthermore, unlike in remote vehicle operation, context is usually stable and understood, so bandwidth is better utilized by improving black and white resolution at the expense of colour.

Head-slaved helmet-mounted display systems (e.g., Pretlove & Asbery, 1995; Tharp et al., 1990), “virtual window” telepresence systems (e.g., Cole, Merritt, Coleman, and Ikehara, 1991; Tharp, Hayati, & Phan, 1994), and a gaze controlled microscope (Charlier, Sourdille, Behague, & Buquet, 1991) have been developed. No implementations of foveated display systems for remote manipulation have been found, possibly because of the smaller fields of view required in telemanipulation than in remote piloting of vehicles and in flight simulation.

*Image Filing Systems*. Image filing systems store and index gigantic volumes of images. Compression is required to reduce the size of image files to a manageable level, for both storage and transmission. Sorting through images, especially from remote locations over bandwidth-limited communication channels, is most efficiently achieved via progressive transmission systems, so that the user can quickly recognize unwanted images and terminate transmission early. A predictive foveated image progressive transmission scheme developed by Tsumura et al. (1996) has already been described. A real-time version has been suggested by Bolt (1984) and Levoy and Whitaker (1990). In Bolt’s scheme, the observer’s gaze position is tracked, and the areas where the observer is looking in the scene are transmitted first. Levoy and Whitaker refer to this as the “sweet spot”. As the observer scans across the image, a trail of sweet spots is left behind, and if the observer fixates on
one spot, the sweet spot grows until it encompasses the entire image. Bolt's method is more appropriate than the predictive scheme of Tsumura et al. for images in which the gaze area is not known. This method could also be applied to video if there are large regions of the image which are relatively static.

*Video teleconferencing.* Video teleconferencing is the audio and video communication of two or more people in different locations. The video sent in teleconferencing is highly structured (Maeder, Diederich, & Niebur, 1996) in that usually the transmitted image consists of a face or head-and-shoulders, and the moving parts of the image are the eyes and mouth, which along with the nose, comprise the most looked-at areas of pictures of faces (Spoehr, 1982). Typically there is only one user at a time at each node. Often, especially when used by the average person, teleconferencing will occur over standard low-bandwidth ISDN communication links (64 or 128 Kb/sec). Thus teleconferencing is a prime candidate for either real-time or predictive foveated display schemes. Predictive schemes are more feasible for widespread use because they require no eye tracking equipment and they allow multi-user operation, and because most of the information relevant to viewers is in the face area, especially the eyes, nose, and mouth. These areas can be transmitted in more detail and updated more frequently than the rest of the image. A real-time gaze contingent foveated display scheme has already been successfully implemented for still images of faces (Kortum & Geisler, 1996), as has a predictive foveated progressive transmission scheme (Tsumura et al., 1996). However no foveated displays have actually been implemented for teleconferencing.

*Medical.* Medical applications have been saved for last because despite paucity of foveated display implementation in this field, they offer rich possibilities of helping to unlock the vast untapped potential of such exciting new fields such as telemedicine, computer-assisted surgery, and picture
archiving and communications systems (PACS). Requirements for computer imagery in medicine vary widely from application to application.

Telemedicine is the delivery of health care through a combination of telecommunications and multimedia technology with medical expertise (Cabral & Kim, 1996). Improved access to medical expertise can be achieved by allowing doctors to remotely examine patients. Some of the roadblocks to telemedicine have been legal and economic. There is also the basic problem of infrastructure. The areas which can most benefit from telemedicine, rural and remote areas, often have a severe lack of high-bandwidth telecommunications networks. There are many types of telemedicine, each with their own communications requirements. Teleconsultation allows rural doctors to get a "second opinion" from a remote specialist to confirm a diagnosis or help the rural doctor to arrive at one. Basic synchronous videoconferencing is required, and some loss of video quality is acceptable (Cabral & Kim, 1996). Telediagnosis is similar to teleconsultation except that the primary diagnosis is made by the remote doctor. This distinction means that no significant loss of video quality is acceptable.

Medical video teleconferencing could potentially enjoy the same benefits from foveated displays as in normal video teleconferencing, as discussed earlier. For images of patients used for diagnosis, real-time foveated display techniques, especially progressive transmission (e.g., Bolt, 1984), could potentially be used to reduce bandwidth to useful levels.

Up until very recently, predictive foveated display techniques were neither feasible nor acceptable for transmission of patient images for telemedicine, because of the difficulty involved in selecting the regions of interest to transmit. However, a group of Greek researchers (Panagiotidis, Kalogeras, Kollias, & Stafylopatis, 1996) have developed a neural network which can be trained to select the regions of interest (ROIs, which will be coded with maximum precision) and the background regions (which will be compressed using a tolerably lossy scheme) with tested accuracies
of 96-98%. The system can support more than two levels of interest, and it lets the user specify the reconstructed image quality for each level of interest to maximize either image quality or compression ratios. Obtained compression ratios for their system, called the Region of Interest JPEG (ROI JPEG) system, vary with image quality settings and range from about 10:1 to 15:1, compared to baseline JPEG compression rates of about 7:1 to 12:1. The system has been tested for a telemedicine system for dermatology examinations, using conventional low-bandwidth telephone lines between remote health care centers and central hospitals. Regions of interest in dermatology images include bruises, burns, and tumours.

Computer-assisted surgery includes telesurgery, surgical simulation and planning, augmented surgery, and surgery education. Telesurgery involves remote manipulation of the patient. This includes laparoscopy, in which the surgeon operates on the patient through tiny incisions to remove internal organs. Visual feedback is provided by a video monitor, and the surgeon holds the handles of long instruments but cannot see the tips of the instruments, so in essence, the surgeon is performing teleoperation. The principles remain the same whether the surgeon is present in the room or is thousands of miles away. With the help of state-of-the-art telecommunications links and teleoperated robots, intercontinental telesurgical laparoscopic techniques were performed in 1993 (Rovetta et al., 1993; Rovetta et al. 1996), just four years after the first laparoscopic gall bladder operation in 1989. The implications are enormous, for both the teleoperated and minimally invasive nature of this surgery. Surgeons of the future may be able to routinely operate on patients in remote areas, limited only by the accessibility of the patient by mobile units such as ambulances or trucks equipped with telesurgical robots. This application may be a long way off, but already telecommunication technology is at a level where surgeon performing an operation can collaborate with a colleague in another city on the same operation through teleconferencing. Minimally invasive procedures have
already produced huge patient benefits, including a reduction of morbidity, patient recovery and rehabilitation time, and hospital length of stay (Blumenfeld, 1995), and in more human terms, no scars, essentially no pain, immediate return home and return to work in a week or less (Satava, 1995).

Perhaps the most visionary seer of these exciting new possibilities in medicine, Richard Satava (1995), sees the video display, as the interface to vast amounts of information, as the focal point of all these new technologies. Telemedicine and collaboration during surgery have been described. One of the most exciting possibilities, however, is the ability to fuse many different digital images, such as the patient’s computerized tomography (CT) or magnetic resonance imaging (MRI) scans with real time video images, effectively giving the surgeon “x-ray vision”. Yoshida et al. (1995a), suggested that one method of accomplishing such fusion is to present CT, MRI, or ultrasound scans inside gaze-contingent insets, with the “real” image in the background. Another exciting application combines surgical simulation with computer-assisted surgery: before performing a surgical procedure, a surgeon could plan a strategy and practice on a virtual patient, then flip a switch and begin operating on the real patient with the same workstation. Already there are many such implementations in place. Satava gives several examples: 3-D MRI scans can be fused with video images of real patient brains, to allow the surgeon to see the tumor in the context of the actual brain, CT scans of a child with bony facial deformities have been used in a surgical simulator to correctly rearrange the bones of the face to allow repeated practice of a difficult procedure.

