

Copyright
by
Abtine Tavassoli
2007

**The Dissertation Committee for Abtine Tavassoli certifies that this is the approved
version of the following dissertation:**

Discovery and Representation of Human Strategies for Visual Search

Committee:

Alan C. Bovik, Supervisor

Lawrence K. Cormack, Supervisor

Michael F. Becker

Wilson S. Geisler

Edward J. Powers

Discovery and Representation of Human Strategies for Visual Search

by

Abtine Tavassoli, Diplôme d'Ingénieur, M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2007

Dedication

To my parents Marie and Davoud,
to my brother Armin,
to my nephew David,
and to my niece Sasha.

Acknowledgements

I would like to express my deepest gratitude to Prof. Alan Bovik and Prof. Lawrence Cormack for their guidance, encouragement, and trust. I have been most fortunate to conduct research with two great advisors and friends, who not only have inspired me in my work but have had an immense influence on my life. I cannot express enough my appreciation for all their support.

I would like to thank Prof. Michael Becker, Prof. Wilson Geisler, Prof. Edward Powers, Prof. Eyal Seidemann, and Prof. Jeremy Wolfe for their valuable support. I would especially like to express my sincere gratitude to Prof. Becker for his constant support and trust since I started as an exchange student at UT.

I am appreciative to all professors and staff in my department, the Electrical and Computer Engineering, and in the department of Psychology. It has been a great honor to be part of these departments. I would like to extend my gratefulness and appreciation to all my school teachers and professors throughout my studies since I first started coloring with crayons.

I would like to show my gratitude to all my friends who have supported me. It would take pages to list them all. I would especially like to thank my friends Alex Sutton, Alejo Perez, and Virat Kapur who dedicated long hours to complete my experiments as subjects. Furthermore, I would like to thank all my lab mates including Umesh

Rajashekar, Ian van der Linde, Russell Jackson, Mehul Sampat, Kalpana Seshadrinathan, James Monaco, Joonsoo Lee, Yang Liu, Sina Jahanbin, Rana Jahanbin, Sumohana Channappayya, Farooq Sabir, Raghu Raj, Thayne Coffman, Hamid Sheikh, Hyohoon Choi, and Shalini Gupta. I would especially like to thank Ian and Umesh for all their support and friendship.

I would like to thank Annie Elderbroom, Tany Norwood, Becky Carreon, Veronica Cantu, Barbara Thompson, and Terri Lavorgna for their friendship and helpful guidance while I was an officer for various student organizations such as the UT Senate and the Graduate Engineering Council.

Finally, I would like to thank my parents Marie and Davoud, my brother Armin, my nephew David and niece Sasha Tavassoli for their love, support and great advice.

Discovery and Representation of Human Strategies for Visual Search

Publication No. _____

Abtine Tavassoli, Ph.D.

The University of Texas at Austin, 2007

Supervisors: Alan C. Bovik

Lawrence K. Cormack

Visual search can simply be defined as the task of looking for an object of interest in a visual environment. Due to its foveated nature, the human visual system succeeds at such task by making many discrete fixations linked by rapid eye movements called *saccades*. However, very little is known about how saccadic targets (*fixation loci*) are selected by the brain in such naturalistic tasks. Discoveries to be made are not only invaluable to the field of vision science but are very important in designing automated vision systems, which to this day lag in performance vis-à-vis human observers.

What I have sought to accomplish in this dissertation has been to reveal previously unknown saccadic targeting and target selection strategies used by human observers in naturalistic visual search tasks. My driving goal has been to understand how the brain selects fixation loci and target candidates upon fixation, with the objective of using these findings for automated fixation selection algorithms employed for visual search.

I have proposed a novel and efficient technique akin to psychophysical reverse correlation to study human observer strategies in locating low-contrast targets under a variety of experimental conditions. My technique has successfully been used to study saccadic programming and target selection in various experimental conditions, including visual searches for targets with known characteristics, targets whose orientation attributes are not known *a priori*, and targets containing multiple orientations. I have found visual guidance in saccadic targeting and target selection under all experimental conditions, revealed by observers' selectivity for spatial frequencies and/or orientations of stimuli close to that of the target. I have shown that under uncertainty, observers rely on known target characteristics to direct their saccades and to select target candidates upon foveal scrutiny. Moreover, I have demonstrated that multiple orientation characteristics of targets are represented in observer search strategies, modulated by their sensitivity / selectivity for each orientation. Some of my findings have been applied towards applications for automated visual search algorithms.

Table of Contents

List of Tables	xi
List of Figures	xii
Chapter 1. Introduction	1
1.1 Motivation.....	1
1.2 Contributions.....	2
1.3 Outline.....	5
Chapter 2. Background	6
2.1 Visual Search	6
2.1.1 Deployment of Visual Attention.....	7
2.1.2 Guidance of Eye Movements.....	9
2.1.3 Low-level vs. High-level Mechanisms	12
2.2 Classification Image Paradigm for Visual Search	13
2.3 Machine Vision.....	17
2.3.1 Search in Passively Acquired Images	18
2.3.2 Active Vision	21
Chapter 3. A Novel and Efficient Technique to Study Visual Search.....	24
3.1 Proposed Method	24
3.1.1 Observers	25
3.1.2 Apparatus	25
3.1.3 Stimuli.....	27
3.1.4 Procedure	28
3.1.5 Analysis Method	30
3.2 Results.....	34
3.2.1 Average/Classification Images	34
3.2.2 Control Experiments	40
3.2.2.1 Implementation without a Grid.....	40
3.2.2.2 Implementation with White Noise	42
3.2.3 Performance Measures.....	45

3.2.3.1 Performance Over Location and Time.....	46
3.2.3.2 Observer Dwell Times	49
3.3 Discussion and Conclusions	50
Chapter 4. Spatial Frequency and Orientation Selectivity in Visual Search	52
4.1 Motivation.....	52
4.2 Methods.....	54
4.2.1 Observers	55
4.2.2 Visual Stimuli	55
4.2.3 Procedure	56
4.2.4 Analysis Method.....	58
4.3 Results.....	59
4.4 Discussion.....	66
4.5 Conclusions.....	69
Chapter 5. Visual Search under Uncertainty Conditions	70
5.1 Motivation.....	70
5.2 Methods.....	72
5.2.1 Observers	72
5.2.2 Visual Stimuli	72
5.2.3 Procedure	73
5.2.4 Analysis Method.....	75
5.3 Results.....	75
5.4 Discussion.....	79
5.5 Conclusions.....	81
Chapter 6. Towards a Better Understanding of Search for Complex Targets	82
6.1 Methods.....	82
6.1.1 Observers	83
6.1.2 Visual Stimuli	83
6.1.3 Procedure	84
6.1.4 Analysis Method.....	84
6.2 Results.....	85

6.3 Discussion.....	91
6.4 Conclusions.....	93
Chapter 7. Towards Applications for Automated Visual Search.....	94
7.1 Some Insights.....	94
7.2 Applications to Visual Search for Objects.....	95
7.3 Conclusions.....	100
Chapter 8. Conclusions and Future Work.....	101
8.1 Conclusions.....	101
8.2 Future Work.....	101
Bibliography	104
Vita	113

List of Tables

Table 3.1:	The nonfoveal noise tile taxonomy.....	33
Table 3.2:	The foveal noise tile taxonomy.....	33

List of Figures

- Figure 2.1: Three examples of stimuli used for high threshold visual search tasks are shown representing cases of (A) pop-out, (B) single-feature, and (C) conjunction of features.....8
- Figure 2.2: Eye movement recordings during search. The top left panel shows a configuration of toys placed in a crib. The top right panel shows the recorded eye movements of an observer searching for the butter target on a dining table during a trial. Endpoints of the first, second and third saccades are shown in the bottom panels, the squares indicating the possible locations where objects appeared10
- Figure 2.3: Demonstration of the classification image technique applied to a vernier acuity task15
- Figure 2.4: An example of scan paths is shown in (a); discrimination images are shown in (b)16
- Figure 2.5: Model for saliency-based attention (from Itti & Koch, 2000)20
- Figure 2.6: Example of an active vision system (from Giefing *et al.*, 1992)22
- Figure 3.1: Targets used in the trials: (A) triangle, (B) dipole. Additional shapes used in the analysis: (C) bowtie, (D) circle, and (E) star28
- Figure 3.2: An example stimulus (with a higher SNR than used during the experiment). The target, a triangle, is in the tile immediately below the central one28
- Figure 3.3: Examples of scanpaths and tile categories. The signal-to-noise ratio has been increased for illustration purposes.....32

Figure 3.4: The expected image from randomly sampling tiles: (A) raw, (B) contrast stretched, and (C) low-pass filtered (using a 3×3 pixel Gaussian mask, with pixel).....34

Figure 3.5: The average images for (a) triangle and (b) dipole target search are shown for 3 observers. Columns labeled A contain the raw average images collectively scaled to a single common grayscale color map, and columns labeled B contain the raw images after low-pass filtering and individual contrast enhancement.....35

Figure 3.6: Classification images for (a) triangle and (b) dipole target search are shown for 3 observers. (c) Foveal and nonfoveal classification images combined across observers. (d) Classification images combined across foveal and nonfoveal categories and across observers. Columns labeled A contain the raw images collectively scaled to a single common grayscale color map, and columns labeled B contain the raw images after low-pass filtering and individual contrast enhancement37

Figure 3.7: Zero-lag 2-D correlation coefficients showing the structural similarity (a) between the classification images for the triangle search and each of the test shapes, and (b) between the average classification image for the dipole search and each of the test shapes. Error bars show the standard errors of the correlations across observers and categories (foveal and non-foveal).....39

Figure 3.8: Classification images for triangle target search in $1/f$ noise, without a stimulus grid, are shown for 3 observers (panel (a)), and the combined classification images are also presented (panel (b)). Columns labeled A contain the raw classification images after individual contrast enhancement, and columns labeled B contain the raw images after low-pass filtering and individual contrast enhancement.....41

Figure 3.9: Zero-lag 2-D correlation coefficients showing the structural similarity between the classification images for the triangle search and those for each of the test shapes, comparing the main experiment with the no-grid control experiment. Error bars show the standard errors of the correlations across observers41

Figure 3.10: The average images for the triangle target search in white noise are shown for 3 observers. Columns labeled A contain the raw average images collectively scaled to a single common grayscale color map, and columns labeled B contain the raw images after low-pass filtering and individual contrast enhancement.....43

Figure 3.11: Classification images for the triangle target search in white noise are shown (a) for 3 observers, (b) combined across observers for foveal and nonfoveal categories, and (c) combined across foveal and nonfoveal categories and across observers. Columns labeled A contain the raw images collectively scaled to a single common grayscale color map, and columns labeled B contain the raw images after low-pass filtering and individual contrast enhancement.....43

Figure 3.12: Combined classification images for uniform noise (A) after being pinkened and (B) after being pinkened, low-pass filtered, and contrast stretched	45
Figure 3.13: Zero-lag 2-D correlation coefficients showing the structural similarity between the classification images for the triangle search and each of the test shapes, comparing the main experiment with the (pinkened) white noise control experiment. Error bars show the standard errors of the correlations across observers	45
Figure 3.14: Graph of observer performance over time measured as cumulative number of hits	47
Figure 3.15: Box plots of the success rates across observers for four different eccentricity regions are shown for (a) triangle and (b) dipole search with $1/f$ noise stimuli with superimposed grid	48
Figure 3.16: Dwell-time distribution for each observer: (a) non-final and (b) final fixations.....	50
Figure 4.1: Targets used in our fifteen separate experiments. Gabor patches of spatial frequency 2 (first row), 4 (second row), and 8 c/deg (third row), oriented anticlockwise from the vertical at (a) 0 deg, (b) 20 deg, (c) 45 deg, (d) 70 deg, and (e) 90 deg were used as targets	55

Figure 4.2: Stimulus creation, data capture, and data analysis. (a) A Gabor patch used as a target (b) was added to a randomly selected tile of the $1/f$ noise grid. Observer eye movements were recorded while they searched for the target. A representative scan path is shown for a trial in which the observer did not find the target, located in the center of the leftmost column. (c) Fixated tiles that did not contain the target constitute our nonfoveal false alarm category, and (d) a subset of these tiles, which were mistakenly selected at the end of the trials as the target by the observer, constitute our foveal false alarm category. (e and f) Average difference spectra were computed by averaging the amplitude spectra of noise tiles in each category and subtracting the spectral bias (see text)57

Figure 4.3: Average difference spectra for three observers in the (a) 2, (b) 4, and (c) 8 c/deg Gabor search experiments. Average difference spectra, smoothed and contrast-stretched for visual enhancement, are shown for sets of 700 trials for visual searches for Gabor targets oriented anticlockwise from the vertical at 0 (first column), 20 (second column), 45 (third column), 70 (fourth column), and 90 deg (fifth column). For each observer and each set of trials, the spectra were created using about 210 and 2,800 noise tiles respectively for the foveal and nonfoveal categories. Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively. Regions in green show frequency components close to the bias. We have indicated the spatial frequency of the search target (for the horizontal and vertical orientations).....61

Figure 4.4: Complementarity effect for three observers. The average of aligned average difference spectra are shown for (a) 2 c/deg, (b) 4 c/deg, and (c) 8 c/deg Gabor search experiments for three observers. Regions in red indicate frequency components having amplitudes above the bias and regions in blue specify those components having amplitudes below the bias65

Figure 4.5: Complementarity effect for all three observers combined. The average of aligned average difference spectra are shown for (a) 2 c/deg, (b) 4 c/deg, and (c) 8 c/deg Gabor search experiments for all observers combined. Regions in red indicate frequency components having amplitudes above the bias and regions in blue specify those components having amplitudes below the bias66

Figure 5.1: Stimulus creation, data capture, and data analysis. (a) A Gabor patch was used as a target and (b) its orientation was randomly selected from the set $\{0, 1, 2 \dots 179\}$ deg. (c) The target was added to a randomly selected tile of the $1/f$ noise grid and observer eye-movements were recorded while they searched for the target. An example of scan path is shown for a trial in which the observer did not find the target, located in the center of the leftmost column. (d) Fixated tiles that did not contain the target constitute our non-foveal false alarm category, and (e) a subset of these tiles, which were mistakenly selected at the end of trials as the target by the observer, constitute our foveal false alarm category. (f and g) Average difference spectra were computed by averaging the amplitude spectra of noise tiles in each category and subtracting the spectral bias (see text).....74

Figure 5.2: Average difference spectra and performance plots for four observers. Average difference spectra, smoothed and contrast-stretched for visual enhancement, are shown for the first set of 700 trials (first column), the second set of 700 trials collected approximately one month later (second column), and for all the 1,400 trials (third column). For each observer and each set of trials, the spectra were created using about 210 and 2,800 noise tiles respectively for the foveal and non-foveal categories. Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively. Regions in green show frequency components close to the bias. In the fourth column, we have cropped and enlarged the results from the third column to better visualize the spectral structures, and we have indicated the spatial frequency (8 c/deg) of the search target (for the horizontal and vertical orientations). Observers performance (correct target detection rate, on a scale 0 to 1) are shown as a function of the orientation of the Gabor patches (pooled into 15 deg bins and averaged, each bin containing about 117 trials). We indicate in red the average performance of each observer across all orientations (this is close to the 68%). Regions in yellow and gray indicate performance above and below the observer's average performance, respectively.....77

Figure 5.3: Test for sequential bias. Rotated average difference spectra, smoothed and contrast stretched for visual enhancement, are shown. Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively. Regions in green show frequency components close to the bias. See text for details.....79

Figure 6.1: Targets used in our two experiments. Plaids created from the sum of two orthogonally oriented Gabor patches, one oriented at 0 deg and the other at 90 deg. The three targets used had component spatial frequencies of (a) 2, (b) 4, and (c) 8 c/deg and bandwidths of 0.25 octaves. Gabor mosaics created using various configurations ((d) “X”, (e) “O”, and (f) “V”) of two pairs of Gabors of spatial frequency 8 c/deg and bandwidth 0.5 octaves, oriented at – 45 and 45 deg were used as targets in our second experiment83

Figure 6.2: Results for three observers in Experiment 1. Average difference spectra, smoothed and contrast-stretched for visual enhancement, are shown for sets of 700 trials for visual searches for targets with spatial frequency components centered at (a) 2 c/deg, (b) 4 c/deg, and (c) 8 c/deg. The first column shows results for search for a Gabor at 0 deg. The second column results for search for a Gabor at 90 deg. The last column shows the results for search for the sum of the two Gabors (indicated as 0 + 90 deg). Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively. Regions in green show frequency components close to the bias. For each observer and each set of trials, these images were created using about 210 and 2,800 noise tiles respectively for the foveal and non-foveal categories86

Figure 6.3: Results for two observers in Experiment 2. Average difference spectra, smoothed and contrast-stretched for visual enhancement, are shown for sets of 600 trials for visual searches for Gabor mosaic targets with “X” (first column), “O” (second column), and “V” (third column) configurations, as described in the text. Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively. Regions in green show frequency components close to the bias. For each observer and each set of trials, these images were created using about 180 and 2,400 noise tiles respectively for the foveal and non-foveal categories90

Figure 6.4: Results for two observers in Experiment 2. Localized average difference spectra, smoothed and contrast-stretched for visual enhancement, are shown for sets of 600 trials for visual searches for Gabor mosaic targets with “X” (first two columns), “O” (second two columns), and “V” (third two columns) configurations, as described in the text. Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively. Regions in green show frequency components close to the bias. For each observer and each set of trials, these images were created using about 180 and 2,400 noise tiles respectively for the foveal and non-foveal categories.91

Figure 7.1: Targets used to test our models: (a) star, (b) triangle, (c) wrench, and (d) banana.95

Figure 7.2: Gabor filter bank. (a) Even- and (b) odd-symmetric Gabor filters of spatial frequencies of {2,4,8} c/deg, orientations of {0, 22.5, ..., 157.5}deg, bandwidth of 1 octave, and aspect ratio of 2) are shown.96

Figure 7.3: Envelope responses obtained for the star. Regions in red and blue indicate high and low values in the responses, respectively. Note that each individual envelope has been normalized for visual purposes .97

Figure 7.4: : Examples of correct detection rates, as a function of the target contrast, obtained for Model 1 for search for the (a) triangle and (b) star.98

Figure 7.5: Examples of correct detection rates, as a function of the target contrast, obtained for Model 2 for search for the (a) triangle and (b) star... ...99

CHAPTER 1

Introduction

1.1 Motivation

Efficient visual search strategies have been vital for the survival of many species, including humans, particularly in locating food, mates, and predators. Search appears in our daily lives in many tasks such as looking for lost keys in an apartment, searching for a friend in a crowd, finding an empty seat on a bus, locating green apples at the grocery store, or simply while walking in the street. Despite the complexity of many of these tasks, we conduct search with great ease unrivaled to this day by any artificial system. For researchers, it has been an ongoing challenge to understand and model human strategies in search tasks. Such knowledge could provide powerful insights for improving machine performance in similar tasks, especially for active vision systems (e.g. robots). Many applications such as unmanned vehicle navigation, image/video database search, automatic tumor detection, and security/surveillance systems could potentially flourish by integrating human-based strategies into their automated search models.

The main objective of my research has been to discover previously unknown low-level fixation strategies employed by human observers during visual search tasks for various types of targets and experimental conditions. More specifically I have contributed to our understanding of saccadic targeting and target selection in naturalistic visual search tasks, i.e. what attracts fixations while human observers perform search tasks and how observers select target candidates upon fixation. I have shown that observers use target attributes such as spatial frequency and orientation in saccadic programming and in selecting the target candidate upon fixation, and this under a great variety of experimental

conditions. I have used some of the findings towards applications for automated visual search.

1.2 Contributions

The contributions of this dissertation are as follows:

1. I have created a new and efficient technique akin to psychophysical reverse correlation and have used stimuli that emulate the natural visual environment to examine observers' ability to locate low-contrast targets under various experimental conditions. I have sought to address through my framework two intriguing questions in active visual search: what attracts human eye fixations during search tasks and how are target-candidates selected upon fixation? With my classification taxonomy, I am able to provide insight into foveal and peripheral processes employed in visual search tasks.
2. I have demonstrated visual guidance in saccadic target selection in a series of 15 separate visual search experiments where Gabor targets (2, 4, and 8 c/deg spatial frequencies at 0, 20, 45, 70, and 90 deg orientations) were used. This was shown by observers' selectivity for spatial frequency and orientation characteristics close to the search target. Additionally, I have shown that observers exhibit inaccuracies and biases in their estimates of target features. Furthermore, complementary type frequency responses were observed, with peaks occurring at frequencies close to that of the sought target and valleys at nearby frequencies (similar to findings in physiology and psychophysics; e.g. Ringach (1998) showed that observers' tuning for orientation generally

presented a “Mexican hat” distribution peaking at orientations close to the orientation observers had to report, with valleys at either side of the peak). Finally, an unusual phenomenon is observed whereby distracters containing close-to-vertical structures are fixated in searches for non-vertically oriented targets. My results provide evidence for the involvement of band-pass mechanisms along feature dimensions (orientation and spatial frequency) during visual search.

3. I have successfully addressed a more general problem in visual search where the orientation of the target is not known to the observer a priori. Such an experimental procedure is more consistent with real-world search environments, in which the orientation of an object is largely uncertain, except that it may be influenced by gravity or its proximal interaction with other objects and planes. I have used my efficient experimental search framework to study the behavior of humans seeking a randomly oriented Gabor of spatial frequency 8 c/deg embedded in noise. Interestingly, I have found that observers seem to rely on invariant target features to perform such search tasks; in particular, the spatial frequency characteristics of the sought target appeared to provide guidance in saccadic targeting. Curiously, despite having no previous knowledge of each target's orientation, observers presented clear biases in orientation selectivity during saccadic programming. These biases persisted into observers' decision-making process upon fixation and showed asymmetries between clockwise and anticlockwise orientations. Moreover, it appears that these biases are idiosyncratic to observers.

4. With the objective of moving towards understanding observer search strategies for complex targets, I have used my experimental search framework to examine how observers search for low-contrast targets created from Gabor summations (Experiment 1) and mosaicing (Experiment 2). I have presented several key findings. First, I showed a strong presence of visual guidance in saccadic programming in search for such complex targets, demonstrated by selectivity for spatial frequencies and (in some cases) orientations close to the characteristics of each target. Second, multiple orientation attributes of the targets were shown to be represented in saccadic targeting and target selection in most cases, modulated by the observer's sensitivity/selectivity for each orientation. Third, different configurations of the Gabor mosaicing produced distinct tunings in orientation, but visibly idiosyncratic to each observer (Experiment 2). Moreover, a localized analysis was performed. Fourth, a curious presence of close-to-vertical structures was observed in fixated distracters, although the search targets did not contain vertically-oriented structures (Experiment 2).

5. I have provided some ideas and insights inspired from what I have discovered in my experiments that could be integrated into automated visual search. A few of these ideas have been implemented in simple search frameworks. In my proposed models, the non-foveal selection of fixation loci is done using only spatial frequency and orientation attributes of the target, without the use of the phase information. The foveal selection uses local feature attributes of stimuli, so as to include some phase information.

1.3 Outline

The subsequent chapters in this dissertation are organized as follows. In Chapter 2, an overview is given on various approaches to the study of visual search. Furthermore, a technique known as classification images, introduced for the study of visual psychophysics about twenty years ago, and its extensions to visual search are discussed. Additionally, various implementations of visual search both passive and active are presented. In Chapter 3, a novel technique for studying saccadic targeting and target selection in visual search is presented. The technique is then used to examine observers' ability in locating low-contrast targets embedded in $1/f$ noise. In Chapters 4, 5, and 6 we have employed our framework to the study visual search for targets of known spatial frequency and orientation attributes, of unknown orientation characteristics, and containing multiple orientations, respectively. In Chapter 7, we have applied some of our findings towards the design of a simple automated visual searcher. Finally, Chapter 8 concludes the dissertation and points to various directions for the future work.

Chapter 2

Background

2.1 Visual Search

Visual search can simply be defined as the task of looking for an object of interest in a cluttered visual environment. Due to its foveated nature, the human visual system succeeds at such tasks by making many discrete fixations linked by rapid eye movements called *saccades*. The goal is to direct the highest resolution region of the retina, the *fovea*, onto various locations of a scene in quest of potential target candidates. It has been suggested that the location of these saccades is far from being random (Yarbus, 1967) and not until recently has there been increasing interest in understanding how the brain decides where to make saccades and perform fixations. This delay in studies of saccadic targeting is partly due to the impact of visual search theories such as the one proposed by Treisman and Gelade (1980) where much emphasis was made on the role of internal attention (Liversedge & Findlay, 2000). Simple search tasks were claimed to be done by pre-attentive processes and that more complex ones were done by serial shifts of attention from one target candidate to another, generally assuming *covert* attention which is mental scanning as opposed to eye movements. Many of these earlier studies of search have only evaluated human performance by measurements such as reaction time (amount of time necessary for the observer to find a target) and accuracy (percentage of correct), neglecting somewhat the vital question of how search is actually performed. It has been through eye tracking that a smaller number of researchers have been able to study the role of *overt* attention, which takes into account eye movements.

