EXAMINING VISUAL MASKING EFFECTS ON TARGET ACQUISITION USING
TWO-DIMENSIONAL FOURIER ANALYSIS TECHNIQUES

Gary S. Olacsi

Dissertation submitted to the faculty of
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in
Industrial and Systems Engineering

Dr. Robert Beaton, Chair
Dr. Albert Prestrude
Dr. Tom Dingus
Dr. Grayson Quclock-Knopp
John Merritt

July, 2001
Blacksburg, Virginia

Keywords: visual masking, military target acquisition, spatial frequency, vision modeling, two-
dimensional Fourier analysis, filter transfer function, digital image processing
EXAMINING VISUAL MASKING EFFECTS ON TARGET ACQUISITION USING TWO-
DIMENSIONAL FOURIER ANALYSIS TECHNIQUES

Gary S. Olacsi

ABSTRACT

Visual masking is a psychophysical phenomenon that occurs when “noise” or background (i.e., non-information bearing) objects degrade an observer’s ability to perceive target (i.e., task-relevant) objects. Masking occurs because the human eye-brain system processes background features in a manner that degrades (masks) the processing of target features. One example of this phenomenon in military operations is camouflage. Camouflage decreases target visibility by masking target structure and intensity.

Psychophysical visual masking studies often employ simple (non-real world) targets and masking stimuli, such as one-dimensional spatial frequency patterns (Wilson, 1995; Wilson, McFarlane & Phillips, 1983; Yang & Stevenson, 1998). A one-dimensional spatial frequency pattern usually consists of a sine- or square-wave grating pattern; that is, the luminance variations in the stimuli oscillate across one dimension of spatial extent. These grating patterns are hypothesized to excite the human visual mechanisms responsible for initial encoding and processing of visual scenes. In this manner, the grating patterns represent a simplification of real-world visual scenes, which are at least two-dimensional (i.e., left-right and up-down dimensions) in spatial (and spatial frequency) content.

Past psychophysical research has justified the use of one-dimensional spatial frequency patterns on the basis that more realistic two-dimensional patterns require extensive computational resources. However, with today’s affordable computing machines, researchers can implement methodologies readily to explore and exploit the two-dimensional nature of visual imagery, especially the perception of digital images.

To begin investing two-dimensional visual processes in target acquisition, it is necessary to establish the existence and functional characteristics of two-dimensional masking phenomena. This dissertation first discusses a preliminary effort to establish a methodology to glean some information on two-dimensional masking effects. Specifically, Experiment 1 provided direct evidence for the existence of masking in the two-dimensional spatial frequency domain. Experiment 2 then demonstrated some functional effects on real-world target perception due to deliberate suppression of selected two-dimensional spatial frequency
structure. Lastly, Experiments 3 and 4 extended the findings of the first two experiments using real-world targets and backgrounds.

The findings of this dissertation extend existing knowledge on visual masking phenomena into the realm of two-dimensional spatial frequency targets and masking fields, as well as provide a foundation for designing and interpreting more advanced studies of two-dimensional spatial frequency masking effects that may moderate visual target acquisition performance.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>II</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>2</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>3</td>
</tr>
<tr>
<td>BACKGROUND</td>
<td>5</td>
</tr>
<tr>
<td>APPLICATION OF FOURIER TECHNIQUES TO SPATIAL VISION</td>
<td>5</td>
</tr>
<tr>
<td>EVIDENCE FOR VISUAL PROCESSING OF FOURIER COMPONENTS</td>
<td>6</td>
</tr>
<tr>
<td>SPATIALLY TUNED VISUAL CHANNELS</td>
<td>10</td>
</tr>
<tr>
<td>VISUAL MASKING LITERATURE</td>
<td>12</td>
</tr>
<tr>
<td>VISUAL CHANNEL MODELS FOR SPATIAL VISION</td>
<td>22</td>
</tr>
<tr>
<td>EVIDENCE FOR TWO-DIMENSIONAL SPATIAL FREQUENCY FILTERING</td>
<td>23</td>
</tr>
<tr>
<td>MILITARY TARGET ACQUISITION MODELING</td>
<td>25</td>
</tr>
<tr>
<td>TWO-DIMENSIONAL VISUAL PROCESSING</td>
<td>30</td>
</tr>
<tr>
<td>EXPERIMENT 1</td>
<td>35</td>
</tr>
<tr>
<td>METHOD</td>
<td>35</td>
</tr>
<tr>
<td>Participants</td>
<td>35</td>
</tr>
<tr>
<td>Apparatus</td>
<td>35</td>
</tr>
<tr>
<td>Stimuli</td>
<td>36</td>
</tr>
<tr>
<td>Experiment Design</td>
<td>39</td>
</tr>
<tr>
<td>Procedure</td>
<td>41</td>
</tr>
<tr>
<td>RESULTS</td>
<td>42</td>
</tr>
<tr>
<td>Target Detection</td>
<td>42</td>
</tr>
<tr>
<td>Target Recognition</td>
<td>45</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>47</td>
</tr>
<tr>
<td>EVIDENCE FOR TWO-DIMENSIONAL SPATIAL FREQUENCY FILTERING</td>
<td>48</td>
</tr>
<tr>
<td>EXPERIMENT 2</td>
<td>52</td>
</tr>
<tr>
<td>METHOD</td>
<td>53</td>
</tr>
<tr>
<td>Participants</td>
<td>53</td>
</tr>
<tr>
<td>Apparatus</td>
<td>54</td>
</tr>
<tr>
<td>Stimuli</td>
<td>54</td>
</tr>
<tr>
<td>Target Processing</td>
<td>55</td>
</tr>
<tr>
<td>RESULTS</td>
<td>62</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>72</td>
</tr>
<tr>
<td>EXPERIMENT 3</td>
<td>75</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>METHODS</td>
<td>80</td>
</tr>
<tr>
<td>Participants</td>
<td>80</td>
</tr>
<tr>
<td>Apparatus</td>
<td>81</td>
</tr>
<tr>
<td>Stimuli</td>
<td>81</td>
</tr>
<tr>
<td>Tasks</td>
<td>86</td>
</tr>
<tr>
<td>Experiment Design</td>
<td>87</td>
</tr>
<tr>
<td>Procedure</td>
<td>88</td>
</tr>
<tr>
<td>RESULTS AND DISCUSSION</td>
<td>89</td>
</tr>
<tr>
<td>EXPERIMENT 4</td>
<td>96</td>
</tr>
<tr>
<td>METHODS</td>
<td>96</td>
</tr>
<tr>
<td>Participants</td>
<td>97</td>
</tr>
<tr>
<td>Apparatus</td>
<td>97</td>
</tr>
<tr>
<td>Stimuli</td>
<td>97</td>
</tr>
<tr>
<td>Tasks</td>
<td>105</td>
</tr>
<tr>
<td>Experiment Design</td>
<td>105</td>
</tr>
<tr>
<td>Procedure</td>
<td>107</td>
</tr>
<tr>
<td>RESULTS AND DISCUSSION</td>
<td>108</td>
</tr>
<tr>
<td>Target Detection</td>
<td>109</td>
</tr>
<tr>
<td>Target Recognition</td>
<td>118</td>
</tr>
<tr>
<td>MODELING</td>
<td>126</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>132</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>134</td>
</tr>
<tr>
<td>VITA</td>
<td>139</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1. From left to right, one sine wave grating and one square wave grating, similar to those used in Campbell and Robson .................................................................7

Figure 2. Examples of one-dimensional spatial sine-wave patterns. Part A-low frequency. Part B-medium frequency. Part C-high frequency .................................................................9

Figure 3. A plot of contrast sensitivity as a function of spatial frequency in log units for a human observer ..................................................................................................................10

Figure 4. Graphic depicting seven proposed channels derived experimentally that constitute the Contrast Sensitivity Function ........................................................................................................11

Figure 5. Masking of visual contrast sensitivity by a 4.5 cycle per degree sine wave grating as a function of target spatial frequency ................................................................................13

Figure 6. Two radial frequency (lower and upper) and two orientation (clockwise and counterclockwise) thresholds depicted by arc segments about each probe point .......24

Figure 7. An illustration of the cycles-on-target process for perception used in the ACQUIRE model. (Adopted from Biberman, 1973) ......................................................................................27

Figure 8. Graphic showing the sequence of the five modules in the GTV model. (Adopted from Doll, 1998) .....................................................................................................................................29

Figure 9. Examples of two-dimensional radial sine-wave patterns and their Fourier amplitude spectra. Part A-Low, medium, and high spatial frequency radial sine waves ..........33

Figure 10. Illustration of testing room used in Experiment 1 .................................................................................................................................36

Figure 11. Example two-dimensional radial sine wave targets used in Experiment 1. (Part A-2 c/d. Part B-4 c/d. Part C-8 c/d.) ..........................................................................................37

Figure 12. Example two-dimensional radial sine wave masks used in Experiment 1. (Part A-1.5 c/d. Part B-2.5 c/d. Part C-12 c/d.) .....................................................................................38

Figure 13. Examples of two-dimensional target and mask sine-wave patterns used in Experiment 1. (Part A-2 c/d. Part B-4 c/d. Part C-8 c/d.) ..............................................................................39

Figure 14. Average correct detection probabilities for 2 and 4 cycle per degree targets ..........43

Figure 15. Average correct detection probabilities for 8, 16, and 32 cycle per degree radial sine-wave targets as a function of two-dimensional mask spatial frequency ..................44

Figure 16. Combined data shown in Figures 14 and 15 on a common scale .................................................................................................45

Figure 17. Average correct recognition probabilities for 2 and 4 cycle per degree radial sine-wave targets as a function of two-dimensional mask spatial frequency ..................46
Figure 18. Average correct recognition probabilities for 8, 16, and 32 cycle per degree radial sine-wave targets as a function of two-dimensional mask spatial frequency


Figure 20. One-dimensional amplitude profiles for the spatial frequency filters used in Experiment 2.

Figure 21. Example of Abrams Tank image after processing with the two-dimensional radial notch filter centered at 32 cycle per degree used in Experiment 2.

Figure 22. Example of APC image after processing with the two-dimensional radial notch filter centered at 32 cycle per degree used in Experiment 2.

Figure 23. Percent threshold elevation for correct detection of Recovery Vehicle image as a function of two-dimensional filter center frequency and gain.

Figure 24. Percent threshold elevation for correct detection of Army Personnel Carrier image as a function of two-dimensional filter center frequency and gain.

Figure 25. Percent threshold elevation for correct detection of Bradley Fighting Vehicle image as a function of two-dimensional filter center frequency and gain.

Figure 26. Percent threshold elevation for correct detection of Abrams Tank image as a function of two-dimensional filter center frequency and gain.

Figure 27. Percent threshold elevation for correct recognition of Recovery Vehicle image as a function of two-dimensional filter center frequency and gain.

Figure 28. Percent threshold elevation for correct recognition of Army Personnel Carrier image as a function of two-dimensional filter center frequency and gain.

Figure 29. Percent threshold elevation for correct recognition of Bradley Fighting Vehicle image as a function of two-dimensional filter center frequency and gain.

Figure 30. Percent threshold elevation for correct recognition of Abrams Tank image as a function of two-dimensional filter center frequency and gain.

Figure 31. Low frequency masking background scene and its two-dimensional Fourier modulation spectrum.

Figure 32. Medium frequency masking background scene and its two-dimensional Fourier modulation spectrum.

Figure 33. High frequency masking background scene and its two-dimensional Fourier modulation spectrum.

Figure 34. Military ground vehicle targets used in Experiment 3.

Figure 35. Forest tree line scene (No. 1).
Figure 36. Forest tree line scene (No. 2)...
Figure 37. Desert brush scene (No. 1)...
Figure 38. Desert brush scene (No. 2)...
Figure 39. Sand dune scene (No. 1)...
Figure 40. Sand dune scene (No. 2)...
Figure 41. Bradley tank embedded in Forest Tree Line (No. 1) background...
Figure 42. M60 tank embedded in Sand Dune (No. 2) background...
Figure 43. Recovery vehicle embedded in Desert Brush (No. 1) background...
Figure 44. Correct detection probabilities for Abrams Tank embedded in various backgrounds.

Error bars represent +/- 1 standard error of the mean.
Figure 45. Correct detection probabilities for M60 tank embedded in various backgrounds...
Figure 46. Correct detection probabilities for Recovery Vehicle embedded in various backgrounds...
Figure 47. Correct detection probabilities for Bradley Fighting Vehicle embedded in various backgrounds...
Figure 48. Correct detection probabilities for Armored Personnel Carrier embedded in various background...
Figure 49. Military ground vehicle targets used in Experiment 2...
Figure 50. Amplitude profiles of radial notch filters used in Experiment 4...
Figure 51. Example of tree line background scenes used in Experiment, Original (unfiltered) and radial notch filtered at 16 cycle per degree versions are shown...
Figure 52. Example of desert brush background scenes used in Experiment 2. Original (unfiltered) and radial notch filtered at 16 cycle per degree versions are shown...
Figure 53. Example of sand dune background scenes used in Experiment 2. Original (unfiltered) and radial notch filtered at 16 cycle per degree versions are shown...
Figure 54. M60 tank embedded in Desert brush background...
Figure 55. M60 tank embedded in Filtered Tree Line (16 cycle per degree: center frequency of filter) background...
Figure 56. Illustration of target positions used in the study. Actual target position are indicated by the "X" markers in the figure...
Figure 57. Graphic showing a plot of filters as a function of background scene for the Abrams Tank...
Figure 58. Graphic showing a plot of filters as a function of background scene for the Recovery Vehicle...
Figure 59. Graphic showing a plot of filters as a function of background scene for the ...........112
Figure 60. Graphic showing a plot of filters as a function of background scene for the Bradley
    Fighting Vehicle. ..................................................................................................................113
Figure 61. Effects of filter center frequency and background scene conditions on percent
    change in detection probabilities for the M1 Abrams tank target. ....................................115
Figure 62. Effects of filter center frequency and background scene conditions on percent
    change in detection probabilities for the M60 tank target. ...............................................116
Figure 63. Effects of filter center frequency and background scene conditions on percent
    change in detection probabilities for the Bradley fighting vehicle target. ........................117
Figure 64. Effects of filter center frequency and background scene conditions on percent
    change in detection probabilities for the recovery vehicle target. ....................................117
Figure 65. Graphic is a plot of filters as a function of background scene for the Recovery
    Vehicle. ..............................................................................................................................120
Figure 66. Graphic is a plot of filters as a function of background scene for the M60 Tank. ...121
Figure 67. Graphic is a plot of filters as a function of background scene for the Bradley Fighting
    Vehicle. ..............................................................................................................................120.
Figure 68. Graphic is a plot of filters as a function of background scene for the Abrams Tank
.............................................................................................................................................122
Figure 70. Effects of filter center frequency and background scene conditions on percent
    change in recognition probabilities for the M1 Abrams tank target. ...............................123
Figure 71. Effects of filter center frequency and background scene conditions on percent
    change in recognition probabilities for the M60 tank target. ............................................124
Figure 72. Effects of filter center frequency and background scene conditions on percent
    change in recognition probabilities for the Bradley fighting vehicle target. ........................124
Figure 73. Effects of filter center frequency and background scene conditions.......................125
# LIST OF TABLES

Table 1. Summary of Similarities and Differences Among Masking Studies Designed to Estimate Channel Bandwidths ...........................................................................................................21

Table 2. Spatial Frequency of Masks used in Experiment 1 ...........................................................................................................38

Table 3. Design Summary for Experiment 1 .................................................................................................................................41

Table 4. Summary of similarities and differences among masking studies designed to estimate channel bandwidths and Experiment 1 ..............................................................................50

Table 5. Two-Dimensional Spatial Frequency Filter Parameters used in Experiment 2 ........................................................................57

Table 6. Linear Regression Models for Effects of Filter Gain on Threshold Detection ........................................................................66

Table 7. Linear Regression Models for Effects of Filter Gain on Threshold Recognition ........................................................................70

Table 8. Summary of Experiment 1 .................................................................................................................................................87

Table 9. Analysis of Variance for Detection Judgments ..................................................................................................................89

Table 10. Newman-Keuls Results for the Main Effect of Target on Detection Judgment ..........................................................90

Table 11. Newman-Keuls Results for the Main Effect of Scene on Detection Judgment ...........................................................90

Table 12. Parameters for Two-Dimensional Radial Notch Filter Used in Experiment 4 ........................................................................98

Table 13. Summary of Experiment 4 .................................................................................................................................................106

Table 14. Analysis of Variance for Detection Judgments ..................................................................................................................109

Table 15. Newman-Keuls Results for the Main Effect of Target on Detection Judgments ..........................................................110

Table 16. Newman-Keuls Results for the Main Effect of Scene on Detection Judgments ...........................................................110

Table 17. Newman-Keuls Results for the Main Effect of Filter on Detection Judgments .............................................................110

Table 18. Analysis of Variance for Recognition Judgments ..............................................................................................................118

Table 20. Newman-Keuls Results for the Main Effect of Target on Recognition Judgments ..........................................................119

Table 21. Newman-Keuls Results for the Main Effect of Scene on Recognition Judgments ...........................................................119

Table 22. Newman-Keuls Results for the Main Effect of Filter on Recognition Judgments .............................................................119

Table 23. Calculation of Visual Channel Areas ..............................................................................................................................129
ACKNOWLEDGMENTS

The author would especially like to thank Dr. Robert J. Beaton, Director of the Displays and Controls Laboratory; without whose valuable technical expertise, mentoring, and patience, the work described in this dissertation would not of been possible. In particular, the author is grateful for his development work on the image manipulation apparatus used in this dissertation. In addition, the author would like to thank Dr. Grayson Quiclock-Knopp, Dr. Albert Prestrude Dr. Tom Dingus, and John Merritt for their guidance and support during the past two years.

I would like to thank my parents and two sisters, Lori and Lisa, for their support throughout my graduate student career. Their love and encouragement truly sustained me throughout the course of the work described in this dissertation.

Lastly, I would like to thank some of the friends I've made in recent years for their support and much needed commiseration. In particular, I would like to thank Teresa Coalson, Jason Saleem, Jeff Lancaster, Hope Johnson, Jennifer Thompson, Ryan Urquhart, Mitch Woods, Angela Austin, and Erik Olsen for their companionship and friendship.
INTRODUCTION

Spatial vision refers to a general class of perceptual capabilities that allows people to see the structure of objects within an image or visual scene. For example, spatial vision underlies people’s abilities to discriminate important objects (or so-called “targets”) from less important “background” objects. Spatial vision capabilities, however, are not fixed or unaffected by the physical characteristics of the scene objects. Objects with similar shape, color, or size, for example, are more difficult to perceive than objects differing physically from one another. Likewise, targets surrounded by many physically similar background objects are more difficult to perceive than targets silhouetted against uniformly intensity and uncluttered scene areas (Toet, 1996).

The capabilities and limitations of visual target perception have been challenging to study scientifically because the number of variations in targets and backgrounds is large, if not intractable. Thus, the determination of causal relationships between physical properties of objects and human perception of those objects has been limited to relatively simple phenomena, such as the minimum size and luminance contrast needed for visual detection. This dissertation was directed at establishing ways to improve the prediction of target perception performance.

Specifically, this research was motivated by the idea that contemporary theories of spatial vision, as well as contemporary digital image analysis techniques, may provide a unified means of classifying the physical properties of targets and backgrounds in real-world scenes. If a unified schema exists or can be derived, the functional relationships between perceptual capabilities and the myriad combinations of target and background properties may be understood and predicted better than that allowed by extant visual psychophysical theories. With this objective, the present work begins the examination of using the quantitative stimulus descriptions of visual masking paradigms as a way to develop a framework for understanding target perception; specifically, target detection and recognition of objects in real-world scenes, such as those relevant to military target acquisition.

Visual masking is a psychophysical phenomenon that occurs when “noise” or background (i.e., non-information bearing) objects degrade an observer’s ability to perceive target (i.e., task-
relevant) objects. Masking occurs because the human eye-brain system processes background features in a manner that degrades (masks) the processing of target features. One example of this phenomenon in military operations is camouflage. Camouflage decreases target visibility by masking target structure and intensity.

Psychophysical visual masking studies often employ simple (non-real world) targets and masking stimuli, such as one-dimensional spatial frequency patterns (Wilson, 1995; Wilson, McFarlane & Phillips, 1983; Yang & Stevenson, 1998). A one-dimensional spatial frequency pattern usually consists of a sine- or square-wave grating pattern; that is, the luminance variations in the stimuli oscillate across one dimension of spatial extent. These grating patterns are hypothesized to excite the human visual mechanisms responsible for initial encoding and processing of visual scenes. In this manner, the grating patterns represent a simplification of real-world visual scenes, which are at least two-dimensional (i.e., left-right and up-down dimensions) in spatial (and spatial frequency) content.

Past psychophysical research has justified the use of one-dimensional spatial frequency patterns on the basis that more realistic two-dimensional patterns require extensive computational resources. However, with today’s affordable computing machines, researchers can implement methodologies readily to explore and exploit the two-dimensional nature of visual imagery, especially the perception of digital images.

To begin investing two-dimensional visual processes in target acquisition, it is necessary to establish the existence and functional characteristics of two-dimensional masking phenomena. This dissertation first discusses a preliminary effort to establish a methodology to glean some information on two-dimensional masking effects. Specifically, Experiment 1 provided direct evidence for the existence of masking in the two-dimensional spatial frequency domain. Experiment 2 then demonstrated some functional effects on real-world target perception due to deliberate suppression of selected two-dimensional spatial frequency structure. Lastly, Experiments 3 and 4 extended the findings of the first two experiments using real-world targets and backgrounds.

The findings of this dissertation extend existing knowledge on visual masking phenomena into the realm of two-dimensional spatial frequency targets and masking fields, as well as provide a
foundation for designing and interpreting more advanced studies of two-dimensional spatial frequency masking effects that may moderate visual target acquisition performance.

BACKGROUND

Observers visual abilities to detect and discriminate target objects from their surrounding background are affected by numerous physical properties of the visual scene. For example, the luminance contrast of target objects relative to their background is known to limit visual detection and discrimination (Boff & Lincoln, 1988; Braje, Tjan, & Legge, 1995; Goldstein, 1989). Likewise, the physical size of target objects, as well as background objects, and their spatial proximity to one another is known to limit visual detection and recognition (Boff & Lincoln, 1988). These effects, however, refer to factors in the physical realm of the visual scene; that is, to spatial domain factors. Considering the variety of realizable levels and combinations of spatial factors in visual scenes, it is easy to appreciate that a boundless number of unique conditions exists, rendering the problem of scientifically understanding human visual detection and discrimination difficult, if not intractable, to study. One can adopt the approach for describing stimuli used in the field of perception, both visual and auditory, to address this problem, namely, Fourier analysis.