Medical education can also reap huge benefits from surgical simulation. Just as pilots can perform hundreds of take-offs and landings in simulators before actually climbing into a real aircraft, simulators can also allow surgical residents to practice a surgical procedure hundreds of times before they see their first patients. Simple laparoscopic surgery simulators have already been developed for
training. Currently the graphics are primitive and cartoon-like, but surgical simulation is in its infancy. As the field develops, the graphical needs will increase to a point where foveated display techniques become useful and perhaps even necessary. Levoy and Whitaker (1990) have already demonstrated the utility of gaze contingent volume rendering of medical data sets, namely MRI data sets. To conclude, medical applications of advanced computer displays continue to grow, and it may not be long before foveated displays form an important part of the health care system, just as they do in aviation.

**Human Factors of Foveated Displays**

Human factors and psychophysics work on foveated displays can be divided into two categories: the first involves basic overall acceptability or noticeability of individual foveated display schemes, perhaps with some manipulations of a few key parameters of that display, such as the field of view, level of resolution, delay, and size of the high resolution insert in AOI displays. The second involves evaluation of the effects of display parameters on performance on some task, such as a visual search task or a flight simulation task. The first category is required to demonstrate the utility of the foveated display scheme in general, and the second is required to optimize it. Most work of the second type has been done on AOI systems, which is not surprising because they are the only foveated displays which have been in widespread use for a long time.

**Acceptability of Display.** The simplest way to demonstrate the acceptability of a foveated display is to simply show viewers a foveated display and ask them to rate how acceptable it is. For example, evaluation of Kortum and Geisler's foveated image coding scheme (1996), as described earlier, consisted of simply asking whether viewers were aware of the variable resolution in the image. Stelmach and Tam (1994) had participants judge the subjective image quality of video
sequences that were processed using a gaze contingent or a non-gaze contingent method and reported only modest benefits of the gaze contingent method.

Tsumura et al. (1996) asked participants to evaluate the image quality of seven photographs that were blurred inside the gazing area of the image, outside the gazing area of the image, and across the entire image. For all but two photographs, the version that was blurred outside of the gazing area of the image was rated most positively. The two exceptions were pictures of fruit and a chart. These images had a lot of detail spread across the image, and many areas to gaze at. Tsumura and Endo also evaluated their coding scheme based on the performance of viewers in an image recognition task, and found that viewers were able to identify mismatches faster using their foveated progressive transmission scheme than using the conventional method.

Psychophysical techniques yield more objective quantitative data. Geri and colleagues (Geri & Zeevi, 1995; Geri, Zeevi, & Vrana, 1994) processed photographs using a function that could yield a set of images with a range of central and peripheral blurring, and gradients of blurring from the centre to the periphery. Images were presented to viewers so that one half of the image was blurred and the other half was not. Presentation times were extremely short (167 msec) to prevent eye movements. The central 16° of the photographs were removed because the display was not capable of the resolutions required at the fovea. A two-alternative forced choice procedure was used to determine if viewers could identify the blurred half of the image. If viewers performed better than chance level (50%), it was concluded that they could perceive the blur. The beauty of their technique is that it allowed them to assess perception of central blur, peripheral blur, and blur gradient. Unfortunately they did not give a measure of display resolution, so it is impossible to interpret their results in absolute terms, but they were able to demonstrate that it is possible to generate a set of blur functions which are indistinguishable from the original image. Thus it provides strong evidence that
resolution can be degraded in the periphery at no perceptual cost. This study should be repeated with other foveated display techniques and with more central vision included.

As a point of interest, there do not appear to be any studies which have compared a foveated display with a nonfoveated display requiring the same bandwidth, which would appear to be an appropriate comparison given that the bandwidth problem is the reason for implementing foveated displays in the first place.

Turner (1984) tested the acceptability of the eye-slaved AOI system using performance on a task similar to those used in flight-simulation as the main dependent variable. His task was designed so that high resolution would not add any performance benefits. Thus if the AOI causes any performance decrements, they would not be masked by improvement due to the increased resolution within the area of interest. He had subjects follow a ground path through a canyon under three different conditions: with an eye-slaved high resolution AOI in a medium resolution background, a medium resolution background with no AOI, and a low resolution background with no AOI. The medium and low resolution background-only conditions were used to ensure that the task was resolution independent, and participants’ performance, as measured by deviation from the ground path, confirmed this. Turner found that the AOI caused no decrements in performance on the path-following task, nor did the participants subjectively feel that the AOI hindered their ability to perform the task.

Browder (1989), compared pilot performance on a simulated dive-bombing task using the VDRT, which, as described earlier, is an AOI dome system which can be either eye-slaved or head-slaved. The task was to fly a curved and level path until the run-in line, and the pilot rolled into the dive-bomb line and used his gunsight to determine the bomb release point. Browder reported no significant differences between pilot performance on the task using the two displays. Furthermore,
there were no significant differences between eye-slaved and head-slaved operation (Browder did not include an eye- and head-slaved control mode) using the VDRT, although individual pilots tended to perform better in one mode or the other. There was wide inter-subject variability with regard to the helmet and preference for eye-tracked or head-tracked mode of operation.

Some flight simulators have the capability for a target-slaved AOI inset. These are based on the idea that the target will always be the focus of interest for the pilot. These are generally not well accepted because of the artificial nature of the target-tracked AOI. For example, target detection is never an issue because it is “highlighted” at all times (Spooner, 1982). Furthermore they prevent effective practice in higher-level tactical skills and judgments (Turner, 1984).

Evaluation of AOI parameters. The focus of this section will be on the evaluation of AOI parameters, as these form the bulk of research on parameters of foveated displays. A number of parameters of AOI displays have been identified as important to simulator performance. Instantaneous field of view, AOI size/AOI resolution, background resolution, image generator throughput delay (time from head- or eye-movement to update of display), size of the blending region, and “popping”. Popping has been defined differently by different researchers. Spooner (1982) and Berbaum (1984) describe popping in terms of an abrupt change in the level of detail of an object. Recall that AOI insets can have a higher level of detail than the background. If a house is in the background at a low level of detail (for example, a featureless block), and the AOI suddenly moves on top of it, it could suddenly have a higher level of detail, and windows and doors will suddenly “pop” into view. Others, such as Hodgson, Murray, and Plummer (1993) define popping as the sudden appearance of the AOI at the fixation point after a delay following a saccade that takes the eye outside of the AOI. Instantaneous field of view has already been discussed in the section on flight simulators.
AOI size and AOI resolution tend to be identified as the most important aspects of AOI displays. They also tend to be grouped together because they form a mutually antagonistic pair, just as they do in the instantaneous field of view. A finite amount of computer image generator power is available for the AOI inset, and thus increasing one necessarily entails a decrease in the other. This relationship can also make it difficult to experimentally manipulate one parameter independently of the other, because the combination of a wide field of view, high resolution inset is simply not possible with the display equipment.

Turner (1984) performed one of the first empirical tests of the effects of AOI parameters on flight simulation tasks. He measured the effects of AOI size/resolution and delay on a path-following task and a concurrent visual search task. He used three levels of AOI: 12 deg, 18 deg, and 28° in diameter, with corresponding resolutions of 2.0, 3.0, and 4.75 arc min/line pair. For the 18° AOI, he also used five levels of delay, 130 msec, 180 msec, 230 msec, and 280 msec, and nominal (which he does not define or plot in his graphs). Participants deviated less from the path and missed less targets in the 18° (medium resolution) condition than in the other two conditions. The smaller AOI had increased resolution, but was also more visible, as evidenced by participant comments such as "I like being able to see the targets more clearly but the edges of the AOI are more visible." Turner did not report whether the size of the blending ring was scaled to the size of the AOI inset. If so, consider this: the blending ring for the small AOI already had the steepest blending gradient between the high and low resolution areas (2.0 to 11.0 arc min/line pair from inset to background). If the blending ring was also scaled to the angular size of the AOI, then the blending ring would also be the smallest for the small AOI. Compound this with the fact that the retinal location of the edges of the small AOI have the highest acuity, and it becomes quite understandable that the edges of the small AOI were more visible than in the other conditions. This raises the question of whether it is really the size of
the AOI which impaired performance in the 12° AOI condition, or actually the increased visibility of
the edges (assuming that the increased resolution can only help performance). In any case, the 18°
AOI appears to be a better mix of size and resolution than the other two AOIs, because the 12° AOI
is too small and/or has highly visible edges, while the 28° AOI has poor resolution. Turner also
reported consistent progressive decrements in both path following and target identification tasks with
increasing levels of throughput delay, that most subjects were aware of the delay by 230 msec, and
that there were slight degradations of performance in both tasks with 180 msec delays and distinct
performance degradations at 230 msec. However, according to his data, the performance
decrements did not actually reach statistical significance until 280 msec.