2.1.1 Deployment of Visual Attention

The main theme, in many earlier visual search studies, has been to determine whether a search task is executed by human observers in a parallel or serial fashion, referring to attentive processes (see Wolfe, 1994; Wolfe, 1998; Kim & Cave, 1999; Palmer *et al.*, 2000; and Yang *et al.*, 2002 for detailed reviews). In laboratory environments, observers were typically asked to determine whether a target was present amongst a set of distracters or not. Targets and distracters would differ in single or multiple attributes such as color, size, orientation, spatial frequency, and so on. Figure 2.1 shows some examples of stimuli shown to observers. One of the most used measures to study visual search tasks has been reaction time as a function of set size (the number of items displayed), i.e. the time required for observers to provide a response on whether a target is present or absent. The rationale used by many researchers was to determine the (reaction time \times set size) slope: if the slope was close to flat (i.e. the reaction time appeared independent of the set size) then the search was said to be parallel, and if the slope was increasing then it was said to be serial (Schneider & Schiffrin, 1977; Kinchla, 1992; Bundesen, 1996). Often times, search tasks have been compared based on the slopes obtained for each task. If task A had a greater slope than task B, then task A was said to be more difficult than task B. Accuracy as a function of set size was also used to speculate on the difficulty of a task. Through these various experimentations, processes of visual attention have been shown to be guided by attributes such as color, size, spatial frequency, and others (see Wolfe, 2004 for a review).

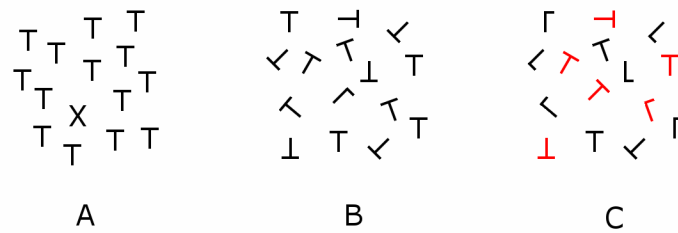


Figure 2.1: Three examples of stimuli used for high threshold visual search tasks are shown representing cases of (A) pop-out, (B) single-feature, and (C) conjunction of features.

Furthermore, many models originated from these studies of visual attention. One of the most prominent theoretical models was the feature integration theory (FIT), proposed by Treisman and Gelade (1980), where search was split into two stages: an initial parallel stage where the visual scene is registered along independent dimensions including color, spatial frequency, orientation, and motion; and a final stage where these dimensions are combined together to represent a single object, performed serially with focal attention. FIT suggested that parallel search with no attention limits occurred in most cases where the target differed by one single feature from the distracters, and that serial search occurred in all the other cases. However, the latter claim has been challenged by many studies (Theeuwes, 1995; Eckstein, 1998). Eckstein (1998) showed that the performance of human observers in search experiments with conjunction of orientation and contrast features could not be predicted by a serial search model. Subsequent search models have been more successful in explaining search behaviors in the context of feature conjunctions (Duncan & Humphreys, 1989; Wolfe, 1994). However these models have avoided considering saccadic eye movements and fixation selections although this is how naturalistic search is performed in many cases.

2.1.2 Guidance of Eye Movements

Recent research has put emphasis on the importance of incorporating eye movements in visual search studies (see Findlay, 2004; Findlay & Gilchrist, 2003 for a review). In fact many researchers have challenged earlier theories and studies that had omitted eye movements (Zelinsky & Sheinberg, 1997; Liversedge & Findlay, 2000; Eckstein *et al.*, 2001; Findlay & Gilchrist., 2003). Zelinsky and Sheinberg have argued that much of the work had been mainly to estimate the relative difficulty of tasks compared to one another. They further stressed that such studies had shrunk a highly complex spatial and temporal behavior into a simple response time measure. In fact simple measures such as reaction times and accuracy overlook more valuable information such as the fixation patterns, dwell times, saccade lengths and so on. Note that, it has even been demonstrated in a few tasks that observers opt to perform eye movements even when such a strategy is not optimal (Findlay, 1997; Findlay & Gilchrist, 1998).

A fundamental problem that many studies using eye movements have attempted to address has been to determine whether saccades are guided and where they land. It has been shown that saccadic eye movements are not random (Yarbus, 1967). However, there is some discord on the fixation loci during visual search tasks (Findlay and Gilchrist, 2003; Hooge and Erkelens, 1999; Motter and Belky, 1998b; Findlay, 1997; Zelinsky, 1996). In some experiments where eye movements were allowed, a group of researchers claimed that saccades, mainly the initial one, were directed to the center-of-gravity of elements in the display (Zelinsky *et al.*, 1997; McGowan *et al.*, 1998). For example, Zelinsky *et al.* tracked the eye movements of six human observers searching for objects placed in what they called pseudorealistic scenes (i.e. toys in a crib, tools on a workbench, and food-related objects on a dining table). 360 trials were run for each observer, with each trial having a unique configuration of objects and positions on a

surface. Three different set sizes were used. A search target was shown to the observer before each trial and the observer had to decide whether the target was present during the trial. In Figure 2.2, the top left panel shows an example of the placement of the toys in a crib, and the top right panel shows recorded eye movements of an observer searching for the butter target amongst two distracters on a dining table. The results for the first, second and third saccades across all trials for one naïve observer are shown in the three bottom panels of Figure 2.2, the squares representing the locations where objects appeared. They found that the first saccade appeared to be directed to the center-of-gravity of the group (see bottom left panel in Figure 2.2). On the other hand, some researchers have argued that saccades are made to elements in the display and not to the blank spaces (Motter & Belky, 1998; McGowan *et al.*, 1998). For instance, Motter and Belky discovered in their tilted bar experiments that saccades landed within 1 deg of the center of the target or distracters having similar features as the target. To address this issue of saccadic targeting, McSorely and Findlay (2003) performed two experiments with Gabor patches and showed that the center-of-gravity effect decreased with increasing number of distracters. Interestingly, they found in their experiments that the search performance improved with a larger number of distracters. One explanation that has been proposed for such phenomenon is that perceptual grouping of similar objects can result in enhanced search speed and accuracy.

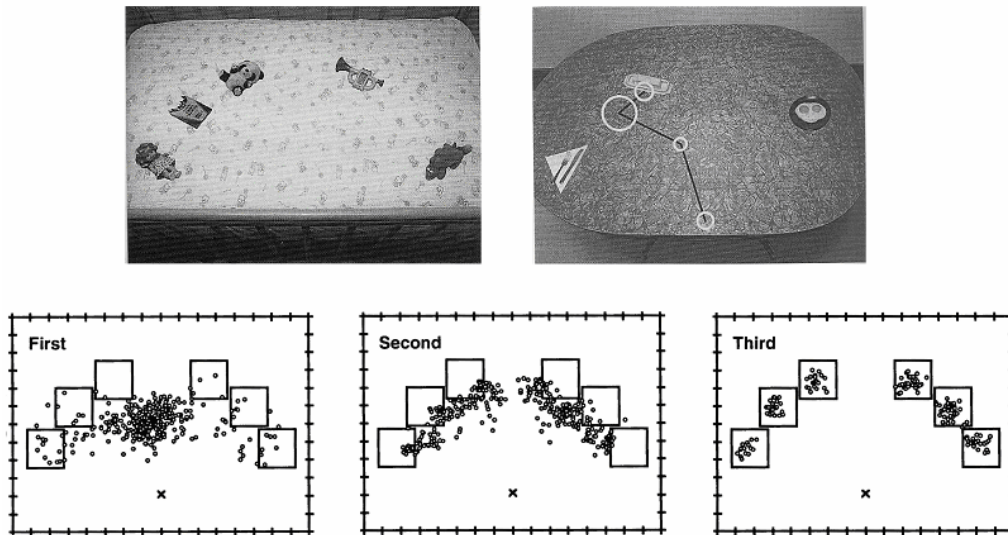


Figure 2.2: Eye movement recordings during search. The top left panel shows a configuration of toys placed in a crib. The top right panel shows the recorded eye movements of an observer searching for the butter target on a dining table during a trial. Endpoints of the first, second and third saccades are shown in the bottom panels, the squares indicating the possible locations where objects appeared (from Zelinsky *et al.*, 1997).

A natural question one could then ask is what stimuli features guide eye movements. Many studies have qualitatively compared the relative weighting of features in the visual guidance of saccadic targeting. For instance, Williams found that observers have a strong inclination to direct saccades to elements of the display having the same color as the target, while information on target size and shape were weakly used (Williams, 1967 from Eckstein *et al.*, 2001 and Williams & Reingold, 2001). Scialfa and Joffe (1998) discovered in their tilted bars experiments, where the target and distracters differed in either contrast (black or white) or orientation (± 45 deg), that observers were more likely to direct their saccades to distracters that shared similar contrast as the target. Rajashekar *et al.* (2002) went further by looking at the statistics at the point of gaze, when observers were searching for a target embedded in noise. They found that, on average, saccades were made to regions of the display containing non-random structures

and showing some resemblance to the target. We have detailed these findings in section 2.2.

2.1.3 Low-level vs. High-level Mechanisms

One of the important issues when studying visual search has been whether observer performance and behavior should be attributed to high-level or low-level processes. This has been a source of some conflict between the cognitive and the “bottom-up” factions of the vision community. Most research studies have tried to isolate one type of process by limiting the other; for instance, low-level studies have used well-trained observers and simplistic stimuli in order to reduce the effects of cognitive mechanisms. Geisler and Chou (1995) suggested that in numerous complex tasks both high-level and low-level mechanisms influence observer performance and that many studies have not been successful in showing the weighting between the two. They introduced a technique to separate the influence of both low and high-level processes, and they further demonstrated in two complex tasks that low-level factors had the highest influence. This was shown by comparing the performance of observers in a visual search task to performance in a well-constrained discrimination task, both experimental setups using the same stimuli. In a controversial paper, Henderson *et al.* (2006) have argued that visual saliency map models obtained by using low-level features fail to account for observer fixations in naturalistic search environments. They showed that observer fixations and the regions rated highly informative by Koch and Itti’s visual saliency model (2000) were weakly correlated. They demonstrated this claim by comparing observer fixations when counting the number of people in a scene to the saliency map obtained by the model. One could counter argue that such a finding is most likely related to their task. In fact, if one had asked observers to count the number of red cars in a

parking lot, the result would probably be much different, with the low-level feature color having a great role in the search process. One could even go further to speculate that many additional red objects other than cars in that parking lot may also attract observer fixations.

Understanding the cognitive aspects of observer visual search strategies can be very useful in real world environments but is somewhat limited to specific objects of interest and may be more difficult to extend to search for other objects. Intuitively speaking, humans possess different high-level approaches to finding faces as compared to finding cars. For example, one may be looking for the wheels when searching for cars in a scene, an approach that may not be extended to finding people in the same scene. By contrast, a low-level approach could help build the foundation of visual search to discover and understand what features are common between various search tasks.

2.2 Classification Images and Extensions to Visual Search

A technique known as the classification image paradigm (Ahumada, 1996; Beard & Ahumada, 1998) can be of particular interest in revealing stimuli features used by observers in various tasks including visual search, in contrast with simple measures of reaction time or accuracy. The classification image paradigm initially originated from work in auditory yes-no detection experiments (Ahumada, 1971; Ahumada, 2002). In this earlier framework, two sound tracks, one containing a continuous noise masker and the other containing marker tones, were presented to human observers who had to decide whether a tone was present or not. The noise masker was then analyzed to see whether correlates with observer responses could be obtained. Ahumada later extended this technique to visual yes-no psychophysical tasks and introduced the concept of classification images (Ahumada, 1996; Beard & Ahumada, 1998). In the classification

image paradigm, observers judge the presence or absence of a target embedded in relatively high-amplitude noise, and properties of the noise that cause correct and incorrect responses reveal the mechanisms responsible for detection.

To demonstrate the various steps constituting the construction of the classification images, let us consider the vernier acuity task where a human observer has to make judgments on the alignment of two bars (see Figure 2.3) over several trials, usually several thousands. During each trial, a stimulus is first constructed by adding random noise to one of the two arrangements of the bars chosen at random. The stimulus is then presented to the observer who has to decide on configuration of the bars (left- or right-aligned). The signal-to-noise ratio of the stimulus is set to influence observer's decision but without completely changing the decision rule. During each trial, the noise image, the stimulus configuration (S_r or S_l) and the observer's response (R_r or R_l) are recorded. The noise images are then classified into one of the four categories $S_r R_r$, $S_l R_r$, $S_r R_l$, and $S_l R_l$, based on observer's response during each trial. For example, if the left-aligned configuration is presented to the observer but the observer decides that the alignment is to the right, then the noise image is classified as $S_l R_r$. Noise images within each category are then averaged and the resulting images combined across categories to create the classification image: $S_r R_r + S_l R_r - S_r R_l - S_l R_l$ (Ahumada, 1996). The obtained result provides insight into how the observer is weighing stimulus features to make a decision. In the vernier acuity example, Ahumada and Beard argued that the features obtained by the classification images disprove a strategy based on the contrast sensitivity of the highest cortical unit response (single even-symmetric Gabor filter) or on the difference of two even-symmetric Gabor filters oriented on either side of the target. Instead, they supported a third strategy consisting of a two odd-symmetric Gabor filters since they are also consistent with non-abutting bars (Beard & Ahumada, 1998).

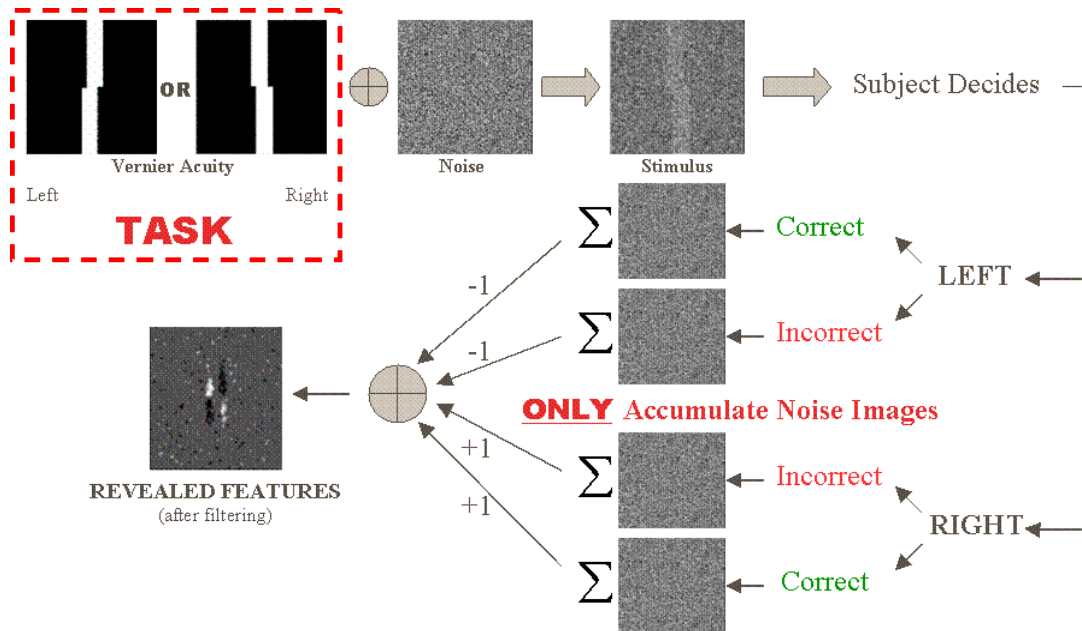
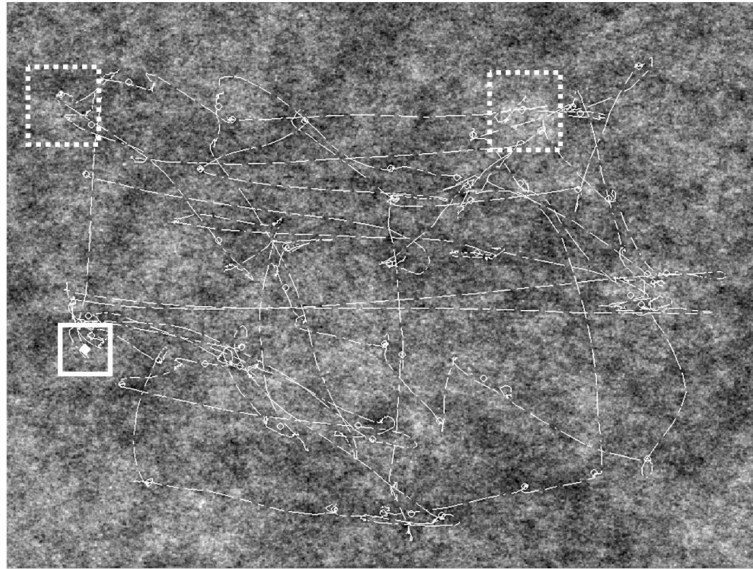


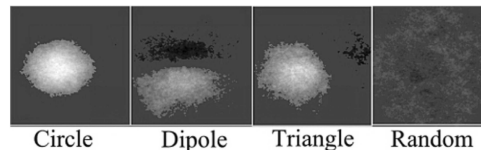
Figure 2.3: Demonstration of the classification image technique applied to a vernier acuity task.

The classification images technique has been extended to include eye movements during visual search (Rajashekar *et al.*, 2002 & 2004; Eckstein *et al.*, 2007). Rajashekar and colleagues recorded observer eye movements while they searched for a target embedded in $1/f$ noise. Assuming that gaze would be drawn to points in the stimulus bearing some resemblance to the target, the noise at all fixations made during a trial was captured, and a large volume of data could thus be gathered in a short time. Observer visual search strategies were analyzed by presenting stimuli that consisted of 640×480 pixel $1/f$ noise images in which a 64×64 pixel target was embedded. A “region of interest” (ROI) of 128×128 pixels around each of the observer fixations was defined and is shown by the dashed boxes in Figure 2.4a. These noise patches were then averaged together and filtered to obtain what they referred to as *discrimination images*. These images are shown in Figure 2.4b for three targets (circle, dipole, and triangle) and are compared to the discrimination images obtained from random fixations. The results were

interpreted as representing features that draw gaze in the periphery and trigger a closer inspection.



(a)



(b)

Figure 2.4: An example of scan paths is shown in (a); discrimination images are shown in (b) (from Rajashekar *et al.*, 2002 & 2004).

Classification images have been applied to an assortment of visual perception problems such as illusory contours (Gold *et al.*, 2000), image feature detection and identification (Neri & Heeger, 2002), stereo (Neri *et al.*, 1999), visual attention (Eckstein *et al.*, 2002), Gabor detection (Ahumada & Beard, 1999; Solomon 2002), face and facial expression discrimination (Sekuler *et al.*, 2004; Kontsevich & Tyler, 2004), and “superstitious” perception (Gosselin & Schyns, 2003). A few researchers have also

attempted to extend the technique to color stimuli (Ahumada & Krebs, 2000; Bouet & Knoblauch, 2004; Hansen & Gengenfurtner, 2005).

One of the main drawbacks of the classification image paradigm in its original form is the need to accumulate a large number of data (close to several thousands of trials per human observer). Furthermore, the extension proposed by Rajashekar *et al.* has also its limitations, mainly lacking in spatial specificity due to the fact that the noise pixels being averaged across trials are not always perfectly aligned (i.e. observer fixations do not always land at the center of the target candidates, see Rajashekar *et al.*, 2006). In the next chapter, I will present a technique addressing both these issues that we have developed to reveal human observer strategies and behaviors during visual search.

2.3 Machine Vision

Perhaps one of the most important tasks in visual scene analysis is searching for objects. It has been of central interest for decades to computer vision and pattern recognition researchers, who have proposed various algorithms and techniques attempting to solve this complex problem. Many of the suggested methods have been developed for full resolution images of a scene, acquired passively by cameras without any type of analysis or any knowledge of the scene to be studied. Not until recently has there been growing interest in active vision systems, where various parameters of the camera(s), including gaze control, are guided by prior knowledge of the scene and/or by decision stages in the search algorithms.

2.3.1 Search in Passively Acquired Images

A wide range of algorithms have been proposed to carry out search tasks such as finding cars, people, and so on in passively acquired images of scenes (see Shivani & Agarwal, 2004; Anuj & Mohan, 2001 for reviews). These methods can be grossly divided into template-matching, feature-based detection, and learning-based approaches. Image subtraction and correlation are operations commonly used in template-matching, with the goal of minimizing the distance between the object being searched (i.e. the template) and a region of the image (Brunelli & Poggio, 1997). In feature-based techniques, an object is represented by various attributes such as color, orientation, and motion. During the detection process, features are extracted from the image being searched and matched to those of the object. In learning-based approaches, images are also represented by features but a learning stage is used to find regions of the feature spaces that correspond to the object class. Learning is done using a training set. It is not until the mid-80s, with advances in physiological research on human vision, that more powerful, biologically inspired methods were developed to tackle the visual scene analysis problem (Milanese, 1998). We will mainly focus in this section on biologically inspired search algorithms.

Possibly the most prominent of such biological vision systems is the saliency-based attention/search algorithm introduced by Koch and colleagues (Koch & Ullman, 1985; Itti & Koch, 2000). In their model, shown in Fig. 2.5, a single saliency map is obtained by the combination of 42 maps (7 feature types at 6 scales). Their system takes as input intensity (on/off), orientation (0, 45, 90, and 135 deg), and color (red, green, blue, and yellow color channels). 7 feature types for which there exists biological evidence are then obtained in a center-surround fashion: 1 encoding for on/off intensity contrast, 2 encoding for red/green and blue/yellow double-opponent channels, and 4 encoding for local orientation contrast. 6 feature maps, corresponding to 6 spatial scales,

are computed for each feature type by using combinations of levels of a Gaussian pyramid, obtained by repeatedly low-pass filtering and sub-sampling the input image. The center-surround operations across scales were executed by differencing a fine and a coarse scale for a feature. As an example, the chromatic information was obtained by first normalizing the red, green, and blue channels by the intensity channel. The red/green feature maps were obtained by center-surround differences calculated at 6 different scales by subtracting (red - green) at the center from (green - red) at the surround and taking the absolute value. The same operations were done to construct the blue/yellow feature maps. Separate “conspicuity maps” are created by first normalizing each of the feature maps between 0 and 1, then by iteratively filtering each with a DoG (difference of Gaussians) filter, and finally summing across scales the feature maps obtained for intensity, color, and orientation. This within-feature spatial competition scheme is similar to a winner-take-all strategy (WTA). The three conspicuity maps are subjected to iterative filtering with a DoG. The final saliency map is created by a linear summation of the conspicuity maps. Koch and colleagues tested their model in 3 different visual search tasks (color pop-out, orientation pop-out, and conjunctive search) and on naturalistic stimuli (a military vehicle in a rural environment). They obtained similar performances for their model in psychophysical experiments as the ones predicted by Treisman and Gelade (1980) for human observers: both pop-out experiments produced flat RT slopes and the conjunctive experiment produced an increasing RT slope. They also claimed that their model found the military vehicle in 41 of 44 full resolution rural images based on finding the most salient location.

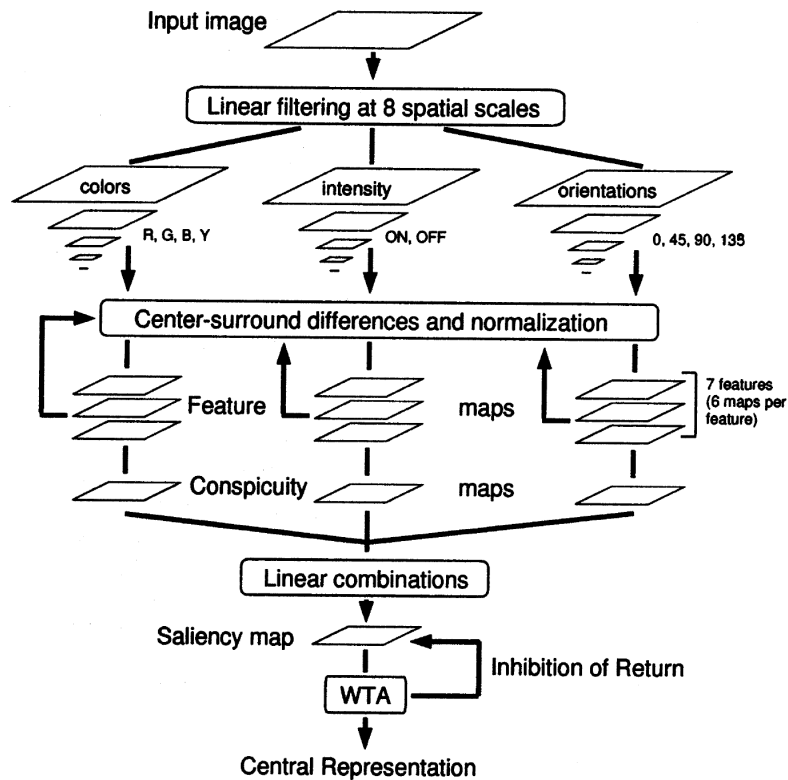


Figure 2.5: Model for saliency-based attention (from Itti & Koch, 2000).