APPLICATION OF FOURIER TECHNIQUES TO SPATIAL VISION

Contemporary theories of human spatial vision overcome some of the difficulties in vision research by theorizing that the eye-brain system processes a scene by decomposing its spatial structure into fundamental elements, which in turn are encoded and transmitted to the brain to form a percept (DeValois and DeValois, 1988). This theoretical framework of perceptual processing of visual scenes vastly reduces the complexities associated with designing scientific experiments to explore spatial vision capabilities and limitations.

The notion that the visual system decomposes scenes is similar to the operation of mathematical transforms; that is, a procedure that casts number functions from one number domain into another. Ideally, a transform is loss-less, meaning that it retains all information in the original domain, and recoverable, meaning it allows the function to be converted back into
the original domain. There are many types of mathematical transforms, ranging from simple scalar operations, such as a logarithm, to complex multidimensional-matrix operations, such as orthogonal basis function operators. For human vision work, the human eye-brain system is envisioned to encode a scene in terms of its spatial frequency domain composition, which is a decomposition described well by the mathematics of Fourier transforms (Bracewell, 1965). A Fourier transform casts a function into a unique set of sine-wave basis functions or components. Each Fourier domain sine-wave component has a unique spatial frequency value, as well as specific intensity (luminance) amplitude and phase values.

Mathematically, Fourier analysis can cast any realizable function into a spectrum of sine waves. Once in the Fourier domain, the sine waves can be combined through the Principle of Superposition to recover the original function—a process called Fourier synthesis (Bracewell, 1965). A Fourier transform of a visual scene identifies the sine-wave components that comprise the scene. For example, detailed parts of a scene are represented by higher frequency components, whereas the coarser and larger parts are represented by lower frequencies (Schiffman, 1996). This information, combined with knowledge on how the human visual system perceives those individual sine-wave components, provides a convenient and universal approach to investigating the relationships among physical properties of an image and its perception judgments by observers.

**EVIDENCE FOR VISUAL PROCESSING OF FOURIER COMPONENTS**

Cambell and Robson (1968) were among the first researchers to show that the eye-brain system may perform a Fourier analysis on visual stimuli. In their landmark experiment, "Application of Fourier Analysis to the Visibility of Gratings," they demonstrated that the eye-brain system might respond to visual stimuli by performing a Fourier analysis on the light entering the eye. Their study used two gratings placed side by side, each having a fundamental frequency (i.e., equal bar widths) of 26 cycles per degree (c/deg). One grating was a sine wave and the other a square-wave grating, (see figure 1). Campbell and Robson showed that the two gratings are indistinguishable to the eye at a certain distance because the eye may perform a Fourier analysis on the two gratings. Specifically, the general appearance of the sign-wave grating is the same from any distance because Fourier analyzing it results in seeing a single sine wave. The square-wave, on the other hand, would be broken down into sine waves with
frequencies of 26 c/deg (the fundamental), 78 c/deg (the third harmonic), and 130 c/deg (the fifth harmonic). These harmonic frequencies are what give the bars on the square wave grating their sharp edges. Because the highest frequency that can be resolved by the eye is approximately 60 c/deg, the harmonic spatial frequency components of the square wave cannot be seen. It is important to note that as an observer moves closer to two gratings such as these, the harmonic spatial frequencies of the square wave grating will become visible (i.e., fall below 60 c/deg), rendering it distinguishable. Put differently, the frequency components subtend larger and larger angles on the blanket of receptors on the retina as the observer advances toward the gratings, causing them fall within the visual pass band.

Figure 1. From left to right, one sine wave grating and one square wave grating, similar to those used in Campbell and Robson (1968).

Campbell, Nachmias, and Jukes (1970) further found that the capacity to discriminate between two above threshold gratings of equal contrast is influenced greatly by the ratio of their spatial frequencies. The objective of the investigation was to determine the spacing between the visual channels that prior researchers have described using similar gratings. Measuring neuron sensitivity to gratings of varying spatial frequency determined these channels. A more complete discussion of spatial frequency channels is reserved for the next subsection of this document, on grating detection investigations. Campbell, et al. demonstrated that just-discriminable frequency ratios increased steadily with increasing spatial frequency until they doubled for gratings whose frequencies were proximal to 20 c/deg. Put differently, higher spatial frequency grating pairs needed to have larger separations in frequency than lower spatial frequency pairs to be discriminable. Watson and Robson (1981) obtained similar results. Specifically, they
varied the spatial frequency of gratings set at threshold contrast and showed that just-discriminable frequency ratios can more than double for grating pairs whose frequencies are in the neighborhood of 25-30 c/deg.

In addition to evidence gleaned from experiments using simple discrimination tasks, investigators have compiled strong support for visual processing of Fourier components employing detection tasks. In detection studies, the observer is required to determine simply whether or not a pattern is present. Specifically, an observer is asked to decide whether a field of uniform luminance (i.e., no target) or a field containing a target stimulus has been presented on any given trial (Boff, Kaufman, and Thomas, 1986). Detection studies use lower contrast patterns (e.g., sine wave gratings) to identify the minimum amount of contrast necessary to perceive a given pattern, that is, distinguish it from an empty field.

Perception research often employs sine-wave patterns (representing individual Fourier components of complex stimuli) to probe various aspects of human spatial vision. Sine-wave patterns can be described in terms of their luminance variations in the spatial domain, as given by

\[ L(x) = b_0 + b_1 \sin\left(\frac{2\pi f x}{D} + \Theta\right) \]  

(Eq. 1)

in which

- \(L(x)\) denotes the luminance of the sine-wave at spatial position \(x\),
- \(D\) denotes the total spatial extent of the sine wave,
- \(b_0\) denotes the luminance offset or bias of the sine wave,
- \(b_1\) denotes the luminance amplitude or gain of the sine wave,
- \(f\) denotes the spatial frequency of the sine-wave, and
- \(\Theta\) denotes the phase angle (0 to \(2\pi\) radians) of the sine wave.

Figure 2 shows several sine-wave grating patterns representing relatively low, medium, and high spatial frequencies. Since luminance variations occur only across the horizontal extent (luminance is constant at any given vertical position), these patterns are one-dimensional sine waves.
Figure 2. Examples of one-dimensional spatial sine-wave patterns are shown: part A—low frequency, part B—medium frequency, and part C is high frequency.

Contrast sensitivity studies represent one class of experiments that establish the eye-brain system as being differentially sensitive to Fourier components, namely, spatial frequency and modulation. The magnitude of luminance variations in a sine-wave pattern is an important stimulus property for spatial vision work. This physical property is known popularly as “contrast,” but it is more correct to use the term “modulation.” Modulation is an index of the peak amplitude of a sine wave relative to its average amplitude level, as given by

\[
M = \frac{L_{\text{max}} - L_{\text{min}}}{L_{\text{max}} + L_{\text{min}}}
\]

(Eq. 2)

in which

\(L_{\text{max}}\) and \(L_{\text{min}}\) refer to maximum and minimum luminance of the sine wave, respectively.

A strong perceptual relationship exists between spatial frequency and modulation. The visibility of sine-wave patterns depends directly on their spatial frequency and modulation values. That is, the human eye-brain system is more sensitive to modulation at some spatial frequencies than at others. For normal human vision, maximum sensitivity to sine wave patterns occurs at spatial frequencies between about 3 and 6 cycles per degree of visual angle. The human visual system requires greater modulation (is less sensitive) to sine wave spatial frequencies above and below this range.
Campbell and Robson (1968) conducted an experiment to determine the threshold contrast necessary for detecting a sine wave grating as a function of spatial frequency. By measuring contrast thresholds for gratings spanning a wide range of spatial frequencies, they derived a curve (Contrast Sensitivity Function or CSF) that describes the eye-brain system’s sensitivity to contrast (see Figure 3). In other words, the CSF defines the “transfer function” of the human visual system, that is, it measures the minimum modulation required to detect spatial sine-wave gratings.

![Figure 3. Plots of contrast sensitivity as a function of spatial frequency in log units for a human observer (Adapted from Campbell and Robson, 1968).](image)

**SPATIALLY TUNED VISUAL CHANNELS**

Contemporary theories of human visual processing are based on the idea that spatial perception is the result of operations performed on the outputs of multiple linear filters (channels), each of which is tuned to different spatial properties (Olzak & Thomas, 1988). Researchers believe that the outputs of these channels provide information about scene elements that permit pattern processing algorithms in higher cognitive centers to detect objects and discriminate them from their backgrounds (Cannon, 1995). The CSF represents the
workings of at least six or seven spatially tuned filters, or spatial frequency channels, each of which are sensitive to a narrow range of spatial frequencies (Wilson, et al. 1983; Ginsburg, 1981) (see Figure 4).

Figure 4. Graphic depicting seven proposed channels derived experimentally that constitute the Contrast Sensitivity Function (Adapted from Ginsburg, 1981).

Spatial pattern vision is said to begin when the cornea and lens focus an optical image of the stimulus on the blanket of photoreceptors in the retina. Consequently, the receptors activate many neural channels that work in collaboration to process the visual information. The concept of spatial tuning refers to the notion that an individual channel will respond to a given stimulus only if it has certain spatial characteristics. For example, a channel may react only if the pattern is comprised of contours or other linear components that lie at or close to a particular orientation in relation to the vertical (Olzak & Thomas, 1988). Other channels, on the other hand, may respond to stimulus components at other orientations. This type of exclusiveness is called orientation tuning. Spatial frequency tuning, which involves luminance fluctuations across a pattern, is perhaps the most important type of spatial tuning.
Researchers typically describe each channel in terms of its sensitivity to different spatial frequency components. The tuning characteristics of a channel usually are defined by evaluating the response of the channel to sinusoidal grating patterns. Three dimensions along which gratings can differ are contrast (luminance difference between light and dark bars), spatial frequency (number of light-dark pairs, or cycles, per degree of visual angle), and orientation (alignment of strips with respect to the vertical). The tuning properties of a channel are defined by specifying which combinations of the three dimensions are required to activate the channel (Olzak & Thomas, 1988). The most important thing to realize about these early vision mechanisms is that the outputs of neighboring channels interact in important ways. For example, it is understood that target backgrounds often mask the detectability and recognizability of real-world military targets.

VISUAL MASKING LITERATURE

It is known that visual masking can occur when the spatial frequency components of target and background stimuli excite the same visual channel(s). When a masking stimulus is presented simultaneously with a target, the modulation necessary to detect the target increases if the spatial frequency of the mask is within about one octave of the target frequency (Legge & Foley, 1980). Moreover, the modulation needed to detect a target increases as the spatial frequency of a masking stimulus approaches (becomes more similar) that of the target.

Figure 5 shows a typical visual masking effect on contrast sensitivity (Tolhurst & Barfield, 1978). Specifically, this experiment used a single masking grating (4.25 cycles per degree) with various spatial frequency targets (ranging from 1 to 25 cycles per degree). The largest loss of contrast sensitivity occurred when the spatial frequency of the target matched that of the mask; that is, 4.25 cycles per degree. Contrast sensitivity improved when the spatial frequency of the target and mask differed by about one octave.
Figure 5. Masking of visual contrast sensitivity by a 4.5 cycle per degree sine wave grating as a function of target spatial frequency (Adapted from Tolhurst and Barfield, 1978).

**Masking Methodology.** The primary objective of a masking study is to determine the degree to which the visibility of one stimulus is affected by a second stimulus. The presence of an interaction, indicated by a change in detectability, provides evidence that the two stimuli share one or more processing channels. In contrast, the absence of an interaction is taken as evidence that the stimuli are processed by separate and independent channels (De Valois & De Valois, 1988). For example, investigators have found that interactions indeed occur when the spatial frequency components of a target and mask differ by less than an octave.

As mentioned above, the visibility of a pattern is often attenuated by the presentation of a second pattern. If this hampering occurs immediately following the onset of the second stimulus, the effect is called masking. Masking is the interference in perception that occurs for a target stimulus as caused by the presence of a mask stimulus that is close in time and space (Olzak & Thomas, 1988). For any given masking study, the mask may be presented, before, after, or simultaneously with the target. These stimulus presentation scenarios are called forward, backward, and simultaneous masking, respectively. The strength of forward and backward
masking is determined by the length of the interval between the starting times of the mask and target. Masking studies also differ in the type of spatial elements that constitute the masking stimuli. If the masks contain spatially random elements (random dots), then the procedure is called masking by noise. When the mask shares spatial structure (e.g., grating pattern masks) with the target, it is termed masking by structure (Henning, Hertz, & Hinton, 1981).

**Psychophysical Tasks.** In addition to the sequence of stimulus exposure and mask construction, studies vary in the type of psychophysical method used to measure the extent of masking. The majority of the masking studies in the literature use a variation on the psychophysical method of limits. For simultaneous masking, this method requires that masker spatial frequency and/or contrast levels be manipulated in either an ascending or descending series. When the series is ascending, the experimenter begins by selecting a subthreshold target (e.g., spatial frequency of target and mask match). On each subsequent trial, the spatial frequency or contrast of the mask is changed by a small amount until the participant reports detection. When the series is descending, the spatial frequency of the mask starts out an octave or more away from that of the target, and then approaches the target spatial frequency in successive increments until the target is no longer detectable. The threshold is obtained by averaging the changeover points received from a number of ascending and descending trials (Gescheider, 1985). The temporal forced choice procedure and staircase methods are the two particular variations on the method of limits frequently adopted in the masking literature.

The psychophysical task that is most often used in the masking literature is the **two-alternative forced choice** procedure (2AFC). For this method, a typical trial requires a participant to make two sequential observations and then determine which of the two contained the target stimulus. The masking pattern appears in both observations, while the target pattern appears in only one (Nachmias, 1993). The spatial frequency and/or contrast level that corresponds to a specified performance level, i.e., two correct detection responses on consecutive trials, defines the threshold (Gescheider, 1985).

Legge and Foley (1980) used the two-alternative forced choice procedure to study the effects of mask contrast and mask spatial frequency on target contrast thresholds. Specifically, they used a 2AFC, simultaneous masking procedure to measure contrast thresholds for a 2.0 c/deg target grating as a function of 7 mask frequencies and 11 mask contrasts. In general, Legge and Foley (1980) found that contrast thresholds were raised (sensitivity was lowered) when the spatial
frequency of the mask approached that of the target; as well as when the contrast of the mask was increased. Thresholds peaked when the spatial frequency of the mask matched the spatial frequency of the target. Tolhurst and Barfield (1978) also used the 2AFC method and reported similar results, except they held mask spatial frequency constant and varied target spatial frequency.

The *staircase method* is another variation on the method of limits frequently adopted by researchers who study masking phenomena (Cornsweet, 1962). For this procedure, an experimenter begins by manipulating masker spatial frequency and/or contrast levels in either an ascending or descending series. When the series is ascending, the experimenter begins by selecting a subthreshold target (e.g., spatial frequency of target and mask match). On each subsequent trial, the spatial frequency or contrast of the mask is changed by a small amount until the participant reports detection. Once detection occurs, the stimulus value is recorded and the direction of the stimulus sequence is reversed from ascending to descending. When the target is no longer detectable, the stimulus sequence again reverses. The method continues until a predetermined number of response-transition points are recorded. The average of the transition points defines the threshold (Cornsweet, 1962).

Nachmias (1993) used the staircase method to investigate target contrast sensitivity as a function of target and mask contrast. In particular, target contrast thresholds were measured for a 10 c/deg target superimposed on a 10 c/deg mask with a three-up, one-down staircase procedure. Nachmias reported that contrast thresholds were lowered (sensitivity was raised) when either the contrast of the target was increased or the contrast of the mask was lowered. Similarly, Swift and Smith (1983) studied target contrast sensitivity as a function of mask frequency and contrast. In addition to finding that higher mask contrast produced higher thresholds, they reported that thresholds peaked when the spatial frequency of the mask matched that of the target (in agreement with Tolhurst and Barfield findings).

**Stimulus Variables.** Masking studies establish the conditions that limit visual performance by measuring target detectability as a function of changes in one or more target and/or mask attributes. Spatial frequency is by far the most important of these stimulus attributes; therefore, it is the one most often varied. Numerous studies have shown that target contrast sensitivity decreases when the mask grating spatial frequency is within an octave of target grating spatial frequency (Legge & Foley, 1980; Tolhurst and Barfield, 1978; Wilson, et al. 1983; Ross, Speed,
and Morgan, 1993). These studies indicate that spatial frequency components that differ by more than an octave are detected by separate vision channels.

Another important stimulus attribute that is known to limit visual performance and that is often manipulated in masking experiments is orientation; that is, the orientation of the target and/or mask grating in relation to the vertical. Researchers have found considerable evidence that gratings which differ in orientation by 15-20 degrees are processed by separate vision channels (Foley, 1994; Phillips and Wilson, 1984; Ross, et al., 1993; Thomas and Gille, 1978; Wilson, et al., 1983). For example, Phillips and Wilson (1984) used a staircase procedure and simultaneous masking paradigm to measure target contrast sensitivity. In particular, they recorded contrast sensitivity as a function of target spatial frequencies and orientation. In general, Phillips and Wilson found that high spatial frequency targets had significantly narrower bandwidths than low spatial frequency targets. Additionally, they reported similar orientation bandwidths for vertically and horizontally orientated gratings.

**Detection studies designed to estimate channel bandwidths.** Blakemore and Campbell (1969) approximated channel bandwidths using adaptation; that is, observers viewed gratings various frequencies for prolonged periods of time in order to reduce contrast sensitivity to subsequent grating presentations. Moreover, grating adaptation was examined psychophysically by determining contrast thresholds before and after adaptation grating presentations. In general, they found that extended viewing of high contrast adaptation gratings reduced contrast sensitivity for low contrast test gratings with similar spatial frequency. Both high and low contrast test gratings were used. Observer viewing distance was 114 inches and the stimulus fields subtended 1.5 degrees of visual angle at the retina. Participants used a potentiometer to set the contrast of the grating to their own threshold (i.e., method of adjustment psychophysical procedure) before and after adapting to a high contrast grating. Once they had viewed an adapting pattern for 60 seconds, the experimenter reduced the contrast of the grating and instructed the participant to set the contrast to threshold again.

Blakemore and Campbell (1969) found threshold elevation for a spectrum of frequencies with a bandwidth of slightly more than an octave at half amplitude, centered on the adapting frequency. Specifically, the strength of the effects and resultant bandwidths were determined to be similar for adapting frequencies between 3 c/deg and 14 c/deg. In addition, they showed that higher frequency bandwidths were a little narrower than lower frequency bandwidths. At lower
adapting frequencies, the effect was strongest for 3 c/deg gratings. Blakemore and Campbell concluded that neurons may exist in the eye-brain system that are selectively sensitive to spatial frequency and size. Moreover, they speculate that these neurons may be located in the visual cortex.

De Valois (1977) also used adaptation to estimate channel bandwidths. As in the Blakemore and Campbell experiment, participants adjusted the contrast (method of adjustment) of a given grating using a potentiometer set to various sensitivity ranges. In particular, observers viewed a 4.3 x 5.5 degree grating monocularly from a distance of 85 cm. The stimuli were sine wave gratings whose values covered a the frequency range between 0.59 to 22.63 c/deg in 1/4 octave steps. Adaptation stimuli were presented for 5 minutes at 95% contrast. At the end of the adaptation stimulus presentation, a test grating was introduced. If participants could not find their threshold in five seconds time, the adaptation grating was presented once more for 10 seconds. They were allowed as many stimulus presentations as necessary to set the contrast for each of the test frequencies.

In general, De Valois (1977) found that prolonged viewing of a high contrast sine wave grating creates a temporary, band-limited reduction in contrast sensitivity centered about the adaptation grating’s frequency. De Valois determined that the reduction was narrower and more symmetrical than Blakemore and Campbell’s findings suggested. She estimated that the effect disappears at an octave above or below the adaptation grating’s frequency. Interestingly, De Valois discovered that contrast sensitivity enhancement occurs for test grating frequencies that are roughly 2.75 to 3.0 octaves away from the adaptation grating frequency.

Stroymeyer and Julesz (1972) used a masking by noise technique to estimate channel bandwidths. Specifically, masking functions were measured “while varied grating frequency relative to various one-octave-wide bands of noise” (Stroymeyer and Julesz, 1972). Overall, these functions approximated the adaptation curves derived by Blakemore and Campbell (1969). In addition, Stroymeyer and Julesz measured masking as a function of the width of a band of noise centered on test grating frequency. Stroymeyer and Julesz found that masking continues to increase as the band of noise widens up to roughly ± 1 octave. After ± 1 octave, they demonstrated that masking may level off.

The stimuli used consisted of vertical sine wave gratings viewed binocularly with or without
masking noise from 4 meters whose field size was 1 degree high and 2.5 degrees wide. The noise took one of two forms, uniform broadband noise or noise that was filtered in certain ways (e.g., high passed or low passed). Using the method of adjustment psychophysical procedure, participants adjusted gratings to threshold with a one dB-step attenuator and were instructed to bracket the adjustments; that is, to set the contrast of the grating too high and too low before reaching the final adjustment. Stroymeyer and Julesz’s (1972) results suggested that a grating is masked exclusively by noise whose spatial frequencies are close to the test grating’s frequency. Their findings provide strong support for the notion that the visual system contains channels that are selectively tuned to different spatial frequencies.

Henning, Hertz, and Hinton (1981) inferred visual channel bandwidths by assessing the detectability of sine wave gratings superimposed in noise grating backgrounds with the same orientation (i.e., vertical). Test gratings were vertical sinusoidal gratings that occupied a 6 degree square. These gratings were presented against noise gratings consisting of vertical patterns of light and dark stripes of random width and contrast. Noise gratings were either high-pass filtered or low-passed filtered. The high-pass noise gratings consisted of spatial frequencies that had the same mean contrast and were above a certain frequency. The low-pass noise contained a band of frequencies below a particular level and also had equal mean contrasts.

Henning et al. (1981) used a two-interval, forced-choice grating-detection experiment to determine psychometric functions relating the percentage of correct responses to test grating contrast. Each trial consisted of two 1 second observation intervals separated by a 600-ms pause. Participants were required to indicate which of the two intervals contained the test grating during a 750-ms answer interval. Specifically, test gratings of 1, 3, and 6 were used. Henning et al. estimated that the bandwidth for these gratings is approximately 1.15 octaves.

Sachs, Nachmias, and Robson (1971) used a masking approach to determine the width of spatial frequency channels in the visual system. Specifically, psychometric functions were determined for the detection of gratings with and without masks from 100 observations per point. Using the method of constant stimuli, test gratings were presented in random order for 760 ms in a binocularly viewed, 2.25 degree square from a distance of 240 cm. Trials that did not contain gratings (i.e., catch trials) also were included. It should be noted that the participant’s task was a simple detection task. They only had to say whether they saw a shift in
constant luminance, not which stimulus was presented.