Wells and Venturino (1990) examined the effects of inset size on performance on a search and
shoot task, and head movements. The inset was positioned at the centre of the instantaneous field of
view, and both were head-slaved. Strictly speaking, the insets, which ranged from 20° square to 120°
(horizontal) x 60° (vertical; the entire instantaneous field of view) were not AOI insets because
resolution and level of detail were the same inside and outside the inset; the difference was that target
and threat stimuli were visible only in the inset. Targets were spread randomly across the gaming
area and turned into threats, one at a time. The task was to find and shoot (by aligning a cross
centered in the inset) the threat before it shot the participant. Reducing inset size resulted in
decreased performance on the task, especially when the task was complex (when there were many
targets), increased head movements and decreased head velocity.

Warner, Hubbard, & Serfoss (1992) measured the effect of AOI size/resolution on threshold
detection distances for artificial stimuli (cylinders and stripes) head-slaved, LFOVD system. They
used two AOI conditions, a small AOI condition, characterized by a 26.5° (horizontal) x 21.5°
(vertical) field of view, a resolution of 1.5 arc min/pixel, and a 2.5° wide blending ring, and a large
A0I condition, characterized by a 40° (horizontal) x 30° (vertical) field of view, a resolution of 2.5 arc min/pixel, and a 5° wide blending ring. They found that detection thresholds were greater for the smaller (higher resolution) AOI. This finding was confirmed in a follow-up experiment by Warner et al. (1993) which also demonstrated larger head movements and more direction changes in head movements with the larger AOI. The former finding is opposite to those reported by Wells and Venturino (1990), which the authors attribute to an artifact of the method and the relationship between size and resolution of the AOI.

To find studies on AOI size independent of resolution, one must look to papers by authors outside of the flight simulation field. Saida and Ikeda (1979) artificially restricted participants’ field of view to a square “window” centred on the fixation point, in order to study the “useful visual field”, or the visual field beyond which no useful information for picture recognition can be extracted. They found that recognition performance for pictures increased with increases in field of view up to a certain critical size beyond which no further improvements occurred. Their experiment suggested that the critical size was not a fixed visual angle, but rather a function of the size of the picture being viewed. Shioiri and Ikeda (1989) improved upon the technique by degrading rather than blanking out the area outside of the window; in other words, they used an AOI technique. They measured the effects of window size, the level of peripheral degradation, and exposure duration on scene encoding and eye movements. They found that the critical window size, or the window size beyond which no improvement in picture processing occurs, increases with the level of peripheral degradation. Useful resolution, or the image resolution that is used at each eccentricity for picture recognition, decreased monotonically with eccentricity. They also found that saccades were shorter when the eye-contingent window was smaller than the critical window size. Similar methods could be used in flight simulation...
to determine the peripheral resolution that the pilot is able to use, and to optimize the size of the AOI.

Stark et al. (1992) and Zhou, Ezumi, and Stark (1993) investigated the effects of field of view size on visual search patterns. Using head- or mouse-controlled windows, they found that with fields of view restricted to small rectangular areas, participants used highly regular, systematic patterns whose coarseness or fineness was determined by window size. When participants had large or unrestricted fields of view, they switched to irregular search patterns. Thus window size affects the search patterns viewers adopt.

Hodgson, Murray, and Plummer (1993) investigated the effects of AOI size, delay, and saccade size on eye fixations in a simple oscillating saccade task. Three AOI conditions were used, a 50° square AOI, a 30° square AOI, and a control condition in which the entire screen was in focus. Three delay conditions (30, 120, and 230 msec), and three saccade sizes (7.5°, 15°, and 30°) were used. Subjects simply made lateral saccades back and forth between two fixed targets in time to a buzzer, and the length of the fixation on the second target was measured. They found that fixation lengths increased with delay, but were not affected by the size of the AOI or the saccade length. They were concerned with popping, but the largest saccades, which would take the eye outside of the AOI, did not show any increases in fixation time compared with the shorter saccades.

Browder (1989) tested the effects of several factors on a concurrent air pursuit and visual search task in the VDRT, including eye-slaved vs. head-slaved operating mode (an eye- and head-slaved mode was not included), AOI blend ring size (2° or 5°), and computer image generator throughput delay (142 or 192 msec). Eight experienced pilots were given the task of pursuing a simulated F-18 lead aircraft while simultaneously detecting and identifying a series of square or circular targets. Targets were detectable in the background, but identifiable only in the AOI.
Browder reported no "overall" significant effects of any of the factors. What this means exactly is unclear, because Browder does not provide any data. The wording "Also unaffected from a practical viewpoint were the trials with the extra 50 milliseconds of time delay..." suggest there may have been some trend toward performance decrements, which would agree with Turner's (1984) data, which showed nonsignificant decrements in performance at 180 msec and 230 msec but showed a consistent trend which reached significance at the 280 msec delay level.

Finally, a pair of studies at the Naval Training Equipment Center investigated the properties of the blending rings at the borders of the AOI. The first, by Baldwin (1981) used a slide projector with a variable resolution filter to project imagery with a central high resolution region set in a low resolution background, and with a blending region in between. The width of the blending ring was modified by exchanging the filter. The filter was slaved to the viewer's horizontal eye movements so that the AOI was always centred at the viewer's fixation point, and various delays could be inserted between the eye movement and scene update. Most observers reported that small or non-existent blend regions were very distracting and objectionable. Baldwin reported that an inset width of 25° with a 5° wide smoothly varying blending region with a delay of 80 msec and an eye-tracker accuracy of ± 2.5° would cause noticeable, but not objectionable, perception of the borders of the AOI inset. Since the periphery was merely blurred in this experiment, it does not address the question of level of detail and popping.

Berbaum (1984) performed a set of psychophysical experiments on the perceptibility of popping. Popping can be described in psychophysical terms as the abrupt temporal onset and offset of high level of detail stimuli. By measuring contrast sensitivity to different sizes of gratings and temporal rates of change in contrast, Berbaum was able to obtain measures of popping threshold and ratings of popping. Based on the data, the author made two main recommendations for suppression
of the perception of popping: that a $5^\circ$ wide contrast ramp (blending region) is used between the AOI and the background, and that restrictions be placed on image database scene modeling so that external contours remain constant and internal contours (high levels of detail) are added or subtracted. For example, in going from a low level of detail to a high level of detail for a house, it is not advisable to suddenly add a roof or a yard, but it is acceptable to add windows and doors.

**AOIs May Interference with Target Detection**

The use of blending rings in flight simulator AOI displays has driven by user reports of objectionability of displays without such blending rings, and by an attempt to curb the popping phenomenon. Participants in Turner’s (1984) study reported preference for larger AOIs because of the small AOIs had more visible edges. If the edges of the AOI are salient, they could interfere with performance of peripheral tasks. Holmes, Cohen, Haith, and Morrison (1977) demonstrated that the mere presence of a foveal item which subjects are instructed to ignore resulted in a decrement in the performance of a peripheral task. The authors interpret the finding as a general interference effect; the foveal item draws the attention of the observer and thus interferes with processing of other stimuli in the visual field. The finding could also be interpreted as a “tunnel vision” effect, in which the foveal item somehow restricts the size of the visual field. Either way, this finding has strong implications for AOI displays. If the edges attract the attention of the observer (as they evidently did in Turner’s study), a decrement in performance can be expected for other stimuli in the visual field.