Other attention / visual search algorithms have been proposed and are detailed in Milanese's review (Milanese, 1993). These models propose a similar parallel search approach as Koch and colleagues: Chapman's model, Ahmad's VISIT, and Wolfe and Cave's implementation of the Guided Search. All three algorithms create a number of bottom-up feature maps such as color and orientation by processing the input image. VISIT and Guided Search additionally integrate top-down features in their models. In the Guided Search model, bottom-up activation maps are created by measuring how unusual an item is in comparison to its surrounding (ex. a vertical orientated line surrounded by horizontally oriented lines), hence guiding attention to distinctive items in a scene. But bottom-up activation alone would not guide attention to the target if the target-distracters similarity increases. Therefore, top-down activation, where knowledge about the target is

introduced, becomes helpful in selecting what features better distinguish the target from the distracters. One of the main setbacks of these methods is that features are combined linearly using simple averaging across saliency or activation maps.

2.3.2 Active Vision

As Bajcsy (1998) puts it, “perceptual activity is exploratory, probing, searching; percepts do not simply fall onto sensors as rain falls onto ground.” If one compares the artificial systems presented in section 2.3.1 to the human visual system, one of the key ingredients missing is gaze control combined with foveation, which could provide control over the acquired images of a scene. Often times, images have been passively acquired at full resolution, preceding any type of analysis or any knowledge of the scene to be studied. Besides being counterintuitive vis-à-vis to the biological vision system where various tasks including search are performed in an active fashion, one of the main drawbacks of such an approach has been the computational load for performing tasks (due to the processing of very large images at full resolution) and the lack of adaptability to various environments. In fact, systems should be adaptable to the environment in which a task is performed. If we consider the acquisition stage of our own visual system, our eyes can for instance adjust to various illuminations, correct the focus, and change the view. In addition, foveation also provides a considerable data reduction since peripheral image data are lost in the process.

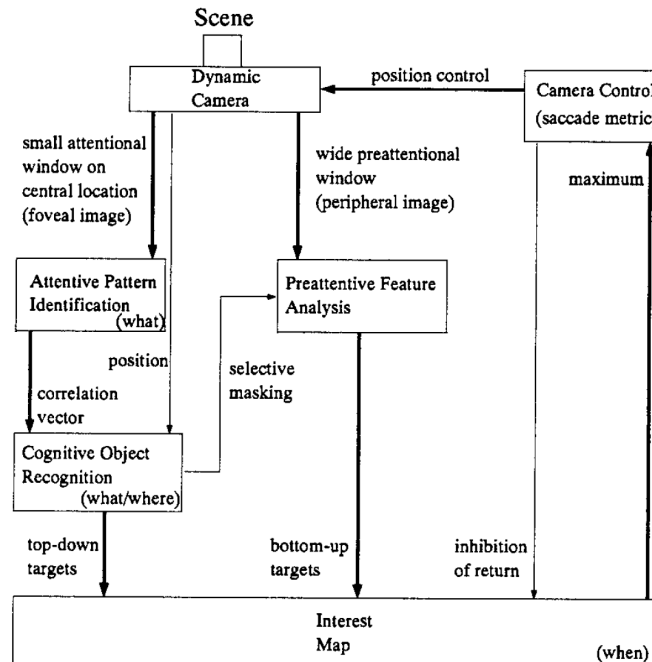


Figure 2.6: Example of an active vision system (from Giefing *et al.*, 1992).

Active vision systems have attempted to provide more rapid and less computationally intensive solutions for performing various scene analysis tasks. The guiding line has been to mimic many aspects of biological visual systems incorporating gaze control, behavior and attention (see Swain *et al.*, 1991; Abbott, 1991 for reviews). Fig. 2.6 shows an example of an active vision system (Giefing *et al.*, 1992) where various stages contribute to accomplishing an object search task. The implementation of gaze control has been possible with the use of a controllable camera that can actively scope complex scenes. Here the dynamic camera was mainly used for gaze shifts but in general parameters such as camera position, orientation, focus, aperture, zoom, and vergence (obtained with 2 cameras, see Klarquist and Bovik, 1998) may be varied. Feedback from decision making stages controls these parameters. In the example, for instance, the interest map provided decisions to guide the camera. Note that both top-down and

bottom-up approaches are used to make a decision whether to visit another location or stop the search. The foveal image was analyzed for specific targets whereas the peripheral image was processed for various features such as edges, corners, and so on.

Chapter 3

A Novel and Efficient Technique to Study Visual Search

The human visual system is remarkably adept at finding objects of interest in cluttered visual environments, i.e. at visual search. It accomplishes this by making many discrete fixations linked by rapid eye movements or saccades. In such naturalistic tasks, we know very little about how the brain selects saccadic targets (the *fixation loci*). Our initial objective was to develop a framework to study saccadic targeting and target selection in naturalistic visual search tasks. In this chapter, we propose a novel and efficient technique (Tavassoli *et al.*, 2007a) akin to psychophysical reverse correlation and stimuli that emulate the natural visual environment to measure observers' ability to locate a low-contrast target, extending earlier techniques (Eckstein *et al.*, 2007; Rajashekar *et al.*, 2006; Ahumada, 1996). We will discuss in detail our method in the next section and how we have successfully tested it for 2 simple shapes, a triangle and a dipole.

3.1 Proposed Method

In our method, a $1/f$ noise mask is divided into discrete tiles (note that other noise types may be used), and the target is embedded in one of them selected at random. The eye movements of observers are then recorded while they search for the target to determine the sequence of tiles fixated during the search. $1/f$ noise has an amplitude spectrum of the form $1/f^a$, where a is near 1, which is similar to the amplitude spectra of natural images (Field, 1987), is used due to its appeal in simulating a naturalistic search environment. The additional power at low spatial frequencies (relative to white

noise) results in rapid emergence of features in the classification image with our method, at the scale of reasonably sized targets, without the requirement for post-processing. Several aspects of our technique allow it to rapidly reveal classification images. First, the use of eye tracking allows a high volume of data to be collected in a given time as compared to traditional psychophysical methods. Second, the use of discrete tiles makes the method more robust to saccadic inaccuracy, the tendency for observers to fixate different parts of the target, and the limited accuracy and precision of the eye movement recordings, all of which would ultimately result in loss of spatial precision (or blur) in the final classification images. Third, our novel classification taxonomy provides several new categories for off-line analysis, allowing us to differentiate foveal and non-foveal aspects of the search process (see section 3.1.6)¹.

3.1.1 Observers

Three male observers (aged 26 through 30), of whom two were experienced (AT and IVDL) and one naïve (AJS), were tested in our experiments, each with normal/corrected-to-normal vision. Each observer completed 400 trials (2 sets of 200 trials, each set with a different target).

3.1.2 Apparatus

An SRI/Fourward Generation V Dual Purkinje eye tracker (Fourward Technologies Inc., Buena Vista, VA) was used to record eye movements. This device has accuracy of better than 10 min of arc, a precision of about 1 min of arc, and a response time of about 1 msec (though we would like to note that a principal advantage of our

¹ We use the term “foveal” to refer to a central patch one degree of visual angle across, and “non-foveal” to refer to regions outside this patch.

methodology is that it permits the use of a considerably less accurate tracker). A bite bar and forehead rest were used to minimize head movements. The continuous output voltage of the eye tracker was first passed through a hardware Butterworth low-pass filter (Krohn-Hite Corp., Brockton, MA) with a 100-Hz cutoff to eliminate extraneous high frequency noise in the recording environment, and then sampled by the host computer at 200 Hz with a National Instruments data acquisition card (National Instruments Corp., Austin, TX).

A calibration routine was run at the beginning of each session and after every 25 trials during a session to establish the linear relationship between output voltage and monitor coordinates. For the calibration, the observer fixated each of nine points in a 3×3 grid spanning a visual angle of $7^\circ \times 7^\circ$ on the display. The average horizontal and vertical voltages were then fit (separately) to the 3 unique horizontal and vertical screen positions (corrections were performed for the small amounts of cross-talk). Afterward, a dot was superimposed on the computed gaze position in real-time so the observer could immediately verify that calibration was successful. In addition to the mandatory recalibration every 25 trials, the calibration was automatically checked at the beginning of each trial. This was done by requiring that the computed fixation be within $\pm 0.25^\circ$ of the center of the fixation mark for 500 msec at the beginning of each trial. If 5 sec elapsed before this requirement was met, recalibration was automatically initiated.

The observers viewed the stimuli on an Image Systems 21-in. grayscale monitor (Image Systems Corp., Minnetonka, MN) driven by a Matrox Parahelia graphics card (Matrox Graphics Inc., Dorval, Québec, Canada) at a screen resolution of $1,024 \times 768$ pixels, a grayscale resolution of 8 bits per pixel, and a refresh rate of 60 Hz. The screen was placed 134 cm from the observer and subtended a visual angle of $16^\circ \times 12^\circ$, giving approximately 1 min of arc per screen pixel. The luminance output was linearized by

putting the inverse of the monitor's measured gamma function in the display look-up table. The ambient illumination in the laboratory was kept constant for all observers, and there was a minimum of 5 minutes to adapt to the ambient illumination and screen luminance while the eye tracker was calibrated.

The experimental software was written in MATLAB (Mathworks Inc., Nantick MA) and the stimulus presentation itself was controlled using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Gaze positions were calculated in real time so that feedback could be provided after each trial. Fixation points and the intervening saccades were discriminated offline, based on spatio-temporal properties of human eye movements, by using an adaptation of an ASL fixation detection algorithm (Applied Science Laboratories, Bedford, MA). This three-stage algorithm was robust with respect to small drifts, blinks, and micro-saccades.

3.1.3 Stimuli

The stimulus consisted of a single 64×64 pixel target embedded in a 7×7 mosaic of 64×64 pixel tiles containing $1/f$ masking noise, where $a = 0.8$. The two targets used are shown in Figures 3.1A and 3.1B (the shapes shown in panels C–E were used in data analysis, but not in the experiment per se; see below). One hundred mosaics were generated offline by creating one hundred 544×544 pixel $1/f$ noise images and then superimposing the 12 pixel wide gray borders. On each trial the target was added to a randomly selected tile in the noise mosaic. An example stimulus in which a triangle is embedded in the tile immediately below the center one is shown in Figure 3.2. The signal-to-noise ratio (SNR) in this example is somewhat higher than those used in the actual experiment, where the SNR was determined at the beginning of each session as described below.

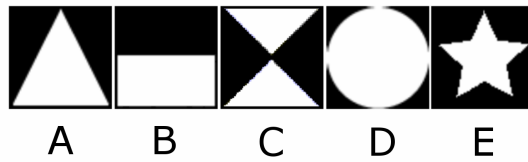


Figure 3.1: Targets used in the trials: (A) triangle, (B) dipole. Additional shapes used in the analysis: (C) bowtie, (D) circle, and (E) star.

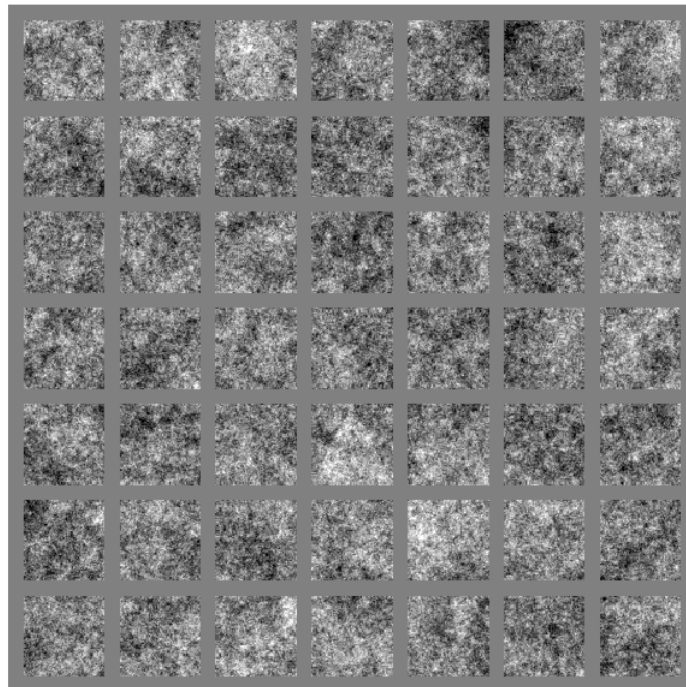


Figure 3.2: An example stimulus (with a higher SNR than used during the experiment). The target, a triangle, is in the tile immediately below the central one.

3.1.4 Procedure

Each observer ran four sessions for the main experiment: two sessions of 100 trials for each of the two target types. Before every session, the SNR yielding 68%² correct target detection was determined using the QUEST adaptive procedure (Watson &

² This percentage has been arbitrarily chosen for all of our experiments, the main idea being to have observers make enough mistakes over several trials.

Pelli, 1983). Note that this is effectively a contrast threshold, but we covaried the contrast of the target and of the noise so that the entire grayscale was used but never exceeded. This SNR threshold was determined using the same procedure as that in the experiment itself. In other words, a trial during the threshold determination was exactly the same as a trial during the experiment, except that, in the former, the SNR was varied to find the 68% correct point, whereas in the latter, the SNR was fixed at that point. Since the first several trials of the QUEST are necessarily done at a relatively high SNR, these trials served to familiarize the observers with the task.

At the beginning of each trial, a fixation mark appeared at the center of the display for a maximum of 5 sec. As was described earlier, if the observer's computed fixation was within our error tolerance, the trial continued. Next, the fixation mark was replaced by the stimulus for 5 sec, and the observer searched for the target with the goal of having his fixation on the correct tile when the trial ended. The computer provided audio feedback ("correct" or "incorrect") after each trial.

The use of a common initial fixation point and a fixed, 5-sec trial duration ensured a somewhat consistent strategy and criterion across observers that yielded several fixations per trial. To wit, if we had used a very short duration, the experiment would effectively become a 49-alternative forced choice yielding few fixations per trial. If we had used long or unlimited durations, different response criteria could have resulted in very different strategies including exhaustive search. We chose 5 sec as a compromise allowing observers to visit several (five to six, on average) likely tiles without the search becoming exhaustive (resulting in fixations on very unlikely tiles). Post hoc analyses (see Results section) suggested that the compromise was an acceptable one. It would also be possible, of course, to use a variable payoff matrix (for example), instead of imposing a time limit, but we chose the simpler option in order to demonstrate our basic method. The

small number of fixations that fell between the tiles in the stimulus grid were not included in our analysis.

3.1.5 Analysis Method

Classification Taxonomy. In a *yes-no* detection experiment, responses can be categorized into hits, misses, false alarms, and correct rejections, depending on the observer’s response and whether the target was actually present. In the psychophysical classification image paradigm, the stimulus noise is averaged within each category, and these averages are combined to form the classification image. For example, the average of the hits and false alarms can be subtracted from the average of the misses and correct rejections, under the assumption that if a given pixel inclines the observer to say “target present” when bright (say), it should also incline the observer to say “target absent” when dark (Ahumada, 1996). The fidelity of the image from each category will actually depend upon the observer’s sensitivity and bias, but the fidelities seem to be about equal in the simple psychophysical situation, so combining the averages with equal weight is close to optimal (Ahumada, 2002).

In this study, we simply extended the above categorization to accommodate eye movements. Consider that each fixation (excluding the initial fixation at stimulus onset) involves two decisions: the decision to fixate a certain tile (and not the others) and the subsequent decision to either remain on that tile or continue searching. The presumption in defining our taxonomy is that the former is based primarily on nonfoveal information and the latter is based primarily on foveal information. Consider the left panel of Figure 3.3; the first fixation is to a tile on the far right, which does not contain the target. This tile can thus be labeled a *nonfoveal false alarm* (\bar{f}_{FA}), since the incorrect decision that the target was in that tile was (presumably) based on peripheral information. Also, each

tile except the central one and the one containing the target can be labeled a *nonfoveal correct rejection* (\bar{f}_{CR}) since the correct decision that the target was not in those tiles was also based on peripheral information, and the tile actually containing the target can be labeled as a *nonfoveal miss* (\bar{f}_{Miss}). Finally, when the eye moves to the subsequent tile (in the lower left), the tile at the first fixation can be labeled a *foveal correct rejection* (f_{CR}), since the decision to reject this tile and continue searching was based on foveal information. Later in the trial, the observer actually fixates the tile containing the target, making that tile an \bar{f}_{Hit} , but then continues searching, so that tile also becomes an f_{Miss} . If the observer had decided to remain on the tile containing the target instead of continuing his search, this tile would have become an f_{Hit} . Trials in which the observer quickly finds the target and in which the observer never fixates the target are shown in the center and right panels, respectively.

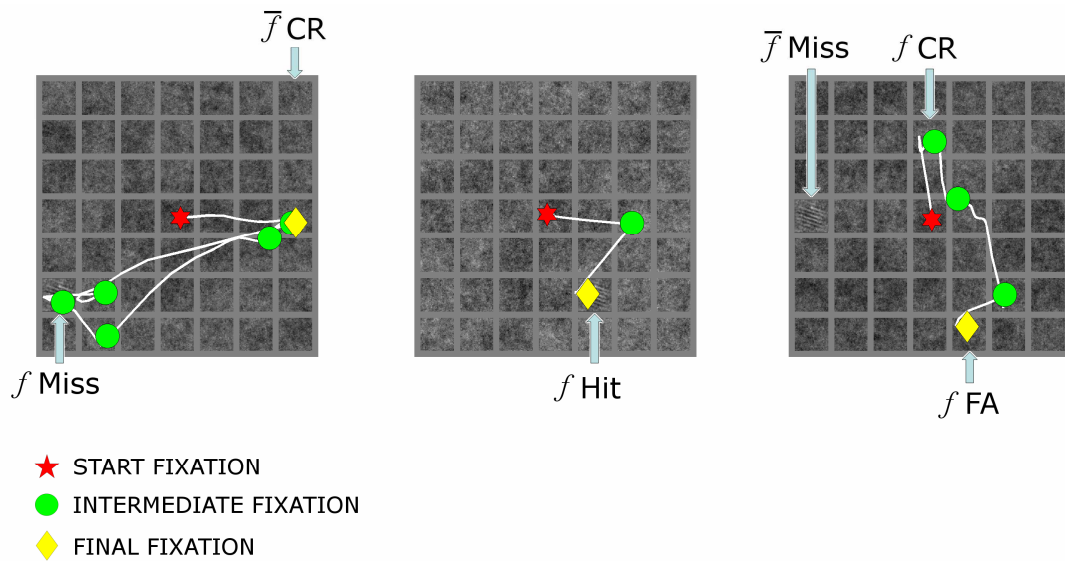


Figure 3.3: Examples of scanpaths and tile categories. The signal-to-noise ratio has been increased for illustration purposes.

Tiles were categorized postexperiment according to Table 3.1 and 3.2 for analysis. Note that for a given trial, each tile can belong to more than one category. As is shown in the tables, each fixated tile was classified as an \bar{f}_{Hit} or an \bar{f}_{FA} , depending on whether the tile contained the target or not. The tile was then additionally classified as one of the foveal categories depending on the observer's response: either maintaining fixation on the tile, indicating he thought the target was there, or continuing the search, indicating that he thought the target was elsewhere. Tiles not fixated were classified as \bar{f}_{Miss} (target present) or \bar{f}_{CR} (target absent).

ALL TILES			
Target?	Attracted?	Class	Max Number of Tiles Possible per Trial
PRESENT	YES	\bar{f}_{Hit}	1
PRESENT	NO	\bar{f}_{Miss}	1
ABSENT	YES	\bar{f}_{FA}	48
ABSENT	NO	\bar{f}_{CR}	48

Table 3.1: The nonfoveal noise tile taxonomy.

ALL FIXATED TILES			
Target?	Observer's Decision?	Class	Max Number of Tiles Possible per Trial
PRESENT	MAINTAIN FIXATION	f_{Hit}	1
PRESENT	CONTINUE SEARCH	f_{Miss}	1
ABSENT	MAINTAIN FIXATION	f_{FA}	1
ABSENT	CONTINUE SEARCH	f_{CR}	(Num of Fixated Tiles - 1)

Table 3.2: The foveal noise tile taxonomy.

Generating the average images and the classification images. Pixel-by-pixel averaging of images within each category was used to obtain the average noise images corresponding to that category. It is important to keep in mind that only the noise patches are used as input to this process and *not* the target. Any structure revealed through these methods therefore originates from the influence of particular samples of noise on the observers' responses.

The average noise tiles were combined in the usual manner (Hit + FA – Miss – CR; Ahumada, 2002) to create the classification images, but this was done separately for our foveal and non-foveal categories:

$$\bar{f}_{CI} = \bar{f}_{Hit} + \bar{f}_{FA} - \bar{f}_{Miss} - \bar{f}_{CR} \quad (3.1)$$

$$f_{CI} = f_{Hit} + f_{FA} - f_{Miss} - f_{CR} \quad (3.2)$$

Because we used a finite number of noise tiles ($49 \times 100 = 4900$), the expected average image that would result by randomly sampling tiles is not uniformly zero but, rather, is the average of all the tiles. This expected image, corresponding to a null hypothesis that an observer does not use spatial structure in the tiles to select fixation points, is shown in Figure 3.4A. As one might expect, it is very flat (with a standard deviation of just 0.0015 on a 0-to-1 scale) but does contain some spatial structure, which can be made clearer by contrast stretching (Figure 3.4B), and blurring (Figure 3.4C). This overall average can be thought of as the bias each pixel has as a result of using a finite number of noise samples. Although the spatial structure in this overall average does not closely resemble the search targets (and we have quantified this assertion by calculating comparative 2D correlation coefficients for each of our experimental targets; see Figure 3.7), we must be aware that any average noise image or classification image resembling this expected image does not possess significant structure of its own.

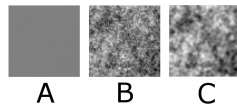


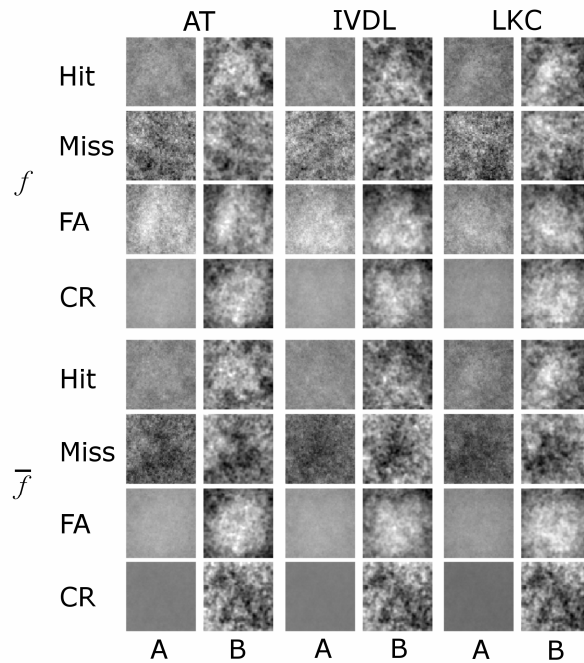
Figure 3.4: The expected image from randomly sampling tiles: (A) raw, (B) contrast stretched, and (C) low-pass filtered (using a 3×3 pixel Gaussian mask, with $\sigma = 0.9$ pixel).