In one of the Sachs, et al. (1971) experiments, psychometric functions were determined for a 14 c/deg test grating presented in masking fields whose frequency ranged from 2.8 to 28 c/deg. In order to show that multiple channels exist, they applied a $\chi^2$ test of independence to their results. Moreover, trials on which a participant responded “no” were categorized according to whether a test grating was present and for instances in which it was present, its frequency modulation. They determined that the channel centered at 14 c/deg must be relatively narrow because it does not respond significantly to spatial frequencies differing by a factor of 2 or more. Subsequent experiments with similar results were performed for a broad frequency range of test gratings, each keeping the ratio of the masker and test grating frequencies equal to 2. As in the case of Stroymeyer and Julesz’s study, Sachs and his colleagues found powerful evidence for the notion that the visual system contains channels that are selectively tuned to different spatial frequencies.

In 1980, Legge and Foley studied contrast masking psychophysically using the method of constant stimuli. Specifically, contrast thresholds were determined for a 2.0 c/deg sign wave grating in the presence of various masking gratings. Detection thresholds were measured for 11 masker contrasts and seven masker frequencies spanning $\pm$ one octave from the frequency of the test grating. For their study, stimuli were viewed binocularly at a distance of 114 cm and subtended 10 degrees horizontally and 6 degrees vertically of visual angle on observer’s retinas. A two-alternative, forced-choice (2AFC) method of adjustment procedure was used to estimate psychophysical thresholds. Before a given block of trials, participants used a logarithmic attenuator to adjust grating contrast to a value just above threshold. Trials consisted of two 200 ms stimulus image presentations separated by 750 ms pause. Only one of the images contained the test grating. As a matter of course in Legge and Foley’s experiment, three successful detection responses at one contrast level were followed by a constant decrease in contrast, and one unsuccessful response was followed by an increase in contrast. They defined threshold or the “0.79 proportion correct contrast level” as the mean of the first six contrast peaks and valleys. Legge and Foley (1980) found that channel bandwidths were $\pm$ 1.8 octaves, centered on the frequency of the test grating.

Watson (1982) used a masking approach to estimate the width of spatial frequency channels within the eye-brain system. He determined contrast sensitivity for three stimuli concurrently;
that is, the mask and test gratings first by themselves and then combined. Watson’s used a 1 c/deg test grating and one of seven different masks (1.189, 1.297, 1.414, 1.682, 2.000, 2.378, or 2.828 c/deg) for the first experiment of his study. In a second study, he used a 16 c/deg test grating and masks that were either 20.75, 22.60, 26.90, or 32.00 c/deg. Stimuli for studies were viewed binocularly from 114 cm and 401 cm, respectively. Contrast thresholds were determined using a two-alternative, forced-choice staircase method of adjustment. Watson (1981) defined contrast threshold as “the contrast at which 82% of the responses are correct.” His findings suggest that visual channels are separated in frequency by less than an octave and that the fovea may include at least seven different frequency-selective detectors.

Wilson, McFarlane, and Phillips (1983) used a visual masking paradigm to assess channel bandwidths for a range of test grating frequencies. Specifically, changes in test grating contrast threshold were measured as a function of the spatial frequency of high-contrast masks, oriented at 14.5 degrees relative to the vertical. In general, threshold elevations were observed for test gratings superimposed in masks whose spatial frequency were close (i.e., ± one octave) to that of the test grating. Fourteen separate masking experiments were performed, one for each of 14 test frequencies covering the range 0.25 - 22.0 c/deg in 0.5 octave steps. Overall, the threshold elevation curves that Wilson et al. (1983) determined suggest that there are six channels or spatial frequency filters at work in the fovea.

The masking patterns employed by Wilson et al. were cosine gratings set at 40% contrast (i.e., high-contrast). A randomized, double staircase method of adjustment was used to determine contrast thresholds. Participants viewed circular fields, 4 degrees in diameter, for frequencies 1.0 c/deg and above and 8.0 degrees for lower frequencies. Stimulus observations were monocular, 1 second presentations. Unlike the other detection studies mentioned above, Wilson et al. discovered that visual channel bandwidths appear to be frequency dependent. Specifically, they found that the low frequency channels centered on 0.75 and 1.5 c/deg have bandwidths between 2.0 and 2.5 octaves. In comparison, the bandwidths they found for channels centered on 2.8, 4.4, 8.0, and 16.0 c/deg fell between 1.25 and 1.5 octaves.

Table 1 summarizes some the similarities and differences between the various studies designed to estimate channel bandwidths. Bandwidths in this table represent the full bandwidths at half amplitude of the spatial frequency channels. For all the masking studies, masks were superimposed on test gratings (i.e., simultaneous masking). Although there are dissimilarities in
Table 1, some prevailing conclusions are indicated. One is that spatial frequency pathway estimates range from 0.4 to 2.4 and most estimates lie between 1 and 2 octaves. Channels tuned to high frequencies also seem to have wider bandwidths than pathways tuned to lower frequencies. It is important to note that one advantage of using the two-dimensional Fourier integral is that it can simultaneously extract the sine-wave frequencies and orientations that make up any complex pattern or scene (Kelly, 1994).

Table 1. Summary of Similarities and Differences Among Studies Designed to Estimate Channel Bandwidths

<table>
<thead>
<tr>
<th>Study</th>
<th>Paradigm &amp; Method</th>
<th>Target &amp; Mask Pattern Types</th>
<th>Contrast &amp; Field Size</th>
<th>Number Filters Estimated</th>
<th>Estimated Bandwidth (Octaves)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blakemore &amp; Campbell (1969)</td>
<td>Adaptation adjustment</td>
<td>vertical gratings</td>
<td>high/low-contrast 1.5° square</td>
<td>2 of them: 3 c/deg &amp; 14 c/deg</td>
<td>1.3</td>
</tr>
<tr>
<td>De Valois (1977)</td>
<td>Adaptation adjustment</td>
<td>vertical gratings</td>
<td>high-contrast 4.3 x 5.5° wide</td>
<td>2 of them: 1.19 &amp; 8.0 c/deg</td>
<td>0.7</td>
</tr>
<tr>
<td>Stromeyer &amp; Julesz (1972)</td>
<td>Masking adjustment</td>
<td>vert. grating target both vertical stripe &amp; band of noise masks</td>
<td>high-contrast 1.0 x 2.5° wide</td>
<td>3 of them: 1.77, 5, &amp; 10.0 c/deg</td>
<td>1.0 to 1.5</td>
</tr>
<tr>
<td>Henning, Hertz, &amp; Hinton (1981)</td>
<td>Masking adjustment</td>
<td>vertical gratings</td>
<td>high-contrast 6.0° square</td>
<td>3 of them: 1.0, 3.0, &amp; 6.0 c/deg</td>
<td>1.15 or 2.30</td>
</tr>
<tr>
<td>Sachs, Nachmias &amp; Robson (1971)</td>
<td>Masking constant stim.</td>
<td>vertical gratings</td>
<td>low contrast 2.25° square</td>
<td>3 of them: 2.8, 11.2, &amp; 14.0 c/deg</td>
<td>0.4</td>
</tr>
<tr>
<td>Pantle (1974)</td>
<td>Masking 2AFC constant stim.</td>
<td>vertical square wave gratings</td>
<td>high-contrast 3.2° square</td>
<td>N/A b/c square waves were used</td>
<td>2.4</td>
</tr>
<tr>
<td>Legge &amp; Foley (1980)</td>
<td>Masking 2AFC constant stim.</td>
<td>vertical gratings</td>
<td>high/low-contrast 6.0 x 10.0° wide</td>
<td>1 of them: 2.0 c/deg</td>
<td>1.8</td>
</tr>
<tr>
<td>Watson (1982)</td>
<td>Masking 2AFC constant stim.</td>
<td>vertical gratings</td>
<td>low contrast varied by frequency</td>
<td>2 of them: 1.0 &amp; 16.0 c/deg</td>
<td>0.5</td>
</tr>
<tr>
<td>Wilson, McFarlane, &amp; Phillips (1983)</td>
<td>Masking adjustment</td>
<td>vertical grating target &amp; 14.5° tilted mask</td>
<td>high-contrast circle, 4/ 8° diameter</td>
<td>14 of them: range 0.25 to 22.0 c/deg</td>
<td>1.25 to 2.5</td>
</tr>
</tbody>
</table>

Recognition studies. Just as masking hinders the ability to detect a visual pattern, the ability to recognize a pattern can be reduced by the presence of a masking pattern. For general spatial frequency recognition studies, the observer is required to distinguish between gratings that differ only in spatial frequency. Researchers have found that accuracy depends on contrast. In particular, performance improves as contrast increases for recognition tasks up to a point and
then levels off depending on the spatial frequency of the grating. Performance levels off at higher contrasts for higher spatial frequencies than lower spatial frequencies (Thomas, 1983). Additionally, accuracy is moderated by the number of cycles that the grating contains (Hirsch & Hylton, 1982; Thomas, Gille, & Barker, 1982). For example, Hirsch and Hylton found that performance is reduced when the grating contains fewer than three cycles.

Recognition masking studies in the visual psychophysics literature have focused on stimulus appearance rather than determining channel recognition tuning characteristics for recognition tasks. For example, Harmon and Julesz (1973) used an image quantization procedure to present a band of high frequency noise into an image of President Lincoln. They found that the high spatial frequency noise introduced into the image made the photograph unrecognizable. It is important to note that the use of noise masks only provides an estimate of the critical band. In contrast, masking by a single spatial frequency grating produces a more exact estimate of interactions between particular frequencies (De Valois & De Valois, 1988).

VISUAL CHANNEL MODELS FOR SPATIAL VISION

One convention in vision science is that two visual stimuli are said to be discriminable if an observer can reliably report which of the two has been presented on a given trial 75% (half way between chance and certainty) of the time. Wilson and Bergen (1979) developed a psychophysical technique for measuring the visual mechanisms that filter an image falling on the human retina. Specifically, they used visual masking experiments in which participants had to detect low contrast test stimuli that were superimposed on high contrast masking patterns to model how the eye-brain system processes information about spatial structure and detail.

The masking experiment data led Wilson and his colleagues to a model consisting of six classes of spatial frequency mechanisms in the fovea (Wilson, et al., 1983). The spatial frequency bandwidths of these mechanisms ranged from 2.5 octaves at low spatial frequencies to as narrow as 1.25 octaves at high spatial frequencies. In a later project, Phillips and Wilson (1984) measured masking as a function of the orientation of the masking stimulus relative to the test stimulus. They demonstrated that masking was reduced as the orientation difference between test and mask increased (Wilson, 1995).
A major drawback of Wilson’s model is that it treats the spatial frequency and orientation coding properties of visual scenes as two separate analytic domains. In other words, orientation sensitivity studies typically have been limited to studies concerned with the discrimination of sharply defined bar patterns (Phillips & Wilson, 1984). Spatial frequency sensitivity, on the other hand, has been modeled by filter mechanisms (linear systems analysis) in a one-dimensional manner (Wilson et al., 1983; Yang & Stevenson, 1998).

EVIDENCE FOR TWO-DIMENSIONAL SPATIAL FREQUENCY FILTERING

A review of the visual masking literature (psychophysical experiments) quickly demonstrates that there are both multiple spatial frequency channels and multiple orientation channels. Stated differently, there are mechanisms with band-pass tuning for both orientation and spatial frequency (i.e., two-dimensional spatial frequency tuning). Therefore, the eye-brain system seems to posses the fundamental mechanisms necessary for performing two-dimensional spatial frequency processing on visual information (De Valois & De Valois, 1990).

Some researchers have extended models of discrimination thresholds of sign-wave gratings to the realm of 2-D discrimination thresholds (Caeli, Brettel, Rentschler, and Hilz, 1983). Specifically, Caeli and his associates conducted experiments in which they determined the discriminability between sinusoidal gratings as a function of orientation and spatial frequency differences. Caeli et al. (1983) found that thresholds for orientation are virtually independent of spatial frequency. Specifically, they found highest sensitivity in the horizontal and vertical directions with an average threshold of about ±3° and increasing up to ±7° at the oblique orientations. Additionally, Caeli et al. (1983) found frequency discrimination thresholds to be constant over all orientations and target frequencies: averaging about ±1/8 octave. Their results provide an extensive set of threshold measurements in the two-dimensional spatial domain. Figure 6 shows an illustration of the spatial frequency channels identified in the experiments in the two-dimensional spatial domain.
To date, researchers have used one-dimensional Fourier analysis techniques for sine-wave gratings to model human vision (Tolhurst & Barfield, 1978; Wilson, et al., 1983; Thomas & Gille, 1979). Visual scenes are at least two-dimensional (i.e., left-right and up-down dimensions), and therefore the use of more complex and realistic two-dimensional patterns represents a better approach for determining the band-pass tuning of spatial frequency channels.

**Detection studies.** Kelly and Magnuski (1975) compared the detection of one- and two-dimensional patterns in psychophysical experiments. They examined whether the eye performs a localized two-dimensional Fourier analysis on visual information using two circularly symmetric patterns, produced by a Bessel function and by a circular cosine function. Kelly and Magnuski (1975) found strong evidence in favor of this 2-D filtering. They demonstrated that circular pattern detection could be predicted from their two-dimensional Fourier fundamental amplitudes. Much like detection studies employing 1-D spatial frequency patterns, Kelly and Magnuski (1975) found that the eye is more sensitive to midrange frequencies (4-8 cycles per degree) than to relatively low and high spatial frequency patterns. Mitchel (1976), as well as, Carlson, Cohen, and Gorog (1977) also found that the detection of spatially filtered 2-D patterns
could be predicted effectively from their spatial frequency content.

Further psychophysical evidence that the eye-brain system possesses a set of local two-dimensional spatial frequency channels comes from two masking experiments. Weissten (Weissten, Harris, Berbaum, Tangney, and Williams, 1977; Weissten and Harris, 1980), examined backwards masking using target and background patterns that were dissimilar in appearance but had some common structural components in the 2-D Fourier domain. For example, a spot of light was used for the masking stimulus and a bull’s-eye for the target stimulus that only overlapped a small amount with the mask in one experiment.

There are two critical differences between Weisstein's studies and the experiment performed by the author of this dissertation. First, the Fourier spectra of the spot and bull’s-eye were only partially alike in her study and therefore the degree of masking anticipated could not be determined. In contrast, the target and background stimuli used in the authors’ study were bessel functions, each representing a pure single frequency in the 2-D Fourier domain. This allowed the authors to attribute the degree of observed masking precisely to the Fourier spectra of the targets and masks. Secondly, Weisstein's studies were not concerned with estimating the bandwidths of the 2-D spatial frequency filters in the eye-brain system. In other words, she did not systematically vary the frequency of the mask while holding target frequency constant in order to determine the range of masking frequencies that reduce detection for a given target. Therefore, the results of the author’s study are only directly comparable to the masking studies that have used one-dimensional Fourier analysis techniques for sine-wave gratings to model human vision (Tolhurst & Barfield, 1978; Wilson, et al., 1983; Thomas & Gille, 1979).

**MILITARY TARGET ACQUISITION MODELING**

Early vision models that incorporate knowledge regarding channel interactions, i.e. spatial tuning characteristics, have received broad application in the military and commercial sensor-display engineering communities. The Georgia Tech Vision (GTV) model is perhaps the most popular military vision model available to date. This model outputs predictions of observer search and detection performance for imagery generated by a number of imaging sensors. For example, the model predicts whether a particular sensor-display system will degrade or improve
an observer’s chances of detecting a target embedded in clutter, camouflaged, and/or partially obscured by terrain or vegetation (Doll, 1998).

The U.S. Army ACQUIRE model is a “standard algorithm for modeling the man-in-the-loop search and target acquisition process that is incorporated in combat simulations” (Mazz, 1998). Drawing on a set of inputs that designate target, sensor, and environmental characteristics, ACQUIRE provides a prediction of the probability of acquiring a target. The ACQUIRE model accounts for visual target acquisition performance by a weighting factor, the so-called n50 parameter, which designates the number of cycles-on-target required for 50 percent target acquisition performance (Mazz, 1998). The ACQUIRE n50 parameter stems from a classic image quality parameter known as the Johnson Criteria.

Johnson’s work showed that target acquisition performance can be predicted by calculating the number of resolvable square-wave bar cycles that extend across a target (see Figure 7) (Biberman, 1973). Johnson reported that target acquisition increased with increasing numbers of resolvable bars, and his findings have lead to a commonly used set of criteria for designating the amount of detail needed for target acquisition.

Much like the Johnson Criteria, calculation of the ACQUIRE n50 parameter requires that the number of cycles-on-target be determined for each target. Then, cycles-on-target values are plotted against a standardized data set describing target detection probabilities. A least squares calculation is made to determine the n50 value that best fits the standardized detection data (Mazz, 1998).
Figure 7. An illustration of the cycles-on-target process for perception used in the ACQUIRE model. (Adopted from Biberman, 1973)

Despite its utility as a first-order engineering criteria, there are many theoretical shortcomings associated with the Johnson Criteria, and, therefore, the ACQUIRE model. For example, the Johnson Criteria assumes that target-related spatial information (cycles-on-target) solely determines visual acquisition performance. And it assumes that contrast requirements for near-threshold and supra-threshold targets are constant across all target backgrounds. These types of shortcomings stem from a failure to consider the array of factors that underlie human spatial vision relate, as evident in more contemporary visual psychophysical models.

One important shortcoming of most Johnson Criteria-based models of visual target acquisition is the absence of quantitative means to account for the effects of background “clutter.” Although background clutter has several definitions in the literature, the list of subjective and objective descriptions include: scene complexity, number and density of target-like components, scene busyness, and a metric called signal-to-clutter ratio (SCR). Across these definitions, it is known that increasing levels of background clutter lead to an increasing number of resolvable cycles-on-target needed for target acquisition (Toet, 1996; Mazz, 1998).

Currently, the ACQUIRE model relies on subjective impressions of clutter from expert observers to categorize the type of clutter present in a background and then to adjust the Johnson Criteria (cycles-on-target) accordingly. In one U.S. Army investigation (Mazz, 1998), imagery was categorized into four clutter classifications:

1. no target-competitive clutter,
2. moderate target-competitive clutter,
3. target-competitive clutter in both the background and foreground, and
4. moderate target-competitive clutter with sky in image.

It was found that the ACQUIRE model required considerable adjustment of the n50 parameter in order to account for visual detection across the clutter categories. For example, while it is typical to set n50 at 0.75, prediction of visual detection in high clutter required the n50 parameter to be set at 1.50—a 100% change in the parameter value.

In recent years, studies have shown wide variability in n50 values, often ranging between 0.75
to 4.0. Some researchers have suggested that the variability in the n50 value arises from inconsistent background clutter classifications, which, in turn, arises from an absence of parametric classification schemes for clutter (Mazz, 1998). Thus, there is a need for a unifying description of clutter that could be used to categorize various types of background scenes.

Given the infinite assortment of scene backgrounds encountered in visual target acquisition tasks, it is readily apparent that the development of a universal description of clutter is a difficult, if not seemingly intractable task. Nevertheless, investigations of approaches leading to the development of objective image-based metrics of background clutter is warranted and, perhaps, necessary, for improving the understanding of background clutter effects on target acquisition. A two-dimensional spatial frequency masking framework may be useful for this objective.

Contemporary notions of human spatial vision are based on the idea that perception is the result of operations performed on the outputs of multiple visual channels, each tuned to Fourier components of different spatial frequency and orientation passbands (Olzak and Thomas, 1988). The vision mechanisms responsible in part for the initial processing of a visual stimulus reside in the retina and, thus, their stage of perceptual processing is called early or low-level vision. The information encoded by the early visual mechanisms is transmitted as neural signals to higher-level areas in the brain, where pattern processing algorithms purportedly operate to enable objects to be detected and discriminated from their backgrounds (Cannon, 1995). For my present purposes, one important functional property of the early perceptual mechanisms is that encodings of a target may be masked by encodings of background information, as evidenced by degraded perceptual target detection and recognition performance.

An existing model of human spatial vision that attempts to incorporate the functional properties of visual channels for its accounting of perceptual performance is the Georgia Tech Vision (GTV) model. In general, the GTV model predicts observer visual search and detection performance for imagery generated by various imaging sensors. It provides a prediction of whether a particular sensor-display system degrades or improves observers chances of detecting targets embedded in clutter, camouflaged, and/or partially obscured by terrain or vegetation (Doll, 1998).

There are five major perceptual stages in the GTV model (Figure 8). The first stage is a front-end module that mimics early vision processing in the eye-brain system. This stage addresses
with two aspects of light entering the eye: luminance and color. This first stage of the GTV model accounts for receptor pigment bleaching, pupil dilation, luminance adaptation, and opponent color processes.

Figure 8. Graphic showing the sequence of the five modules in the GTV model. (Adapted from Doll, 1998)

The next two stages are called the preattentive and attentive modules. These stages mimic the feature-extraction processes that occur within the peripheral and foveal (central) fields of vision, and they predict the saliency of targets in those visual regions. It is important to note that these GTV model stages assess “channels of information specific to spatial frequency and orientation, based on the Wilson model” (Doll, 1998). In other words, the GTV model uses image-based Fourier representations (although one-dimensional) of targets and backgrounds to assess their perceptual interactions within the visual channels.

The fourth stage of the GTV model is a selective attention/training module. Using the outputs of the preattentive and attentive modules, it divides scene elements into objects. A discrimination function then is formulated and used to extract candidate targets (so-called “blobs”) from the background clutter.
The final stage of the GTV model calculates measures of visual search and discrimination performance for every identified candidate target. Probabilities of target detection (as well as false alarms) during short duration visual glimpses are provided, along with overall probabilities for detection and false alarm rates for visual search times.

Although quite comprehensive in its coverage of theoretical visual processing issues, the GTV model (and others) is based on a one-dimensional analytical approach for characterizing visual stimuli and their subsequent perceptual processing. This analytic foundation is commonly found in the visual psychophysics literature, where human visual sensitivities to spatial frequency patterns are investigated through detection and discrimination judgments of one-dimensional square and sine-wave patterns (Phillips and Wilson, 1984; Wilson et al., 1983; Yang and Stevenson, 1998). However, since real-world scenes are at least two-dimensional in nature, the predictions of visual models predicated on these one-dimensional gratings studies may not generalize adequately to visual performance with real-world images. In other words, a major drawback of the Georgia Tech Vision model is that it treats spatial frequency and orientation coding properties of visual scenes as two separate analytic domains. One distinct advantage of a two-dimensional analytic approach is that information regarding spatial frequency and orientation (and their interactions) can be obtained simultaneously (Kelly, 1994; Kelly and Magnuski, 1975).