Most work in AOI displays has relied on subjective evaluations of system performance during relatively uncontrolled tasks, or performance measures of rather complex flight simulation-type tasks. Although a few studies have demonstrated effects of AOI size on head movements (i.e., Warner et al., 1992; Warner et al., 1993; Wells & Venturino, 1990), only Hodgson et al. (1993) and Shioiri and Ikeda (1989) have investigated the effects of different AOI parameters on eye movements. Also, the
latter two studies are the only ones to directly compare performance on AOI displays and full-field high-resolution displays.

The terms “window” and “AOI” will be used interchangeably throughout the rest of this thesis. The current experiments investigate the effects of salient window edges on target detection tasks. In the first experiment, the effects of having gaze contingent windows on the detection of moving targets are measured. The second experiment is an investigation of the effects of window size on the detection of stationary targets.
Experiment 1: Moving Targets

The first experiment explores the effects of gaze contingent windows on the detection of moving targets in complex scenes. Because in the standard AOI procedure, the background is lower resolution than the AOI, any decrements in performance which occur as a result of having a window present could be attributed to the degraded quality of the background image. Thus two types of window are used: the standard resolution-defined window which is high resolution in the window and blurred in the periphery, and a luminance-defined window condition in which a window is present, but the quality of the background image is preserved. Two non-window conditions are used, one in which the display is uniformly high resolution, and one in which the display is uniformly blurred. The task is to detect target stimuli moving across the screen. It is hypothesized that the presence both types of the windows will impair performance on this task, indicating that the effect of having a salient window can impair peripheral task performance independently of resolution differences between the two regions. Furthermore, it is hypothesized that the blurring of the background in both the window present and window absent conditions will impair performance, indicating that the degraded quality of the background imagery also impairs detection performance.

Method

Participants

Participants were 60 undergraduate students (22 males and 38 females, ages 19 to 56) at the University of Toronto who received credit in an introductory psychology course for participation. All participants had normal or corrected-to-normal vision. 30 of the participants had previously participated in Experiment 2.
Apparatus

The apparatus consisted of four IBM-compatible computers and headband-mounted cameras, optics, and cables. Gaze position was recorded at 250 Hz using the Eyelink gaze-tracking device (Figure 2), consisting of a headband with adjustable infrared cameras used to monitor pupil position of each eye using the dark pupil method. A headband-mounted camera monitored 4 infrared LED markers mounted on the corners of the display unit, and this information was used to compute head position, which was used to determine gaze direction. The headband and associated cameras and electronics weighed 500 g and was evenly-counterbalanced. Viewing was binocular and data collection was monocular, usually from the left eye (data from the right eye were used occasionally when calibration was significantly better for that eye). The eye tracker host computer, a 100 MHz Pentium PC compatible, tracked the participant’s gaze position and displayed it in real time using gaze cursors on the experimenter’s monitor, out of the participant’s sight. The eye tracker host computer also supplied gaze position information to the display computer. The second and third computers, 66 MHz 486-DX PC compatibles, were used to concurrently play the processed (blurred or high luminance) and unprocessed videoclips, respectively (see Stimuli & Design, below), and feed the video to the display computer. The display computer, a 100 MHz 486-DX PC compatible [check!] controlled stimulus presentation, integrated incoming video signals, displaying one channel as background imagery and part of the other channel as a window at the participant’s point of gaze and displayed combined stimuli in a 17” monitor in front of the participant. The display was positioned at a viewing distance of 60 cm so that the field of view was 30° (horizontal) x 22.5° (vertical). The total system throughput delay (time it takes from the eye movement to a change in the display) was 20 msec. System specifications are available on the World Wide Web (SR Research, 1996).
Stimuli

Stimuli were full-colour videoclips containing a target moving against a moving background. The clips were approximately 3 sec long each, shown at a rate of 30 fps at 320 x 240 pixels. The average luminance was about 60 fL. There were two versions of each clip: non-blurred and blurred. The non-blurred videoclips had an effective resolution of about 11 arc min/line pair, and an average luminance of about 60 fL. The blurred videoclips were produced using a process equivalent to a Gaussian filter (0.5 cycles/deg) to blur both the target and the background. The effective resolution of the blurred videoclips was about 85 arc min/line pair using a -6 dB criterion, and the luminance was unchanged. The blurred target was still discriminable from the background, although sometimes only by target motion.

Resolution-defined windows were created by combining blurred and non-blurred versions of the same clip, running simultaneously and synchronized in time. All windows were roughly circular and 6° in diameter, and centred at the participant’s gaze position, as measured by the Eyelink equipment. The edges of the window remained sharp; there was no blending region between the window and the background. Resolution-defined windows were created by displaying the non-blurred version of the videoclip inside the window, and the blurred version of the videoclip in the background. Luminance-defined windows were created by displaying the non-blurred version of the videoclip across the entire screen, but selectively increasing the luminance inside the window by 20%.

The target, a 1° white ring, moved in a straight line at a constant velocity of approximately 8°/sec. There were four directions of target motion: vertically down the left side, vertically down the right side, diagonally down and to the left, and diagonally down and to the right. The backgrounds were 16 videoclips of mountainous and desert terrains shot from a moving helicopter, some from a forward-looking vantage point, which contained optic flow cues for forward self-motion, and some
from directly above looking down. All background motion was from the top to the bottom of the screen, but was never the same direction and speed as the target motion.

**Design**

A 2 x 2 (Blurring x Windowing) design was used. There were two levels of Blurring, low (blur) or high (non-blur), and two levels of Windowing, window and non-window, for a total of four conditions. The blur window and non-blur window conditions have been previously described as the resolution-defined and luminance-defined windows, respectively. The display was uniformly blurred in the Blur Non-Window condition, and uniformly non-blurred (and normal luminance) in the Non-Blur Non-Window condition.

The two Resolution conditions and the two Visibility conditions were counterbalanced with the four types of target motion for a total of 16 combinations of resolution, visibility, and target motion. Each combination, and each background scene, appeared in random sequence four times per block, for a total of 64 trials per block. Three blocks of trials were used in the experiment.

Before the experiment began, participants were given a practice block of eight trials, in which each combination of Resolution and Visibility appeared twice, and each target motion appeared twice. Following the experiment, many of the participants were asked to comment on various aspects of the experiment.

**Procedure**

*Setup and Calibration.* Participants were seated comfortably and fitted with the Eyelink headgear. The eye cameras were set up under each eye so that the pupil was clearly defined at all eye positions, and headband was adjusted so that the four display LED markers were evenly centred in the display from the headband camera. The calibration procedure was explained to the participant (see Appendix 1 for full instructions): the participant was asked to look at a fixation target on the
screen and follow it wherever it appeared on the screen. They were told it would appear randomly and asked not to try to anticipate where it would appear next, and also to try to minimize head and body movement during the procedure. The fixation target was presented at each of 9 locations on the screen in random sequence, and the position of the pupil was recorded for each screen location. An autocalibration procedure was used so that when the target was fixated steadily for a period of 500 msec, the pupil position was recorded and the next point was presented. A few participants had difficulty maintaining a steady gaze for 500 msec at each target location; a manual calibration procedure was employed under these circumstances so that the experimenter could control the calibration procedure. Participants were told to repeat the procedure in the next phase, which was the validation phase.