3.2 Results

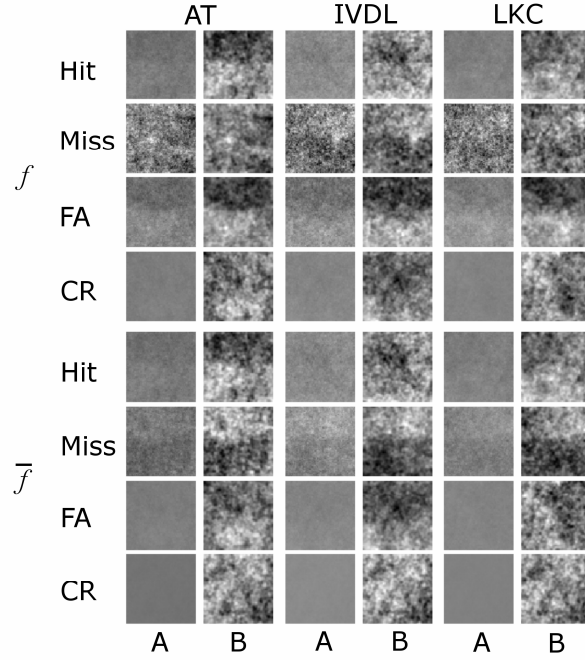
3.2.1 Average/Classification Images

The pixel-by-pixel averages of the noise tiles in each of the eight categories are shown for each observer in Figure 3.5. Columns labeled A contain the raw average images collectively scaled to a single common grayscale color map, and columns labeled B contain the raw images after low-pass filtering (using a 3×3 pixel Gaussian mask with

$\sigma = 0.9$ pixel) and individual contrast enhancement. The former shows the relative fidelity of the average image from each category, and the latter reveals possible structures present in each of the classification images. All the categories presented some target-dependent spatial structure except for \bar{f}_{CR} , which converged to the overall average shown in Figure 3.4. \bar{f}_{Hit} , \bar{f}_{FA} , f_{Hit} , f_{FA} , and f_{CR} all show features associated with the target whereas both \bar{f}_{Miss} and f_{Miss} present features anticorrelated with the target. In the future, more accurate pixel weights could be obtained by applying a foveation algorithm (e.g. Geisler & Perry, 1998; Lee & Bovik, 2003) to the stimuli at each fixation point prior to computing the nonfoveal average images and classification images, to attenuate higher spatial frequencies outside the acuity range of the visual system in a space-variant fashion at each fixation. In this section, however, we will confine ourselves to simple averaging of unfoveated patches in order to illustrate our basic method.



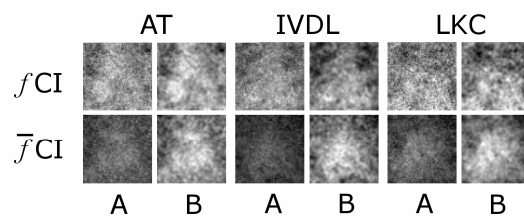
(a)



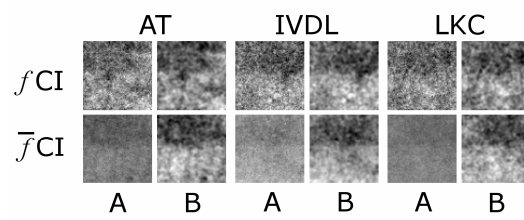
(b)

Figure 3.5: The average images for (a) triangle and (b) dipole target search are shown for 3 observers. Columns labeled A contain the raw average images collectively scaled to a single common grayscale color map, and columns labeled B contain the raw images after low-pass filtering and individual contrast enhancement.

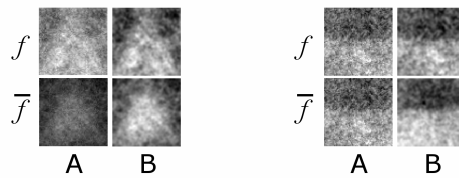
The foveal and nonfoveal classification images, f_{CI} and \bar{f}_{CI} , obtained by linearly combining the average images in the four response categories (defined in our classification taxonomy) in both the foveal and the nonfoveal cases, are shown in Figures 3.6a and 3.6b. Both foveal and nonfoveal classification images were created for each observer and each target. As shown in Figure 3.5, columns labeled A contain the raw average images collectively scaled to a single common grayscale color map, and columns labeled B contain the raw images after low-pass filtering and individual contrast enhancement. These foveal and nonfoveal classification images provide cleaner target-like features.



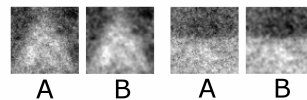
(a)



(b)



(c)



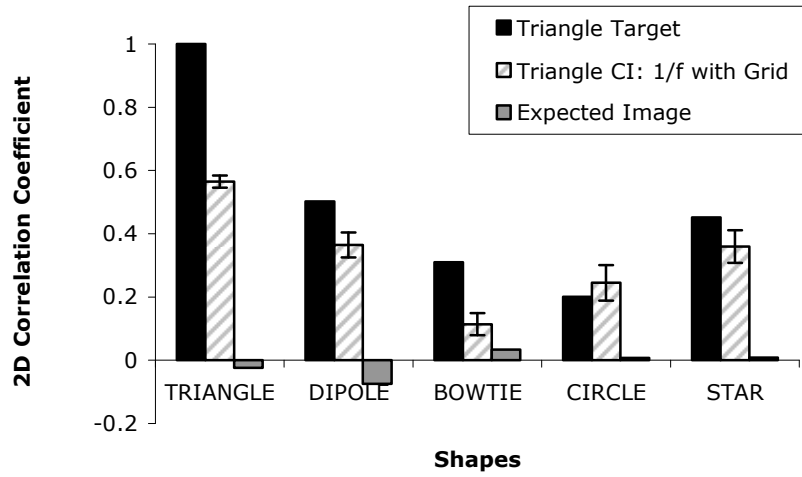
(d)

Figure 3.6: Classification images for (a) triangle and (b) dipole target search are shown for 3 observers. (c) Foveal and nonfoveal classification images combined across observers. (d) Classification images combined across foveal and nonfoveal categories and across observers. Columns labeled A contain the raw images collectively scaled to a single common grayscale color map, and columns labeled B contain the raw images after low-pass filtering and individual contrast enhancement.

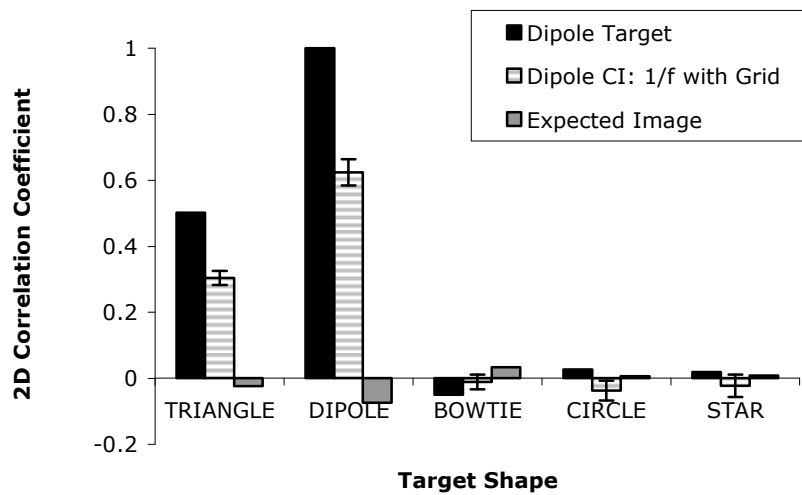
Average images for both target types in the foveal and nonfoveal categories, averaged across all 3 observers, are shown in Figure 3.6c. The combined classification image, obtained by averaging the foveal and nonfoveal classification images across

observers, is shown for each target type in Figure 3.6d. These combined classification images obviously show a strong resemblance to the sought targets.

The level of structural similarity between the classification images (shown in Figures 3.6a and 3.6b) and the search targets was quantified by computing the zero-lag 2-D correlation coefficients between them and the set of shapes in Figure 3.1. The correlation coefficients obtained, averaged across observers and the categories (foveal and nonfoveal), are shown by the hatched bars in Figures 3.7a and 3.7b for both the triangle and the dipole classification images. Also shown are the coefficients obtained by computing the correlation between the search target and each of the shapes (black bars) and the coefficients obtained by computing the correlation between the expected image (shown in Figure 3.4) and each of the shapes (gray bars). The error bars show the standard errors of the coefficients across observers and categories (foveal and nonfoveal). Note that the correlations are highest when computed between a classification image generated from a particular target (the triangle in panel (a) and the dipole in panel (b)) and that target itself. Moreover, the patterns of the experimental correlation coefficients (hatched bars), are virtually identical to those obtained using the targets themselves rather than the classification images (black bars). These results show that our technique produces classification images that rapidly converge to relatively high fidelity representations of the pixel weights used by the observers and, in this case, these weights strongly resemble the actual targets.



(a)



(b)

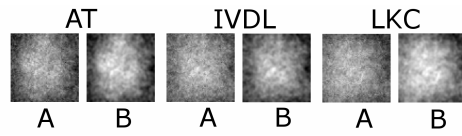
Figure 3.7: Zero-lag 2-D correlation coefficients showing the structural similarity (a) between the classification images for the triangle search and each of the test shapes, and (b) between the average classification image for the dipole search and each of the test shapes. Error bars show the standard errors of the correlations across observers and categories (foveal and non-foveal).

3.2.2 Control Experiments

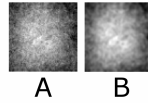
3.2.2.1 Implementation without a Grid

To show the effect of dividing the stimulus into a grid of tiles and using the accompanying taxonomy, we simply repeated the experiment without the grid, as was done in earlier work pioneering the use of eye tracking with classification images (Rajashekar *et al.*, 2002). In this version of the technique, the actual location of each fixation is computed, and the 64×64 pixel patch of the stimulus noise surrounding each fixation is sampled and stored. The resulting set of noise patches is then simply averaged to form the *classification images* for each observer.³ These are shown in Figure 3.8a; columns labeled A contain the classification images after individual contrast enhancement, and columns labeled B contain the raw images after low-pass filtering and individual contrast enhancement. The combined classification image obtained by averaging the classification images across observers is shown for each target type in Figure 3.8b. Although there does appear to be some spatial structure in these images, it seems less specifically triangular than seen in Figure 3.6a. This was confirmed by doing the same correlation analysis as that just described, the results of which are shown in Figure 3.9. The black and hatched bars show the target/shape and raw classification images/shape correlations replotted from Figure 3.7a, and the quilted bars show the correlations obtained using the classification images shown in Figure 3.8. Not only is the pattern of correlations across shapes different, but the actual target used (the triangle) produced a substantially lower correlation with the classification images than did two of the other shapes (the circle and the star).

³ This is not strictly a classification image but can be thought of as the average spatial structure that was fixated by the observer.



(a)



(b)

Figure 3.8: Classification images for triangle target search in $1/f$ noise, without a stimulus grid, are shown for 3 observers (panel (a)), and the combined classification images are also presented (panel (b)). Columns labeled A contain the raw classification images after individual contrast enhancement, and columns labeled B contain the raw images after low-pass filtering and individual contrast enhancement.

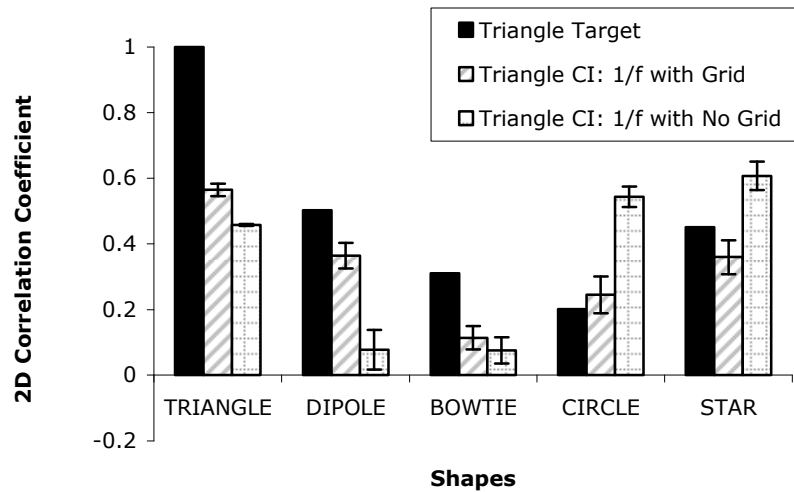


Figure 3.9: Zero-lag 2-D correlation coefficients showing the structural similarity between the classification images for the triangle search and those for each of the test shapes, comparing the main experiment with the no-grid control experiment. Error bars show the standard errors of the correlations across observers.

3.2.2.2 Implementation with White Noise

$1/f$ noise approximates the spectral distribution of natural scenes, making it a valuable tool for probing search behavior within a statistically natural visual environment. Despite this important benefit, the presence of spatial correlation in $1/f$ noise leads to classification images that do not correctly estimate the linear *independent* contribution of each pixel to an observer's behavior, since the noise itself is already spatially correlated. In this control experiment, we show that because information actually determining the observer's behavior exists predominately at low spatial frequencies (presumably), the classification images converge to a similar degree regardless of whether $1/f$ noise is used or whether another noise type (such as white noise) is postprocessed to amplify lower frequencies after the experiment has been completed.

To compare the classification images derived from $1/f$ noise to those derived using white noise, we simply repeated our procedure using uniform white noise,⁴ with 200 trials and the same 3 observers. Figure 3.10 shows the resulting data in the same format as Figure 3.5. Visual comparison of the two figures indicates an apparent lack of spatial structure in the white noise case when processed for viewing as before, with low-pass filtering and contrast enhancement. The foveal and nonfoveal classification images for each observer are shown in Figure 3.11a and averaged across observers in Figure 3.11b, and the combined classification image is shown in Figure 3.11c. Features of the triangle target are present but comparatively faint in the average images. Some spatial structure emerges in the combined classification image, but it is unclear without further processing (see below).

⁴ We used uniform, rather than Gaussian, noise because of higher RMS contrast; at a 68% correct SNR in our task, Gaussian noise would have been substantially clipped at the tails.

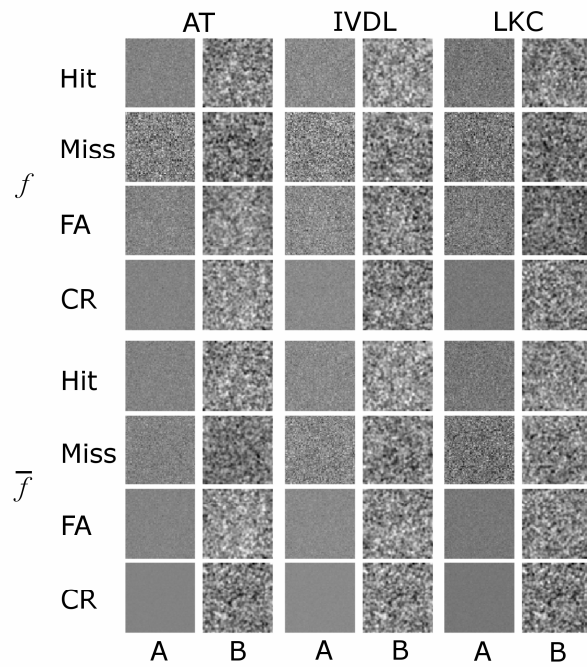
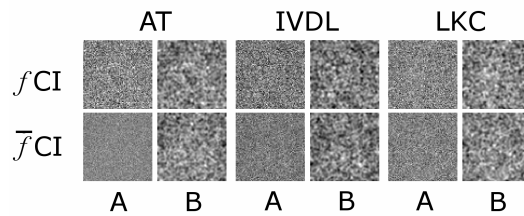
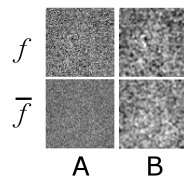


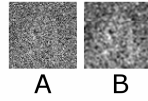
Figure 3.10: The average images for the triangle target search in white noise are shown for 3 observers. Columns labeled A contain the raw average images collectively scaled to a single common grayscale color map, and columns labeled B contain the raw images after low-pass filtering and individual contrast enhancement.



(a)



(b)



(c)

Figure 3.11: Classification images for the triangle target search in white noise are shown (a) for 3 observers, (b) combined across observers for foveal and nonfoveal categories, and (c) combined across foveal and nonfoveal categories and across observers. Columns labeled A contain the raw images collectively scaled to a single common grayscale color map, and columns labeled B contain the raw images after low-pass filtering and individual contrast enhancement.

To effect a fairer comparison, we pinkened our white noise stimuli and recalculated the classification images. We then compared these pinkened classification images with those obtained directly from the $1/f$ noise stimuli. The pinkening procedure was derived from the computation of the unbiased estimate described by Abbey and Eckstein (2002). This procedure involves multiplying each noise image by the covariance matrix of the $1/f$ noise (computed to within an arbitrary scaling factor) given by $B * B^T$, where B represents the $1/f$ blurring filter and T the matrix transposition. The classification images combined across foveal and nonfoveal categories and across observers are shown in Figure 3.12. Column A shows the raw result obtained with the pinkened white noise, and column B shows the low-pass filtered and contrast stretched version. Again, our results indicate that preblurring (using $1/f$ noise stimuli) or postblurring (pinkening uniform noise post hoc) produces closely comparable results, evidenced by the correlation analysis shown in Figure 3.13. The black and hatched bars show the target/shape and raw classification images/shape correlations replotted from Figure 3.7a, and the striped bars show the correlations obtained using the classification images shown in Figure 3.11.

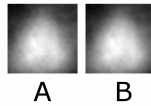


Figure 3.12: Combined classification images for uniform noise (A) after being pinkened and (B) after being pinkened, low-pass filtered, and contrast stretched.

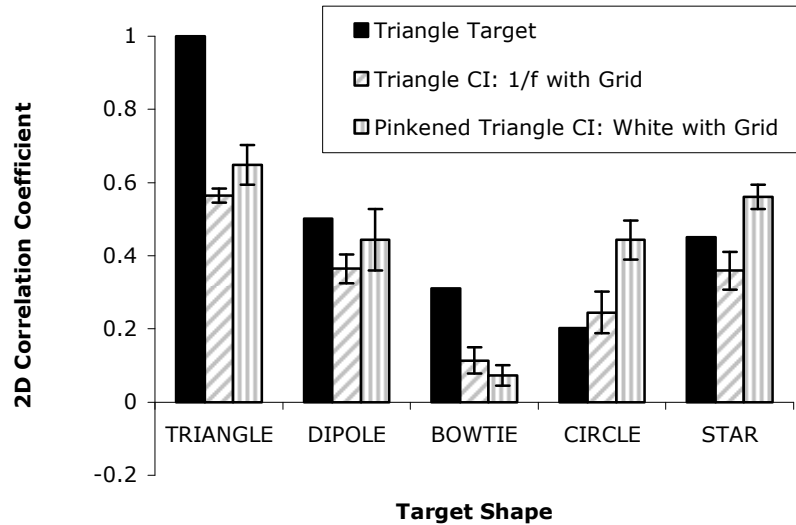


Figure 3.13: Zero-lag 2-D correlation coefficients showing the structural similarity between the classification images for the triangle search and each of the test shapes, comparing the main experiment with the (pinkened) white noise control experiment. Error bars show the standard errors of the correlations across observers.

3.2.3 Performance Measures

In general, classification images are valuable insofar as observers do the same thing in each trial. If an observer switches back and forth between two strategies, say, the pixel weights will reflect the linear combination of the two with no way to disentangle them. Our task is slightly more complicated than traditional psychophysics. We therefore wanted to ensure that observer's performance remained roughly constant across trials and did not depend on the target location (i.e. the initial target eccentricity). Although this is

not a direct measure of strategy, a change in strategy would probably be accompanied by a change in performance.

3.2.3.1 Performance over Location and Time

Figure 3.14 shows the cumulative number of hits as a function of the trial number, obtained for the 3 observers with two sets of 100 trials for each of the two search tasks (triangle and dipole target search) in the basic $1/f$ noise, with grid, experiment. The mean cumulative hit number is represented by the thick black curve, and it reaches the 68% rate sought during the QUEST procedure at the final trial. These performances are compared with that of a perfect observer (dashed curve labeled as perfect) and with that of a random observer (dashed curve labeled as chance).

Each set of 100 trials yielded a slope between roughly 0.5 and 0.8, and for each set, this slope was roughly constant throughout. Again, this is not direct evidence that observers were not changing strategies, but it does indicate a constant level of performance was maintained within a session.

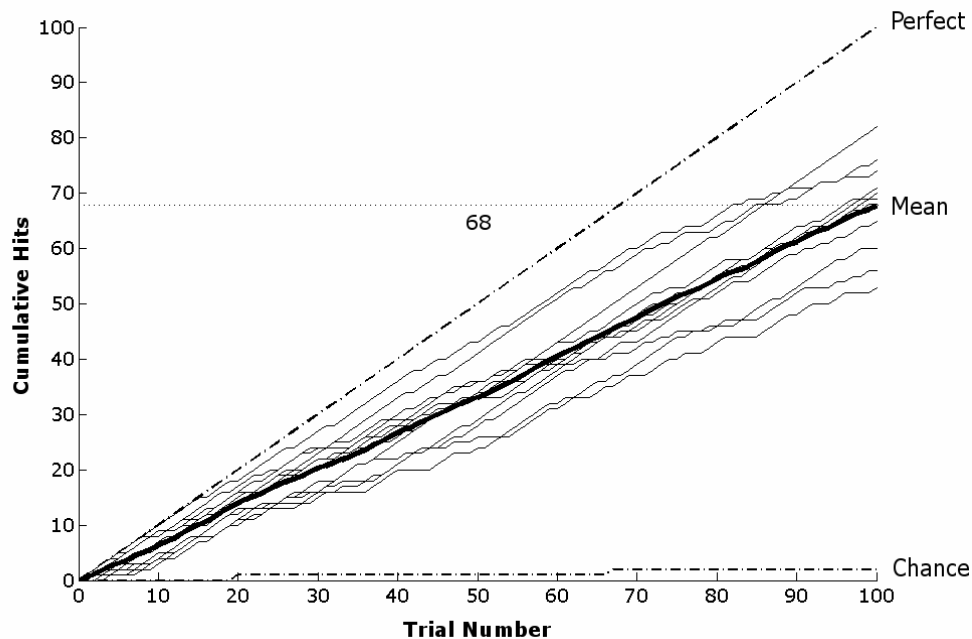
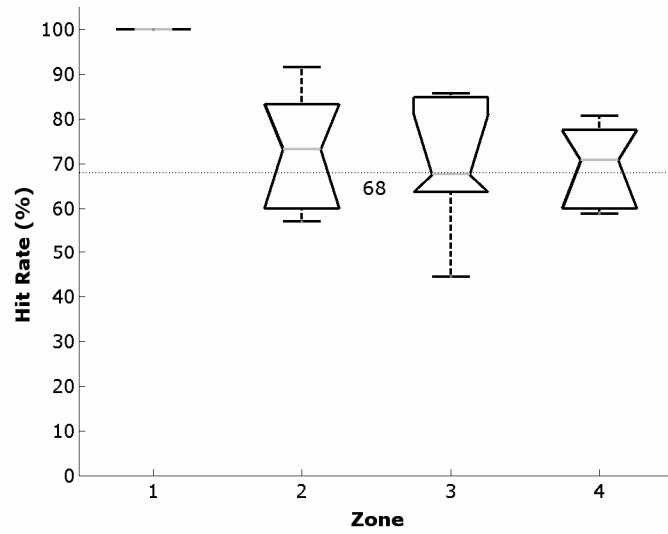
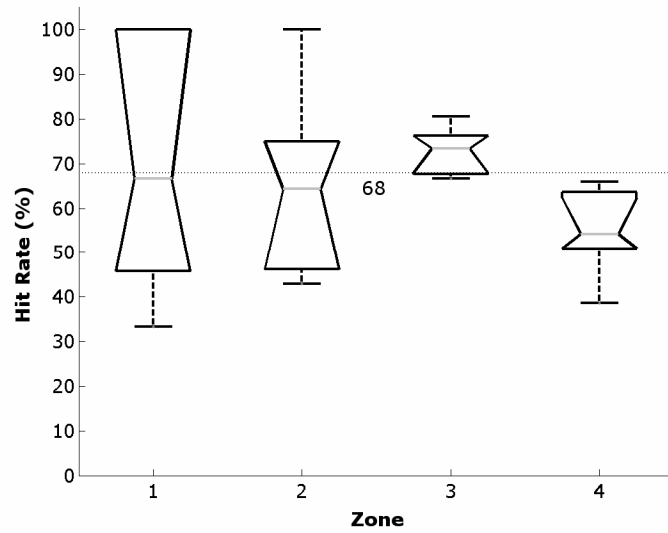


Figure 3.14: Graph of observer performance over time measured as cumulative number of hits.