**TWO-DIMENSIONAL VISUAL PROCESSING**

The present work was motivated by an interest to improve the understanding of perceptual processes involved in real-world visual target acquisition, particularly those processes underlying the visual detection and recognition of targets located within real-world scenes (backgrounds). This area of interest, therefore, inherently involves issues associated with two-dimensional spatial vision, since real-world scenes represent luminance variations over at least horizontal and vertical spatial extents of a scene. It is appreciated that real-world scenes may consist of additional energy-distribution dimensions, such as depth and time, but our focus is restricted to extending knowledge about spatial vision by generalizing one-dimensional spatial vision considerations directly into the two-dimensional spatial frequency domain.
As discussed elsewhere in this document, contemporary notions of human spatial visual are based on the idea that perception is the result of operations performed on the outputs of multiple visual channels, each tuned to Fourier components of different spatial frequency and orientation passbands (Olzak and Thomas, 1988). The vision mechanisms responsible for initially processing any stimulus reside in the retina and, thus, their stage of perceptual processing is called early or low-level vision. The information encoded by the early visual mechanisms is transmitted as neurological signals to higher-level processes in the brain, where pattern processing algorithms purportedly operate to enable objects to be detected and discriminated from their backgrounds (Cannon, 1995). For our present purposes, one important functional property of the early perceptual mechanisms is that the encoding of target information may be masked by encodings of background information, as evidenced by degraded perceptual target detection and recognition performance.

Some existing models of human spatial vision have attempted to incorporate the functional properties of visual channels for their accounting of perceptual performance. For example, the Georgia Tech Vision (GTV) model predicts observer search and detection performance for imagery generated by various imaging sensors. The GTV model predicts whether a particular sensor-display system degrades or improves observers chances of detecting a target embedded in clutter, camouflaged, and/or partially obscured by terrain or vegetation (Doll, 1998). One drawback, however, of current spatial vision models is that they treat the two-dimensional properties of a scene (i.e., Fourier domain spatial frequency and orientation) as separate analytic issues, because they largely employ one-dimensional approaches to characterizing visual stimulus and the perceptual processing of those stimuli. For example, orientation sensitivity typically is investigated in studies concerned with the discrimination of sharply defined one-dimensional bar patterns (Phillips and Wilson, 1984). Spatial frequency sensitivity, on the other hand, typically is studied by assessing the sensitivities of the filter mechanisms to one-dimensional gratings (Wilson et al., 1983; Yang and Stevenson, 1998). Since real-world scenes are at least two-dimensional in nature, we purport that an opportunity to improve our understanding of spatial vision may exist through the use of two-dimensional targets. One advantage of a two-dimensional approach is that information about spatial frequency and orientation is obtained simultaneously (Kelly, 1994; Kelly and Magnuski, 1975).

A two-dimensional real-world scene can be considered in the Fourier domain as having many sine-wave components with viewable modulation located at many spatial frequency and
orientation values. The visibility of these sine-wave components depends on their encoding by the visual channels. If a target within the two-dimensional real-world scene is proximal to background objects in the scene, it is reasonable to suspect that masking phenomena can degrade its perception. Hence, the present work investigates the existence and functional properties of visual masking with two-dimensional stimuli.

Existing knowledge about spatial visual mechanisms primarily relates to determination of the number of visual channels, as well as the spatial frequency and orientation selectivity of those visual channels. Far more information is available about the number of visual channels and their spatial frequency selectivities than about their orientation response properties (Caeli, et al., 1983). For real-world scenes, it is known that Fourier spectra tend to be symmetrical; meaning that the sine-wave components at any spatial frequency have many orientations. Thus, it is reasonable to investigate two-dimensional visual processes by probing visual channels centered at specific spatial frequency passbands and spanning all possible orientations.

For this research, novel stimuli were employed to investigate the existence of two-dimensional visual masking within specific visual channels. These stimuli were radial sine waves, as shown in Figure 9a. Each sine wave consists of modulation at a specific spatial frequency and at all orientations simultaneously. This property is evident by inspection of the corresponding two-dimensional Fourier amplitude spectra shown in Figure 9b. The radial sine-wave stimuli are defined as

\[
L(x, y) = b_0 + b_1 \sin \left[ 2\pi \sqrt{\left( \frac{f_x(x - x_c)}{N_x} \right)^2 + \left( \frac{f_y(y - y_c)}{N_y} \right)^2} \right]
\]  

(Eq. 3)

in which

\[L(x, y)\] denotes the luminance of the radial sine wave at spatial coordinate \((x, y)\),
\[b_0, b_1\] denote intensity offset and gain parameters, respectively, for the radial sine-wave,
\[x_c, y_c\] denote the center coordinates for radial sine-wave,
\[f_x, f_y\] denote the cardinal axes spatial frequencies for the radial sine-wave, and
\[N_x, N_y\] denote the number of samples along the x and y cardinal axes, respectively.
The radial sine-wave patterns defined by Eq. 3 provide a means to probe two-dimensional perceptual processing in human vision. In particular, these patterns can be used to investigate the existence of two-dimensional masking phenomena. Evidence on the existence of two-dimensional masking would support the notion that target detection and recognition performance in real-world scenes may be affected by proximal background objects. Additional knowledge, then, would be desired on the potential magnitude of two-masking effects in real-
world scenes. This latter information might be useful for designing and interpreting future studies of two-dimensional masking processes in real-world scenes. Given these directions, four psychophysical experiments are reported below that attempt to establish the existence of two-masking effects (using the radial sine-wave stimuli) and the potential magnitude of masking effects in real-world targets.
EXPERIMENT 1

This experiment investigated the existence of visual masking effects with two-dimensional spatial frequency grating stimuli. The objective of the experiment was to assess whether or not visual masking occurs with two-dimensional radial sine-wave targets and, if so, to uncover some information on the functional nature of two-dimensional masking effects. Thus, the present experiment sought to extend classic visual masking findings from simple one-dimensional gratings into the realm of two-dimensional visual stimuli. Visual detection and recognition performance were assessed.

METHOD

Participants
Fourteen individuals (8 females; 21.7 average years of age) were recruited from civil servant and military personnel working at the U.S. Army Aberdeen Proving Grounds to participate in the experiment. Participants were required to pass a screening test for normal spatial vision abilities (Titmus II Vision Tester; 20/20 acuity, corrected or uncorrected) before final selection for the experiment.

Apparatus
The experiment was conducted in a dimly lighted (i.e., < 1 lux) room, which was bisected by a wall containing a square viewing port (29 cm diagonal). The viewing port served to direct participant’s line-of-sight to the stimulus display screen and to eliminate unwanted visual stimulation. One-half of the testing room contained the observer’s station, consisting of a table with a chin rest and a height-adjustable chair. The other one-half of the testing room contained the experimenter station and stimulus display system. The experimenter station consisted of a chair, desk, and low-intensity desk lamp. The experimenter operated the stimulus display system and recorded observer responses while seated at the station during the data collection sessions.

Observers viewed the stimulus display binocularly from 4.57 m. Viewing distance remained constant throughout the data collection session by requiring observers to use the chin rest. The height of the chin rest remained fixed (height adjustable chair) to ensure proper line-of-sight and stimuli size.
A desktop computer (Apple Computers, Model: Power Macintosh 8500/200) was used to generate and present the visual stimuli during the data collection sessions. All stimuli were rendered on a high-quality, color cathode ray tube (CRT) monitor (Mitsubishi, Model: Diamond Pro 87TXM; 17-inch diagonal viewable raster). The CRT monitor was operated at 1024 by 768 by 24-bit pixel addressability and 72 Hz temporal refresh rate. The CRT had a large-area white-field luminance range of 0.05 to 72.5 cd/m\(^2\) and a luminance gamma of 2.3 over the combined 8-bit RGB video channels.

![Illustration of testing room used in Experiment 1.](image)

**Figure 10. Illustration of testing room used in Experiment 1.**

**Stimuli**

Target stimuli were two-dimensional radial sine-wave patterns (see Eq. 3) modulated by a two-dimensional Gaussian envelope (i.e., target luminance profile truncated by multiplication with a Gaussian function) subtending 4.0 degrees of visual angle in diameter (see Figure 11). The
spatial frequency of the target stimuli varied across five levels: 2, 4, 8, 16, and 32 cycles per degree of visual angle.

Masking stimuli also were two-dimensional radial sine wave patterns generated within a 6.0-degree of visual angle square area on the stimulus display system (see Figure 12). The spatial frequency of the masking stimuli varied across nine levels; that is, four spatial frequencies above the target frequency, four spatial frequencies below the target, and one at the spatial frequency of the target stimulus. The only exception to the range of masking frequencies occurred for the 32 cycle per degree target, which was examined with only five masking frequencies (i.e., four below that target frequency and one at the target frequency) because the stimulus display system was not capable of presenting spatial frequency patterns greater than 32 cycles per degree. Table 2 lists the masking frequencies used with each target.

Figure 11. Example two-dimensional radial sine wave targets used in Experiment 1. (Part A-2 c/d. Part B-4 c/d. Part C-8 c/d.)
Figure 12. Example two-dimensional radial sine wave masks used in Experiment 1. (Part A-1.5 c/d. Part B-2.5 c/d. Part C-12 c/d.)

<table>
<thead>
<tr>
<th>Target Frequency</th>
<th>Masking Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.5, 0.7, 0.9, 1.3, <strong>2.0</strong>, 2.8, 4.3, 5.4, 7.2</td>
</tr>
<tr>
<td>4</td>
<td>1.2, 1.8, 2.5, 3.3, <strong>4.0</strong>, 5.0, 6.5, 8.0, 9.6</td>
</tr>
<tr>
<td>8</td>
<td>1.0, 2.4, 3.3, 3.9, <strong>8.0</strong>, 12.7, 14.8, 17.2, 19.6</td>
</tr>
<tr>
<td>16</td>
<td>4.4, 5.3, 8.6, 12.7, <strong>16.0</strong>, 19.0, 23.8, 29.6, 33.5</td>
</tr>
<tr>
<td>32</td>
<td>7.3, 9.3, 14.2, 22.6, <strong>32.0</strong></td>
</tr>
</tbody>
</table>

Table 2. Spatial Frequency of Masks used in Experiment 1

For the stimulus presentations, the target and masking stimuli were combined using an interleaved pixel-mixing technique (see Figure 13). That is, even-numbered pixels within the combined image bit map rows represented the target stimulus, whereas odd-numbered pixel within the bit-map rows represented the masking stimulus. Additionally, the target and masking sine-wave patterns had the same phase. Thus, when the spatial frequency of the masking stimulus matched that of the target stimulus, the appearance of the displayed stimulus largely resembled the masking stimulus, since it had larger spatial extent than the target stimulus.
Masking spatial frequencies were determined through a pilot visual detection experiment with six observers. The pilot study findings provided a range of masking spatial frequencies for each target stimulus. The lowest and highest points for the mask spatial frequency range were selected on the basis that the target in the presence of the mask was detected 90% of the time. Once the two extremes of the masking range were identified, three additional equally spaced points above and three points below the target spatial frequency were selected.

Luminance modulation of the target and masking stimuli was 0.4, as defined by Eq. 3. To ensure that our photometric calibration of the display monitor and stimulus generation program was valid, a 0.7 cycle per degree radial sine-wave test pattern was generated within a 6 degree of visual angle square area centered on the display screen. A luminance-calibrated photometer (Minolta, Model: CS-100) was used to measure the light and dark rings of the radial sine-wave test pattern. All stimuli had the expected modulation.

**Experiment Design**

Data collection was organized by a two-factor, incomplete-blocks design. One factor was target spatial frequency, which varied over five levels (i.e., 2, 4, 8, 16, and 32 cycles per degree of visual angle). The second factor was masking spatial frequency, which varied across nine levels for four targets and across five levels for one target (see Table 3). All participants received all treatment conditions, which were blocked by target spatial frequency. Within a block, the presentation order of the masking stimuli and the replications were randomized uniquely for each observer.
Two measures of human visual performance were recorded during the experiment: target detection and target recognition. Target detection refers to the extent of correctly detecting a target under a particular viewing condition (i.e., target frequency and masking frequency combination). Observers were required to indicate the presence or absence of a target on each trial. Observer’s responses were recorded as either correct or incorrect; thus, the proportion of correct responses across all trials for a particular viewing condition defined the target detection score. Note that the probability of random correct detection (guessing) with the two response alternatives on each trial was $p = 0.50$.

Target recognition refers to the extent of correctly recognizing a particular target, given that correct detection occurred, under each viewing condition. Observers were required to identify (i.e., recognize) the spatial frequency of each target following their detection judgment. As explained below, observers were shown two alternatives for their recognition judgments on each trial. Recognition responses were recorded as either correct or incorrect; thus, the proportion of correct recognition responses across all trials for a particular viewing condition defined the recognition score. The probability of random correct recognition with the two response alternatives on each trial was $p = 0.50$.

Two sampling variables were used during data collection to improve the accuracy of the visual detection and recognition findings. First, measurements of target detection and recognition were replicated four times for each viewing condition and participant. Second, catch trials presenting only masking fields were interspersed randomly with the actual trials—20% of all trials were catch trials.
Table 3. Design Summary for Experiment 1

<table>
<thead>
<tr>
<th>Design Characteristic</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent Variables:</strong></td>
<td></td>
</tr>
<tr>
<td>Target Spatial Frequencies</td>
<td>5</td>
</tr>
<tr>
<td>Mask Spatial Frequencies</td>
<td>8</td>
</tr>
<tr>
<td><strong>Sampling Variables:</strong></td>
<td></td>
</tr>
<tr>
<td>Replications</td>
<td>4</td>
</tr>
<tr>
<td>Catch Trials</td>
<td>8</td>
</tr>
<tr>
<td>Participants</td>
<td>14</td>
</tr>
<tr>
<td><strong>Data Set Characteristics:</strong></td>
<td></td>
</tr>
<tr>
<td># Viewing Conditions</td>
<td>48</td>
</tr>
<tr>
<td># Observations per Participant</td>
<td>192</td>
</tr>
<tr>
<td># Total Observations</td>
<td>2688</td>
</tr>
</tbody>
</table>

**Procedure**

Before beginning the experiment, each participant read and signed a consent form, as well as, read a set of instructions for the experiment. The instructions detailed the method and type of stimulus presentations, the stimulus-response protocol, and trial and rest-break procedures. Participants familiarized themselves with the stimulus-response protocol by viewing example targets and target/mask combinations presented in the instruction packet and on the stimulus display monitor. In terms of practice, each participant was required to identify five randomly presented targets in succession before beginning the experiment. Following a review of the instructions with the experimenter to ensure their accurate understanding of the procedures, the participant began the actual experiment.

Each data collection session involved an observer viewing a series of stimulus presentation trials. At the start of each trial, the observer viewed a uniform gray-colored field on the stimulus display screen. The observer verbally indicated “READY” to initiate the stimulus presentation, which was controlled by the experimenter. Immediately following the stimulus presentation, the display screen returned to the uniform gray-colored field while the observer verbally indicated the target detection and recognition responses and prepared for the next trial.

Each trial consisted of three sequentially presented stimulus fields, temporally separated by uniform gray-colored inter-stimulus fields. The first two stimulus fields contained one unmasked
target each, while the third stimulus field contained the combined target and mask stimulus (or just a mask stimulus on a catch trial). The stimulus and inter-stimulus fields were rendered for two seconds each. Observers made their detection and recognition judgments with respect to the third stimulus field. That is, immediately following the termination of the third stimulus field, observers indicated if a stimulus was detected in that field and, if a stimulus was detected, they indicated if the stimulus had the same spatial frequency as the unmasked target shown in either the first or second stimulus fields. One of the unmasked targets always matched the spatial frequency of the target in the third stimulus field, but the order of its appearance in either the first or second stimulus fields was randomized across trials.

Total time to complete a data collection session averaged 55 minutes per participant (including rests breaks). Within each data collection session, trials were blocked by target spatial frequency. Blocking was counterbalanced across participants using a Latin-square procedure. During a block of trials, masking frequency levels and replications were presented in a unique random order for each observer. Observers were given the option to take a five-minute break after every two blocks of trials.

RESULTS

Target Detection
Correct target detection scores for each viewing condition were calculated by comparing observers “Present” and “Absent” responses to the known target occurrence on each trial. Correct detection judgments of a target were coded as 1, whereas incorrect judgments were coded as 0. The probability (proportion) of correct target detection was computed as the sum of the coded detection judgments divided by the maximum number of observations per viewing condition per observer. These data were averaged across participants.

Prior to analyzing the stimulus data, performed adherence to the trial procedures was assessed by determining the error rate for the catch trials. An error rate of 2.89% was observed, indicating the absence of guessing and a clear understanding of experimental procedures on the part of the participants.
Figure 14 shows the average correct detection probabilities for the 2 and 4 cycle per degree targets. Considering the data trends for both targets, it is clear that correct detection decreases as the spatial frequency of the masking stimulus approaches that of the target stimulus. The greatest suppression of target detection occurs when the masking stimulus has the same spatial frequency as the target stimulus. As the spatial frequency of the masking stimulus increasingly differs from the spatial frequency of the target, average correct detection probabilities improve. These data trends are consistent with previously reported masking effects for one-dimensional stimuli and, thus, provide evidence that visual masking processes occur with two-dimensional sine-wave patterns.

![Figure 14. Average correct detection probabilities for 2 and 4 cycle per degree radial sine-wave targets as a function of two-dimensional mask spatial frequency. (Error bars represent +/- 1 standard error of the mean.)](image)

Correct detection probabilities for the masked 8, 16, and 32 cycle per degree targets are shown in Figure 15. In general, these data trends are similar to those observed for the 2 and 4 cycle per degree targets. Specifically, correct detection decreases as the masking frequency approaches the target frequency.
Figure 15. Average correct detection probabilities for 8, 16, and 32 cycle per degree radial sine-wave targets as a function of two-dimensional mask spatial frequency. (Error bars represent +/- 1 standard error of the mean.)

To further examine the functional properties of the two-dimensional masking effects observed in the present experiment, it is meaningful to consider the magnitude and bandwidth of masking across the target stimuli. Since the target stimuli were selected to tap visual processing by separate visual channels, a comparison of masking effects across targets may provide insights to the functional similarities and differences in the human visual system.
Figure 16. Figure 16 combines the data shown in Figures 14 and 15 on a common scale to facilitate the comparison of masking effects across the visual channels stimulated by the target stimuli.

It is apparent that the relative magnitude of peak masking (i.e., when the target and mask frequency are the same) decreases as target spatial frequency increases. In other words, the masking effects appear to be stronger at lower target spatial frequencies than at higher target spatial frequencies. Additionally, it is apparent from Figure 16 that the bandwidth of the masking increases with increasing target spatial frequencies. This latter observation is consistent with previously reported masking phenomena using simple one-dimensional stimuli; that is, masking appears to occur within at least one octave of the target spatial frequency.

Target Recognition
Correct target recognition scores for each viewing condition were calculated in a manner similar to that used for the detection probabilities above, with the exception that only recognition responses from trials having correctly detected targets were considered. With this treatment,
the effects of the independent variables on recognition performance were indexed after removing variance due to chance detection performance.

Figure 17 shows the average probability of correct recognition for the 2 and 4 cycle per degree targets, whereas Figure 18 shows the same results for the 8, 16, and 32 cycle per degree targets. The data trend for each target shows that correct recognition performance decreases as the masking spatial frequency approaches that of the target. However, when the spatial frequency of the mask matched that of the target, correct recognition performance peaked. This latter observation is expected, since a target masked by the same spatial frequency radial sine wave presents a large and clearly recognizable pattern.

Figure 17. Average correct recognition probabilities for 2 and 4 cycle per degree radial sine-wave targets as a function of two-dimensional mask spatial frequency. (Error bars represent +/- 1 standard error of the mean.)
Additionally, the data trends suggest that the bandwidth of masking on correct recognition performance may be asymmetrical. That is, masking appears to occur with 0.5 octave below the center frequency of the target. However, masking remains noticeable well beyond 0.5 octave above the center frequency of the target, especially for the higher spatial frequency values examined. This observed asymmetry was not expected and may indicate a finding unique to two-dimensional spatial vision.

**DISCUSSION**

Visual masking paradigms in large part have been used to determine the tuning characteristics of spatial frequency channels involved in the detection and recognition processes. In general, the purpose of these experiments is to determine the amount that the detection and/or recognition thresholds for one stimulus are changed by a second stimulus. Evidence of an
interaction among two or more channels is indicated by shifts in observed target detectability or recognizability thresholds. Inversely, the absence of an interaction is taken as evidence that the stimuli are processed by separate and independent channels (De Valois & De Valois, 1988). Ultimately, these studies work to establish the conditions under which performance is reduced for people on visual detection and recognition tasks.

EVIDENCE FOR TWO-DIMENSIONAL SPATIAL FREQUENCY FILTERING

A review of the visual masking literature (psychophysical experiments) quickly demonstrates that there are both multiple spatial frequency channels and multiple orientation channels. Stated differently, there are mechanisms with band-pass tuning for both orientation and spatial frequency (i.e., two-dimensional spatial frequency tuning). Therefore, the eye-brain system seems to possess the fundamental mechanisms necessary for performing two-dimensional spatial frequency processing on visual information (De Valois & De Valois, 1988). To date, researchers have used one-dimensional Fourier analysis techniques for sine-wave gratings to model human vision (Tolhurst & Barfield, 1978; Wilson, et al., 1983; Thomas & Gille, 1979). Visual scenes are at least two-dimensional (i.e., left-right and up-down dimensions), and therefore the use of more complex and realistic two-dimensional patterns represents a better approach for determining the band-pass tuning of spatial frequency channels.

Detection studies. Kelly and Magnuski (1975) compared the detection of one- and two-dimensional patterns in psychophysical experiments. They examined whether the eye performs a localized two-dimensional Fourier analysis on visual information using two circularly symmetric patterns, produced by a Bessel function and by a circular cosine function. Kelly and Magnuski (1975) found strong evidence in favor of this 2-D filtering. They demonstrated that circular pattern detection could be predicted from their two-dimensional Fourier fundamental amplitudes. Much like detection studies employing 1-D spatial frequency patterns, Kelly and Magnuski (1975) found that the eye is more sensitive to midrange frequencies (4-8 cycles per degree) than to relatively low and high spatial frequency patterns. Mitchel (1976), as well as, Carlson, et al., (1977) also found that the detection of spatially filtered 2-D patterns could be predicted effectively from their spatial frequency content.