A fixation target was presented in random sequence at the same 9 locations, and the pupil positions during the calibration phase (estimated gaze position) and the validation phase (actual gaze position) were compared to yield an error measure. If the average error was greater than 0.5 degrees, calibration was repeated. Over the course of the experiment, head or body movements could cause slight shifts in measured gaze position. Rather than performing frequent recalibrations, a process of drift correction was performed before each trial. The participant fixated a central dot, and the measured error was used to compensate for any drifts of calibration for gaze position calculations during that trial. Between trials, the experimenter had the option of validation and recalibration. Recalibration was performed when the participant's gaze cursor consistently deviated from the target cursor, but this rarely occurred.

**Trial Sequence.** A trial sequence began with a fixation dot on a blank screen. The participant fixated the dot, and the experimenter initiated the trial when the gaze cursor stabilized. This initial fixation served the double purpose of ensuring the participant started each trial looking at the centre
of the screen, and to perform a drift correction. The fixation dot disappeared, and after
approximately 0.5 sec the videoclip started. The task of the participants was to locate and fixate the
target as quickly as possible. When the videoclip ended, the fixation dot reappeared, and the next
trial began.

**Measures and Analyses.**

*Operational definitions.* One of the advantages of the Eyelink system is the wealth of data,
and large number of variables it can provide. Several measures of target detection performance were
used: latency to first saccade, latency to target acquisition, error after first saccade, duration of
second fixation, proportion of trials with a second saccade, number of saccades, and average saccadic
distance.

For the purposes of the analysis, a saccade was defined as any eye movement with a peak
velocity over a threshold of 25°/sec, and an amplitude of at least 1°. The latency to first saccade is
the time from video onset to the first saccade (in any direction). The latency to acquisition is the time
from the start of the videoclip (video onset) to the acquisition of the target, defined as the first full 20
msec period that the gaze position keeps within 2° of the target. Error after first saccade is the
distance in degrees from the point of fixation to the target at the end of the first saccade. Number of
saccades is the number of saccades required to acquire the target, and average saccadic distance is
the average magnitude of the saccades in degrees. The proportion of trials with a second saccade is
the proportion of trials in which the subject requires more than one saccade to acquire the target. On
these trials, another measure is used: duration of second fixation is the length of the second fixation,
provided the second fixation does not acquire the target.

*Dropped Trials.* Not included in the analyses were: trials containing a blink up to 100 msec
prior to video onset or target acquisition; trials in which acquisition occurred more than 2 sec
following video onset; trials in which error was greater than 3 deg in the first 100 msec after acquisition; and trials in which the average error was greater than 2 deg after acquisition. Any saccades within 100 msec before or 60 msec after video onset, were judged to be anticipatory saccades, and those trials were also removed. In total, 5.8% of the trials were dropped.

Results

The mean latency to first saccade is plotted by condition in Figure 3. A 2 x 2 (Blurring x Windowing) within-subjects analysis of variance of the latency to the 1st saccade revealed main effects of Blurring, $F(1,59)=245.84$, $p<.001$, and Windowing, $F(1,59)=138.70$, $p<.001$. The main effect of Blurring indicated that participants were significantly slower to make their first saccade when the background was blurred, and this was true both when a window was present, $t(59)=-10.55$, $p<.001$, and when a window was absent, $t(59)=-13.11$, $p<.001$. Furthermore, the amount of slowing of the first saccade produced by blurring was very nearly the same whether a window was present or not (52 and 50 msec, respectively). The main effect of Windowing indicated that participants were slowed by the presence of a window, and this was true for both the blur window $t(59)=-7.946$, $p<.001$ and the non-blur (luminance) window $t(59)=-9.43$, $p<.001$. The amount of slowing produced by the blur and non-blur (luminance) windows was about the same (38 msec and 36 msec, respectively). Mean latency to first saccade is a relatively pure measure of the effects of the window, because unlike the other measures, it is not influenced by the number of saccades required to acquire the target.

Latency to acquire the target, duration of second fixation, and proportion of trials with a second fixation before acquisition, are heavily influenced by the number of saccades required to acquire the target. All of these measures yielded the same pattern of results (see Figures 4, 5, 6, and 7, respectively), so only the latter measure will be reported here. The number of saccades required to
acquire the target is plotted by condition in Figure 7. An analysis of variance of the data showed main effects of Blurring, $F(1, 59) = 36.58, p < .001$, and Windowing, $F(1, 59) = 7.72, p < .01$, and no interaction, $F(1, 59) = 1.75, \text{ns}$. Thus both the presence of the window and the blurring of the background caused increases in the number of saccades required to acquire the target. However, unlike the mean latency to first saccade data, the data were not symmetric: the window effect was far smaller in the non-blur condition than the blur condition, and in fact a t-test revealed that the window effect in the non-blur condition was not large enough to reach significance, $t(59) = -1.31, \text{ns}$.

Average saccadic distance is plotted by condition in Figure 8. An analysis of variance showed main effects of Blurring, $F(1, 59) = 83.08, p < .001$, and Windowing, $F(1, 59) = 60.79$, and a Blurring x Windowing interaction which approached significance, $F(1, 59) = 3.84, p = .055$. The Blurring main effect indicates that blurring the background decreased average saccade size, and this is true for both the window present, $t(59) = 7.98, p < .001$ and window absent, $t(59) = 5.41, p < .001$ conditions. The Windowing main effect indicates that the presence of the window also decreased average saccade size, by about $0.6^\circ$ in the non-blur condition, $t(59) = 4.25, p < .001$, and by about $1.0^\circ$ in the blur condition, $t(59) = 7.29, p < .001$. Notice that the effect of the window was greater in the blur condition than in the non-blur condition.

Error after first saccade is plotted by condition in Figure 9. A $2 \times 2$ within-subjects analysis of variance showed a main effect of Blurring, $F(1, 59) = 23.55, p < .001$, but no main effect of Windowing, $F(1, 59) = .77, \text{ns}$. The Blurring x Windowing interaction approached significance, $F = 3.388, p = .071$. The main effect of Blurring indicates that blurring the background increased the distance of participants' gaze positions from the target after the first saccade, and this was true for both the window conditions, $t(59) = -4.021, p < .001$ and the non-window conditions, $t(59) = -2.524, p < .05$. Error after the first saccade was significantly higher in the blur window condition than the
blur non-window condition, $t(59)=5.24$, $p<.001$, but was about the same in the non-blur window and non-blur non-window conditions $t(59)=.441$, ns.

*Participants' comments.* Almost all of the participants noticed that the windows were gaze contingent. Also, a large number of participants described the windows as looking like “spotlights”.

**Discussion**

The goal of Experiment 1 was to determine what the effects are, if any, of gaze contingent windows on the detection of a moving target. As expected, both the resolution- and luminance-defined windows impaired performance on target detection.

Blurring the background of the clips caused impairments in target detection on several measures, including latency to the first saccade. It is reasonable to attribute increased reaction times and detection latencies on the reduced contrast of the target from the background in the blur conditions. Previous studies have shown that decreasing target contrast (by bringing the luminance intensity level of the target to closer to that of the background) causes increases in saccadic latencies (Brown, 1972; Haegerstrom-Portnoy & Brown, 1979). The symmetry across window and non-window conditions of the impairments produced in the latency to the first saccade suggest that the effect of blurring is independent of any effects of the window. The implication for flight simulators and other foveated display systems is that any impairment in target detection caused by the blurring of peripheral targets will be unaffected by the addition of a foveal window or AOI.

The data from the luminance-defined windows show that the presence of a gaze contingent window alone causes impairments in target detection, independently of impairments caused by degraded resolution in the periphery. Thus it appears that there is something intrinsic to the window itself that causes impairments. The major common element shared by the resolution- and luminance-
defined windows was the presence of well-defined edges at the perimeter of the window, so it is reasonable to attribute the effects of the window to these edges.