We also measured the success rates of observers in four different initial eccentricity regions covering the full stimulus, the center tile (Zone 1) and three concentric square annuli surrounding the center tile (Zones 2-4), to see whether the location of the target had any influence on the performance. Because these zones were square, they include tiles centered at eccentricity ranges of 0° , 1.19° – 1.68° , 2.38° – 3.36° , and 3.56° – 5.04° . Figure 3.15 shows the comparison of box plots of success rates in the four different initial eccentricity regions, for sets of 100 trials performed by the observers for both targets: (a) triangle and (b) dipole. The only obvious aberration in the data is that the dipole target was always detected when presented in the central tile, and this is presumably because this target at this location results in the edge's being presented directly to the foveola. The triangle target was also more difficult to detect when presented in the outermost tiles, but not dramatically so.



(a)



(b)

Figure 3.15: Box plots of the success rates across observers for four different eccentricity regions are shown for (a) triangle and (b) dipole search with $1/f$ noise stimuli with superimposed grid.

3.2.3.2 Observer Dwell Times

As discussed in the Proposed Method section, the observers were given 5 sec to find the target, in order to ensure a fairly consistent strategy across observers, allowing several fixations to be made per trial but precluding the possibility of an exhaustive search. Figure 3.16a shows the distribution of the dwell-times from the main experiment for all the fixations, excluding the initial and the final ones of each trial, for all the observers and both target types. It can be seen that the dwell times are concentrated between 200 and 600 msec, in accordance with previous studies (Jacob, 1995). Figure 3.16b shows the distribution of the dwell times for only the final fixations. Over 83% of the dwell times observed for the final fixations are equal to or longer than 600 msec, the upper bound on typical fixation durations, reaching 95% for cases in which the target is actually found. We interpret this observation as indicating that, in our experiment, search was fairly naturalistic and that there was enough time for observers to deliberately select a single tile as containing the target on most trials. For greater rigor in ensuring that the final fixation categories (f_{Hit} , f_{FA}) do not contain search fixations, one could eliminate those with dwell times below a threshold.

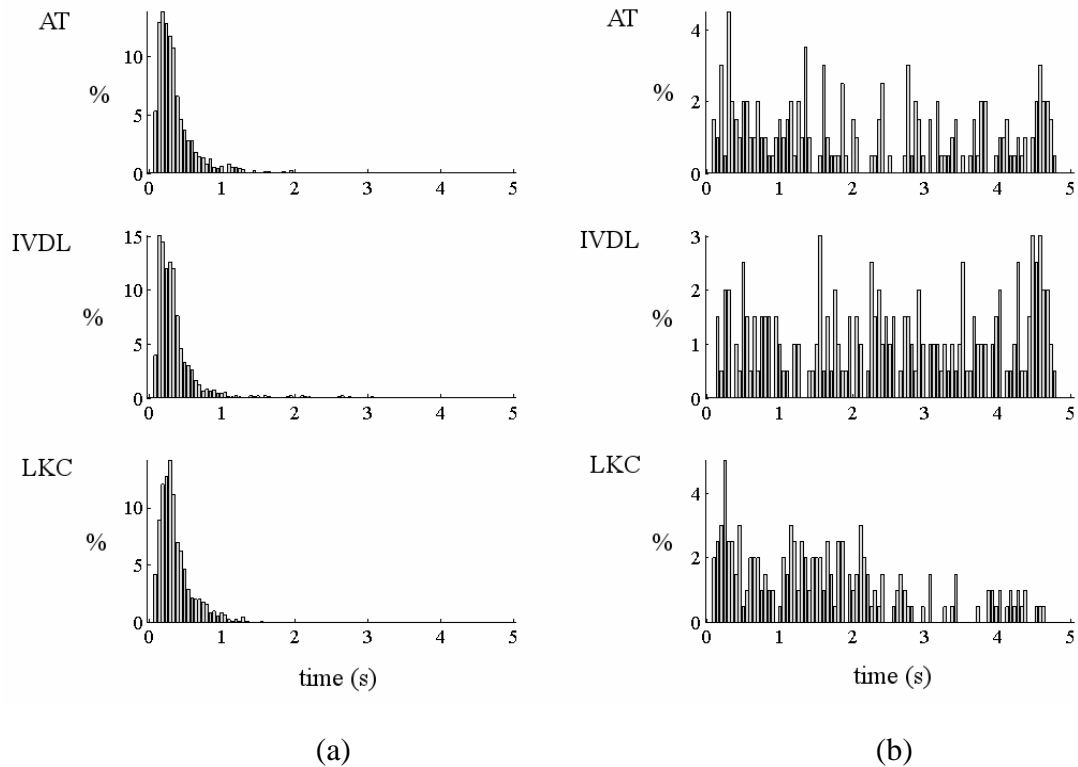


Figure 3.16: Dwell-time distribution for each observer: (a) non-final and (b) final fixations.

3.3 Discussion and Conclusions

In this chapter we have demonstrated a technique for expediting the convergence of classification images in visual search experiments. In fact, for each of the 3 observers and two target types, and with only 200 trials per observer, we see that the classification images obtained with our method closely resemble the target sought (Figures 3.5-3.7). Although the number of tiles falling into many of the categories was small, we still managed to obtain fairly distinctive average images and, hence convincingly robust classification images. Stronger classification images were obtained in comparison to a nongrid control experiment (this claim is supported by both visual inspection of the

results and the strength of the correlation coefficients obtained between the classification images and the targets). The use of naturalistic $1/f$ masking noise was evaluated with a second control experiment in which white noise was used. Visual inspection and correlation coefficients indicate that there is a minimal difference between classification images generated with either noise type if we either pinken white noise tiles and compare to $1/f$ noise tiles, or whiten $1/f$ noise tiles and compare to white noise tiles.

In addition, we have introduced a new taxonomy for the categorization of results from each fixation during a trial. This new taxonomy simply extends the conventional signal detection theory categories to distinguish foveal and nonfoveal processes. However, this extension should allow us and others to characterize the kinds of information used in the fovea and periphery during naturalistic visual search. For instance, Figure 3.5a shows blob-like average images across observers for the nonfoveal category \bar{f}_{FA} , hence characterizing the features that attracted observer fixations to tiles not containing the target. But as outlined in our taxonomy, noise images in the \bar{f}_{FA} category are divided into two foveal categories f_{FA} (corresponding to the observer's final selection of a wrong candidate) and f_{CR} (corresponding to a rejection of a wrong candidate). In fact, f_{FA} presents sharper target-like features compared to f_{CR} and \bar{f}_{FA} . Although preliminary, such results hint at the difference between the foveal and nonfoveal selection process. Moreover, stimuli could be filtered to take into account the eccentricities of the tiles with regard to the fixation points (foveation), prior to averaging, thus eliminating any contribution of spatial frequencies outside the pass band of the visual system at a given eccentricity.

Chapter 4

Spatial Frequency and Orientation Selectivity in Visual Search

In this chapter, we use our novel technique (Tavassoli *et al.*, 2007a), described in the previous chapter, to examine observers' strategies when seeking low-contrast targets of different spatial frequency and orientation characteristics. We present four major findings (Tavassoli *et al.*, in preparation). First, we provide strong support for visual guidance in saccadic targeting, characterized by observers' selectivity for spatial frequency and orientation attributes close to the search target. Second, we show that observers exhibit inaccuracies and biases in their estimates of target features. Third, a complementarity effect is generally observed, indicating the existence of interactions between neighboring spectral components of stimuli. Finally, an unusual phenomenon is observed whereby distracters containing close-to-vertical structures are fixated in searches for non-vertically oriented targets. Our results provide evidence for the involvement of band-pass mechanisms along feature dimensions (orientation and spatial frequency) during visual search.

4.1 Motivation

The existence of neurons along the visual pathway that are selective for the spatial frequency and orientation characteristics of visual stimuli is well established by physiological studies. Selectivity for spatial frequency is present in early stages of the visual pathway and is refined in later ones, i.e. broad tuning at the level of the retina (Enroth-Cugell & Robson, 1966; Kuffler, 1953) and relatively narrower tuning in the

visual cortex (DeValois *et al.*, 1982; Schiller *et al.*, 1976; Campbell *et al.*, 1969). Tuning for orientation is a principal characteristic of cells in the visual cortex; neurons located in earlier stages, including in the lateral geniculate nucleus, have not been found to be orientation tuned (Hubel & Wiesel, 1968, 1962). However, the refinement of orientation tuning is a controversial issue, with various theories proposed, including feedforward (Ferster *et al.*, 1996; Hubel & Wiesel, 1962) and intracortical inhibition mechanisms (Ringach *et al.*, 1997; Ramoa *et al.*, 1986; Morrone *et al.*, 1982).

Evidence for observer selectivity for orientation and spatial frequency has been provided by numerous psychophysical studies, generally using contrast sensitivity (Graham & Nachmias, 1971; Campbell & Robson, 1968), masking (Stromeyer & Julesz, 1972; Campbell & Kulikowski, 1966; Wilson *et al.*, 1983), and spatial adaptation (Snowden, 1992; Tolhurst, 1972; Blakemore *et al.*, 1970; Blakemore & Campbell, 1969) paradigms. For example, Campbell and Robson (1968) conducted detection and discrimination tasks using gratings (e.g. sine-, square-waves, and so on) and showed that observers' contrast thresholds were directly related to the harmonic Fourier components of the gratings. They postulated the existence of independent band-pass mechanisms selective for spatial frequencies. Using stimuli consisting of rapid sequential presentations of sinusoidal gratings at random orientations and spatial phase, Ringach (1998) showed that observers' tuning for orientation generally presented a "Mexican hat" distribution peaking at orientations close to the orientation observers had to report, with valleys at either side of the peak.

An interesting question is how selectivity for spatial frequency and orientation is used in visual search tasks. A number of studies have demonstrated, through measurements such as reaction times as a function of set size, that both feature dimensions can indeed be used in guiding attention in visual search (Wolfe & Horowitz,

2004; Sagi, 1988; Treisman & Gelade, 1980). Search efficiency in many tasks was found to depend on target-distracter discriminability and distracter homogeneity along these feature dimensions (Foster & Ward, 1994; Verghese & Nakayama, 1993; Wolfe *et al.*, 1992). In addition, search asymmetries were observed (Wolfe *et al.*, 1992; Foster & Ward, 1991; Treisman & Gormican, 1988); for instance, the detection of a tilted line amongst vertical lines has been shown to be easier than an otherwise identical search for a vertical line amongst tilted lines.

However, many of these previous visual search studies, despite intending to elucidate visual search, have avoided the analysis of eye movements, either as a result of using short stimulus display times or by instructing observers to keep their eyes still. The importance of incorporating eye movements to study observer strategies in visual search has been emphasized in the recent years (Geisler *et al.*, 2006; Findlay & Gilchrist, 2003; Zelinsky & Sheinberg, 1997). Observers naturally move their gaze when searching for a target, and it has even been demonstrated in a few tasks that observers opt to perform eye movements even when such a strategy is not optimal (Findlay, 1997; Findlay & Gilchrist, 1998).

4.2 Methods

We use our new and efficient experimental search framework to study the behavior of humans seeking gratings of known characteristics embedded in $1/f$ noise. Note that the use of noise tiles as distracters, instead of lines, gratings, and so on as used in many of the previous visual search studies, permits a much larger set of distracting items that possess differences across many feature dimensions.

4.2.1 Observers

Three male observers (aged 26–30) were tested in our experiments, of whom two were experienced (AT and IVDL) and one was naïve to the purpose of the study (AJS), each with normal/corrected-to-normal vision. Each observer completed 10,500 trials (fifteen sets of 700 trials, each set with a different search target) over a period of several months.

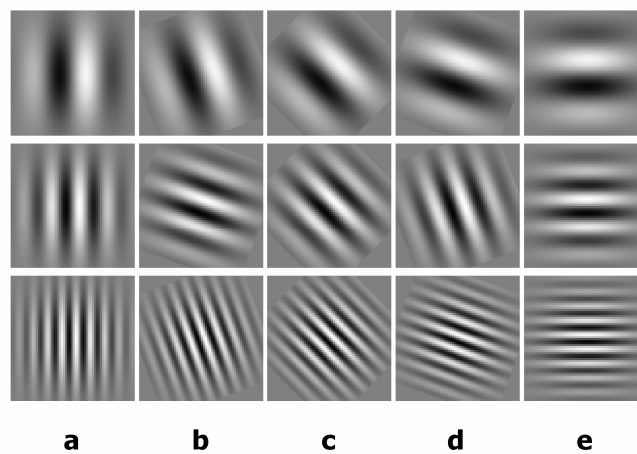


Figure 4.1: Targets used in our fifteen separate experiments. Gabor patches of spatial frequency 2 (first row), 4 (second row), and 8 c/deg (third row), oriented anticlockwise from the vertical at (a) 0 deg, (b) 20 deg, (c) 45 deg, (d) 70 deg, and (e) 90 deg were used as targets.

4.2.2 Visual Stimuli

Our fifteen search targets were 64×64 pixel Gabor patches of spatial frequency 2, 4, and 8 c/deg, oriented anticlockwise from the vertical at 0, 20, 45, 70 and 90 deg (Figure 4.1). We use the same one hundred 7×7 tile mosaics that were generated to test our methodology described in the previous chapter, i.e. obtained offline by creating one hundred 544×544 pixel $1/f$ noise images (with an amplitude spectrum of the form $1/f^a$ with $a = 0.8$) and then superimposing gray borders 12 pixels in width (Figure 4.2b).

On each trial, the Gabor target was added to a randomly selected tile of the $1/f$ noise grid (Figures 4.2a and 4.2b). Observers viewed the stimuli on an Image Systems 21" grayscale monitor (Image Systems Corp., Minnetonka, MN) driven by a Matrox Parahelia graphics card (Matrox Graphics Inc., Dorval, Québec, Canada) at a screen resolution of $1,024 \times 768$ pixels, a grayscale resolution of 8 bits per pixel, and a refresh rate of 60 Hz. The screen was placed 134 cm from the observer and subtended a visual angle of 16×12 deg, giving approximately 1 min of arc per screen pixel. The luminance output was linearized by putting the inverse of the monitor's measured gamma function in the display look-up table. The ambient illumination in the laboratory was kept constant for all observers, and there was a minimum of 5 min to adapt to the ambient illumination and screen luminance while the eye tracker was calibrated.

4.2.3 Procedure

The procedure was the same as described in the Chapter 3. We would like to point out that a measure of dwell-times on the observers final fixations on the tile they believed to contain the target indicated that they were deliberately selecting a single tile as containing the target on most trials (i.e. over 80% of the dwell-times observed for final fixations were equal to or longer than 600 ms, an upper bound on typical fixation durations).

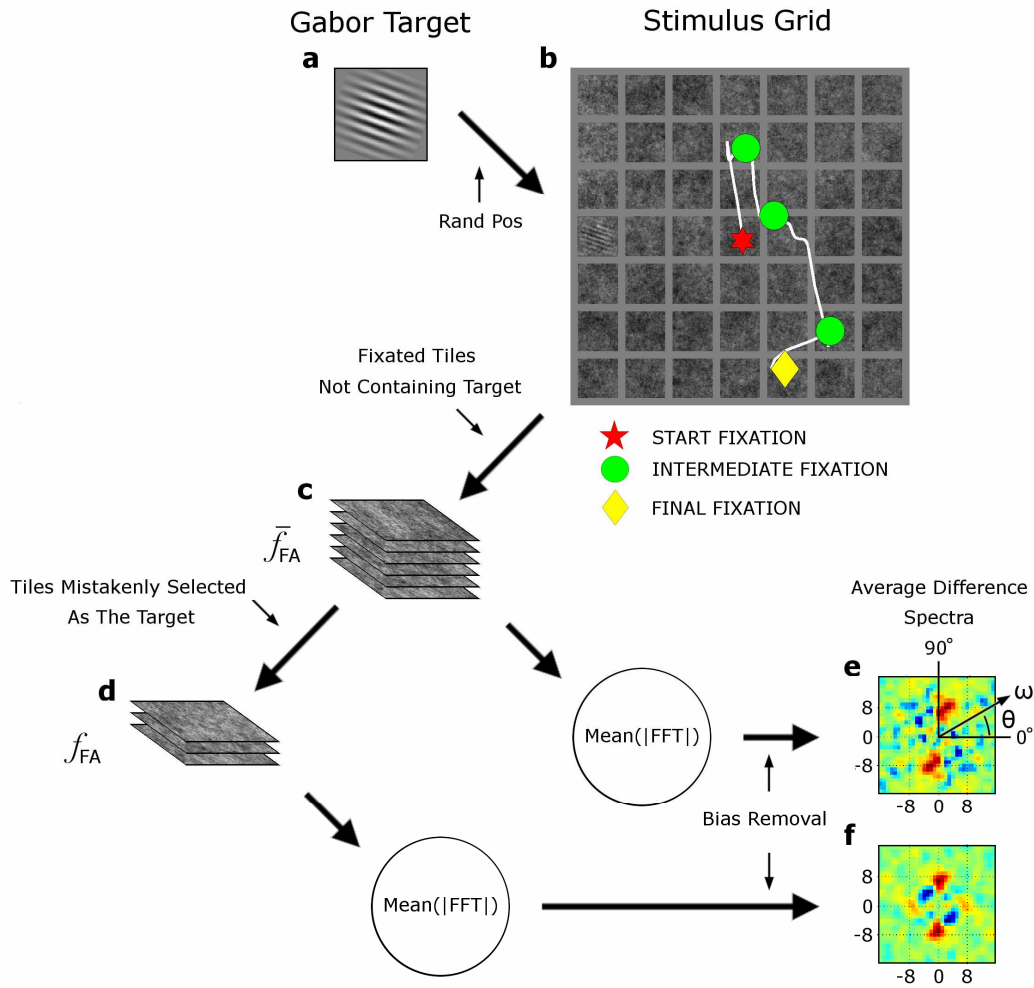


Figure 4.2: Stimulus creation, data capture, and data analysis. (a) A Gabor patch used as a target (b) was added to a randomly selected tile of the $1/f$ noise grid. Observer eye movements were recorded while they searched for the target. A representative scan path is shown for a trial in which the observer did not find the target, located in the center of the leftmost column. (c) Fixated tiles that did not contain the target constitute our nonfoveal false alarm category, and (d) a subset of these tiles, which were mistakenly selected at the end of the trials as the target by the observer, constitute our foveal false alarm category. (e and f) Average difference spectra were computed by averaging the amplitude spectra of noise tiles in each category and subtracting the spectral bias (see text).

4.2.4 Analysis Method

Observers typically performed close to five fixations on average per trial in our experiments (note that we disregarded saccades landing between tiles), hence visiting tiles containing only noise and in some trials selecting one such tile as the target-candidate; an example stimulus grid with representative eye movements for a single observer is shown in Figure 4.2b. We were therefore interested in examining why some noise-only tiles were fixated whereas others were not? And second, why, at the end of some trials, was a noise-only tile mistakenly selected as the tile containing the target?

To answer these questions, we assume that each fixation (excluding the initial fixation at stimulus onset) involves two decisions: the decision to fixate a certain tile (and not the others), and the subsequent decision to either remain on that tile or continue searching. We consider that the former is based primarily on nonfoveal information and the latter is based primarily on foveal information. We therefore stored noise-only tiles that were fixated while en route to the target and labeled them as “nonfoveal false alarms” (\bar{f}_{FA}) (Figure 4.2c). Additionally, noise-only tiles that were mistakenly selected as the target at the end of a trial were labeled as “foveal false alarms” (f_{FA}) - these necessarily being a subset of the nonfoveal false alarms (Figure 4.2d). These signal-absent categories better reflect observer behavior than signal present categories (those composed of tiles that contained the target), since only patterns in the noise, corresponding to visual information that the observer took to imply the presence of a target, are used (Eckstein *et al.*, 2002).

We therefore computed the Fourier transform of each tile and averaged their amplitude spectra within category and observer. Because we used a finite number of $1/f$ noise tiles ($100 \times 7 \times 7 = 4900$) for the experiment, a spectral bias is introduced in these averages, i.e. the expected amplitude spectrum that would be obtained by randomly

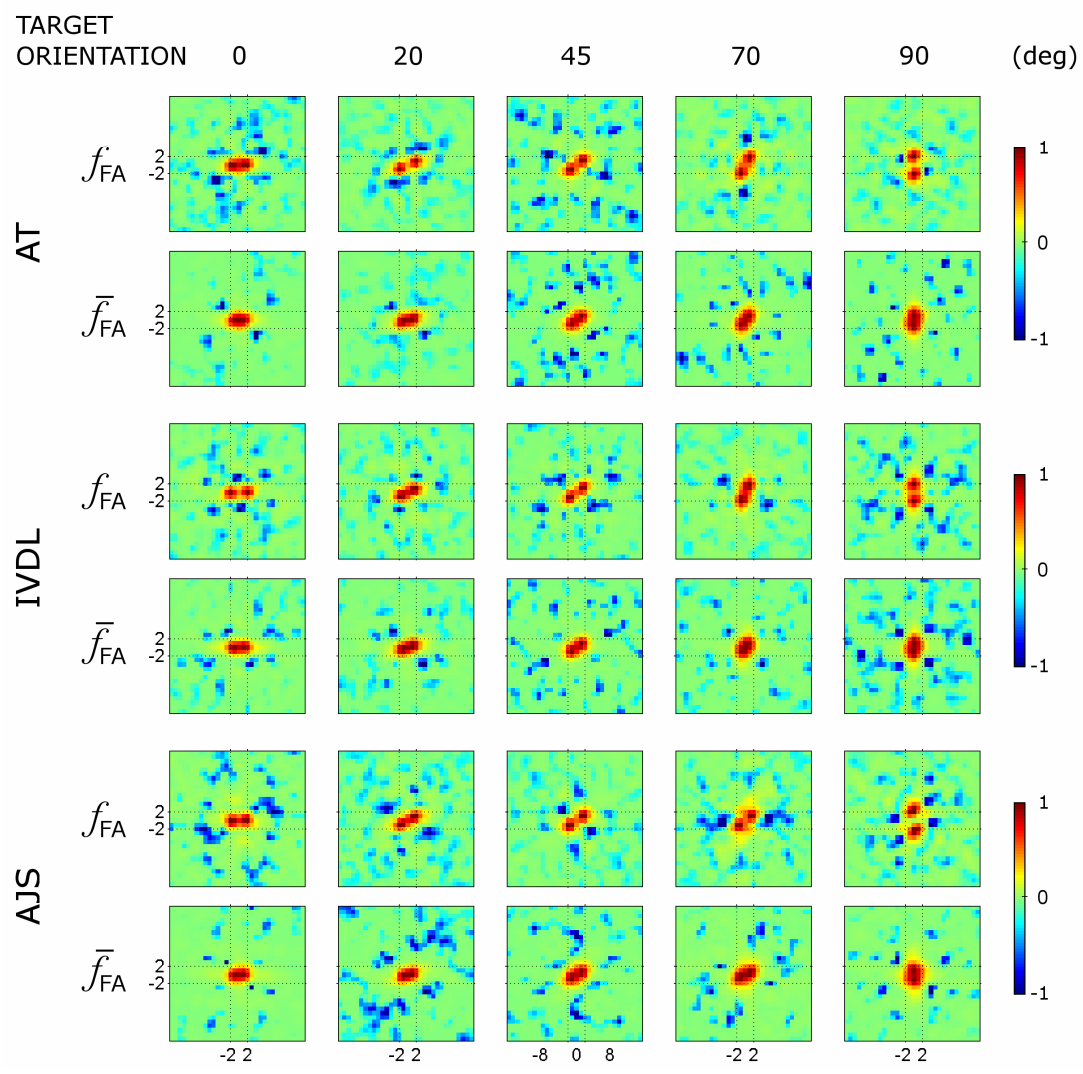
sampling noise tiles would have a shape close to $1/f$. We therefore examined differences between the averages in our categories and the expected bias. We obtained the bias by averaging the amplitude spectra of all the 4,900 noise tiles used to generate our stimuli. We then subtracted the bias from the averages obtained in each category to form what we will refer to as average difference spectra (Figures 4.2e and 4.2f); this process is similar to the amplitude spectrum correction method described by Willmore and Smyth (2003). These average difference spectra represent dominant (relative to the bias) spatial frequency (indicated by the distance from the origin, ω , see Figure 4.2e) and orientation (indicated by the angle, θ , from vertical orientation, 0 deg, see Figure 4.2e of the noise tiles within each category. To improve visualization, we zeroed the DC (in all the average difference spectra) and 1 c/deg (in results obtained for searches for the 4 and 8 c/deg Gabors) components, then smoothed each image with a 3×3 pixel Gaussian mask with $\sigma = 0.9$ pixel. Setting the very low frequencies (DC and 1 c/deg) to zero simply allows the full color map to be used for the more interesting spectral structures in surrounding frequency components.

Note that directly averaging the noise tiles (i.e. in the spatial domain, as described in the previous chapter) within each category and observer produced an effect similar to those reported in earlier psychophysical detection studies (Beard & Ahumada, 1999; Solomon, 2002) whereby the average images contain some target-like structures in the case of search for lower frequency targets but no relevant structure for search for higher frequency targets.