Further psychophysical evidence that the eye-brain system possesses a set of local two-
dimensional spatial frequency channels comes from two masking experiments. Weissten (Weissten et al., 1977; Weissten & Harris, 1980), examined backwards masking using target and background patterns that were dissimilar in appearance but had some common structural components in the 2-D Fourier domain. For example, a spot of light was used for the masking stimulus and a bull’s-eye for the target stimulus that only overlapped a small amount with the mask in one experiment.

There are two critical differences between Weisstein’s studies and the experiment performed by the author of this report. First, the Fourier spectra of the spot and bull’s-eye were only partially alike in her study and therefore the degree of masking anticipated could not be determined. In contrast, the target and background stimuli used in the author’s study were Bessel functions, each representing a pure single frequency in the 2-D Fourier domain. This allowed the authors to attribute the degree of observed masking to the Fourier spectra of the targets and masks. Secondly, Weisstein’s studies were not concerned with estimating the bandwidths of the 2-D spatial frequency filters in the eye-brain system. In other words, she did not systematically vary the frequency of the mask while holding target frequency constant in order to determine the range of masking frequencies that reduce detection for a given target. Therefore, the results of the author’s study are only directly comparable to the masking studies that have used one-dimensional Fourier analysis techniques for sine-wave gratings to model human vision (Tolhurst and Barfield, 1978; Wilson, et al., 1983; Thomas and Gille, 1979).

Table 4 summarizes some of the similarities and differences between the various masking studies and the findings in this report designed to estimate channel bandwidths. Bandwidths in this table represent the full bandwidths at half amplitude of the spatial frequency channels. Although there are dissimilarities in Table 4, some prevailing conclusions are indicated. One is that spatial frequency pathway estimates range from 0.4 to 2.4 and most estimates lie between 1 and 2 octaves. Channels tuned to high frequencies also seem to have wider bandwidths than pathways tuned to lower frequencies. Lastly, the magnitude of masking was substantially stronger for the Bessel function targets than sine-wave gratings for targets above 8 cycles per degree. The suggests that masking studies using 1-D representations of Fourier components may underestimate the strength of masking that occurs for high spatial frequency visual target information in response to higher frequency background information. It is important to note that one advantage of using the two-dimensional Fourier integral is that it can simultaneously extract the sine-wave frequencies and orientations that make up any complex pattern or scene (Kelly,
Table 4. Summary of similarities and differences among masking studies designed to estimate channel bandwidths and Experiment 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Target &amp; Mask Pattern Types</th>
<th>Stimuli Contrast</th>
<th>Bandwidth (in Octaves)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pantle (1974)</td>
<td>vertical gratings</td>
<td>high-contrast</td>
<td>2.4</td>
</tr>
<tr>
<td>Legge &amp; Foley (1980)</td>
<td>vertical gratings</td>
<td>high-contrast</td>
<td>1.8</td>
</tr>
<tr>
<td>Wilson, McFarlane, &amp; Phillips (1983)</td>
<td>vert. grating target &amp; 15 degree tilted mask</td>
<td>high-contrast</td>
<td>1.25 to 1.5</td>
</tr>
<tr>
<td>Henning, Hertz, &amp; Hinton (1981)</td>
<td>vertical gratings</td>
<td>high-contrast</td>
<td>1.15 or 2.30</td>
</tr>
<tr>
<td>Sachs, Nachmias &amp; Robson (1971)</td>
<td>vertical gratings</td>
<td>low contrast</td>
<td>0.4</td>
</tr>
<tr>
<td>Watson (1982)</td>
<td>vertical gratings</td>
<td>low contrast</td>
<td>0.5</td>
</tr>
<tr>
<td>Present Investigation:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment 1</td>
<td>Bessel functions</td>
<td>high-contrast</td>
<td>1.0 to 1.3</td>
</tr>
</tbody>
</table>

**Recognition studies.** For general spatial frequency recognition studies, the observer is required to distinguish between gratings that differ only in spatial frequency. Researchers have found that accuracy depends on contrast. In particular, performance improves as contrast increases for recognition tasks up to a point and then levels off depending on the spatial frequency of the grating. Performance levels off at higher contrasts for higher spatial frequencies than lower spatial frequencies (Thomas, 1983). Additionally, accuracy is moderated by number of cycles that the grating contains (Hirsch and Hylton, 1982; Thomas, et al., 1982). For example, Hirsch and Hylton found that performance is reduced when the grating contains...
fewer than three cycles.

Just as masking hinders the ability to detect a visual pattern, the ability to recognize a pattern can be reduced by the presence of a masking pattern. Recognition masking studies in the visual psychophysics literature have focused on stimulus appearance rather than determining channel recognition tuning characteristics for recognition tasks. For example, Harmon and Julesz (1973) used an image quantization procedure to present a band of high frequency noise into an image of President Lincoln. They found that the high spatial frequency noise introduced into the image made the photograph unrecognizable. It is important to note that the use of noise masks only provides an estimate of the critical band. In contrast, masking by a single spatial frequency grating produces a more exact estimate of interactions between particular frequencies (De Valois and De Valois, 1990). Experiment 1 in this report used single pure 2-D spatial frequency patterns to demonstrate recognition masking.

Interestingly, it was found in Experiment 1 that the probability of correct recognition was substantially higher when the spatial frequency of the mask fell within an octave or so of the targets’ spatial frequency and was appreciably lower when the mask frequency was 2 octaves away. This represents a great departure from the detection results found in Experiment 1 because the probability of detection was significantly lower when the mask frequency fell within an octave of the target every time. This finding does however agree with what Harmon and Julesz (1973) found for the altered image. It was determined that the masking exhibited in their study was caused by spatial frequency components within 2 octaves above the fundamental frequency of the picture. This finding may describe a supra-threshold phenomenon that applies to the discriminability of pure 2-D patterns, that is, a space domain issue rather than a frequency domain issue. Stated differently, perhaps the difference threshold is substantially larger for the 2-D Bessel function than for a 1-D sine-wave grating set at a particular frequency. In the case of the difference threshold for the 2-D pattern, masking rings falling within an octave of the target could have simply bolstered the accuracy of judgements due to a large difference threshold (Olzak & Thomas, 1988).

In general, the findings of Experiment 1 are comparable to existing knowledge on visual masking effects of target detection obtained from studies using one-dimensional spatial frequency patterns (Henning et al., 1981; Legge and Foley, 1980; Wilson et al., 1983; Watson, 1982). Specifically, Experiment 1 demonstrated that detection probabilities are reduced when
the spatial frequency of a masking stimulus approaches that of the target. Thus, the present findings provide compelling evidence for the existence of two-dimensional spatial frequency masking. Moreover, these findings indicate that two-dimensional masking effects on target detection do not differ radically from expectations based on findings using one-dimensional stimuli.

Experiment 1 also demonstrated that masking effects can influence target recognition performance. The most notable similarly between the present findings and those reported previously in the literature is the decrease in correct recognition as the spatial frequency of the target and mask approach one another. However, the present findings show a strong asymmetry in the masking effects on correct recognition performance. That is, correct recognition decreases as the mask frequency approaches about 0.5 octave below the target frequency. But correct recognition decreases as the mask frequency approaches well more than 1.0 octaves above the target frequency. Additionally, this observed asymmetry appears more pronounced for targets of higher spatial frequencies examined. It is suggested that the observed asymmetry may reflect a unique property of two-dimensional spatial vision.

Given this evidence for the existence of two-dimensional spatial frequency masking, as well as the indication that two-dimensional spatial vision may involve previously unreported properties, further investigation of the functional properties of two-dimensional visual masking phenomena is warranted. These issues available for further investigation are beyond the scope of this dissertation. Instead, the author is interested in one very basic question. What is the magnitude of two-dimensional masking upon detection and recognition of real-world targets? This concern received priority from the author since the demonstration of the existence of a two-dimensional masking effect does not establish that it has a substantial impact on real-world target visibility. Therefore, the following experiment examined the potential magnitude of two-dimensional masking on real-world target detection and recognition. Specifically, a demonstration of some functional effects on real-world target perception due to deliberate suppression of selected two-dimensional spatial frequency structures was attempted.

**EXPERIMENT 2**
The purpose of this experiment was to examine the potential effects of two-dimensional spatial frequency masking on real-world target detection and recognition. This experiment represents a preliminary investigation of two-dimensional spatial frequency masking phenomena, which will be continued in subsequent experiments in this dissertation.

It is well known that visual masking occurs when the spatial frequency of a mask falls close to that of a target. Although the precise neurological mechanisms of visual masking are not well understood at this time, the effects of masking on observer performance qualitatively resemble a loss of visual sensitivity to the modulation of spatial frequency components comprising a target image. Said differently, the effects of visual masking on performance resemble a suppression of modulation within the Fourier domain of a target. With this perspective, the effects of two-dimensional masking can be simulated by deliberate suppression of modulation at appropriate Fourier components. Given the approach adopted for Experiment 1, the appropriate Fourier components are represented by visual channels located at specific center spatial frequencies and spanning all possible orientations.

In this experiment, visual detection and recognition of military ground vehicles was examined. The images were prepared digitally using a radial notched filter, centered at the visual channels probed in Experiment 1. The filter was circularly symmetric in the two-dimensional Fourier plane and it was programmable in terms of gain and bandwidth. Using the filter, ground vehicle targets were produced that varied in terms of the amount of modulation suppression within selected visual channels. Thus, this preliminary experiment provides a controlled investigation of the effects of two-dimension spatial frequency suppression in real-world targets on visual detection and recognition performance.

**METHOD**

**Participants**
Three individuals (1 female, 21 to 30 years of age) were recruited from the student population at Virginia Tech to participate in the experiment. Participants were required to pass a screening test for normal spatial vision (Vistech Contrast Test and Landolt C acuity test) capabilities before final selection for the experiment. None of the participants possessed expert knowledge of the military vehicles used as stimuli in the experiment.
Apparatus

The experiment was conducted under nearly identical viewing conditions as used in Experiment 1; however, a different laboratory facility hosted the data collection activities. The experiment was conducted in a dimly lighted (i.e., < 1 lux) room that was bisected by a partition containing a square viewing port (16 cm diagonal). One-half of the testing room contained the observer’s station, whereas the other one-half of the testing room contained the experimenter’s station and stimulus display system. The experimenter operated the stimulus display system and recorded observer responses while seated at the station during the data collection sessions.

Observers viewed the stimulus display binocularly from 7 feet. Viewing distance was maintained throughout data collection by requiring observers to use the chin rest. The height of the chin rest remained fixed (height adjustable chair) to ensure proper line-of-sight and stimuli size. The same desktop computer and stimulus display system reported for Experiment 1 were used in the present experiment.

Stimuli

The stimuli used in this experiment were based on photographs of U.S. military ground vehicles. Photographs of the vehicles were taken under bright, but overcast daylight conditions (in July, 1999) at Aberdeen Proving Grounds, Maryland. All photographs were taken with the same commercial-grade 35-mm single lens reflex camera (Pentax, Model: 100-FX, 55 mm f1.4 lens; automatic internal light meter; Eastman Kodak 100 color print film). The film was developed as 8.5 by 11 matt-finish prints by a commercial Kodak film processing service. The photographic prints then were converted into digital format using a flatbed scanner (ColorScan, Model: 100x). The resulting digital images had 600 RGB pixels per inch, each 24-bits deep, in the horizontal and vertical dimensions. Each digital image was stored as a tagged image file format (tiff) computer file.

Target Images. The target images were U.S. Army ground vehicles (Figure 19). The vehicles were: recovery vehicle, Army personnel carrier (APC), M1 Abrams tank, and Bradley fighting vehicle. These particular vehicles were selected because they possess similar low spatial frequency structures; however, their high spatial frequency structures differ notably.
**Target Processing**

The digital images were converted from full color 24-bit RGB images to 8-bit gray scale images, using a commercial digital image processing software program (Adobe, Model: Photoshop for Macintosh, v5). Next, the target object (i.e., the vehicle) was extracted from each digital scene. Each target object was embedded in the center of a 512 by 512 pixel by 0 bit value uniform field for subsequent numerical processing.

Each target image (i.e., cropped and gray-scale converted version of original vehicle image) was subjected to two-dimension spatial frequency filtering. The filtering operation is characterized as,

\[
\tilde{I}_{F,G} = F^{-}[F^{+}[I_{00}]] \times F^{+}[Q_{F,G}]
\]  

(Eq. 4)

in which

- \( \tilde{I}_{F,G} \) denotes an image filtered at spatial frequency channel \( F \) using a gain of \( G \),
- \( I_{00} \) denotes an original unfiltered image,
- \( Q_{F,G} \) denotes a filter function centered at spatial frequency channel \( F \) with a gain of \( G \),
- \( F^{+}[...] \) denotes the forward two-dimensional Fourier transform of a function [...],
- \( F^{-}[...] \) denotes the inverse two-dimensional Fourier transform of a function [...],
Eq. 4 defines a well-known digital image processing algorithm—convolution of two functions using Fourier domain equivalence properties. The unique aspect of our application of Eq. 6 is the definition of the two-dimensional spatial filter (Q). As indicated above, the present objective was to examine the effects of two-dimensional modulation suppression in the spatial frequency domain on the detection and recognition of real-world targets. This objective was realized by suppressing select regions of spatial frequencies within the two-dimensional Fourier spectrum of the vehicle images. Suppression of the two-dimensional spatial frequency components was implemented using a Gaussian-shaped radial notched filter centered at a visual channel and extending about +/- 0.5 octave in spatial frequency across all orientations. The gain of the filter—or, equivalently, the amplitude of the Gaussian function—controlled the degree of suppression imposed on the image spectrum. Once filtered by multiplying the image and filter spectra, the filtered image was recovered by inverse Fourier transforming the product spectrum. Small shifts in space-averaged intensity in the recovered image also were corrected to match the original, unprocessed image.

The two-dimensional spatial frequency filter (Q) was defined as
\[ Q(f) = G e^{-4 \ln 2 \left( \frac{f - f_F}{\omega} \right)^2} \]  
\hspace{1cm} (Eq. 5)

in which

\( Q(f) \) denotes the value of the two-dimensional filter at spatial frequency \( f \),
\( f \) denotes a two-dimensional spatial frequency in cycles per degree of visual angle,
\( f_F \) denotes the center spatial frequency of the filter,
\( G \) denotes the gain of the filter, and
\( \omega \) denotes the bandwidth of the filter, in cycles per degree units.

Note that two-dimensional spatial frequency is given as

\[ f = \sqrt{f_H^2 + f_V^2} \]  
\hspace{1cm} (Eq. 6)

in which

\( f_H, f_V \) denote the one-dimensional horizontal and vertical cardinal axes spatial frequency coordinates in the two-dimensional Fourier spectrum plane.

Table 5 lists the filter parameters (\( f_F, \omega, \) and \( G \)) used to process the images for the present experiment. And Figure 20 shows the one-dimensional profiles of \( Q \) for the \( f_F \) values, with \( G=1.0 \). As shown, filter bandwidths (\( \omega \)) were selected to minimized overlap between adjacent spatial frequency octaves.

Table 5. Two-Dimensional Spatial Frequency Filter Parameters used in Experiment 2
<table>
<thead>
<tr>
<th>Center Frequency $\left( f_F \right)$</th>
<th>Bandwidth $\left( \omega \right)$</th>
<th>Gain $\left( G \right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>1.3</td>
<td>0.0-1.0 by 0.1</td>
</tr>
<tr>
<td>4.0</td>
<td>1.6</td>
<td>0.0-1.0 by 0.1</td>
</tr>
<tr>
<td>8.0</td>
<td>3.2</td>
<td>0.0-1.0 by 0.1</td>
</tr>
<tr>
<td>16.0</td>
<td>6.4</td>
<td>0.0-1.0 by 0.1</td>
</tr>
<tr>
<td>32.0</td>
<td>12.8</td>
<td>0.0-1.0 by 0.1</td>
</tr>
</tbody>
</table>

Following the filtering operation for a particular center frequency and gain, the target object pixel values were normalized (rescaled) to match the space-average luminance of the unfiltered target and, then, embedded at the center of a 512 by 512 pixel uniformed gray-colored field. The intensity of the background field was set equal to the space-average luminance of the unfiltered target object (using a measured CRT display gamma of 2.3). Figures 21 and 22 present examples of the final processed images used in the experiment.

As explained below, the psychophysical task determined total image modulation required for threshold detection and recognition of each target. Total image modulation was accomplished by weighting the pixel-value histograms of each image by a multiplicative constant, ranging from 0.0 to 1.0 in increments of 0.1. The effect of weighting reduced the luminance contrast of the target images without affecting spatial (or spatial frequency) structure.
Figure 20. One-dimensional amplitude profiles for the spatial frequency filters used in Experiment 2.

Figure 21. Example of Abrams Tank image after processing with the two-dimensional radial notch filter centered at 32 cycle per degree used in Experiment 2.
Experimental Design

Data collection was organized by a three-factor, within-subjects design. One factor was vehicle image, which varied over four levels: Abrams Tank, Bradley Fighting Vehicle, Recovery Vehicle, and Army Personnel Carrier. A second factor was filter center frequency, which varied over four levels: 4, 8, 16, and 32 cycles per degree. The third factor was filter gain, which varied across five levels: 0.0, 0.4, 0.6, 0.8, and 1.0. Note that a filter gain of 1.0 produced no suppression of modulation, whereas a filter gain of 0.0 completely suppressed modulation within the selected center frequency pass band.

Two measures of human visual performance were recorded during the experiment: target detection and target recognition. Target detection refers to the correct detection of a target under a particular viewing condition (i.e., target frequency and masking frequency combination). Observers were required to indicate the presence or absence of a target on each trial. Observer’s responses were recorded as either correct or incorrect; thus, the proportion of correct responses across all trials for a particular viewing condition defined a target detection score. Note, the probability of random correct detection (guessing) with the two response alternatives on each trial was $p = 0.50$. 

Figure 22. Example of APC image after processing with the two-dimensional radial notch filter centered at 32 cycle per degree used in Experiment 2.
Target recognition refers to the correct recognition of a particular target under a viewing condition. Observers were required to identify (i.e., recognize) a specified target from two alternatives shown on each trial sequence. Recognition responses were recorded as either correct or incorrect; thus, the proportion of correct recognition responses across all trials for a particular viewing condition defined a recognition score. The probability of random correct recognition with the two response alternatives on each trial sequence was $p = 0.50$.

Two sampling variables were used during data collection to improve the accuracy of the findings. First, measurements of target detection and recognition were replicated two times for each viewing condition and participant. Second, catch trials presenting only masking fields were interspersed randomly with the actual trials—50% of all trials were catch trials.

**Procedure**

Before beginning the actual experiment, each participant was asked to read and sign a consent form, as well as to read a set of instructions for the experiment task. The instructions detailed the method and type of target presentations, the stimulus-response protocol, and the procedures for trial and rest-break initiation. The participants began the actual experiment after they reviewed the instructions with the experimenter and demonstrated they had an accurate understanding of the task.

Data collection was distributed over two consecutive sessions. In the first session, visual detection responses were collected; whereas, visual recognition responses were obtained in the second session. Observers took a 10 min rest break between sessions, although they did not leave the testing room during this time period.

The psychophysical Method of Adjustment was used to determine the total image modulation required for threshold detection of each target presented under each viewing condition. The stimulus presentation sequence for each viewing condition began with the total target modulation well below threshold. The participant verbally indicated a “Present” or “Absent” response immediately following the 2 s target presentation interval. On subsequent stimulus presentations in the sequence, the modulation weighting of the image was increased by 0.1 units until the participant reported just detecting the target. The procedure used only ascending modulation weight trials, without any staircase presentation procedure. Once the observer correctly detected the target, the experimenter recorded the modulation weight value used on
that presentation and, then, verbally queried the participant to initiate the next stimulus presentation sequence. This procedure was repeated for each image condition (Image Type x Filter Center Frequency x Filter Gain conditions).

After a short rest following the visual detection trials, a similar presentation procedure was used to determine target recognition thresholds. In this case, however, the stimulus presented on each trial contained two vehicles placed side by side at the center of the uniform gray field. At the start of the presentation sequence, the participant was shown an original (unprocessed) of the target vehicle to be recognized in the immediately following stimulus sequence. The stimulus presentation sequence used only ascending trials; that is, the presentations began with a modulation level for both images well below threshold. On subsequent presentations, the modulation weight of both images increased by 0.1 units until the participant correctly recognized the designated target. Once recognition occurred, the experimenter recorded the modulation weight value used on that trial and, then, verbally queried the participant to initiate the next stimulus sequence. This procedure was repeated for each image condition.

Total time to complete the detection and recognition sessions averaged about 100 minutes (including rests breaks). Within each session, trials were blocked by center frequency of the radial notch filter, and the blocks were counterbalanced across participants using a Latin-square procedure. Within a block, vehicle image, filter gain, and replications were presented randomly in a unique order for each observer. Observers were given a five-minute break after every two blocks.

RESULTS

The total image modulation values required for correct threshold detection and correct threshold recognition under each viewing condition were analyzed separately. For the detection data, the average modulation weight of each Vehicle Image x Filter Center Frequency x Filter Gain condition was determined. Then, a threshold change score was computed by comparing the average modulation weight associated with each processed image condition to its corresponding baseline (unprocessed, G=1.0) image condition, given as

$$T_c = 100 \left( \frac{T_{F,G} - T_{0,0}}{T_{0,0}} \right)$$

(Eq. 7)
in which

\( T_c \) denotes the percent threshold modulation change (elevation or suppression) for a filtered image relative to its unfiltered baseline image.

Figures 23-26 present threshold change scores for visual detection of military vehicle images used in Experiment 2. All filtered versions of these images required greater modulation for threshold detection than did their corresponding unfiltered (baseline) image; therefore, the results are plotted in terms of percent threshold elevations.

Two important data trends are consistent across the military vehicle images. First, total image modulation required for threshold detection increases with increasing center frequency of the two-dimensional spatial frequency filter. Second, the total image modulation required for threshold detection increases with increasing gain of the two-dimensional filter. These observed data trends indicate visual detection of the targets decreases with greater amplitude and higher frequency visual channel suppression of image structure. Across the images examined, the change in total image modulation required for threshold detection ranged up to about 29% (for the Abrams Tank image).

Additionally, the loss of image detectability (as indexed by larger threshold change values) appears well behaved across filter gain levels; that is, threshold change values appear to be a linear function of filter gain. One caveat to this observation is that filter center frequencies below 8 cycles per degree had attenuated effects on detection threshold change values. This latter observation is not unexpected, since observers may have utilized information available in higher frequency visual channels for their detection judgments. Nevertheless, these data suggest that threshold shifts due to suppression of image structure (e.g., by masking) may be modeled and predicted easily, and, possibly, accounted for by an intelligent target acquisition system.
Figure 23. Percent threshold elevation for correct detection of Recovery Vehicle image as a function of two-dimensional filter center frequency and gain.