One possibility, especially with the small window sizes used in this experiment, is that the edges are highly salient, and tend to either shrink the visual field (the tunnel vision phenomenon) or attract the viewer's attention, causing decrements in performance for other stimuli in the visual field. This study replicates the finding of Holmes et al. (1977) that the mere presence of a foveal item is sufficient to cause a decrement in the performance of a peripheral task. While the participants in the current experiments were not specifically told to ignore the windows, it was obvious that their task on any given trial was the same whether a window was present or not. The inside-out attentional processing scheme suggested by Holmes et al. suggests that participants' attention is attracted first by the edges of the window, the foveal item, and that only when the participant is finished processing the foveal window can the peripheral target be processed.

A similar idea is that the increased latencies to the first saccade may also reflect a need to first disengage attention from the window in order to fixate the target. The "attention disengagement hypothesis" has been used to explain the "gap effect" phenomenon (e.g., Dorris & Munoz, 1995; Fischer & Weber, 1993), in which the removal of a foveal fixation point shortly before the appearance of a parafoveal or peripheral target enhances saccadic reaction times relative to a condition in which the foveal fixation point remains in view while the target appears. According to the attention disengagement hypothesis, in order to fixate a new target, one must first disengage attention from other stimuli in the visual field, including the fixation point. The removal of the fixation point disengages the viewer's attention prior to the appearance of the target, and if this disengagement occurs before the target appears, reaction time will be reduced because attention is already in a disengaged state. The windows in the current experiment were present for the entire
duration of the trials in which they appeared. Thus it may have been necessary for participants to first disengage attention from the window before they could make a saccade to the target. This possibility seems likely in light of comments made by participants in a pilot study which used only the resolution-defined windows; participants noted that it was difficult to wrench their eyes away from the window to the target because the high clarity and detailed area inside the window was much more interesting than the blurry, indistinct background, despite being fully aware that the window would follow their eye movements.

Before proceeding further, the possibility must be investigated that the increased latencies to the first saccade are due simply to the utilization of more conservative response strategies rather than impairments caused by the blurring and the windows. In other words, participants could be making a speed-accuracy tradeoff, and increasing saccadic accuracy at the expense of response speed. There are two reasons this is not likely. First, the highly symmetric pattern of differences in the latency to first saccade data suggest the operation of two independent, additive, and dissociable processes, blurring and windowing, rather than the shifting of response strategy differentially for different conditions. Second, the error after the first saccade data are not consistent with the speed-accuracy tradeoff interpretation. Error is inversely related to accuracy, so according to this interpretation error should be lower in the blur conditions and the window present conditions, but this was not the case. In fact, error after the first saccade was significantly higher in the blur conditions than the non-blur conditions, and higher in the blur window condition than the blur non-window condition. Thus a speed-accuracy tradeoff cannot explain the pattern of results obtained here.

There is an interesting asymmetry in the error of first saccade data in the effects of the window across blur and non-blur conditions. Namely, the presence of a window resulted in larger errors in the blur condition, but did not affect errors in the non-blur conditions. The latter finding
does not contradict Holmes et al.'s notion that the mere presence of a foveal stimulus can impair peripheral detection, because it is the time course of processing that is affected, not performance in general. That is, the window must be processed first, but once this is completed, other stimuli can be processed normally; the window only delays the detection of peripheral stimuli.

The impairment produced by the presence of the window in the blur condition is more difficult to explain. The most likely explanation is a tunnel vision-type account, in that the increased cognitive load associated with processing a greater amount of foveal and parafoveal detail present in the window condition effectively shrinks the functional visual field. A number of studies have shown that increasing foveal load causes increasingly greater impairments with increasing eccentricity on peripheral tasks (e.g., Mackworth, 1965, Ikeda & Takeuchi, 1975, Williams, 1988). A tunnel vision account would make two predictions: contractions in functional visual fields would decrease average saccade sizes, because saccades greater than double the radius of the functional field of view would result in holes or incomplete coverage of the visual field. Second, partly as a result of the decreased saccade sizes, there would be an increase in the number of saccades required to acquire the target.

The data from this experiment support this account. Saccade sizes were decreased by the addition of a window, especially when the window contained a greater amount of information (i.e., the blur window condition). Furthermore, the presence of a window caused an increase in the number of saccades required to acquire the target in the Blur conditions but not in the Non-Blur conditions. Thus the data support the interpretation that the blur windows caused a contraction of the functional visual field, or a tunnel vision-like phenomenon.

Experiment 1 established that a 6° gaze contingent window with a well-defined edge has detrimental effects on target detection independent of resolution. Some questions remained unanswered. In Experiment 1, sometimes the target was easily discriminable from the background
only due to its motion. Thus motion was a major factor in determining detection performance. It remained to be seen what the effects of gaze contingent windows would be without motion. It also remained to be seen what the effects of varying the size of gaze contingent windows would have on target detection performance.
Experiment 2: Stationary Targets

In the second experiment, the effects of window size on the detection of stationary targets in a static background is investigated. Four window conditions are used: a 3° window, a 6° window, a 9° window, and a no window control condition. Experiment 1 established that the effects of the window edge are independent of resolution. Thus, in order to examine the effects of the edge of the window in isolation from resolution, luminance-defined windows were employed in Experiment 2. It is hypothesized that the presence of the windows will impair detection performance, relative to the non-window condition. Turner's (1984) results support the prediction of larger impairments of performance with smaller windows, because the edges are more salient with the smaller windows. Thus it is also hypothesized that the smaller windows will impair performance more than the larger windows.

Method

Participants

Participants were 30 undergraduate students (9 males and 21 females, ages 19 to 56) at the University of Toronto who received course credit for participation. All participants had normal or corrected-to-normal vision. 30 other participants had previously participated in Experiment 1. Data for these participants were not used because there was a practice effect from Experiment 1.

Apparatus

Apparatus was the same as in Experiment 1, except that subjects were also given a button box for making responses.

Stimuli
Stimuli were static images containing a target against a background scene. Single frames were taken 300 msec from the beginning or end of videoclips selected from Experiment 1, so that the targets were located in each of the four corners of the screen. The screen dimensions, resolution, target and backgrounds were the same as in Experiment 1. Luminance-defined windows were created as in Experiment 1, so that unprocessed and luminance-increased versions of the images were combined based on the participant’s gaze position. Windows were high luminance inside the window and normal luminance in the background. Three different sizes of luminance windows were created, 3°, 6°, and 9° in diameter.

**Design**

The independent variable was Window Size. Three different window sizes: small (3°), medium (6°) and large (9°) plus a no window control condition were used in this experiment. The four window types were counterbalanced with the four possible target locations, for a total of 16 possible combinations. Each combination and each background appeared in random sequence four times per block, and there were three blocks used in the experiment. A practice block of 16 trials was given prior to the experimental trials. Each combination of window condition and target location appeared once in this practice block.

**Procedure**

*Setup and calibration.* Setup and calibration was the same as in Experiment 1. Subject instructions are given in Appendix 1.

*Trial sequence.* A trial sequence began with a fixation dot on a blank gray screen. This initial fixation dot served the double purpose of ensuring the participant started each trial looking at the centre of the screen, and to perform a drift correction before the trial began. The participant fixated the dot, and the experimenter initiated the trial when the gaze cursor stabilized. The fixation dot
disappeared immediately, and after 500 msec the first scene appeared. Participants were instructed to locate and fixate the target in the scene as quickly as possible, and then press a button on the button box. The trial ended when the subject pressed a button, or 3 sec had elapsed, whichever came first. The screen blanked immediately, and after 7 sec, the fixation dot returned and the next trial began.

**Measures and Analyses**

Trials in which participants pressed a button to end the trial before the target had been acquired were discarded. Otherwise, detection performance measures and dropped trial criteria were the same as in Experiment 1. In total, 5.5% of the trials were dropped.