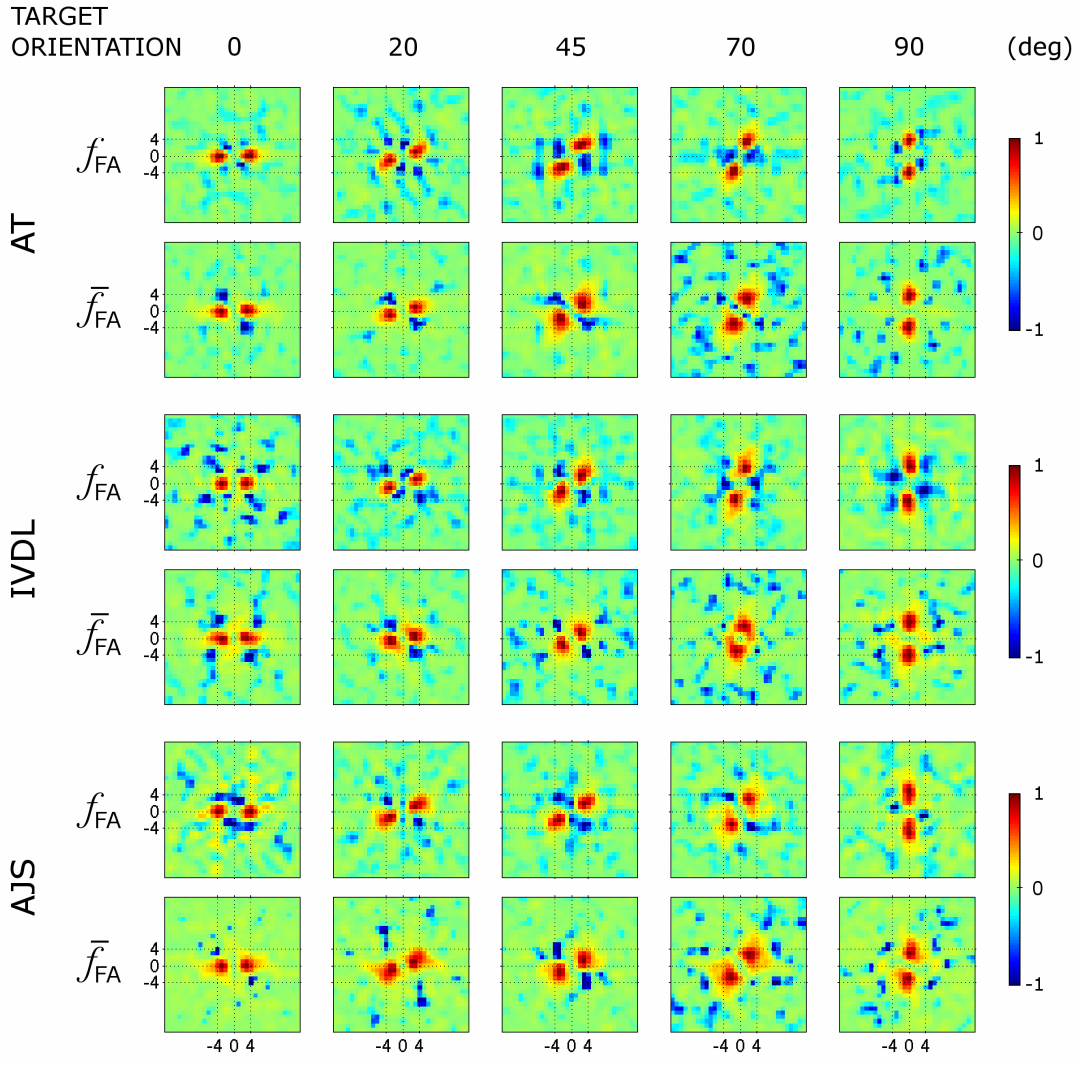
4.3 Results

Figures 4.3a, 4.3b, and 4.3c show the average difference spectra for the two false alarm categories obtained for search experiments using Gabor targets of spatial frequency

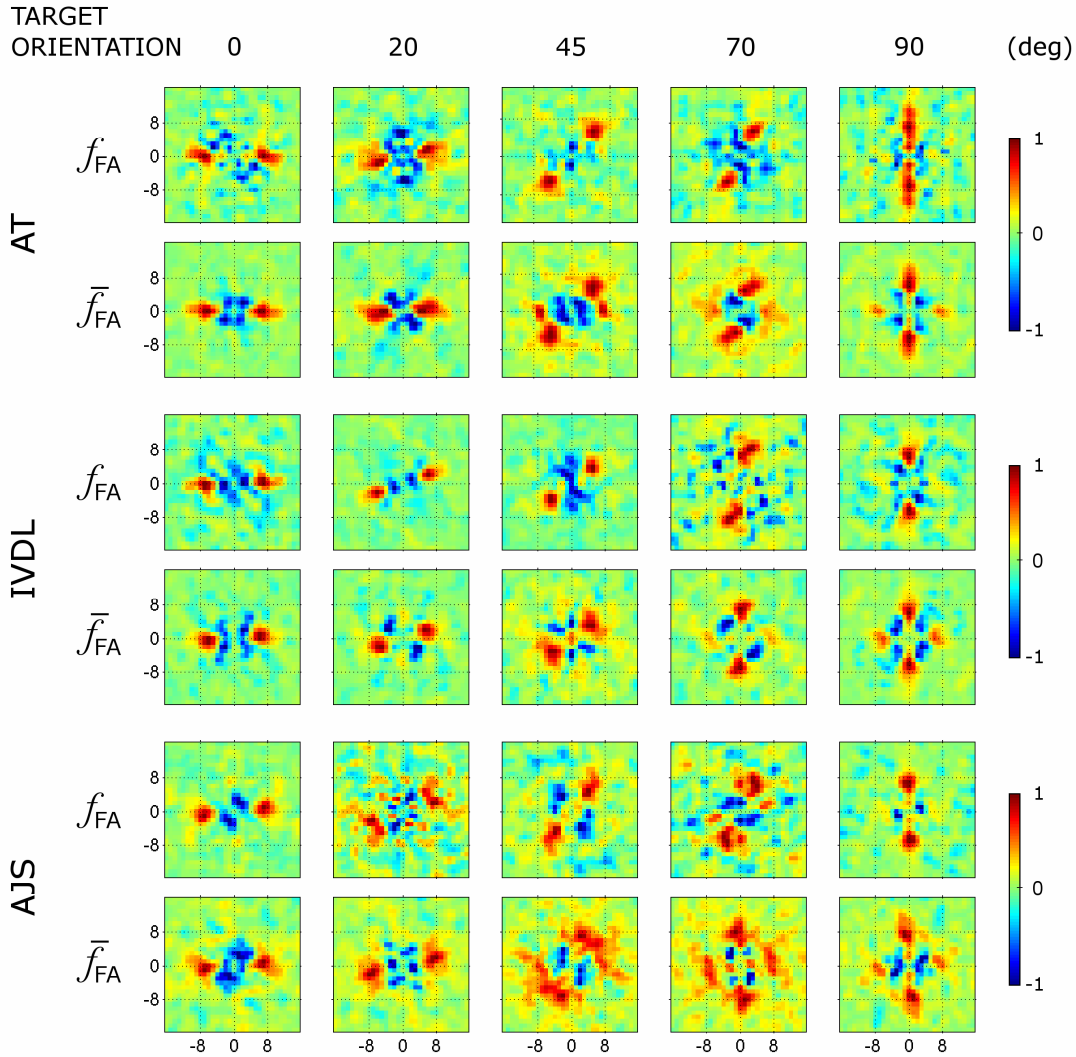
2, 4, and 8 c/deg, oriented anticlockwise from the vertical at 0 (first column), 20 (second column), 45 (third column), 70 (fourth column), and 90 deg (fifth column). For each observer and set of 700 trials, amplitude spectra were created using about 210 and 2,800 noise tiles respectively for the foveal and nonfoveal categories. Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively (i.e. above and below the expected amplitude spectrum for a random observer). Regions in green show frequency components close to the bias.



(a)



(b)



(c)

Figure 4.3: Average difference spectra for three observers in the (a) 2, (b) 4, and (c) 8 c/deg Gabor search experiments. Average difference spectra, smoothed and contrast-stretched for visual enhancement, are shown for sets of 700 trials for visual searches for Gabor targets oriented anticlockwise from the vertical at 0 (first column), 20 (second column), 45 (third column), 70 (fourth column), and 90 deg (fifth column). For each observer and each set of trials, the spectra were created using about 210 and 2,800 noise tiles respectively for the foveal and nonfoveal categories. Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively. Regions in green show frequency components close to the bias. We have indicated the spatial frequency of the search target (for the horizontal and vertical orientations).

Clear spectral structures are obtained in all of the average difference spectra, i.e. peaks close to the spatial frequency and orientation of the search target, flanked by well localized valleys present in the surround, more visibly for the 4 and 8 c/deg target searches. To better illustrate the placement of the valleys relative to the peaks, we first rotated each average difference spectrum by the negative of the estimated orientation of its peaks. The orientation estimate was obtained by fitting each average difference spectrum with the amplitude spectrum of a Gabor whose parameters (spatial frequency, bandwidth, orientation, and aspect ratio) were varied using a simplex search method. Then, for each target spatial frequency condition, we averaged the aligned average difference spectra for each observer (Figure 4.4) and for all observers combined (Figure 4.5).

Figure 4.3c shows a surprising outcome for visual searches for the 8 c/deg Gabors. All three observers have strong peaks close to 0 deg in the nonfoveal category for searches for the non-vertical Gabor targets, in addition to peaks close to that of the search targets. The additional peaks appear to vanish in the foveal category, i.e. once observers fixated the noise tiles. This effect is best illustrated in the last column of Figure 4.3c, where tightly tuned peaks are present at 0 deg in the nonfoveal average difference spectra along with peaks at 90 deg, and then the additional peaks fade away in the foveal average difference spectra.

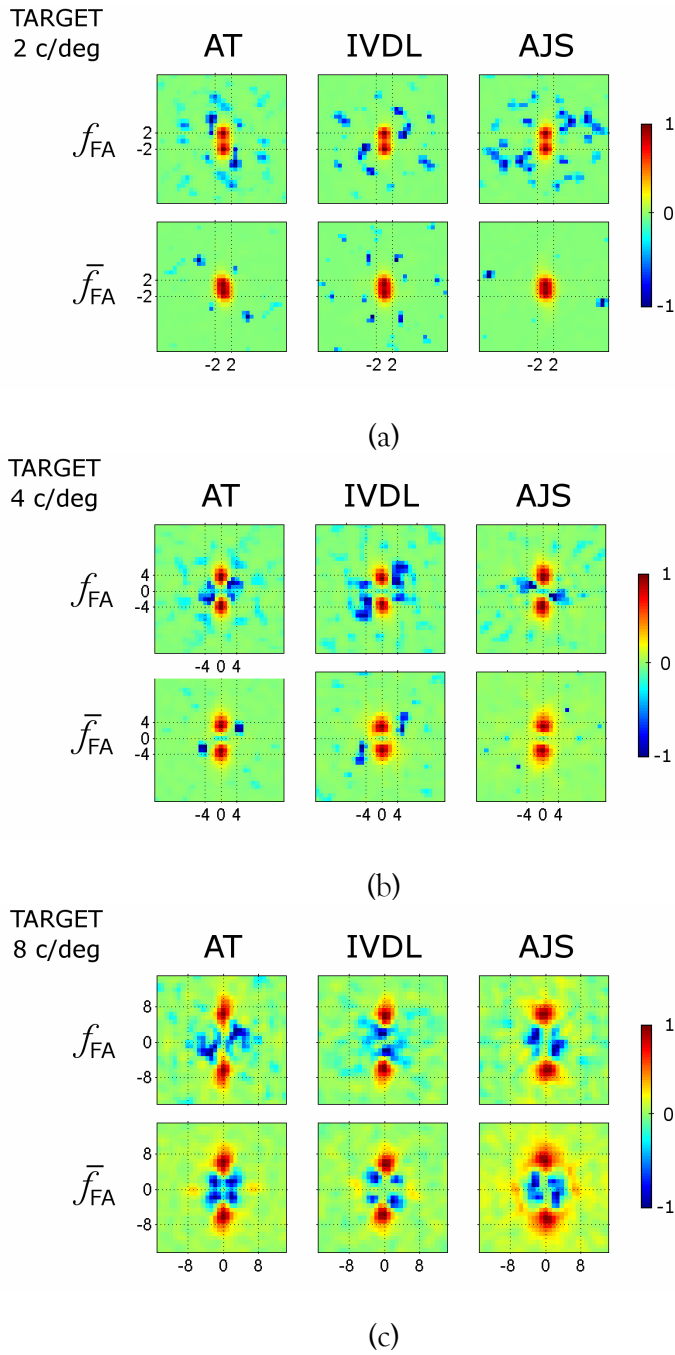


Figure 4.4: Complementarity effect for three observers. The average of aligned average difference spectra are shown for (a) 2 c/deg, (b) 4 c/deg, and (c) 8 c/deg Gabor search experiments for three observers. Regions in red indicate frequency components having amplitudes above the bias and regions in blue specify those components having amplitudes below the bias.

Moreover, we were curious to examine the effects of eccentricity and saccade order on saccadic targeting. We further binned the noise tiles in the nonfoveal category by eccentricity and by order, separately. We found that the structures present in the average difference spectra in each bin, using either binning approaches, were generally similar to ones in the average difference spectra obtained without binning, hence demonstrating the robustness of our results. However, when binning by saccade length, we noticed in the average difference spectra some instances in which peaks for larger eccentricity bins were lower in spatial frequency than for smaller eccentricities. Although the latter effect may be expected due to the falloff of resolution in peripheral vision, however the effect was not reliable perhaps due to the limited number of noise tiles in each bin, especially for larger saccade lengths or saccade numbers.

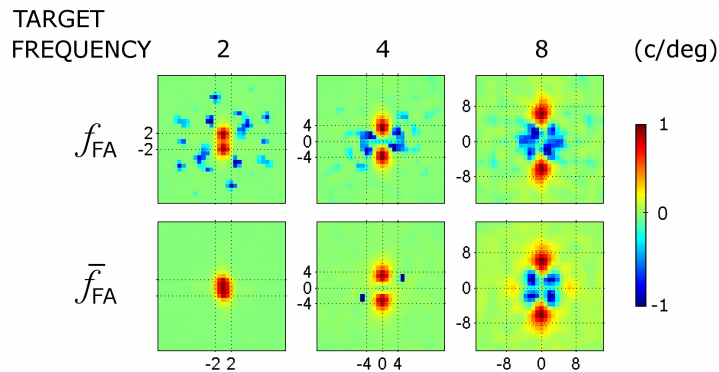


Figure 4.5: Complementarity effect for all three observers combined. The average of aligned average difference spectra are shown for (a) 2 c/deg, (b) 4 c/deg, and (c) 8 c/deg Gabor search experiments for all observers combined. Regions in red indicate frequency components having amplitudes above the bias and regions in blue specify those components having amplitudes below the bias.

4.4 Discussion

The aim of the current study was primarily to explore saccadic targeting and target selection in naturalistic visual search tasks, more precisely when human observers

search for a Gabor target with known characteristics embedded in a grid of $1/f$ noise (which has a similar falloff in amplitude spectrum as natural images). We are interested in understanding what attracts fixations and how target candidates are selected upon fixation.

Our results clearly indicate visual guidance in saccadic targeting, which has been a somewhat contentious issue in previous studies of active visual search (Findlay & Gilchrist, 2003; Hooge & Erkelens, 1999; Motter & Belky, 1998b; Findlay, 1997; Zelinsky, 1996). A *similarity* effect (as defined by Findlay & Gilchrist, 2003) is revealed, showing that saccades are guided, on average, to distracters (here, noise tiles) presenting featural similarities to the Gabor target. In particular, we demonstrate that observers are selective for spatial frequencies and orientations close to the central frequency and orientation of the search target, i.e. the average difference spectra for fixated noise tiles show peaks localized in spatial frequency and orientation close to that of the target (Figures 4.3a, 4.3b, and 4.3c).

Furthermore, observers exhibit inaccuracies in their estimates of target attributes. These errors are revealed by the uncertainties and offsets in the average difference spectra, i.e. elongations in spatial frequency and orientation bandwidths; radial spread corresponding to less selectivity in spatial frequency and rotational smearing to less tuning in orientation. In many cases, offsets in estimates of target features occur nonfoveally but are corrected upon fixation, e.g. observers AJS and IVDL are attracted to noise tiles containing predominantly near-horizontal (close to 90 deg) structures when looking for a Gabor target of spatial frequency 8 c/deg oriented at 70 deg, then foveally select those with prevalent structures close to 70 deg (Figure 4.3c). Inaccuracies in observer estimates have been reported in psychophysical tasks (Ringach, 1998) and appear in psychophysical reverse-correlation data for detection tasks (Solomon, 2002).

Interestingly, a marked *complementarity* effect is found in much of our human data; that is, the absence of various spatial frequencies and orientations appears to influence whether a noise tile is fixated and selected as target candidate upon fixation. In fact, the average difference spectra for fixated noise tiles contain valleys localized at spatial frequencies and orientations neighboring the peaks, more consistently for the 4 and 8 c/deg Gabor target experiments. For example, in Figure 4.3b, most of the average difference spectra in the nonfoveal category for observer AT show distinctive valleys at oblique orientations. The existence of valleys in the results may signify that observers are often disregarding noise tiles containing frequency components surrounding their estimates of the target's orientation and spatial frequency. Alternatively, it could reflect that a reduced presence of particular frequency components may have an enhancing effect in the detection of the components of interest. In general, it appears that the valleys in the average difference spectra tend to be at lower frequencies than the peaks for the higher frequency (8 c/deg) Gabor search experiments and that this tendency is reversed for lower frequency search experiments (Figure 4.5). This observation is consistent with findings in masking experiments where it was found that the most effective masks for low frequency test gratings were at higher frequencies and vice versa (Wilson *et al.*, 1982); and, somewhat comparable to the “Mexican hat” orientation profiles found by Ringach (1998) in psychophysical experiments, although he reported that the effect disappeared for higher frequencies.

Curiously, there is, on average, an unusual presence of close-to-vertical structures in fixated noise tiles for visual searches for the 8 c/deg Gabors for the non-vertical orientation conditions, i.e. the average difference spectra for the nonfoveal category present peaks close to the spatial frequency of the search target but at an orientation of 0 deg. This effect could reflect possible facilitations in detection; for instance, Sillito *et al.*

(1995) showed that responses of many neurons in V1 to their preferred orientation could be enhanced by introducing a surrounding field containing a pattern at significantly different orientation than the center. Alternatively, it could be the consequence of double-orientation tuning in nonfoveal detection; for example, Shevelev *et al.* (1994) demonstrated the existence of neurons in V1 that have a main preferred orientation and an additional preferred orientation. Nevertheless, another possibility may be a windowing or end stopping effect, i.e. the size and width of structures present in noise tiles may be influencing peripheral decision in making saccades. The additional frequency components vanish once observers fixated the noise tiles.

4.5 Conclusions

Our results provide compelling evidence for band-pass mechanisms in saccadic targeting and target selection during visual search, in particular for grating-like targets. Furthermore, selectivity along feature dimensions (here, spatial frequency and orientation) shows inaccuracies, offsets, and curious biases. These errors are to some extent corrected during the foveal decision process. Furthermore, it appears that the presence or absence of various spectral components, other than those close to that of the search target, influence the guidance of saccades. We find that the absence of certain surround frequency components or the presence of near-vertical structures (i.e. components close to 0 deg) in the noise tiles attracts observer fixations.

Chapter 5

Visual Search under Uncertainty Conditions

In this chapter, we use our experimental search framework, presented in Chapter 3, to measure observers' ability to locate a low-contrast target of unknown orientation. We were curious to examine observer strategies under (orientation) uncertainty. We present three main discoveries (Tavassoli *et al.*, 2007b). First, we provide strong evidence for saccadic selectivity for spatial frequencies close to the target's central frequency. Second, we demonstrate that observers have distinct, idiosyncratic biases to certain orientations in saccadic programming, although there were no priors imposed on the target's orientation. These orientation biases cover a subset of the near-cardinal (horizontal/vertical) and near-oblique orientations, with orientations near vertical being the most common across observers. Further, these idiosyncratic biases were stable across time. Third, within observers, very similar biases exist for foveal target detection accuracy. These results suggest that saccadic targeting is tuned for known stimulus dimensions (here, spatial frequency) and also has some preference or default tuning for uncertain stimulus dimensions (here, orientation).

5.1 Motivation

Studies of visual acuity and contrast sensitivity in humans have shown an unequal sensitivity across orientation; generally, these studies find a greater sensitivity to gratings with cardinal (horizontal/vertical) relative to oblique orientations (Berkeley *et al.*, 1975; Campbell *et al.*, 1966). This orientation anisotropy is referred to as the “oblique effect” or “cardinal bias” in the literature (Appelle, 1972). This effect is also found in cats and

macaque monkeys, but not as consistently as for humans (Li *et al.*, 2003). Some have argued that using broadband stimuli, such as $1/f$ noise or natural images filtered in orientation, instead of gratings, could instead yield greater sensitivity for oblique orientations (Hansen & Essock, 2004).

A number of single-cell physiological studies (see Li *et al.*, 2003) of the primary visual cortex (V1) have found variability in populations of orientation tuned cells (*viz.*, a larger number of cells tuned to horizontal and vertical than to oblique orientations), but several other studies were unsuccessful in finding such differences (Finlay *et al.*, 1976; Mansfield, 1974). It has been suggested that causes of these discrepancies are that different studies have used different measurement procedures, and that some have sampled only a small population of cells. More recently, researchers have observed an oblique effect using optical imaging (Coppola *et al.*, 1998) and functional magnetic resonance imaging (Furmanski & Engel, 2000). Overall, the belief in a generic deficit for oblique stimuli remains a source of some contention, and its incidence is likely to vary with the stage in the visual pathway measured and the experimental technique employed.

A question of great interest is how visual search might be affected by anisotropies in the perception of orientation. Studies have reported search asymmetries in tasks where human observers seek an oriented target amongst a set of distracters (Wolfe, 1998; Carrasco *et al.*, 1998; Wolfe *et al.*, 1992; Foster & Ward, 1991; Treisman & Gormican, 1988). For instance, the detection of a tilted line amongst vertical lines has been found to be easier than search for a vertical line amongst tilted lines. In this paper, we address a more general problem in visual search where the orientation of the target is not known to the observer *a priori*. Such an experimental procedure is similar to many real-world search tasks, in which the orientation of an object is largely uncertain, though it may be influenced by gravity or its proximal interaction with other objects and planes.

5.2 Methods

We use our experimental search framework to study the behavior of humans seeking a randomly oriented grating embedded in $1/f$ noise.

5.2.1 Observers

Four male observers (aged 26 through 30), of whom two were experienced (AT and IVDL) and two naïve (AEP and AJS), were tested in our experiments, each with normal/corrected-to-normal vision. Each observer completed 1,400 trials (2 sets of 700 trials, separated by a period of about 1 month).

5.2.2 Visual Stimuli

Our search target was a 64×64 pixel Gabor patch of frequency 8 c/deg and bandwidth 0.25 octaves (Figure 5.1a). We use the same one hundred 7×7 tile mosaics that were generated to test our methodology described in the previous chapter, i.e. obtained offline by creating one hundred 544×544 pixel $1/f$ noise images (with an amplitude spectrum of the form $1/f^a$ with $a = 0.8$) and then superimposing gray borders 12 pixels in width (Figure 5.1c). On each trial, the orientation of the Gabor was randomly selected from the set $\{0, 1, 2 \dots 179\}$ deg (Figures 5.1a and 5.1b) and this Gabor was then added to a randomly selected tile of the $1/f$ noise grid (Figure 5.1c). As a convention, angles ascended from 0 deg (vertical bars) in an anticlockwise direction. Observers viewed the stimuli on an Image Systems 21" grayscale monitor (Image Systems Corp., Minnetonka, MN) driven by a Matrox Parahelia graphics card (Matrox Graphics Inc., Dorval, Québec, Canada) at a screen resolution of $1,024 \times 768$ pixels, a grayscale

resolution of 8 bits per pixel, and a refresh rate of 60 Hz. The screen was placed 134 cm from the observer and subtended a visual angle of 16×12 deg, giving approximately 1 min of arc per screen pixel. The luminance output was linearized by putting the inverse of the monitor's measured gamma function in the display look-up table. The ambient illumination in the laboratory was kept constant for all observers, and there was a minimum of 5 min to adapt to the ambient illumination and screen luminance while the eye tracker was calibrated.