Figure 24. Percent threshold elevation for correct detection of Army Personnel Carrier image as a function of two-dimensional filter center frequency and gain.
Figure 25. Percent threshold elevation for correct detection of Bradley Fighting Vehicle image as a function of two-dimensional filter center frequency and gain.

Figure 26. Percent threshold elevation for correct detection of Abrams Tank image as a function of two-dimensional filter center frequency and gain.
To examine the effects of two-dimensional modulation suppression on threshold detection more closely, Table 6 lists the least squares fit parameters for linear regressions computed for the threshold change values as a function of filter gain shown in Figures 18-21. The linear regression model was

\[ T_c = kG \]  

(Eq. 8)

in which \( k \) denotes the least-squares regression coefficient, with units of percent threshold change per filter gain range.

Inspection of Table 6 reveals that linear regressions provide reasonable first-order descriptions for the effects of filter gain on image detectability—\( R^2 \) statistics range from 0.740 to 0.990 across all models. Moreover, the regression coefficient (\( k \)) increases with increasing center frequency of the two-dimension filter across all images, albeit at slightly different rates among the images examined.

### Table 6. Linear Regression Models for Effects of Filter Gain on Threshold Detection

<table>
<thead>
<tr>
<th>Center Frequency</th>
<th>k</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recovery Vehicle</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>19.12</td>
<td>0.992</td>
</tr>
<tr>
<td>16</td>
<td>12.27</td>
<td>0.927</td>
</tr>
<tr>
<td>8</td>
<td>6.72</td>
<td>0.848</td>
</tr>
<tr>
<td>4</td>
<td>1.17</td>
<td>0.975</td>
</tr>
<tr>
<td>2</td>
<td>1.01</td>
<td>0.983</td>
</tr>
<tr>
<td><strong>Army Personnel Carrier</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>18.93</td>
<td>0.984</td>
</tr>
<tr>
<td>16</td>
<td>13.46</td>
<td>0.909</td>
</tr>
<tr>
<td>8</td>
<td>6.94</td>
<td>0.970</td>
</tr>
<tr>
<td>4</td>
<td>5.35</td>
<td>0.961</td>
</tr>
<tr>
<td>2</td>
<td>3.26</td>
<td>0.937</td>
</tr>
</tbody>
</table>
Recognition data from Experiment 2 were evaluated in a similar way to the detection data. That is, average modulation weights for each participant and viewing condition (Vehicle Image x Filter Center Frequency x Filter Gain) were determined and then used to compute a threshold change score using Eq. 7. These data were used to index the effects of two-dimensional filter center frequency and gain on recognition performance on an image-dependent basis.

Figures 27-30 show the threshold change scores for visual recognition of military vehicle images used in Experiment 2. As with the visual detection findings reported above, all filtered versions of a vehicle image required greater total modulation for threshold recognition than did the corresponding unfiltered (baseline) image.

Inspection of Figures 27-30 suggest that the two data trends noted in the detection data also exist in the recognition data; that is, total image modulation required for threshold recognition increases with increasing center frequency and gain of the two-dimensional spatial frequency filter. Moreover, these two data trends appear more pronounced in the recognition data than in the detection data. Thus, the effects of suppressing modulation at selected spatial frequency passbands in the military vehicle images examined appear relatively greater on visual recognition than on visual detection performance. Across the images examined, the change in total image modulation required for threshold recognition ranged up to about 74% (for the Bradley Fighting Vehicle image).
Also, the loss of image recognizability appears well behaved; that is, for filter center frequencies at or above 8 cycles per degree of visual angle, threshold change appears to be an increasing linear function of increasing filter gain. As with the detection findings, the two lowest filter center frequencies examined had attenuated, even minimal, effects of recognition performance. This latter observation likely stems from the fact that overall target form is reflected in the lower spatial frequency visual channels, whereas the higher frequency channels encode smaller features and details of the image. Logically, then, it is possible that recognition performance depends more crucially upon higher spatial frequency components. This reasoning gains some additional justification when one considers that the military targets examined had similar low frequency structures; that is, the vehicle images were of similar overall size and shape, but they differed substantially in their high frequency structures.

![Graph](image)

**Figure 27.** Percent threshold elevation for correct recognition of Recovery Vehicle image as a function of two-dimensional filter center frequency and gain.
Figure 28. Percent threshold elevation for correct recognition of Army Personnel Carrier image as a function of two-dimensional filter center frequency and gain.

Figure 29. Percent threshold elevation for correct recognition of Bradley Fighting Vehicle image as a function of two-dimensional filter center frequency and gain.
Figure 30. Percent threshold elevation for correct recognition of Abrams Tank image as a function of two-dimensional filter center frequency and gain.

To examine the effects of modulation suppression on threshold recognition more closely, Table 7 lists the results of least squares linear regressions computed for each data trend shown in Figures 22-25, using Eq. 8. Inspection of Table 7 reveals that linear regressions provide excellent first-order descriptions for the effects of filter gain on image recognition—$R^2$ statistics range from 0.855 to 0.996 across all models. Additionally, the regression coefficient ($k$) increases with increasing center frequency of the two-dimension filter across all images, although at slightly different rates among the images examined.

Table 7. Linear Regression Models for Effects of Filter Gain on Threshold Recognition

<table>
<thead>
<tr>
<th>Center Frequency</th>
<th>$k$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recovery Vehicle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 c/d</td>
<td>55.84</td>
<td>0.996</td>
</tr>
<tr>
<td>16 c/d</td>
<td>40.37</td>
<td>0.982</td>
</tr>
<tr>
<td>8 c/d</td>
<td>24.34</td>
<td>0.994</td>
</tr>
</tbody>
</table>
The findings of Experiment 2 show that modulation suppression in selected bands of spatial frequency components comprising real-world military vehicle images can degrade visual detection and recognition. The degradation of visual detection occurs more strongly with suppression of modulation at higher spatial frequency visual channels; that is, visual channels centered at 8 cycles per degree of visual angle or more. Likewise, degradation of visual detection increases with increasing amounts of modulation suppression. The loss of detectability, as indexed by a threshold change relative to an original image, approached a maximum of 29%. A similar pattern of results was observed for the effects of modulation
suppression on image recognition, although the degradation in recognition performance appeared more pronounced, approaching a maximum of 72% threshold change relative to an original image.

DISCUSSION

Contemporary theories of human spatial vision embrace the notion that the eye-brain system decomposes a scene into separate neurological pathways or so-called visual channels of information. Evidence indicates that visual channels originate in the early (retinal) visual system and have receptive fields tuned (selective) to particular sizes and orientations of features in a scene. Functionally, the visual channels appear to encode structural elements of a scene in a manner similar to some types of mathematical transforms; that is, casting the holistic spatial structure of a scene into separate bands of information representing unique sizes and orientations of scene features. From this theoretical view of perceptual processing, human spatial vision has been described, examined, and modeled in terms of Fourier transforms. That is, the visual channels are conceptualized as encoding independent passbands of Fourier domain basis components, which are sine waves having particular ranges of spatial frequencies and orientations.

Many modern psychophysical studies use sine-wave stimuli to investigate the functional properties of the visual system. For example, studies of visual masking, which is a perceptual phenomena where the visibility of target objects is degraded by temporally and spatially proximal background objects, use sine-wave targets and masking stimuli to assess the spatial frequency and orientation bandwidths of visual channels. One limitation, however, of some previous studies is that visual mechanisms have been probed with one-dimensional sine-wave patterns. This methodological approach inherently separates the joint properties of visual channels (i.e., spatial frequency and orientation selectivity) into separate analytic domains.

Since visual scenes consist of at least two spatial dimensions, it is reasonable to consider the study of human spatial vision using two-dimensional stimuli. In the present work, a two-dimensional radial sine-wave pattern was employed to examine visual masking by simultaneously activating the joint properties for the visual channels. Experiment 1
demonstrated that two-dimensional visual masking exists under the conditions examined and that it can degrade both visual detection and visual recognition performance. Moreover, some information on the functional properties of two-dimensional visual masking phenomena were obtained; for example, two-dimensional visual masking degrades visual detection in a manner similar to one-dimensional visual masking, but degrades visual recognition in broadband and asymmetrical manner. Although precise characterizations of these two-dimensional masking effects were beyond our present purposes, the findings of Experiment 1 provide strong evidence for the existence of two-dimensional masking effects, and, therefore, provide justification for further investigations in this area of human spatial vision.

Experiment 2 represents a preliminary examination of two-dimensional masking effects on the visual detection and recognition performance with real-world targets; specifically, military ground vehicles. A database of digitally processed ground vehicle images was prepared and used to assess observers’ abilities to correctly detect and recognize the vehicles. A novel two-dimensional digital filter was developed to suppress Fourier domain modulation in select bands of sine-wave spatial frequencies. The effects of the digital filtering were intended to simulate visual masking of sine-wave image components detected by the visual channels. The findings of Experiment 2 demonstrate that the potential magnitude of two-dimensional visual masking effects on real-world target detection and recognition may be substantial—threshold modulation for visual detection increased up to 27% and for visual recognition up to 74%. Moreover, the effects of simulated visual masking appear to be predictable, suggesting that further efforts to model and predict two-dimensional masking may be straightforward.

Experiments 1 and 2 represent an initial investigation into the validity of two-dimensional visual masking phenomena. Interest in this topic area arose from a desire to improve the extant understanding of human spatial vision, particularly in regard to psychophysical processes underlying visual detection and recognition of real-world targets. The findings of this work provide evidence for the existence of two-dimensional masking and show that its potential impact on real-world target detection and recognition may warrant further investigation.

In summary, this experiment implemented a deliberate suppression of modulation in specified passbands of spatial frequencies through digital image processing technique. The intention of this manipulation was to simulate the effects of two-dimensional masking in real-world image detection and recognition tasks. From the findings reported above, these effects have been
demonstrated; albeit simulated and not confounded by extraneous factors that may influence actual two-dimensional masking effects in real-world images. Nevertheless, the magnitude of the simulated masking effects is substantial and adds justification for subsequent experiments to this dissertation. The experiments that follow will continue this line of investigation by examining the effects of two-dimensional visual masking by real-world scene backgrounds on real-world target detection and recognition.
EXPERIMENT 3

**Background.** A primary objective of this dissertation was to demonstrate the visual masking effects using two-dimensional imagery and Fourier domain representations. And then attempt to categorize the observed masking effects in terms of image backgrounds. For the purpose of examining the effects of two-dimensional visual masking by real-world scene backgrounds on real-world target detection and recognition, the following organizational scheme for background scenes was implemented. While it is recognized that diverse target and background scenarios are encountered in military contexts, a convenient categorization scheme is available through consideration of visual psychophysical principles and Fourier representations of images. This new organizational scheme is described below.

Given that the human visual system possesses distinct neurological channels to encode and process stimuli, it is reasonable to develop a categorization scheme around the notion that image structure is handled in distinct ways by one or more visual channels. Moreover, it is reasonable to expect that a categorization scheme will account for the notion that spatial frequency components affect processing efficiencies within particular visual channels. In other words, a premise for categorizing image backgrounds is to consider their potential masking influence on particular visual channels. This approach toward a categorization scheme fits well with Fourier representations of images, since those mathematics provide unique descriptions of scenes in spatial frequency terms. An arbitrary scene could be described in terms of its potential masking influences on specific spatial frequency channels of vision.

For the author’s purposes here, three categories of background images were identified: low, medium, and high masking frequencies. These categories refer to the bandwidth or range of spatial frequencies present in the background scenes. Figures 31 through 33 present examples of images and their Fourier spectra for these three categories.
Figure 31. Low frequency masking background scene and its two-dimensional Fourier modulation spectrum.
Figure 32. Medium frequency masking background scene and its two-dimensional Fourier modulation spectrum.
Figure 33. High frequency masking background scene and its two-dimensional Fourier modulation spectrum.
The categorization scheme for target backgrounds was sparse at this stage of development and is revisited after Experiment 4 of this dissertation. Nevertheless, it is considered applicable to a broad range of military targets viewed in operational relevant context. For example, a military observer attempting to detect and recognize a ground vehicle within an arbitrary terrain must be able to receive, encode, and process the target spatial frequency components within one or more of the visual channels. To the extent that the spatial frequency components relevant for target detection and recognition fall outside of the pass band of the visual channels, then the ground vehicle will not be perceived, encoded, and processed by the observer. Similarly, if the spatial frequency components of the ground vehicle fall within the observer’s visual pass band, then the probability of correctly detecting and recognizing the target depends upon the available target modulation within one or more of the visual channels. To the extent that target modulation is limited within a channel, that channel may fail to adequately receive, encode, and process that component of the image. The lack of adequate modulation may arise from the properties of the target per se or arise from masking influences of competing spatial frequency components. In either case, the observer may fail to receive, encode, and process the stimulus components efficiently or at all. Consequentially, the probability of correctly detecting and/or recognizing the target stimulus is expected to decrease.

The perceptual process above suggests that an algorithmic approach may be employed to model the masking potential of background scenes upon a particular target. Specifically, using the theoretical notion that the human visual system analyzes information in separate visual channels, the amount of target modulation relative to background modulation within each of the visual channels may be computed as a metric of masking potential. Formally, the masking potential of a target and background pair is computed as

$$MP = \sum_{f} F(M_{T,f}, M_{B,f})$$

Eq. 9

in which

- $MP$ denotes the Masking Potential metric value,
- $f$ denotes the center spatial frequency of a visual channel,
- $M_{T,f}$ denotes the target modulation within the visual channel centered at $f$,
- $M_{B,f}$ denotes the background modulation within the visual channel centered at $f$,
- $F(\ldots)$ denotes a function based on $M_{T,f}$ and $M_{B,f}$.
The mathematical form of $F(\ldots)$ can be determined empirically. However, based on available visual psychophysical knowledge, it may be reasonable to expect $F(\ldots)$ to resemble a signal-to-noise ratio, Weber’s (power) Law, or simple differential characteristic. Regardless of its forms, Eq. 9 suggests that the inclusion of background masking effects into models of target acquisition may be approached by comparing the modulation of a background scene with that of a particular target object across each of the visual channels. This approach is similar to psychophysical models used in other human perceptual modalities (e.g., Articulation Index used in audition).

It is recognized that the dependence of perception on spatial frequency information within visual channels varies across the visual pass band. That is, low spatial frequency channels are known to encode overall form and shape information for a target, whereas high spatial frequency channels encode finer details and edges of a target. Thus, the extent to which an observer is able to detect and recognize a target should vary by masking the spatial frequency information across the visual pass band. Thus, for our present purposes, the application of two-dimensional visual masking on target detection is likely to be served best by higher frequency influences and dependencies. Two visual psychophysical experiments have been designed to examine this assertion.

Experiment 3 used the same data collection procedures as Experiment 1, but the stimuli were real-world background scenes (desert brush, sand dunes, forest tree line) with foreground military vehicle target (tanks, armored personnel carrier, Bradley fighting vehicle). The stimuli were constructed digitally from original 35-mm film pictures. Moreover, the background scenes were selected on the basis that their spatial frequency spectra contained components at appropriate spatial frequencies to mask the targets.

**METHODS**

**Participants**

Five participants (2 female; 25 to 34 years of age) were recruited from the general academic community at Virginia Polytechnic Institute and State University, Blacksburg, Virginia. All participants were required to pass a screening test for normal spatial vision capabilities using a Landholt C acuity test. All participants were naive with respect to the psychophysical testing
procedures and, in particular, to the specific hypothesis examined in the experiment.

**Apparatus**

The experiment was conducted in a dimly lighted (i.e., < 1 lux) room consisting of an experimenter’s station, observer’s station, and a display system. The experimenter’s station consisted of a chair, desk, and a low-intensity desk lamp. The experimenter operated the display system and recorded subject responses while seated at the station during the experiment sessions. The stimulus display was viewed binocularly from 7 ft (2.13 m) with the aid of a chin rest. The monitor pixel addressability was about 82 pixels per inch (32.28 pixels per cm) horizontally and vertically, using a 1024 by 768 pixel format.

**Stimuli**

The stimuli used in the experiment were created digitally by fusing each of five targets with each of seven backgrounds. The stimuli were digital images derived from original 35-mm single reflex, lens camera photographs. The photographs were taken using conventional photographic procedures.

The target images consisted of military ground vehicles: Armored personnel carrier, recovery vehicle, M1 Abrams Tank, Bradley Fighting Vehicle, and M60 A3 Tank (Figure 34). As discussed below, the target images were extracted from digital versions of original photographic scenes.
Figure 34. Military ground vehicle targets used in Experiment 3.

The background scenes consisted of three diverse terrain environments: forest tree line, desert brush, and sand dunes; as well as a synthetic uniform field equal to the space-averaged luminance of the backgrounds (Figures 35 to 40). The background scenes also were extracted from digital versions of original photographs.

Figure 35. Forest tree line scene (No. 1).

Figure 36. Forest tree line scene (No. 2).
Figure 37. Desert brush scene (No. 1).

Figure 38. Desert brush scene (No. 2).
The real-world images were collected using a single-reflex lens camera and commercial grade 35mm film (Model: Canon AE-1, Canon, f1.4 lens; Kodak 100 24x exposure print film). Exposure settings for each image were made using the internal light meter of the camera. Focusing for each image was accomplished using the internal split image viewer of the camera.

Film developing was conducted by a commercial service, with no special film processing.
instructions. All images were developed, enlarged, and printed with commercial Kodak processes and materials. Original prints from each image were prepared in 8.5 x 11 inch (21.6 x 27.9 cm) format on matt-finish paper. Next, the original color prints were scanned into digital format with a flat bed scanner (Colorscan, Model: 100-FX). The photographic prints were scanned at 600 dpi in the horizontal and vertical dimensions, and the resulting digital images were stored on computer disk as *tagged image file format* (tiff) files. The tiff files contained 8 bit pixels per *red, green, and blue* (RGB) color component.

The target and background images were subjected to image processing prior to the experiment. Specifically, the digital images were converted from 24-bit RGB color to 8-bit gray scale images. Then, the target objects and background regions were extracted from their scanned images and stored on computer disk as *raw* 8-bit binary files. The image processing was performed with a commercial software package (Adobe Systems, Model: Photoshop, v5.0.2 for Macintosh). Subsequently, the extracted target and background images were normalized in terms of global image statistics; that is, mean and variance of the 8 bit values. This image processing step was necessary to eliminate luminance contrast differences across the target conditions, and it was performed with a commercial software package (Wolfram Research, Model, Mathematica, v4.1 for Macintosh). For the final image processing step, each target image was overlaid on each background scene, using commercial software (Adobe Systems, Model: Photoshop, v5.0.2 for Macintosh). Four versions of each target-background combination were produced, each containing the target at a different pre-selected location within the background scene. Figures 41 to 45 show examples of target-in-background images.
Figure 41. Bradley tank embedded in Forest Tree Line (No. 1) background.

Figure 42. M60 tank embedded in Sand Dune (No. 2) background.

Figure 43. Recovery vehicle embedded in Desert Brush (No. 1) background.

Tasks
Each participant in the experiment received all viewing conditions. A psychophysical measure of human visual performance (i.e., 50% target detection) was recorded on each trial under each
condition. The dependent variable recorded during the experiment is described below.

50% Target Detection Probability. This dependent variable refers to the correct detection of a target under a particular viewing condition. The 50% target detection probability is a statistical quantity computed for each observer under each viewing condition (i.e., target type, target position, and replications) in the experiment. Observers were required to detect the presence of a target on each trial. The observer’s detection response was either correct or incorrect; thus, the proportion of correct detection responses across all trials under each viewing condition defined the 50% detection probability.

Experiment Design
A two-factor, within-subjects design was used to define the data collection conditions. The factor of target type was manipulated over five levels (i.e., Abrams Tank, Bradley, Recovery Vehicle, M60 Tank, and Armored Personnel Carrier). The second factor of background was manipulated over seven levels (i.e., two versions each of tree line, desert brush, and sand dune, as well as a uniform gray field). The measurement of 50% target detection was replicated four times for each participant and viewing condition. The experiment design is summarized in Table 8 below.

Table 8. Summary of Experiment 1
Two sampling variables were used for data collection in the experiment, as described below.

- Replications -- refers to number of repeated measurements taken for each viewing condition (i.e., target image & masking background combination). The measurement of target detection probability was replicated four times for each participant and viewing condition.

- Catch Trials -- refer to trials containing only background fields. Fifteen percent of all trials were catch trials. Randomly replacing actual trials with catch trials minimized observer response expectations and guessing strategies.

**Procedure**

Before beginning data collection, each participant read and signed an informed consent form, as well as read a set of instructions for the experiment task. The instructions detailed the method and type of target presentations, the stimulus-response protocol, and the procedures for trial and rest-break initiation. Participants familiarized themselves with the stimulus-response protocol by viewing sample stimuli in an instruction packet and on the stimulus display screen.
Following a review of the task protocol with the experimenter, the participant began the actual experiment.

The psychophysical method of constant stimuli was used to determine target detection thresholds. On each trial, a target appeared in a randomly selected location within the background scene. The stimulus presentation lasted for 2 s. Immediately following a stimulus presentation, the observer pressed the left arrow key on the keyboard if a target was detected and the right arrow key if no target was present. Upon pressing one of the two arrow keys, the experiment control program paused for two seconds and then presented the next image to be judged. A uniform gray field having a space-average- luminance equal to that of the stimuli was presented before and after each trial.

The entire data collection session lasted about 1 hr for each participant. Trials in a session were blocked by Target Type. Blocking was counterbalanced across participants using Latin-square matrix. Also, participants viewed an empty adaptation field for two minutes prior to each block of trials. During a block of trials, replications were varied in a unique random order for each participant.

**RESULTS AND DISCUSSION**

Prior to analyzing the stimulus data, performed adherence to the trial procedures was assessed by determining the error rate for the catch trials. An error rate of 3.12% was observed, indicating the absence of guessing and a clear understanding of experimental procedures on the part of the participants.

The detection judgment values (i.e., 0 for incorrect and 1 for correct) were subjected to a two factor, within-subjects analysis of variance (ANOVA) procedure. The main effects of Scene and Target were significant, as was the interaction of Scene x Target (Tables 9). Post hoc test results are shown in Tables 10 and 11.