**Results**

The mean latency to the first saccade is plotted by condition in Figure 10. A one-way analysis of variance of the latency to the first saccade revealed a significant main effect of Window Size, $F(3, 87)=8.33, p<.001$. Two-tailed paired-samples t-tests showed that the latencies to the first saccade in the Non-Window condition were lower than the latencies in the large window condition, $t(29)=3.31, p<.01$, medium window condition, $t(29)=3.94, p<.001$, and small window condition, $t(29)=4.09, p<.001$. None of the window conditions differed significantly from each other in their latencies to first saccade, $t(29)=0.58, ns., t(29)=.93, ns.,$ and $t(29)=1.59, ns.$, for large versus medium, medium versus small, and large versus small windows, respectively. The latency to acquisition measure yielded the same pattern of results (see Figure 11).

Other measures, including duration of second fixation, $F(3, 81)=1.44, ns$, proportion of trials with a second fixation, $F(3, 87)=.32, ns,$ and number of saccades before acquisition, $F(3, 87)=1.78, ns$, (see Figures 12, 13, and 14, respectively) did not yield significant results.

A nonsignificant analysis of variance of error after the first saccade $F(3, 87)=0.49, ns$ shows that
the distance of participants' gaze positions from the target after the first saccade did not differ across conditions (see Figure 15).

**Discussion**

The goal of Experiment 2 was to find out what the effects are, if any, of different sizes of luminance-defined gaze contingent windows on the detection of a stationary target. As expected, the windows impaired performance on target detection. However, the effect of window size was not significant.

These findings replicate the target detection data of Experiment 1, in that the presence of a window slows response times. Again, a reasonable explanation of this phenomenon is that the edges of the windows are salient, and either attract attention or narrow the functional visual field, in either case degrading performance of peripheral tasks. It would seem surprising, then, that the smaller windows did not impair performance more than the larger windows, given that the edges of the smaller windows are closer to the fovea and presumably therefore more salient. The current experiment cannot provide a solution to this problem. It may be that the increased salience of the smaller windows associated with the increased proximity of the edges to the fovea is offset by a decrease in salience caused by the decreased size of the windows.

The failure to find an effect of window size on the number of saccades before acquisition suggests that a contraction of the visual field did not occur. This finding agrees with Experiment 1, in which the luminance-defined window did not significantly increase the number of saccades required to acquire the moving target.
General Discussion

The main findings of these experiments are that the mere presence of gaze contingent windows can impair detection of stationary and moving targets, and that this impairment can be partitioned into two additive and dissociable components: effects due to blurring of the target and the background, and effects of the window edge. Window size has no consistent effect on target detection, at least for windows 3° to 9° in diameter.

The results of this experiment differ from Turner’s (1984) in that the mere presence of a high resolution AOI inset can impair performance on a task even relative to a condition in which the entire screen is blurred. However, the type of task Turner used was not the kind that is likely to be impaired by the presence of a foveal stimulus. It was a path-following task, which likely required only central vision for normal performance. Furthermore, it was not a speeded task. The impairments caused by the windows in the current experiment occurred because the stimuli were located in peripheral vision, and because it is speeded tasks that are affected the most by tunnel vision or general interference due to the presence of foveal stimuli. Turner did include a visual search type task in another experiment that involved detection of peripheral stimuli, but it was ultimately a central vision task because the participants were required to use the AOI inset to distinguish targets from distractors. Turner’s experiment is the only one that directly compares performance with and without an AOI inset, with the background constant.

Turner also found an effect of inset size, which also differs from our data. Caution must be exercised in interpreting Turner’s results because the different inset sizes also differed in resolution, and blending ring size, but there appeared to be an optimal inset size/resolution trade-off. Turner’s finding that the smaller AOI insets had more visible edges, together with the findings of the current experiment that salient edges can cause impairments in detection of peripheral targets, emphasizes the
importance of investigating the parameters of the blending ring. Turner’s experiment suggests that a
2.5° blending ring is objectionable to participants. It is unclear what the effects of the blending ring
were on the path following and target identification tasks.

Browder (1989) included a task involving detection of peripheral targets in which speed of
detection and accuracy of identification were of equal importance. However, Browder only
compared head- and eye-slaved modes of operation, there was no control condition that did not
include an AOI. Browder manipulated, tracking mode, blending ring size, and throughput delay, but
reported no difference in performance due to any of these manipulations. Browder did report that
acquiring peripheral targets was not as good as acquiring more central targets, but this is to be
expected even without invoking a tunnel vision explanation.

At throughput delays (30 msec) similar to those in this experiment, Hodgson et al. (1993)
found no effect of AOI insets on eye fixations. However, fixations grew longer as throughput delays
increased. Consistent with the results of the current studies, no effects of AOI size were found, at
any level of delay. The lack of significant size effects in Experiment 2 and in Hodgson et al.’s
experiment is perplexing. There may be opposing processes working that mask the effects of AOI
size. The possibility that the high salience of the edges of small AOIs may be offset by the low
salience of small stimuli has already been raised. Another possibility is that any gains made by having
available large areas of high detail in large AOI insets are offset by the increased cognitive load
associated with processing those large areas of high detail. Shioiri and Ikeda (1989) however did find
window size effects in a scene recognition task.

The relevance of these studies for implementations of foveated displays and particularly for
AOI flight simulator systems is that it is very important to minimize the perceptibility of the edges of
the high-resolution or high-detail foveal region, perhaps through the use of a blending region between
the high and low resolution regions. Although previous papers have emphasized the importance of having a blending region to prevent popping effects and objectionability of the display to viewers, none have quantitatively demonstrated a disruptive effect on the eye movements of the user. One consideration is the nature of the blending region. In most foveated compression schemes, the blending region is actually a transition region (or series of regions) which has a uniformly mid level resolution or mid-detail. If the edges of these regions are noticeable, they may compound rather than alleviate the problem because each additional transition region means another visible edge will compete for the viewer's attention. In flight simulators however most blending regions do truly blend the background and the inset because they either optically or electronically continuously vary the relative contributions of the background and inset displays across the blending region. However, as the results of Turner's experiment shows, even true blending regions can be salient.

Another relevant aspect emerging from these studies for AOI displays is that a luminance change alone between a foveal region and a background region can cause impairments in target detection. This is an important consideration when determining how the high-resolution inset will be combined with the background image. Even if a blending region gracefully ramps a resolution or level of detail change between the inset and the background, if a luminance step remains, the edges of the inset could still be perceptible and still degrade the detectability of peripheral targets. There are a number of reasons an AOI display could have a high resolution inset which is also higher luminance than the background. Some older systems merely superimposed the inset on top of the background image (Spooner, 1984); the inset image had to be significantly brighter than the background in order to be visible and not result in a "double image" or ghosting phenomenon. Some AOI systems use different types of displays for the inset and background imagery. For example, some systems system combine helmet-mounted display imagery with see-through imagery, such as the helmet mounted
AOI system proposed for the DART (Kelly, Shenker, & Weisman, 1992a, 1992b). Brightness discontinuities between the different types of displays could result in perceptible AOI edges which may impair the detection of peripheral targets. Unfortunately, when specifications for AOI display systems are reported, they usually do not include separate measurements of inset and background luminance.

Incomplete and inconsistent reporting of AOI system specifications is a general problem in the technical papers in this field. More consistency among technical papers of AOI display specifications, particularly the parameters of the inset and the blending ring, is desirable for system evaluation and for comparison of different systems. As a minimum, the following specifications should be reported, where appropriate:

1. Total field of view
2. Mode of inset slaving (eye-, head-, or target-slaving)
3. Throughput delay
4. Display update rate
4. Colour or monochrome
5. Instantaneous field of view (background) size
6. Background resolution
7. Background luminance
8. Background contrast ratio
9. AOI inset size
10. AOI inset resolution
11. AOI inset luminance
12. AOI contrast ratio

13. Blending ring size

The method of blending of the inset and background should always be described. Many of these parameters could affect the usability of the system.