5.2.3 Procedure

The procedure was the same as described in the Chapter 3. We would like to indicate that over 81% of the dwell-times observed for final fixations were equal to or longer than 600 ms, an upper bound on typical fixation durations, indicating that observers were deliberately selecting a single tile as containing the target on most trials.

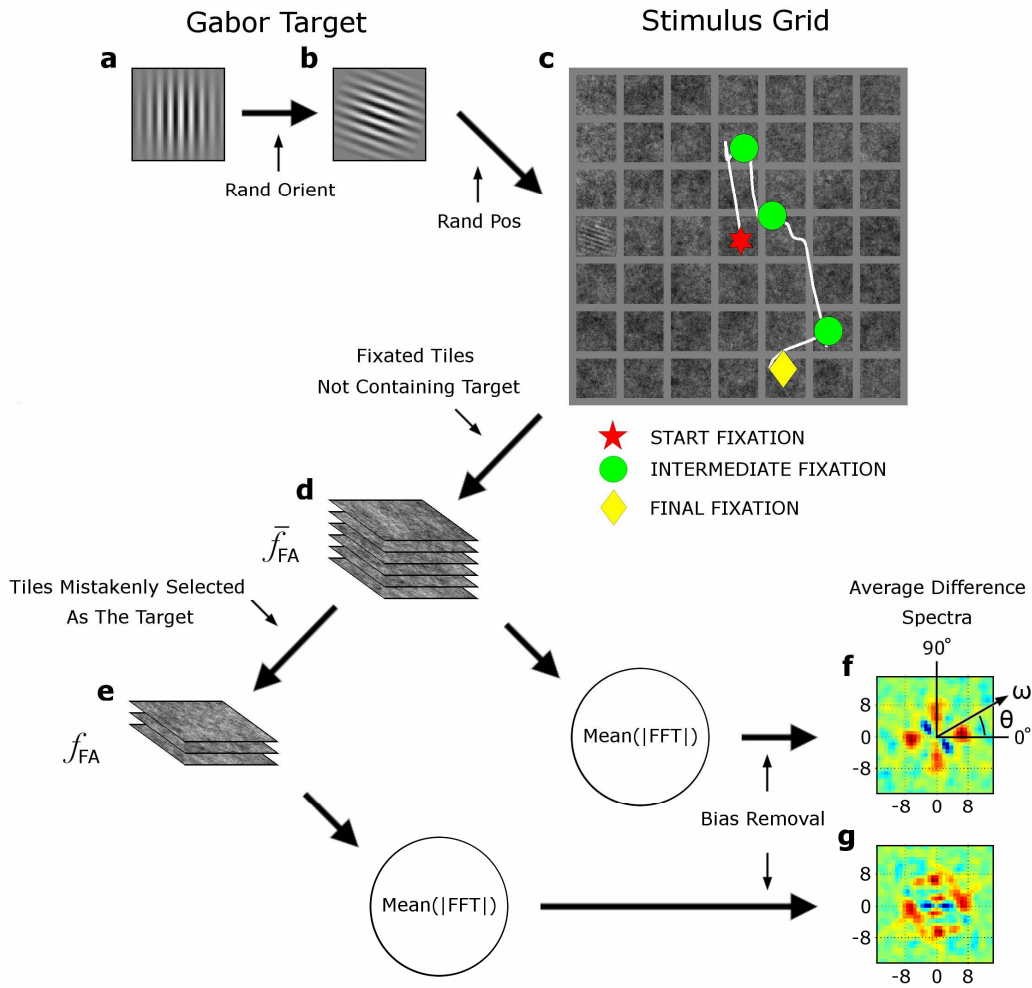


Figure 5.1: Stimulus creation, data capture, and data analysis. (a) A Gabor patch was used as a target and (b) its orientation was randomly selected from the set $\{0, 1, 2 \dots 179\}$ deg. (c) The target was added to a randomly selected tile of the $1/f$ noise grid and observer eye-movements were recorded while they searched for the target. An example of scan path is shown for a trial in which the observer did not find the target, located in the center of the leftmost column. (d) Fixated tiles that did not contain the target constitute our non-foveal false alarm category, and (e) a subset of these tiles, which were mistakenly selected at the end of trials as the target by the observer, constitute our foveal false alarm category. (f and g) Average difference spectra were computed by averaging the amplitude spectra of noise tiles in each category and subtracting the spectral bias (see text).

5.2.4 Analysis Method

Observers typically performed four to five fixations on average per trial in our current experiments, therefore visiting tiles not containing the target (i.e. noise-only tiles) and in some trials selecting one such tile as the target; an example stimulus grid with representative eye movements for a single observer is shown in Figure 5.1c.

We asked the same questions as for earlier experiments: Why were some noise-only tiles fixated whereas the others were not? And why, at the end of some trials, was a noise-only tile mistakenly selected as the tile containing the target? The analysis of the noise tiles was identical to that presented in the previous chapter. Note that no significant patterns were obtained by directly averaging, pixel by pixel in the spatial domain (i.e. retaining the phase information).

5.3 Results

Figure 5.2 shows the average difference spectra for the two false alarm categories obtained for the first set of 700 trials (first column), the second set of 700 trials collected approximately one month later (second column), and all 1,400 trials (third column) for each observer. For each observer and each set of trials, amplitude spectra were created using about 210 and 2,800 noise tiles respectively for the foveal and non-foveal categories. Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively (i.e. above and below the expected amplitude spectrum for a random observer). Regions in green show frequency components close to the bias. Surprisingly, each observer shows an idiosyncratic preference for certain distinct orientations. Further, note the high degree of similarity within observers between each set of 700 trials, particularly in the \bar{f}_{FA} category, which indicates the stability over time of these somewhat curious results. We have quantified

these similarities, using zero-lag correlation between the smoothed average difference spectra of the two sets, for each observer, and we have obtained on average 0.72 (ranging from 0.6 for AJS to 0.8 for AT). In the fourth column of Figure 5.2, we have cropped and enlarged the results from the third column to better visualize the spectral structures, and we have indicated the spatial frequency of the sought target (for the horizontal and vertical orientations). Notice that the peaks are close to the spatial frequency of the sought target (8 c/deg). Note that we did not find any reliable, dramatic effects of saccade length (which might be expected due to the falloff of resolution of the visual system).

We show in the last column of Figure 5.2 observers' performance in finding the Gabor target as a function of its orientation. Performance was pooled into 15 deg bins (12 bins total) and then averaged. Each bin contains about 117 trials. We indicate in red the average performance of each observer across all orientations (which is close to the initial value of 68% correct sought using the QUEST procedure). Regions in yellow and gray indicate performance above and below the observer's average performance, respectively. Notice that peaks in the average difference spectra for the f_{FA} category correspond quite closely to increases in performance at similar orientations. A repeated measures analysis of variance (ANOVA) with orientation and observer as factors showed a significant effect of orientation ($F(11,705) = 5.46$, $p = 2.17 \times 10^{-8}$) on the performance in finding the target, but a marginal effect of observer (note that the QUEST procedure ensured similar average performance for observers).

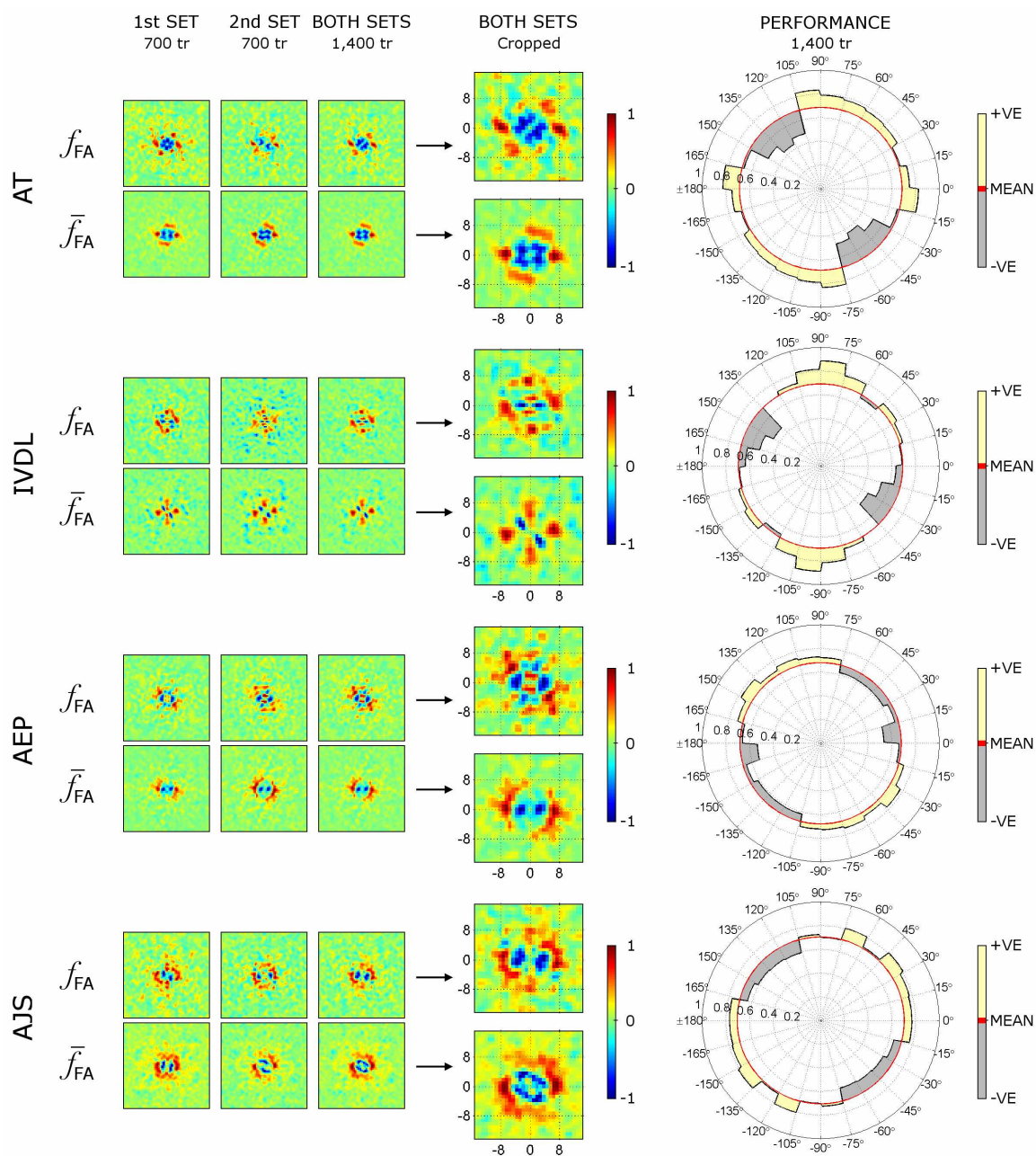


Figure 5.2: Average difference spectra and performance plots for four observers. Average difference spectra, smoothed and contrast-stretched for visual enhancement, are shown for the first set of 700 trials (first column), the second set of 700 trials collected approximately one month later (second column), and for all the 1,400 trials (third column). For each observer and each set of trials, the spectra were created using about 210 and 2,800 noise tiles respectively for the foveal and non-foveal categories. (*continued*)

(*continued*) Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively. Regions in green show frequency components close to the bias. In the fourth column, we have cropped and enlarged the results from the third column to better visualize the spectral structures, and we have indicated the spatial frequency (8 c/deg) of the search target (for the horizontal and vertical orientations). Observers performance (correct target detection rate, on a scale 0 to 1) are shown as a function of the orientation of the Gabor patches (pooled into 15 deg bins and averaged, each bin containing about 117 trials). We indicate in red the average performance of each observer across all orientations (this is close to the 68%). Regions in yellow and gray indicate performance above and below the observer's average performance, respectively.

We have also tested whether behavior in a given trial is affected by the outcome of the preceding trial. Essentially, we wished to establish if the orientation of the target in trial τ_{n-1} affected the outcome of trial τ_n , i.e. if a delusive sequential strategy or bias permeates the observer's results (observers were told that the target orientation was chosen at random for each trial). We introduce the rotated average difference spectra obtained by rotating the noise tiles at trial τ_n by the negative of the orientation of the target at trial τ_{n-1} , then averaging them across trials. For example, if the target at trial τ_{54} has an orientation of 75 deg, we would rotate all the noise tiles in trial τ_{55} by -75 deg before averaging them in the FA categories. This process is designed to highlight dependencies between successive trials. For instance, if on average the observer tends to look for a similar orientation as the target in the preceding trial, then we would expect to see strong increases in amplitude close to the reference orientation (θ_{Ref}), set at 0 deg. If there are no dependencies then we should observe an annulus (denoting an isotropic distribution) of the orientations. Figure 5.3 shows that for all four observers the \bar{f}_{FA} categories present structures close to an annulus, therefore that there appears to be, on average, no significant sequential bias. For observers AT and AEP, the f_{FA} category shows some biases; for example, a wide spread of high amplitude frequency components orthogonal to the reference orientation is observed for AT, possibly suggesting a decrease

in frequency detection accuracy for orientations perpendicular to the previous trial, or a slight tendency to select perpendicular orientations from trial to trial.

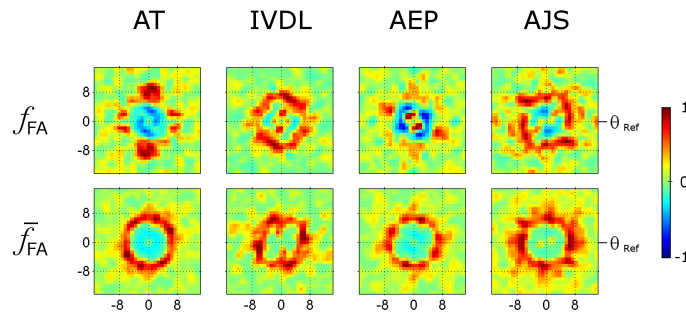


Figure 5:3: Test for sequential bias. Rotated average difference spectra, smoothed and contrast stretched for visual enhancement, are shown. Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively. Regions in green show frequency components close to the bias. See text for details.

5.4 Discussion

The main objective of this study was to investigate saccadic targeting and target selection in a naturalistic visual search task, when observers sought a randomly oriented Gabor target in a grid of $1/f$ noise (which has an amplitude spectrum distinctive to natural scenes). We are interested in discovering what attracts fixations and how target candidates are selected upon fixation, in particular, when observers are uncertain about a target feature (here, the orientation).

The results clearly point to visual guidance in saccadic target selection, in particular, under orientation uncertainty. We demonstrate that observers are selective for spatial frequencies close to the central frequency of the sought target; that is, the average difference spectra for fixated noise tiles show peaks localized in spatial frequency (close to 8 c/deg) but spread across various orientations (see Figure 5.2). Note that, in the previous chapter, we found that observers were selective for both spatial frequency and orientation when the orientation of the target was known (i.e. when we used Gabor

targets of spatial frequency 8 c/deg and fixed orientation {0, 20, 45, 70, and 90} deg in five separate experiments).

Surprisingly, even under conditions of complete orientation uncertainty, observers show pronounced, idiosyncratic biases for certain stimulus orientations in saccadic programming, that is, rotational smearing of the peaks in the average difference spectra for the fixated noise tiles is limited to a subset of orientations. Note that if observers were equally selective for all stimulus orientations, one would expect to obtain peaks spread across all orientations, hence, giving rise to a full annulus structure in the Fourier domain. Interestingly, these preferences are not exclusively limited to the cardinal directions (except for observer IVDL), which is somewhat inconsistent with many physiological and behavioral studies (Li *et al.*, 2003; Berkeley *et al.*, 1975; Campbell *et al.*, 1966), although the strongest preference across all observers appears to be close to vertical (0 deg). Nor are the biases solely reserved to the oblique orientations, as might be expected given the results of behavioral studies using more naturalistic stimuli (Hansen & Essock, 2004). Instead, we demonstrate preferences for a subset of orientations that encompass, in part, cardinal and oblique orientations.

These biases are also present in observers' performance data, and, within each observer, they are remarkably similar to the biases seen in average difference spectra for noise tiles selected as target candidates upon fixation (although more trials would be needed to examine the fine structure, if any, of this similarity). We show that the performance in finding the target is dependent on its orientation and that an asymmetry exists between clockwise and anticlockwise orientations; AJS, AT, and IVDL have a preference for anticlockwise oriented stimuli whereas AEP has a bias towards clockwise. Although mysterious in origin, such asymmetries have been reported in physiological studies of macaque monkeys (Finlay *et al.*, 1976) and appear in earlier behavioral data

(Boltz, Harwerth, & Smith, 1979). This observation may be a consequence of unequal populations or firing strength of orientation-tuned cells involved in the task, or may result from the assembly of search filters tuned such that certain orientations are amplified in sensitivity at the expense of others, in a dynamically reconfigurable, task-dependent manner. One may also speculate that these orientation biases could be related to observers' daily interactions with their unique environments, although further investigation would be required to substantiate these possibilities.

5.5 Conclusions

Our results offer insight into observer behavior in visual search tasks under uncertain stimulus conditions. In our experiment, the spatial frequency was held constant while the orientation varied. We found that the observers relied on an invariant target feature, namely spatial frequency structure similar to the sought target. Surprisingly, despite having no previous knowledge of each target's orientation, observers showed clear idiosyncratic biases in orientation selectivity during saccadic programming. These biases were also present in observers' foveal detection data and showed asymmetries between clockwise and anticlockwise orientations. Further examination of the effects of learning (e.g., training to least preferred orientations) may be useful in understanding mechanisms of plasticity in such tasks.

Chapter 6

Towards the Better Understanding of Search for Complex Targets

In the current chapter, we present two directions for further understanding observer strategies in visual searches for textures and complex targets. We use our experimental search framework, presented in Chapter 3, to examine how observers search for low-contrast targets created from Gabor summations (*Experiment 1*) and mosaicing (*Experiment 2*). We present several key discoveries. First, we show a strong presence of visual guidance in saccadic programming in search for such complex targets, demonstrated by selectivity for spatial frequencies and (in some cases) orientations close to the characteristics of each target. Second, multiple orientation attributes of the targets are shown to be represented in saccadic targeting and target selection in most cases, modulated by the observer's sensitivity / selectivity for each orientation. Third, different configurations of the Gabor mosaicing produce distinct tunings in orientation, but visibly idiosyncratic to each observer (*Experiment 2*). Moreover, a localized analysis is performed. Fourth, a curious presence of close-to-vertical structures is observed in fixated distracters, although the search targets did not contain vertically-oriented structures (*Experiment 2*).

6.1 Methods

The experimental framework for our experiments was similar to what we described earlier, and additionally contained several methodological extensions that we have highlighted in the next sections.

6.1.1 Observers

Four male observers (aged 27 through 31), of whom two were experienced (AT and IVDL) and two naïve (AEP and AJS), were tested in our experiments, with normal or corrected-to-normal vision. Observers AJS and IVDL were tested in Experiment 1, AEP in Experiment 2, and AT in both experiments. Each observer in Experiment 1 completed 2,100 trials (three sets of 700 trials spread over a period of about 1 month); and each observer in Experiment 2 completed 1,800 trials (three sets of 600 trials spread over a period of about 1 month).

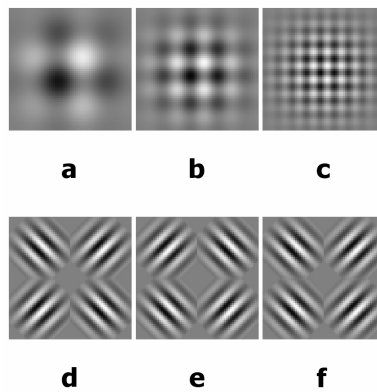


Figure 6.1: Targets used in our two experiments. Plaids created from the sum of two orthogonally oriented Gabor patches, one oriented at 0 deg and the other at 90 deg. The three targets used had component spatial frequencies of (a) 2, (b) 4, and (c) 8 c/deg and bandwidths of 0.25 octaves. Gabor mosaics created using various configurations ((d) “X”, (e) “O”, and (f) “V”) of two pairs of Gabors of spatial frequency 8 c/deg and bandwidth 0.5 octaves, oriented at -45 and 45 deg were used as targets in our second experiment.

6.1.2 Visual Stimuli

In Experiment 1, our search targets were 64×64 pixel plaids (or compound gratings), created from the sum of two orthogonally oriented Gabor patches, one oriented

at 0 deg (vertical) and the other at 90 deg (horizontal). The three targets used had component spatial frequencies of 2, 4, and 8 c/deg (Figures 6.1a, 6.1b, and 6.1c).

In Experiment 2, we used 64×64 pixel targets created from mosaics of Gabor patches of size 32×32 pixels, using various configurations (that we will refer to as “X”, “O”, and “V”) of two pairs of Gabors of spatial frequency 8 c/deg and bandwidth 0.5 octaves, oriented at -45 and 45 deg (Figures 6.1d, 6.1e, and 6.1f).

6.1.3 Procedure

The procedure was identical to that described in the Chapter 3. We would like to point out that over 80% of the dwell-times observed for final fixations were equal to or longer than 600 ms, an upper bound on typical fixation durations).

6.1.4 Analysis Method

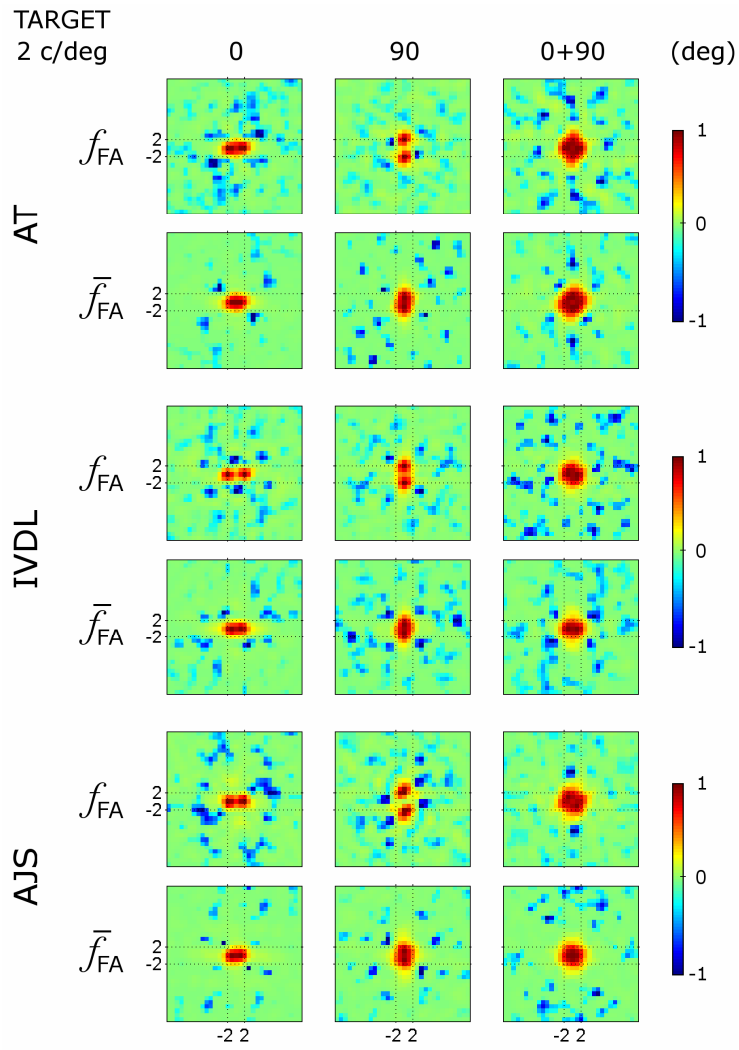
We pursued a similar philosophy as described in the previous chapters; that is, discovering why some noise-only tiles were fixated and why on some trials a noise-only tile was selected as a target-candidate, when observers performed our search tasks. Typically in the past experiments we have averaged the noise spectra in the foveal and non-foveal categories, and then removed the spectral bias to obtain what we have referred to as the *average difference spectra*. Here, we have adopted the same analysis for both our experiments, and additionally we have introduced a new local-based spectral analysis for Experiment 2. The latter analysis consists of first dividing each noise tile (of size 64×64 pixels) in each category into four equal-sized adjacent 32×32 pixel quadrants (i.e. the center of each quadrant is the same as the center of each of the Gabors in the target). We then compute the amplitude spectrum of each noise quadrant, therefore obtaining for each noise tile four local amplitude spectra. We then average all the first quadrant spectra of

the noise tiles in each category and subtract the noise bias corresponding to the first quadrant of all the 4,900 noise tiles. We perform the same operations for the three other noise quadrants to obtain what we will call the *localized average difference spectra*. With this analysis we can examine separately how different parts of the stimuli are treated during search tasks.