**Table 9. Analysis of Variance for Detection Judgments**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
</table>

89
Source | DF | SS | MS | F  | p  
--- | --- | --- | --- | --- | --- 
**Between** 
Subject | 4 | 3.091 | 0.773 | 9.27 | 0.0001 
**Within** 
Scene | 6 | 69.194 | 11.532 | 138.39 | 0.0001 
Scene * Subject | 24 | 12.348 | 0.514 | 6.17 | 0.0001 
Target | 4 | 5.991 | 1.497 | 17.97 | 0.0001 
Target * Subject | 16 | 3.380 | 0.211 | 2.53 | 0.0009 
Scene * Target | 24 | 17.948 | 0.747 | 8.97 | 0.0001 
Scene * Target * Subject | 96 | 15.580 | 0.162 | 1.95 | 0.0001 
**Total** | 170 | 127.532 | 

Tables 10 and 11 show

**Table 10. Newman-Keuls Results for the Main Effect of Target on Detection Judgment**

(Note: Alpha= 0.05  df= 16  MSE= 0.21125)
(Means with the same letter are not significantly different)

<table>
<thead>
<tr>
<th>SNK Grouping</th>
<th>Mean</th>
<th>TARGET</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.750</td>
<td>APC</td>
</tr>
<tr>
<td>B</td>
<td>0.578</td>
<td>Abrams</td>
</tr>
<tr>
<td>B</td>
<td>0.521</td>
<td>M60</td>
</tr>
<tr>
<td>B</td>
<td>0.514</td>
<td>Recovery</td>
</tr>
<tr>
<td>B</td>
<td>0.500</td>
<td>Bradley</td>
</tr>
</tbody>
</table>

**Table 11. Newman-Keuls Results for the Main Effect of Scene on Detection Judgment**

(Note: Alpha= 0.05  df= 24  MSE= 0.514524)
(Means with the same letter are not significantly different)
ANOVA and Newman-Keuls results show that backgrounds and targets had significant effects on detection. For the author’s purpose, the significant scene effects are of utmost importance (see Table 11). As can be seen from the Newman-Keuls results, the backgrounds with associated medium and high spatial frequency content (i.e., Desert Brush and Tree Line) both significantly reduced correct detection from the no background or baseline condition. The two Sand Dune backgrounds produced significantly better detection results than the two Desert brush and Tree line backgrounds. Interestingly, the two Sand dune backgrounds were not significantly different from the empty background. Lastly, the Tree line 2 background significantly lowered detection more than the Tree line 1, significantly lowered detection performance more than any of the other backgrounds.

Further data reduction included the computation of probabilities for correct target detection under each viewing condition. The probability of correct target detection was calculated by comparing an observer’s target detection responses (i.e., “Yes” vs. “No”) to the known target condition on each trial. Correct detection judgments were coded as 1; whereas, incorrect judgments were coded as 0. Probabilities of correct detection were computed as the sum of the coded judgments divided by the maximum number of observations per viewing condition per participant.

Figure 44 shows the correct detection probabilities for the Abrams Tank image embedded within each of the seven backgrounds. The probabilities of correct target detection were lowest for the tree line backgrounds and highest for the sand dune backgrounds and uniform field. The desert brush backgrounds lead to an intermediate level of correct target detection performance.

It is important to note that the background scenes were equated in terms of global luminance
properties. Thus, the observed changes in correct target detection can not be attributed to
global luminance contrast effects. Rather, the observed changes in detection performance are
consistent with local influences of the backgrounds on the targets. Given that the observed
detection probabilities appear to change in a regular pattern with the spatial complexity of the
backgrounds, it is reasonable to suspect that local spatial frequency processes underlie these
results. Thus, these findings support the notion of two-dimensional spatial frequency masking in
real-world scenes.
Figure 44. Correct detection probabilities for Abrams Tank embedded in various backgrounds. Error bars represent +/- 1 standard error of the mean.

Similar results were obtained for each of the remaining four ground vehicle targets used in the experiment (see Figures 45 - 48). Across these findings, some noticeable deviations in detection trends exist; that is, the effect of backgrounds on correct detection vary with target image. This observation is expected in a two-dimensional spatial frequency masking paradigm, since the spatial frequency content of the targets vary, as does the critical bands of their spatial frequencies necessary for visual detection. Thus, it is likely that knowledge of the spatial frequency composition of the target and the background will play an essential role in quantitative models of two-dimensional spatial frequency masking phenomena. Experiment 5 will attempt to provide strong evidence for this notion by suppressing the background spatial frequency components of the real-world backgrounds to determine if doing so reduces the likelihood of visual masking effects on real-world targets.
Figure 45. Correct detection probabilities for M60 tank embedded in various backgrounds. Error bars represent +/- 1 standard error of the mean.

Figure 46. Correct detection probabilities for Recovery Vehicle embedded in various backgrounds. Error bars represent +/- 1 standard error of the mean.
Figure 47. Correct detection probabilities for Bradley Fighting Vehicle embedded in various backgrounds. Error bars represent +/- 1 standard error of the mean.

Figure 48. Correct detection probabilities for Armored Personnel Carrier embedded in various backgrounds. Error bars represent +/- 1 standard error of the mean.
EXPERIMENT 4

The findings of Experiment 3 provide evidence for the existence of two-dimensional spatial frequency masking in real-world scenes. Indeed, the observed findings of Experiment 3 are consistent with the hypotheses established in Experiment 1 using two-dimensional grating patterns. Nevertheless, there remains a need to provide a direct test of the hypothesized two-dimensional masking effects in real-world scenes.

Given the purported two-dimensional spatial frequency masking of targets by real-world backgrounds, it is reasonable to expect that direct manipulations of background spatial frequency content should affect visual target acquisition performance. That is, if the visual channels carrying stimulus information critical for visual detection and recognition are masked by background spatial frequency information in those same channels, then manipulations of background spatial frequency content can be expected to lead to changes in detection and/or recognition performance. The present experiment was designed to examine this hypothesis.

Experiment 4 used a different data collection procedure than that used in Experiment 3 and in Experiment 1. In Experiment 4, participants were required to indicate the position and the name of a real-world target on each trial; thus, the participants provided a visual detection and recognition judgments on each trial. Additionally, Experiment 4 used specially processed background scenes. That is, while the stimuli resembled those used in Experiment 3 (i.e., real-world scenes containing one of several military ground vehicles embedded in backgrounds of varying terrains), the background scenes were notch-filtered in the two-dimensional spatial frequency domain. These radial filters were programmable to be centered at the spatial frequency of any selected visual channel and allowed the amplitude of spatial frequency components within that channel to be attenuated. Thus, the filters allowed the extent of two-dimensional spatial frequency masking within a visual channel to be controlled, thereby allowing a direct examination of the two-dimensional masking hypothesis.

METHODS
Participants
Fourteen participants (6 females; 20 to 42 years of age) were recruited from the student and staff population at Virginia Polytechnic Institute and State University. All participants were required to pass a screening test for normal spatial detection capabilities using a Landholt C acuity test. All participants were naive with respect to psychophysical testing procedures, and, in particular, to the specific experiment hypothesis under investigation.

Apparatus
The same test facility, laboratory equipment, and viewing environment used in Experiment 1 were employed for this experiment.

Stimuli
The stimuli were digitally constructed images of U.S. military ground vehicles located in natural terrains, as described for Experiment 1. The stimuli were constructed from original photographs of four ground vehicle targets and three terrain backgrounds. The vehicle targets were: recovery vehicle, M1 Abrams tank, Bradley Fighting Vehicle, and a M60 A3 tank (see Figure 49). Terrain background scenes were: tree line, desert brush, and sand dune.

![Military ground vehicle targets](image)


*Figure 49. Military ground vehicle targets used in Experiment 2.*

Unlike Experiment 1, each background image was subjected to two-dimensional spatial frequency filtering. The filtering operation is characterized as,
\[ \tilde{I}_{F,G} = F^{-}[F^{+}[I_{0,0}] \times F^{+}[Q_{F,G}]] \]  

(Eq. 9)

in which

\( \tilde{I}_{F,G} \) denotes an image filtered at spatial frequency channel \( F \) using a gain of \( G \),

\( I_{0,0} \) denotes an original unfiltered image,

\( Q_{F,G} \) denotes a filter function centered at spatial frequency channel \( F \) with a gain of \( G \),

\( F^{+}[...] \) denotes the forward two-dimensional Fourier transform of a function \([...]\),

\( F^{-}[...] \) denotes the inverse two-dimensional Fourier transform of a function \([...]\),

Eq. 9 defines a well-known digital image processing algorithm—convolution of two functions using Fourier transforms. Our unique use of Eq. 9 is the definition of the two-dimensional spatial filter \( Q \). As indicated above, our objective was to examine the effects of suppressing modulation of masking frequencies on the detection and recognition of real-world targets. This objective was realized by suppressing selected regions of spatial frequencies within the two-dimensional Fourier spectrum of the background images. Suppression of the two-dimensional spatial frequency components was implemented using a Gaussian-shaped radial notched filter centered at a visual channel and extending about +/- 0.5 octave in spatial frequency across all orientations in the spatial frequency domain. The gain of the filter—or, equivalently, the amplitude of the Gaussian function—controlled the degree of suppression imposed on the image spectrum. For Experiment 2, the gain was set to 100% and thus fully attenuated the modulation with the selected visual channels. Once filtered by multiplying the image and filter Fourier spectra, the image was recovered by inverse Fourier transforms. Small shifts in space-averaged luminance in the recovered image were corrected by scaling the image to match the mean and variance of the original, unprocessed image.

The two-dimensional spatial frequency filter \( Q \) was defined as

\[ Q(f) = Ge^{-\frac{4 \ln 2 (f - f_f)^2}{\omega^2}} \]  

(Eq. 10)

in which
Q(f) denotes the value of the two-dimensional filter at spatial frequency f,
f denotes a two-dimensional spatial frequency in cycles per degree,
f_F denotes the center spatial frequency of the filter,
G denotes the gain of the filter, and
ω denotes the bandwidth of the filter, in cycles per degree units.

Note that two-dimensional spatial frequency is given as

\[ f = \sqrt{f_h^2 + f_v^2} \] (Eq. 11)

in which

\[ f_h, f_v \] denote the one-dimensional horizontal and vertical spatial frequency coordinates in the two-dimensional Fourier spectrum plane.

Table 9 lists the filter parameters (f_F, ω, and G) used to process the background scenes for the present experiment. And Figure 50 shows the one-dimensional profiles of Q for the \( f_F \) values, with G=1.0. As shown, filter bandwidths (ω) were selected to minimized overlap between adjacent spatial frequency channels.

Table 12. Parameters for Two-Dimensional Radial Notch Filter Used in Experiment 4

<table>
<thead>
<tr>
<th>Center Frequency (f_F)</th>
<th>Bandwidth (ω)</th>
<th>Gain (G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>1.4</td>
<td>1.0</td>
</tr>
<tr>
<td>4.0</td>
<td>1.6</td>
<td>1.0</td>
</tr>
<tr>
<td>8.0</td>
<td>3.2</td>
<td>1.0</td>
</tr>
<tr>
<td>16.0</td>
<td>6.4</td>
<td>1.0</td>
</tr>
<tr>
<td>32.0</td>
<td>12.8</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Figure 50. Amplitude profiles of radial notch filters used in Experiment 4.

Following the filtering and reconstruction operation, the background image was normalized (rescaled) to match the space-average luminance of the unfiltered background. Then, a target was embedded at the various locations of a 450 X 450 pixel image. Figures 51 to 55 present examples of the final processed images used in the experiment.
Figure 51. Example of tree line background scenes used in Experiment 2. Original (unfiltered) and radial notch filtered at 16 cycle per degree versions are shown.
Figure 52. Example of desert brush background scenes used in Experiment 2. Original (unfiltered) and radial notch filtered at 16 cycle per degree versions are shown.
Figure 53. Example of sand dune background scenes used in Experiment 2. Original (unfiltered) and radial notch filtered at 16 cycle per degree versions are shown.
Figure 54. M60 tank embedded in Desert brush background.

Figure 55. M60 tank embedded in Filtered Tree Line (16 cycle per degree: center frequency of filter) background.
Tasks
Each participant received all viewing conditions in Experiment 4. Two psychophysical measures of visual performance (i.e., probability of correct detection and probability of correct recognition) were recorded on each trial under each condition. The dependent variables recorded during the experiment are described below.

*Probability of Correct Target Detection.* This dependent variable refers to the correct detection of a target under a particular viewing condition. This probability is a statistical quantity computed for each observer under each viewing condition in the experiment. Observers were required to detect the presence of a target in one of eight positions. The observer’s detection response was either correct or incorrect; thus, the proportion of correct detection responses across all trials under a viewing condition defined a 50% detection probability. It should be noted that the probability of random correct detection is defined as one out of eight possible target positions per trial, or \( p = 0.125 \).

*Probability of Correct Target Recognition.* This dependent variable refers to the correct recognition of a target under a particular viewing condition. This probability is a statistical quantity computed for each observer under each viewing condition that yielded a correct detection response (i.e., conditional probability). Observers were required to recognize each detected target as one of the four types of ground vehicles. The observer’s recognition response was either correct or incorrect; thus, the proportion of correct recognition responses across trials under a viewing condition defined a 50% recognition probability. It should be noted that conditional probability of random correct detection is defined as one out of four possible target types per trial multiplied by the change detection probability, or \( p = (1/4) \times 0.125 = 0.031 \).

Experiment Design
A three-factor, within-subjects design defined the data collection conditions. The first factor of *target type* was manipulated over four levels (i.e., Abrams Tank, Bradley, Recovery Vehicle, and M60 Tank). The second factor of *background* was manipulated over three levels (i.e., tree line, desert brush, sand dune). The third factor of *filter frequency* was manipulated over six levels (i.e., No Filter, 2.0, 4.0, 8.0, 16.0, and 32.0 cycles per degree). Target detection and recognition measurements were replicated four times for each viewing condition and participant. The
experimental design is summarized in Table 13.

Table 13. Summary of Experiment 4

<table>
<thead>
<tr>
<th>Design Characteristic</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent Variables:</strong></td>
<td></td>
</tr>
<tr>
<td>Targets</td>
<td>4</td>
</tr>
<tr>
<td>Backgrounds</td>
<td>3</td>
</tr>
<tr>
<td>Filter Frequency</td>
<td>6</td>
</tr>
<tr>
<td><strong>Sampling Variables:</strong></td>
<td></td>
</tr>
<tr>
<td>Replications</td>
<td>4</td>
</tr>
<tr>
<td>Catch Trials per Participant</td>
<td>36</td>
</tr>
<tr>
<td>Participants</td>
<td>14</td>
</tr>
<tr>
<td><strong>Data Set Characteristics:</strong></td>
<td></td>
</tr>
<tr>
<td>No. of Viewing Conditions</td>
<td>72</td>
</tr>
<tr>
<td>No. of Observations per Participant</td>
<td>324</td>
</tr>
<tr>
<td>Total No. of Observations</td>
<td>4,536</td>
</tr>
</tbody>
</table>

Two sampling variables were used for the data collection in this experiment, as described below.

- **Replications.** Replications refer to the number of repeated measurements taken for each viewing condition (i.e., target image & masking background combination). The measurement of target detection probability was replicated four times for each participant and viewing condition.

- **Catch Trials:** Catch Trials refer to trials that only contained background fields. Eleven percent of all viewing conditions were catch trials for this experiment. Randomly replacing actual trials with catch trials minimized observer response expectations and guessing strategies.
Procedure

Before the experiment session, each participant was asked to read and sign a consent form, as well as to read a set of instructions for the experiment task. The instructions detailed the method and type of target presentations, the stimulus-response protocol, and the procedures for trial and rest-break initiation. Participants familiarized themselves with the stimulus-response protocol by viewing sample stimuli in an instruction packet and on the stimulus display screen during the practice session.

Participants were required to complete a practice session before beginning the actual data collection trials. During practice, observers viewed 16 sample stimuli, consisting of four example ground vehicle targets on uniform gray backgrounds and presented randomly at four of the clock-face positions on the display screen (see below). Practice continued until the participant correctly detected and recognized two successive presentations of the 16 sample stimuli without errors.

The psychophysical method of constant stimuli was used to determine target detection and recognition scores under each viewing condition. On each trial, a target appeared in one of eight randomly selected clock-face locations within the background scene (see Figure 56). The stimulus presentation lasted for 2 s. Immediately following a stimulus presentation, the observer verbally reported that a target was detected by indicating its clock-face position (e.g., “12:00”, “1:30”, etc.) and if it was recognized by indicating its target type (e.g., “Abrams”, “Recovery Vehicle”, etc.). Similar to the experiment control program used in Experiment 3, the subject controlled the pace of the experiment using an input device. Specifically, they initiated each trial by clicking the mouse button. Participants verbally reported target position and type at the end of each trial and the experimenter recorded these responses manually.
Figure 56. Illustration of target positions used in the study. Actual target position are indicated by the "X" markers in the figure.

An entire data collection session lasted about 1.5 hrs for each participant. Within a session, the presentation order of the stimulus conditions was randomized uniquely for each participant. Observers were provided with 5 min rest breaks after every two blocks of 54 trials. Following a rest break, observers viewed a uniform gray adaptation field for 2 min prior to beginning the next block of 54 trials.

RESULTS AND DISCUSSION

Prior to analyzing the stimulus data, performed adherence to the trial procedures was assessed by determining the error rate for the catch trials. An error rate of 3.12% was observed, indicating the absence of guessing and a clear understanding of experimental procedures on the part of the participants.
Target Detection

The detection judgment values (i.e., 0 for incorrect and 1 for correct) were subjected to a three factor, within-subjects analysis of variance (ANOVA) procedure. The main effects of Scene, Target, and Filter were significant, as were the interactions of Scene x Target, Scene x Filter, Target x Filter, and Scene x Target x Filter (see Table 14). *Post hoc* analysis results are shown in Tables 15, 16, and 17.

**Table 14. Analysis of Variance for Detection Judgments**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>13</td>
<td>32.674</td>
<td>2.513</td>
<td>20.83</td>
<td>0.0001</td>
</tr>
<tr>
<td><strong>Within</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scene</td>
<td>2</td>
<td>54.951</td>
<td>27.475</td>
<td>227.68</td>
<td>0.0001</td>
</tr>
<tr>
<td>Scene * Subject</td>
<td>26</td>
<td>16.463</td>
<td>0.633</td>
<td>5.25</td>
<td>0.0001</td>
</tr>
<tr>
<td>Target</td>
<td>3</td>
<td>32.316</td>
<td>10.772</td>
<td>89.26</td>
<td>0.0001</td>
</tr>
<tr>
<td>Target * Subject</td>
<td>39</td>
<td>12.00</td>
<td>0.307</td>
<td>2.55</td>
<td>0.0001</td>
</tr>
<tr>
<td>Scene * Target</td>
<td>6</td>
<td>23.601</td>
<td>3.933</td>
<td>32.60</td>
<td>0.0001</td>
</tr>
<tr>
<td>Scene * Target * Subject</td>
<td>78</td>
<td>10.052</td>
<td>0.128</td>
<td>1.07</td>
<td>0.3225</td>
</tr>
<tr>
<td>Filter</td>
<td>5</td>
<td>22.337</td>
<td>4.467</td>
<td>37.02</td>
<td>0.0001</td>
</tr>
<tr>
<td>Filter * Subject</td>
<td>65</td>
<td>12.753</td>
<td>0.196</td>
<td>1.63</td>
<td>0.0013</td>
</tr>
<tr>
<td>Scene * Filter</td>
<td>10</td>
<td>20.955</td>
<td>2.095</td>
<td>17.36</td>
<td>0.0001</td>
</tr>
<tr>
<td>Scene * Filter * Subject</td>
<td>130</td>
<td>22.252</td>
<td>0.171</td>
<td>1.42</td>
<td>0.0016</td>
</tr>
<tr>
<td>Target * Filter</td>
<td>15</td>
<td>14.813</td>
<td>0.98</td>
<td>5.38</td>
<td>0.0001</td>
</tr>
<tr>
<td>Target * Filter * Subject</td>
<td>195</td>
<td>21.950</td>
<td>0.112</td>
<td>0.93</td>
<td>0.7349</td>
</tr>
<tr>
<td>Scene * Target * Filter</td>
<td>30</td>
<td>9.366</td>
<td>0.312</td>
<td>2.59</td>
<td>0.0001</td>
</tr>
<tr>
<td>Scene * Target * Filter * Subject</td>
<td>390</td>
<td>55.499</td>
<td>0.142</td>
<td>1.18</td>
<td>0.0130</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1007</td>
<td>361.982</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 15. Newman-Keuls Results for the Main Effect of Target on Detection Judgments

(Note: Alpha= 0.05  df= 26  MSE= 0.633216)
(Means with the same letter are not significantly different)

<table>
<thead>
<tr>
<th>SNK Grouping</th>
<th>Mean</th>
<th>SCENE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.750</td>
<td>Sand dune</td>
</tr>
<tr>
<td>B</td>
<td>0.578</td>
<td>Desert brush</td>
</tr>
<tr>
<td>c</td>
<td>0.521</td>
<td>Tree line</td>
</tr>
</tbody>
</table>

Table 16. Newman-Keuls Results for the Main Effect of Scene on Detection Judgments

(Note: Alpha= 0.05  df= 39  MSE= 0.307732)
(Means with the same letter are not significantly different)

<table>
<thead>
<tr>
<th>SNK Grouping</th>
<th>Mean</th>
<th>TARGET</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.946</td>
<td>Recovery</td>
</tr>
<tr>
<td>B</td>
<td>0.856</td>
<td>Bradley</td>
</tr>
<tr>
<td>C</td>
<td>0.734</td>
<td>Abrams</td>
</tr>
<tr>
<td>C</td>
<td>0.728</td>
<td>M60</td>
</tr>
</tbody>
</table>

Table 17. Newman-Keuls Results for the Main Effect of Filter on Detection Judgments

(Note: Alpha= 0.05  df= 65  MSE= 0.196213)
(Means with the same letter are not significantly different)

<table>
<thead>
<tr>
<th>SNK Grouping</th>
<th>Mean</th>
<th>Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.941</td>
<td>8</td>
</tr>
<tr>
<td>B</td>
<td>0.880</td>
<td>16</td>
</tr>
<tr>
<td>B</td>
<td>0.839</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>0.763</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0.758</td>
<td>32</td>
</tr>
<tr>
<td>C</td>
<td>0.729</td>
<td>2</td>
</tr>
</tbody>
</table>
ANOVA and Newman-Keuls analyses show that backgrounds, filters, and targets had significant effects on detection. For the author’s purpose, the significant filter effects were of utmost importance (see Table 17). As can be seen from the Newman-Keuls results, the filters associated with medium to high spatial frequency suppression (i.e., 4, 8, and 16 center frequency filters) significantly reduced correct detection from the no filter or baseline condition. In addition, the 8 and 16 filters improved detection performance significantly more than the 2 and 32 filters. Figures 57-60 illustrate the three-way interaction between Scene, Target, and Filter. Each graph plots filters as a function of scene for each target.