Future Directions

Because these experiments reveal a previously unknown impairment caused by gaze contingent AOI-type insets, they raise more questions than they answer, some of them interesting from a theoretical perspective and others from a purely practical viewpoint. First and foremost, how general is this phenomenon? Will it also apply to larger displays, larger inset sizes, and inset shapes, especially the range of display and inset sizes and rectangular-shaped insets commonly used in flight simulators? Will it apply when there is a blending ring? What are the parameters of the blending ring that could bring about this impairment?

Various aspects of the task parameters should also be investigated. In this experiment, the task was somewhat artificial in that all three components of the image, the background, the window, and the target, all appeared and disappeared together. It could be argued that this inflates the effect, because no preprocessing of the window edges (the foveal item) could occur. What would happen in a more realistic task in which the background and window imagery remained on throughout an entire block of trials? It is conceivable that at least some preprocessing of the edges could occur, so that when the targets appear, they can be detected more rapidly. One would expect that the effect would still occur but be smaller. However it is conceivable that with continued exposure, users could habituate to the window edges, or at least learn to ignore them with practice, so that the impairment entirely disappears.
Another task parameter which is of both theoretical and practical interest is target eccentricity. In these experiments, the targets always appeared peripherally, at the edges of the display. Thus there is no direct test to determine if a tunnel vision-like phenomenon or a general interference phenomenon is occurring (Holmes et al., 1977). A tunnel vision interpretation would predict a smaller effect for targets closer to the fovea. There are also two special cases to be investigated: one is when targets appear inside the window, and the other is when targets appear at the edge of the window. According to the fovea-first scheme of Holmes et al., one might expect the effect under either of these conditions to be either reduced or disappear altogether if the target onset and window onset occur together, as they did in this experiment. Things become considerably more interesting if the window onset precedes the target onset. If the edges of the window have already engaged the viewer’s attention, then targets that are close to the edge should be detected faster than targets that are farther away. The possibility emerges that parafoveal or peripheral targets close to the window edge may be detected faster than targets closer to the fovea, but more distant from the window edge. This produces a strong test of the idea that it is the window edges that are responsible for impairment of target detection, rather than the window as a whole.

Effects of target eccentricity are useful to know in flight simulation. Although targets commonly appear from the side and move across the visual field, as in the current experiment, targets can also appear centrally. For example, distant objects become visible as they come within visible range. Occluded objects can move into the line of sight. Lights, jet plumes from other planes or missiles, or even glints of sunlight, can suddenly appear within central vision.

Related to the target eccentricity question are the resolution requirements at different eccentricities, especially peripheral resolution in AOI displays. Work has been done on the effects of resolution on a variety of tasks for whole-field or central vision, (e.g., Kennedy et al., 1988; Turner,
1984) but few studies have investigated the effects of varying peripheral resolution in AOI displays (e.g., Shioiri & Ikeda, 1989), although insufficient peripheral resolution has been a complaint of users of at least one AOI system, the LFOVD display (Warner, Serfoss, Baruch, & Hubbard, 1993). The results of Experiment 1 demonstrate that if degradation of peripheral resolution impairs target detection performance, the addition of a high-resolution inset will not improve detection performance.

At least one head-slaved AOI system features a targeting reticle, or crosshairs, which is always visible in the centre of display (Wells & Venturino, 1990). A targeting reticle could cause a similar impairment of target detection. Finally, the effects of other parameters of AOIs on eye movements need to be investigated, particularly throughput delay and the level of resolution or detail required in the inset and the background.
References


List of Figures

Figure 1. Acuity Function of the Human Visual System

Figure 2. Eyelink apparatus

Figure 3. Latency to First Saccade as a function of Blurring and Windowing, Experiment 1

Figure 4. Latency to Acquire the Target as a function of Blurring and Windowing, Experiment 1

Figure 5. Duration of Second Fixation as a function of Blurring and Windowing, Experiment 1

Figure 6. Proportion of Trials with a Second Fixation Before Acquisition as a Function of Blurring and Windowing, Experiment 1

Figure 7. Number of Saccades Required to Acquire the Target as a function of Blurring and Windowing, Experiment 1

Figure 8. Average Saccadic Distance as a function of Blurring and Windowing, Experiment 1

Figure 9. Error after First Saccade as a function of Blurring and Windowing, Experiment 1

Figure 10. Mean Latency to First Saccade as a function of Window Size, Experiment 2

Figure 11. Latency to Acquisition of target as a function of Window Size, Experiment 2

Figure 12. Duration of Second Fixation as a function of Window Size, Experiment 2

Figure 13. Proportion of Trials with a Second Fixation as a function of Window Size, Experiment 2

Figure 14. Number of Saccades Before Acquisition of target as a function of Window Size, Experiment 2

Figure 15. Error after First Saccade as a function of Window Size, Experiment 2
Figure 1. Acuity Function of the Human Visual System. The fovea is +/- 0.5 deg from 0 and the parafovea is +/- 2.5 deg from 0.
Figure 2. Eyelink eye-tracking system.
Figure 3. Latency to first saccade as a function of Windowing and Blurring conditions for Experiment 1.
Figure 4. Latency to acquisition of target as a function of Blurring and Windowing in Experiment 1.
Figure 5: Duration of second fixation as a function of Blurring and Windowing, Experiment 1.
Figure 6: Proportion of trials with a second fixation as a function of Blurring and Windowing, Experiment 1.
Figure 7. Number of saccades required to acquire the target as a function of Blurring and Windowing, Experiment 1
Figure 8: Average saccadic distance as a function of Blurring and Windowing. Experiment 1.
Figure 46. Error after first saccade as a function of Blurring and Windowing.
Figure 10: Latency to first saccade as a function of window size, Experiment 2.
Figure 11: Latency to Acquisition of target as a function of Window Size, Experiment 2.
Figure 12: Duration of Second Fixation as a function of Window Size, Experiment 2.
Figure 13: Proportion of Trials with a Second Fixation as a function of Window Size, Experiment 2.
Figure 14: Number of Saccades Before Acquisition as a function of Window Size, Experiment 2.
Figure 15: Error After First Saccade as a function of Window Size, Experiment 2.
Appendix: Subject Instructions

Calibration

Notice the black dot with the white centre. That dot will jump around to different locations on the screen. I would like you to look at the dot and focus on the white centre, wherever it appears on the screen. You are NOT being timed; please try not to anticipate where it will appear next. Also please do not move your head or move around in your chair while you are following the dot.

Experiment 1

Look at the black dot in the middle of the screen. It is very important that you focus on the white dot in the centre of the black dot every time it appears. It will disappear, and after a few seconds a video clip will start. It will contain a moving target. Please find the target as quickly as you can and track it until the end of the clip. After a few seconds, the black dot will reappear -- again, please focus on the white dot in the centre, and prepare to look for the target in the next clip.

The first 8 clips will be practice. Please do not to move your head or move around in your chair during the clips. Do you have any questions before we begin?

Okay, now we will begin the actual experiment. Remember that it is important that you focus on the white dot in the centre of the black dot whenever it appears. Also please do not move your head or move around in your chair. If you must move, or ask a question, please do so between clips.

Experiment 2

Look at the black dot in the middle of the screen. It is very important that you focus on the white dot in the centre of the black dot every time it appears. It will disappear, and after a few seconds a static scene will appear with a target in one of the four corners of the screen. Please find it as quickly as possible, look at it, then press any one of the buttons on the game pad to end the trial.

It is important that you do not press the button to end the trial until you have first looked at the target. After a few seconds, the black dot will reappear -- again, please focus on the white dot in the centre until it disappears, and prepare to look for the target in the next scene.

The first 16 scenes will be practice. Please do not to move your head or move around in your chair during the scenes. Do you have any questions before we begin?

Okay, now we will begin the actual experiment. Remember that it is important that you focus on the white dot in the centre of the black dot whenever it appears. Also please do not move your head or move around in your chair. If you must move, or ask a question, please do so between scenes.