![Figure 57](image)

**Figure 57. Graphic showing a plot of filters as a function of background scene for the Abrams Tank.**
Figure 58. Graphic showing a plot of filters as a function of background scene for the Recovery Vehicle.

Figure 59. Graphic showing a plot of filters as a function of background scene for the M60 Tank.
Figure 60. Graphic showing a plot of filters as a function of background scene for the Bradley Fighting Vehicle.
Further data reduction involved the computation of probabilities for correct target detection and correct target recognition under each viewing condition. The probability of correct target detection was calculated by comparing an observer’s target detection responses (i.e., target position judgments) to the known target location on each trial. Correct detection judgments were coded as 1; whereas, incorrect judgments were coded as 0. Probabilities of correct detection were computed as the sum of the coded judgments divided by the maximum number of observations per viewing condition per participant.

Probability of correct target recognition was calculated in a similar manner; that is, by comparing an observer’s target recognition responses (i.e., vehicle type judgments) to the known target type on each trial. Correct recognition judgments were coded as 1 if the observer correctly detected the target (i.e., correct position) on that trial, as well as, correctly recognized the target; otherwise, the judgments were coded as 0. Probabilities of correct recognition were computed as the sum of the coded judgments divided by the maximum number of observations per viewing condition per participant.

Additionally, because the experiment objective was to assess changes in visual target acquisition performance across levels of background spatial frequency content, the detection and recognition probabilities were expressed as percent change values relative to the No Filter (unfiltered) conditions. In this manner, the effects of background spatial frequency content, as well as the effects of the various filter passbands, on visual target acquisition performance were indexed readily.

Figure 61 shows the effects of filter center frequencies on the percent change in correct detection probabilities for the M1 Abrams tank target across the backgrounds examined. It should be appreciated that percent change in correct detection probabilities are relative to the corresponding unfiltered image conditions. Thus, a percent change value less than zero indicates that detection probabilities decreased for a filtered image relative to its unfiltered image, whereas a percent change value greater than zero indicates that detection performance improved. Moreover, since visual masking is associated negatively with target acquisition performance, positive percent change values suggest decreasing masking (relative to the unfiltered image condition) and negative percent change values suggest increasing masking.
Figure 61. Effects of filter center frequency and background scene conditions on percent change in detection probabilities for the M1 Abrams tank target.

As shown in Figure 61, several of the filter conditions examined produced positive percent change in correct detection probabilities, suggesting a reduction in visual masking relative to the unfiltered background image condition. The largest percent change values were associated with the 4.0, 8.0, and 16.0 cycle per degree filter center frequencies. It is noteworthy that the improvement in correct detection often exceeded 25% compared to the unfiltered image condition. More importantly, these data provide strong evidence for the two-dimensional masking hypothesis underlying this line of research. By reducing the amplitude of background spatial frequency components in select visual channels, correct detection of targets improves; sometimes dramatically.

Additionally, it is evident in Figure 61 that the effects of filtering on percent change in correct detection probability were dependent on background scenes. Specifically, filtering at the mid-frequency visual channels (i.e., 4.0, 8.0, and 16.0 cycles per degree) had the greatest effect on percent change values with the tree line background scene, followed next by the desert brush scene. Filtering had no appreciable effect on percent change values with the sand dune background. These findings are consistent with a two-dimensional spatial frequency masking paradigm. That is, the three background scenes had substantially different spatial frequency
compositions: the tree line image possesses a broadband spectrum, the desert brush background consists of a somewhat lower, predominantly mid-frequency spectrum, and the sand dune background is composed mainly of low spatial frequency components. Therefore, one would not expect filtering of these background spectra to produce equivalent effects on visual performance.

A similar pattern of results was obtained for the remaining three background scene conditions (see Figure 62-64). In particular, the trends in percent change values for the three tank images (Figure 61, 61, and 63) show dramatic improvements in detection performance for the mid-frequency filter conditions. And in general, the improvements in detection performance are most evident for background scenes with broadband spectra (i.e., tree line and desert brush).

Figure 62. Effects of filter center frequency and background scene conditions on percent change in detection probabilities for the M60 tank target.
Figure 63. Effects of filter center frequency and background scene conditions on percent change in detection probabilities for the Bradley fighting vehicle target.

Figure 64. Effects of filter center frequency and background scene conditions on percent change in detection probabilities for the recovery vehicle target.
Target Recognition

The recognition judgment values (i.e., 0 for incorrect and 1 for correct) were subjected to a three factor, within-subjects analysis of variance (ANOVA) procedure. The main effects of Scene, Target, and Filter were significant, as were the interactions of Scene x Target, Scene x Filter, Target x Filter, and Scene x Target x Filter (Tables 18). Post hoc analysis results are shown in Tables 19, 20, and 21.

Table 18. Analysis of Variance for Recognition Judgments

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>13</td>
<td>72.424</td>
<td>5.571</td>
<td>30.34</td>
<td>0.0001</td>
</tr>
<tr>
<td><strong>Within</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scene</td>
<td>2</td>
<td>83.583</td>
<td>41.791</td>
<td>227.58</td>
<td>0.0001</td>
</tr>
<tr>
<td>Scene * Subject</td>
<td>26</td>
<td>7.829</td>
<td>0.301</td>
<td>1.64</td>
<td>0.0218</td>
</tr>
<tr>
<td>Target</td>
<td>3</td>
<td>22.095</td>
<td>7.365</td>
<td>40.11</td>
<td>0.0001</td>
</tr>
<tr>
<td>Target * Subject</td>
<td>39</td>
<td>10.0562</td>
<td>0.257</td>
<td>8.66</td>
<td>0.0001</td>
</tr>
<tr>
<td>Scene * Target</td>
<td>6</td>
<td>9.544</td>
<td>1.590</td>
<td>8.97</td>
<td>0.0001</td>
</tr>
<tr>
<td>Scene * Target * Subject</td>
<td>78</td>
<td>18.249</td>
<td>0.233</td>
<td>1.27</td>
<td>0.0546</td>
</tr>
<tr>
<td>Filter</td>
<td>5</td>
<td>22.014</td>
<td>4.402</td>
<td>23.98</td>
<td>0.0001</td>
</tr>
<tr>
<td>Filter * Subject</td>
<td>65</td>
<td>13.089</td>
<td>0.201</td>
<td>1.10</td>
<td>0.2797</td>
</tr>
<tr>
<td>Scene * Filter</td>
<td>10</td>
<td>30.955</td>
<td>3.095</td>
<td>16.86</td>
<td>0.0001</td>
</tr>
<tr>
<td>Scene * Filter * Subject</td>
<td>130</td>
<td>24.212</td>
<td>0.186</td>
<td>1.01</td>
<td>0.440</td>
</tr>
<tr>
<td>Target * Filter</td>
<td>15</td>
<td>14.813</td>
<td>0.98</td>
<td>5.38</td>
<td>0.0001</td>
</tr>
<tr>
<td>Target * Filter * Subject</td>
<td>195</td>
<td>33.932</td>
<td>0.174</td>
<td>0.95</td>
<td>0.6844</td>
</tr>
<tr>
<td>Scene * Target * Filter</td>
<td>30</td>
<td>8.438</td>
<td>0.281</td>
<td>1.53</td>
<td>0.0324</td>
</tr>
<tr>
<td>Scene * Target * Filter * Subject</td>
<td>390</td>
<td>72.896</td>
<td>0.186</td>
<td>1.02</td>
<td>0.4005</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1007</td>
<td>444.1292</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 19. Newman-Keuls Results for the Main Effect of Target on Recognition Judgments

(Note: Alpha = 0.05  df = 16  MSE = 0.21125)
(Means with the same letter are not significantly different)

<table>
<thead>
<tr>
<th>SNK Grouping</th>
<th>Mean</th>
<th>SCENE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.721</td>
<td>Sand dune</td>
</tr>
<tr>
<td>B</td>
<td>0.432</td>
<td>Desert brush</td>
</tr>
<tr>
<td>B</td>
<td>0.395</td>
<td>Tree line</td>
</tr>
</tbody>
</table>

Table 20. Newman-Keuls Results for the Main Effect of Scene on Recognition Judgments

(Note: Alpha = 0.05  df = 24  MSE = 0.514524)
(Means with the same letter are not significantly different)

<table>
<thead>
<tr>
<th>SNK Grouping</th>
<th>Mean</th>
<th>TARGET</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.641</td>
<td>Bradley</td>
</tr>
<tr>
<td>B</td>
<td>0.497</td>
<td>Recovery</td>
</tr>
<tr>
<td>B</td>
<td>0.465</td>
<td>M60 Tank</td>
</tr>
<tr>
<td>B</td>
<td>0.456</td>
<td>Abrams</td>
</tr>
</tbody>
</table>

Table 21. Newman-Keuls Results for the Main Effect of Filter on Recognition Judgments

(Note: Alpha = 0.05  df = 24  MSE = 0.514524)
(Means with the same letter are not significantly different)

<table>
<thead>
<tr>
<th>SNK Grouping</th>
<th>Mean</th>
<th>FILTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.639</td>
<td>8</td>
</tr>
<tr>
<td>B</td>
<td>0.857</td>
<td>16</td>
</tr>
<tr>
<td>B</td>
<td>0.537</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>0.456</td>
<td>32</td>
</tr>
<tr>
<td>C</td>
<td>0.450</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0.437</td>
<td>2</td>
</tr>
</tbody>
</table>

ANOVA and Newman-Keuls analyses show that backgrounds, filters, and targets had significant effects on detection. The significant filter effects were of utmost importance to the author.
Table 18). Similar to the results found for detection values, post hoc analysis showed that the filters associated with medium to high spatial frequency suppression (i.e., 4, 8, and 16 center frequency filters) significantly reduced correct recognition from the no filter or baseline condition. In addition, the 8 and 16 filters improved detection performance significantly more than the 2 and 32 filters. Figures 65-68 illustrate the three-way interaction between Scene, Target, and Filter. Each graph plots filters as a function of scene for each target.

Figure 65. Graphic is a plot of filters as a function of background scene for the Recovery Vehicle.
Figure 66. Graphic is a plot of filters as a function of background scene for the M60 Tank.

Figure 67. Graphic is a plot of filters as a function of background scene for the Bradley Fighting Vehicle.
Figure 68. Graphic is a plot of filters as a function of background scene for the Abrams Tank.

Figure 70 shows the effects of filter center frequencies on the percent change in correct recognition probabilities for the M1 Abrams tank target across the backgrounds examined. It is apparent that the spatial frequency filtering exerted a mixed effect on the percent change in recognition performance; some filter conditions lead to substantially improved recognition over the unfiltered condition, whereas other filter conditions, especially those applied to the tree line background, decreased recognition performance. It is noted, however, that these percent change values are based on conditional recognition probabilities; that is, a target had to be detected correctly in order for its correct recognition to be counted in the dataset. Consequently, far fewer targets were recognized as compared to detected and, therefore, the percent change in recognition probability values were subject to greater numerical variations.
Figure 69. Effects of filter center frequency and background scene conditions on percent change in recognition probabilities for the M1 Abrams tank target.

Figures 71 and 72 present the percent change in recognition values for the remaining two tank images, while Figure 70 shows the same percent change values for the recovery vehicle. A similar pattern of results is exhibited across these images, that is, target recognition clearly improved for the mid-frequency filter conditions applied to the desert brush background scene. And filtering exerted little effect on recognition performance when applied to the sand dune background. These findings are consistent with the notion that masking depends upon the spatial frequency composition of the target and background image components.
Figure 70. Effects of filter center frequency and background scene conditions on percent change in recognition probabilities for the M60 tank target.

Figure 71. Effects of filter center frequency and background scene conditions on percent change in recognition probabilities for the Bradley fighting vehicle target.
Figure 72. Effects of filter center frequency and background scene conditions on percent change in recognition probabilities for the recovery vehicle target.
MODELING

The objective of the present research program was to investigate visual target acquisition processes involved in real-world scenes, especially visual acquisition of military ground vehicle targets. This work stemmed from a desire within the U.S. Army to seek improvements to quantitative models of visual target acquisition performance. As indicated above, the approach taken here to address the objective was based on a generalization of visual masking knowledge in the psychophysics literature. Our generalization of the visual masking knowledge base fills a void in the current quantitative approaches used to model target acquisition performance. Moreover, the framework established by our generalization of the visual masking knowledge is amenable to incorporation into existing and new quantitative models. Example uses of our findings for this purpose are described below.

The findings of our work provide strong support for the existence of two-dimensional spatial frequency masking effects in real-world scenes. In particular, the findings suggest that spatial frequencies of the local area surrounding a target can mask those of a target. Obviously, if the masked target spatial frequencies are important for visual acquisition, observer target acquisition performance will be hindered. Thus, an important new approach toward modeling visual acquisition is suggested by these observations. Specifically, the two-dimensional spatial frequency spectrum of a target can be compared quantitatively to that of its immediate background. From this type of comparison, it should be possible to index the likelihood of masking effects on the target.

To simplify this modeling approach, one can rely on the fact that the human visual perception system is most sensitive to spatial frequencies in the 2.0 to 8.0 cycles per degree range. Thus, it is reasonable to purport that this spatial frequency range provides information critical to visual acquisition performance. Therefore, to index the likelihood of masking effects on a target, it may be sufficient to understand just the spatial frequency composition of the background scene.

Also as mentioned above, previous work to characterize the effects of background clutter on visual performance have been hindered by the intractable diversity of spatial structures in real-world scenes. However, with the theoretical notion that the human visual system behaves functionally like a Fourier transform operator, the problems associated with modeling diverse spatial structures can be reduced, if not eliminated. That is, background images of any spatial
complexity can be represented in a common format by their two-dimensional modulation (normalized Fourier amplitude) spectra. Moreover, the two-dimensional modulation spectra can be characterized according to the amount of modulation they contain in each of the visual channels. Through this metric approach, any background scene can be categorized by its potential to exert masking influences on targets.

To illustrate this latter approach, consider the following. Let the two-dimensional modulation spectrum of a background scene be given as

\[ M_B(u,v) = \left| \frac{\iint_{x,y} I(x,y) e^{-j2\pi(xu+yv)} dx dy}{\iint_{x,y} I(x,y) dx dy} \right| \]

(Eq. 12)

in which

\( I(x,y) \) denotes the luminance of the background at spatial coordinate \( x,y \),

\( M_B(u,v) \) denotes the modulation of the background at spatial frequency \( u,v \), and

\( ||...|| \) denotes a complex-number absolute value operation.

Note that Eq. 12 can be translated between rectangular \( x,y \) (or \( u,v \)) coordinates and polar \( r,\theta \) coordinates using the following relationships:

\[ x = r \sin(\theta) \]  \hspace{1cm} (Eq. 13)
\[ y = r \cos(\theta) \]  \hspace{1cm} (Eq. 14)

To determine the amount of background modulation within a visual channel, we need a quantitative representation of visual channels in the two-dimensional spatial frequency domain. Although the psychophysical literature indicates that visual channels are selective to spatial frequency and orientation (see, for review, Beaton and Olacsi, 2001), it is known that human vision is much less sensitive to orientation than to frequency—typically, contrast sensitivities change only 4% across all orientations. Therefore, we propose a simplified representations of visual channels by integrating over all orientations in the two-dimensional spatial frequency domain, as given by
\[ C_i = \int_\theta \int_r Q(r, w, \Theta) \, dr \, d\Theta \]  
\[ C_i = 2\pi \int_r Q(r, w) \, dr \]  
(Eq. 15)

in which 
\[ Q(r, w, \Theta) \] denotes a two-dimensional function describing the shape of the \( i \)th visual channel in polar coordinates,
\[ Q(r, w) \] denotes a one-dimensional function describing the cross-sectional shape of the \( i \)th visual channel as a function of \( r \) in polar coordinates,
\( r \) denotes the center frequency of the visual channel, and
\( w \) denotes the bandwidth of the visual channel.

For computing the amount of background modulation within a visual channel, we desire to use Eq. 15 as a multiplicative filter applied to Eq. 16 to extract (integrate) the modulation within each two-dimensional visual channel. It is convenient, therefore, to consider the cross-sectional shape of \( Q() \) as a rectangular pulse, such as

\[ Q(r^* : r, w) = \begin{cases} 1, & \text{iff } (r - \frac{w}{2}) \leq r^* \leq (r + \frac{w}{2}) \\ 0, & \text{otherwise} \end{cases} \]  
(Eq. 16)

in which 
\( r^* \) denotes any observed \( r \) in polar coordinates.

Substituting Eq. 15 into Eq. 16, the total surface area of the \( i \)th visual channel in the two-dimensional spatial frequency plane is computed as

\[ C_i = 2\pi w_i f_{c_i} \]  
(Eq. 17)

in which 
\( f_{c_i} \) denotes the center frequency of the \( i \)th visual channel and
\( w_i \) denotes the spatial frequency bandwidth of the \( i \)th visual channel.
Table 11 lists representative values of Eq. 17 for several visual channels. The $C_i$ values indicate the total surface areas of each visual channel in the two-dimensional spatial frequency domain. Equivalently, since modulation values range between $\{0,..,1\}$, the $C_i$ listed indicate theoretical maximum amounts of modulation within each visual channel.

Table 22. Calculation of Visual Channel Areas

<table>
<thead>
<tr>
<th>$i$</th>
<th>$f_{ci}$</th>
<th>$w_i$</th>
<th>Lower Channel Limit</th>
<th>Upper Channel Limit</th>
<th>$C_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.25</td>
<td>0.75</td>
<td>1.571</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.75</td>
<td>0.6</td>
<td>1.375</td>
<td>4.712</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1.5</td>
<td>1.25</td>
<td>2.75</td>
<td>18.850</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2.5</td>
<td>5.5</td>
<td>75.398</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>11</td>
<td>301.593</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>12</td>
<td>10</td>
<td>22</td>
<td>1206.372</td>
</tr>
<tr>
<td>7</td>
<td>32</td>
<td>24</td>
<td>20</td>
<td>44</td>
<td>4825.486</td>
</tr>
</tbody>
</table>

Using Eqs. 15, 16, and 17, it is straightforward to formulate an index for the total amount of masking modulation potentially contributed to a visual channel by a background scene. In other words, the masking potential of a background scene can be modeled by a profile indexing the amount of its modulation within each of the visual channels, as given by

$$P_i = \frac{2\pi \int M_B(f_{c_i}, w_i)C_i(f_{c_i}, w_i) \, dr}{C_i}$$  \hspace{1cm} (Eq. 18)

in which,

$P_i$ denotes the masking potential from background scene modulation in the $i^{th}$ visual channel.

It is noted that Eq. 18 expresses the amount of channel modulation as a proportion relative to the theoretical maximum amount of channel modulation (see Table 11). This expression is
justified on the grounds that the visual channel areas increase with increasing center frequency. Therefore, the relative index of channel modulation allows for meaningful comparisons across channels.

Figures 73 shows the masking potential profiles for the unfiltered background scenes used in Experiment 4. It is evident that the sand dune image contained less modulation across the visual channels than did the desert brush and tree line images. This quantitative observation supports our human performance findings that filtering had little effect on target detection when applied to the sand dune image. It now is apparent that the lack of filtering effects associated with this image stemmed from the fact that the sand dune image contained little masking modulation at the mid- to high- spatial frequency visual channels.

![Figure 73. Masking potential profiles for background scenes used in Experiment 2.](image)

These initial modeling notions indicate that the masking potential index may be a novel and useful metric to incorporate into existing target acquisition models. Additionally, the masking
potential index could be used to revise other image quality metrics, such as those based on
perceptual-weighted signal-to-noise ratios and the image-dependent modulation transfer
function area (MTFA) (see, for reviews, Beaton, 1983; Beaton and Farley, 1991).
CONCLUSIONS

The research presented in this dissertation was based on two hypotheses. First, two-dimensional visual masking may degrade observers’ ability to acquire real-world targets embedded in real-world background scenes. This hypothesis was based on the notion that visual masking arises from spatial frequency components of the background that interfere with similar spatial frequency components associated with a target. Second, suppression of the real-world background spatial frequency components reduces the likelihood of visual masking effects on real-world targets. The hypothesis provided a direct test of the two-dimensional masking effects purported in this research project. The findings of this dissertation provide strong support for both hypotheses.

Additionally, the present findings demonstrate that two-dimensional Fourier concepts can be used to describe and model the visual masking effects exerted by real-world backgrounds upon military ground vehicle targets. Several novel implications from the two-dimensional Fourier framework were formulated. Specifically, a quantitative approach toward categorizing real-world background scenes was developed, based on the amount of modulation present in critical spatial frequency bands (i.e., visual channels). And a quantitative approach toward the development of a masking potential metric was described, which compares the modulation of a real-world target to that of its immediate scene background in a signal-to-noise ratio format. The masking potential metric should have direct use in improving extant visual target acquisition models, as well as in formulating new image quality metrics.

The present findings have implications for refinement to existing Johnson Criteria-based models of visual target acquisition, such as the U.S. Army ACQUIRE model. As stated elsewhere in this dissertation, the ACQUIRE model presumes that target-related spatial information (cycles-on-target) is responsible for acquisition performance; although it is well known that the unacceptable levels of prediction accuracy occur for targets viewed in cluttered background. This dissertation suggests a parametric approach for developing empirical relationships for the effects of background clutter, which, in turn, could be used to extend and improve the existing ACQUIRE model.

Finally, a reasonable future application of the present findings may be an intelligent target
acquisition system that predicts in real-time the visibility of selected targets within arbitrary backgrounds. Given the physical characteristics of a target (i.e., its size and potential viewing range), then the critical spatial frequency components for visual acquisition (i.e., visual detection and/or recognition) can be estimated readily. The visibility characteristics of the target then could be assessed in real-time by computing the masking potential metric from a live image of the terrain environment.
REFERENCES


VITA

Gary Olacsi was born in the city of Pacoima, Southern California, in 1970. He attended Bishop Alemany High school in Mission Hills California. Gary received his B.A. degree in Psychology from California State University Northridge in 1994. His strong interest in Human Factors Engineering lead him to Virginia Polytechnic Institute and State University, where he received a Masters of Science degree in Industrial Systems Engineering, Human Factors and Ergonomics Option, in 1998. While at Virginia Tech, Gary worked in the Displays and Controls Laboratory as a graduate research assistant. During this time he was an Eastman Kodak and Army Research Laboratory internship recipient. In addition, he was provided an opportunity to work on corporate sponsored research projects for various companies such as LG Electronics and 3M Optical. While working as an Eastman Kodak intern, Gary evaluated a new flat-panel display technology. As an ARL intern, he investigated visual masking phenomena using two-dimensional Fourier analysis techniques and developed general guidelines for the effective use of color in military displays. Upon completing his degree, Gary moved out west to complete his job search for a Human Factors Engineer (hardware/software usability specialist) position in industry.