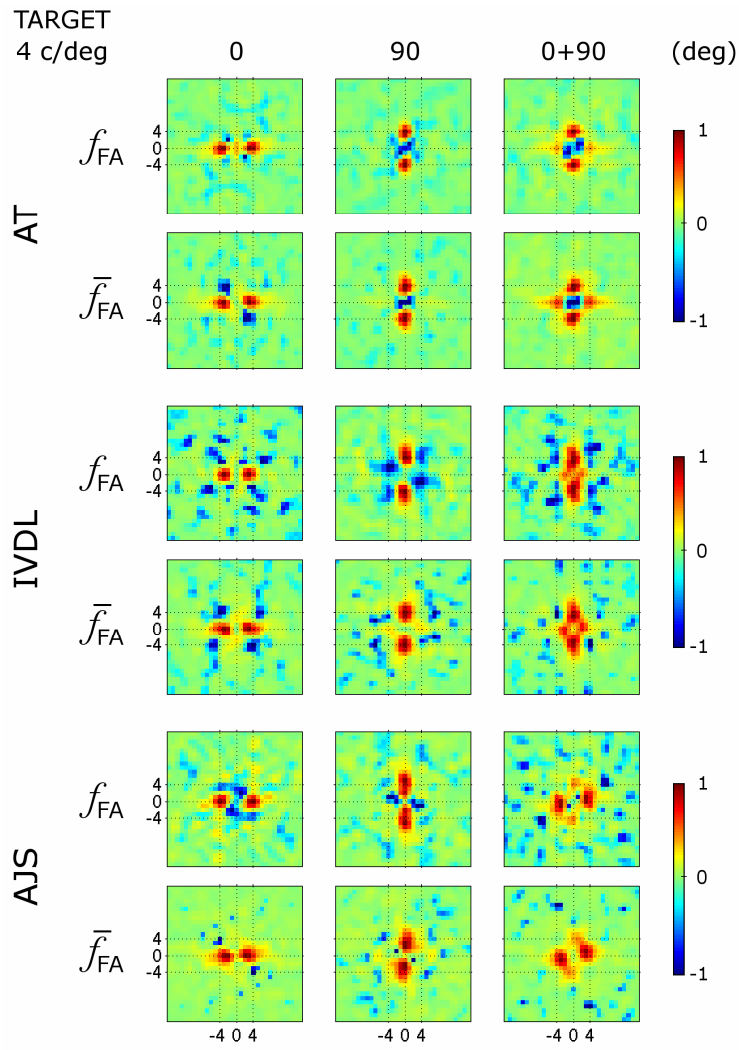
Note that directly averaging the noise tiles (i.e. in the spatial domain) within each category and observer produced an effect similar to that observed for search tasks using single Gabor targets described in Chapter 4, whereby the average images contain some structures in the case of search for lower frequency targets but non-reliable for searches for higher frequency targets.

6.2 Results

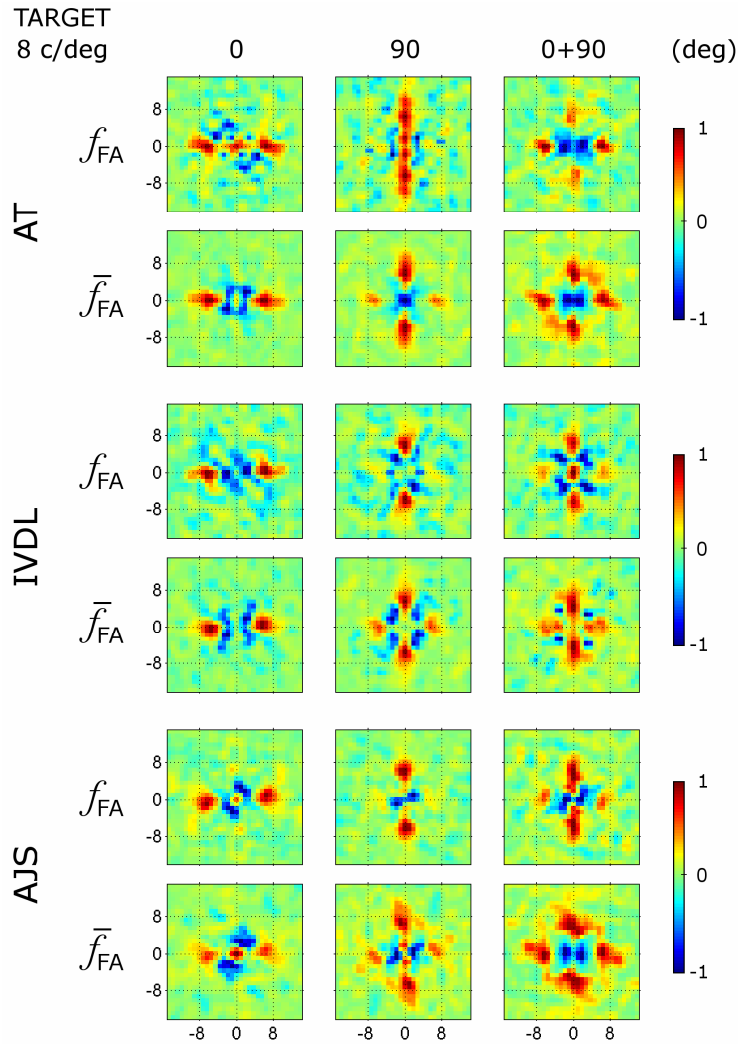
Figures 6.2a, 6.2b, and 6.2c show the average difference spectra for the two false alarm categories obtained for Experiment 1 using plaids of spatial frequency 2, 4, and 8 c/deg (third columns, denoted “0 + 90”). As a comparison we have shown the results from previous experiments (see Chapter 4) when single Gabor targets oriented anticlockwise from the vertical at 0 (first columns) and 90 deg (second columns) were used as targets. For each observer and set of 700 trials, amplitude spectra were created using about 210 and 2,800 noise tiles respectively for the foveal and nonfoveal categories. Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively (i.e. above and below the expected amplitude spectrum for a random observer). Regions in green show frequency components close to the bias.



(a)



(b)



(c)

Figure 6.2: Results for three observers in Experiment 1. Average difference spectra, smoothed and contrast-stretched for visual enhancement, are shown for sets of 700 trials for visual searches for targets with spatial frequency components centered at (a) 2 c/deg, (b) 4 c/deg, and (c) 8 c/deg. The first column shows results for search for a Gabor at 0 deg. The second column results for search for a Gabor at 90 deg. The last column shows the results for search for the sum of the two Gabors (indicated as 0 + 90 deg). Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively. Regions in green show frequency components close to the bias. For each observer and each set of trials, these images were created using about 210 and 2,800 noise tiles respectively for the foveal and non-foveal categories.

Note that the average difference spectra for fixated noise tiles show peaks close to both spectral components of the sought target, when observers searched for plaids with component spatial frequencies of 4 and 8 c/deg. Moreover, the peaks in the nonfoveal average difference spectra for observers AJS and AT show a spread in orientation. The two previous observations are less evident in the average difference spectra obtained when observers searched for the 2 c/deg plaid target; nevertheless, the results obtained for this particular target are clearly low-pass.

Figure 6.3 shows the average difference spectra for the two false alarm categories obtained for Experiment 2 using Gabor mosaic targets with “X” (first column), “O” (second column), and “V” (third column) configurations. For each observer and set of 600 trials, amplitude spectra were created using about 180 and 2,400 noise tiles respectively for the foveal and nonfoveal categories. Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively (i.e. above and below the expected amplitude spectrum for a random observer). Regions in green show frequency components close to the bias. Clear band-pass spectral structures are present in the nonfoveal average difference spectra; i.e. peaks are close to the component spatial frequencies of the sought target (8 c/deg) and spread in orientation, including orientations close to those characteristic of the sought target (i.e. at -45 and 45 deg).

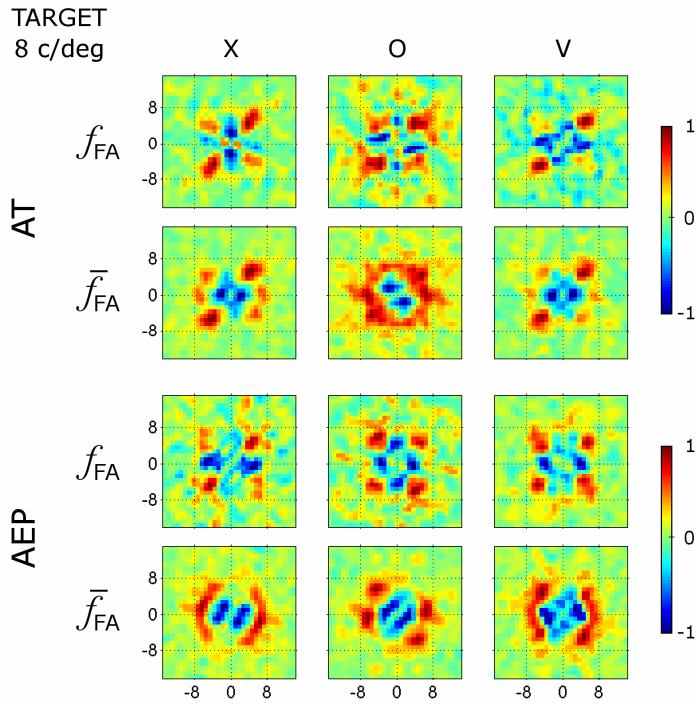


Figure 6.3: Results for two observers in Experiment 2. Average difference spectra, smoothed and contrast-stretched for visual enhancement, are shown for sets of 600 trials for visual searches for Gabor mosaic targets with “X” (first column), “O” (second column), and “V” (third column) configurations, as described in the text. Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively. Regions in green show frequency components close to the bias. For each observer and each set of trials, these images were created using about 180 and 2,400 noise tiles respectively for the foveal and non-foveal categories.

The structural similarity between certain average difference spectra in Figure 6.3 (e.g. the first and third column results for AEP) may lead one to think that analogous approaches may be used by observers in visual searches for distinct targets. We address this issue and whether observers make use some phase information with our local-based analysis described in the previous section. Figure 6.4 shows the localized average difference spectra for the two observers for the foveal false alarm category. Notice that the localized average difference spectra show differences for various configurations of

the Gabor mosaics. Also in some cases, the peaks are close to the local frequency components of the target, hence showing some use of the spatial configuration of the target components in the target-candidate selection process. However, there is some leakage between local regions.

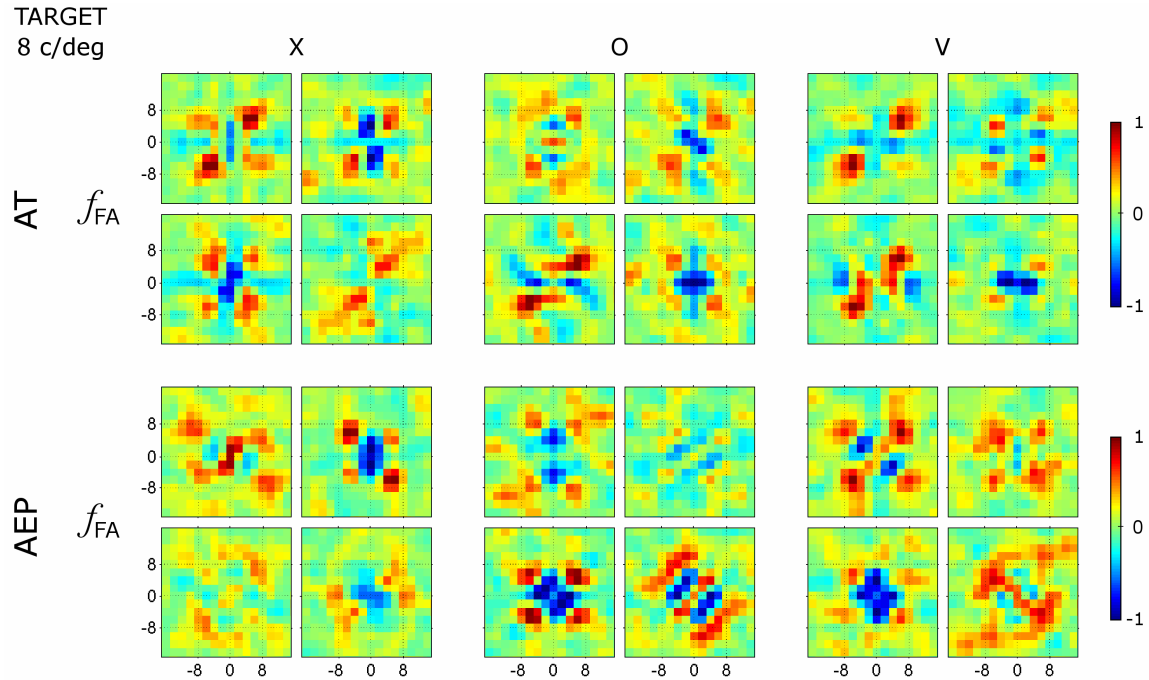


Figure 6.4: Results for two observers in Experiment 2. Localized average difference spectra, smoothed and contrast-stretched for visual enhancement, are shown for sets of 600 trials for visual searches for Gabor mosaic targets with “X” (first two columns), “O” (second two columns), and “V” (third two columns) configurations, as described in the text. Regions in red and blue indicate frequency components having amplitudes above and below the spectral bias, respectively. Regions in green show frequency components close to the bias. For each observer and each set of trials, these images were created using about 180 and 2,400 noise tiles respectively for the foveal and non-foveal categories.

6.3 Discussion

Our main goal in this chapter was to take a step forward in the further understanding of how observers search for complex targets in naturalistic visual search

tasks. To this end, we have designed search targets created from Gabor summations (Experiment 1) and mosaicing (Experiment 2) to measure observers' aptitude in finding such targets embedded in $1/f$ noise. Similarly to our experiments in the past two chapters, we were curious in discovering what attracts fixations and how target candidates are selected upon fixation.

The results underscore, once again, the prevalence of visual guidance in saccadic targeting, in particular during search for complex targets. We show that observers are selective for spatial frequencies and (in some cases) orientations close to those characteristic of the target; that is, the average difference spectra for fixated noise tiles present peaks localized in spatial frequency and sometimes in orientation, close to the spectral components of the sought target (Figures 6.2 and 6.3).

Furthermore, we find that multiple orientation attributes of the target are generally represented in saccadic targeting and target selection, demonstrated by selectivity for orientations close to both component orientations of the target (0 and 90 deg, in Experiment 1; - 45 and 45 deg, in Experiment 2). However, the representations of the oriented components are not equally weighted and appear to be modulated by observer's sensitivity / selectivity for each orientation. For instance, the average difference spectra for observer AT, in the "X" and "V" target search experiments (Figure 6.3), contain well localized peaks close to - 45 and 45 deg, but unequally weighted.

Interestingly, different configurations of the Gabor mosaicing produce distinct tunings in orientation, but visibly idiosyncratic to each observer (Experiment 2). For instance, the average difference spectra for both observers, obtained when the search target had the "O" configuration, is structurally quite different from the average difference spectra obtained for the "V" and "X" target configurations. Note that the

Gabor mosaic targets with the “X” and “O” configurations have the exact same Fourier amplitudes.

Intriguingly, there is, on average, an unusual presence of close-to-vertical structures in fixated distracters, although the search targets did not contain vertically-oriented structures in Experiment 2. Several possible explanations can be brought forward, some analogous to those discussed in Chapter 4 for a similar phenomenon when single Gabors were used as targets. One explanation could be that the presence of vertical components may be facilitating the detection of other orientations present in the noise tiles. Alternatively, it could simply reflect that observer’s gaze are drawn to noise tiles containing close-to-vertical orientations but having dominant spatial frequencies close to that of the target. These vertical components vanish in the foveal average difference spectra, i.e. for noise tiles selected as target candidates.

6.4 Conclusions

Our results provide convincing evidence for band-pass mechanisms and multiple orientations in saccadic targeting and target selection during visual search, targets. Furthermore, selectivity along feature dimensions (here, spatial frequency and orientation) shows inaccuracies, offsets, and curious biases. These errors are to some extent corrected during the foveal decision process. Furthermore, it appears that the presence or absence of various spectral components, other than those close to that of the search target, influence the guidance of saccades. We find that the absence of certain surround frequency components or the presence of near-vertical structures (i.e. components close to 0 deg) in the noise tiles attracts observer fixations.

Chapter 7

Towards Applications for Automated Visual Search

The main objective of this chapter is to present some ideas and insights inspired from what we have discovered in our experiments that could be integrated into automated visual search. A few of these ideas have been implemented in simple search frameworks. Although a template matching approach could produce better performance in some of these tasks, the intent here was to build an approach that could potentially outperform the latter on the long run. Besides, note that template matching makes the unrealistic assumption that observers have a full high-resolution representation of the target. To convince oneself, look at a nearby object for a few seconds, then just try drawing its contours on a piece of paper without looking at the object.

7.1 Some Insights

One of the key findings in all of our experiments has been the strong presence of visual guidance in saccadic targeting and target selection. Selectivity along feature dimensions (spatial frequency and orientation, in our experiments) plays a crucial role in saccadic programming and foveal scrutiny of target candidates. Furthermore, grouping of our human observer data based on saccade lengths and saccade order exposed the same tendency in each group, hence supporting our findings. Another observation was that in general spatial phase was not well used by human observers, i.e. if observers were strongly selective for spatial phase then one would expect to obtain target-like average noise images for the false alarm categories in our experiments. However, the latter

argument is not meant to fully reject the use of spatial phase, since some form of relative spatial phase or spatial relationships may be used by observers. In fact, we have found some indication of foveal usage of the spatial configuration of stimuli components in some of our experiments; that is, in the case of searches for Gabor mosaics, instances where local spectral components of stimuli presented similar orientation relationships as the components of the target. Moreover, under uncertainty, we found that observers appeared to rely on invariant target features to succeed at visual search tasks.

7.2 Applications to Visual Search for Objects

We propose extending these concepts to visual search for objects. We suggest first extracting the general featural characteristics of stimuli, without spatial relationships between them, for saccadic programming; then, upon fixation extracting the spatial configuration between various components of stimuli. The targets used for testing our models included a triangle, a star, and a wrench as search targets (Figure 7.1). We have used the exact same grid-based setup as described previously, but we intend to expand it to stimuli without grids in the near future. The target for each set of trials was embedded in a randomly selected tile of our $1/f$ noise grid.

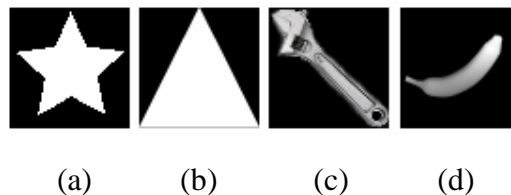
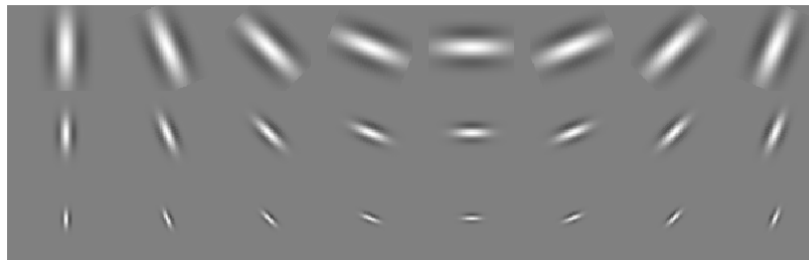
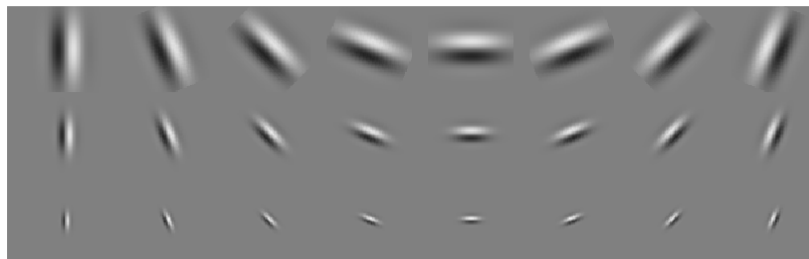


Figure 7.1: Targets used to test our models: (a) star, (b) triangle, (c) wrench, and (d) banana.

To extract the main spatial frequency and orientation characteristics of the target, we have used a bank of complex Gabor filters of various spatial frequencies and orientations; more specifically, we have employed a filter bank of 24 odd- and even-symmetric Gabors (of spatial frequencies of $\{2, 4, 8\}$ c/deg, orientations of $\{0, 22.5, \dots, 157.5\}$ deg, bandwidth of 1 octave, and aspect ratio of 2). We convolved the search target with each filter, shown in Figures 7.2a and 7.2b, to obtain the Gabor coefficients for the target, and then computed the envelope responses for the object (see Figure 7.3 for the case of the star).



(a)



(b)

Figure 7.2: Gabor filter bank. (a) Even- and (b) odd-symmetric Gabor filters of spatial frequencies of $\{2,4,8\}$ c/deg, orientations of $\{0, 22.5, \dots, 157.5\}$ deg, bandwidth of 1 octave, and aspect ratio of 2) are shown.

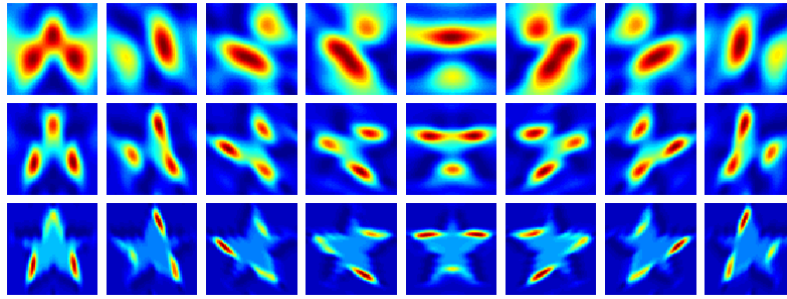


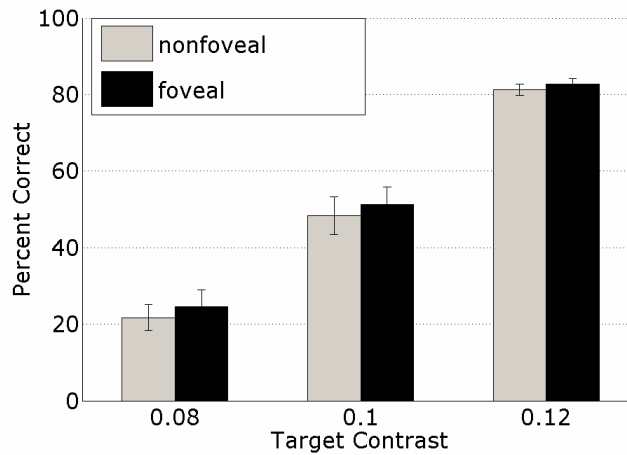
Figure 7.3: Envelope responses obtained for the star. Regions in red and blue indicate high and low values in the responses, respectively. Note that each individual envelope has been normalized for visual purposes.

We propose two models to make use of the Gabor coefficients for visual search: one where we select a fixed number (here $N = 8$ per scale) of peaks without the restriction of one for each envelope response (Model 1), and the other where we use one peak location from each envelope response (Model 2). Both models have a total number of 24 peak locations selected. The search algorithm consists in using only these peak locations found in the target to make nonfoveal and foveal decisions.

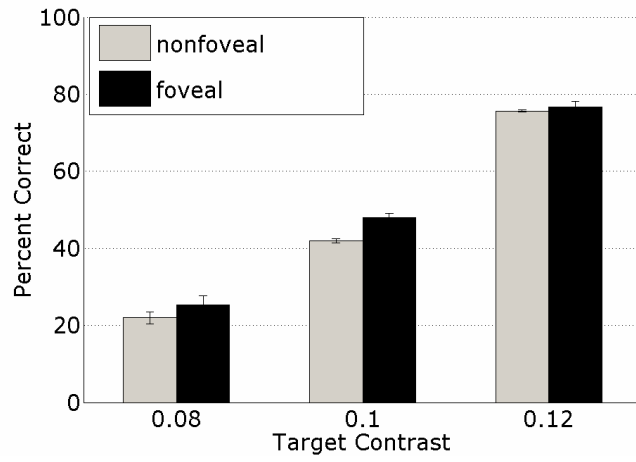
For both models, the nonfoveal decisions are made by matching the amplitude spectra of the stimulus tiles to the amplitude spectrum obtained by adding the amplitude spectra of the Gabors producing peak responses to the target (these can be weighted by the corresponding peak values); in this stage we assume no knowledge of the relative spatial positions of the Gabors. Once a nonfoveal match is found, the algorithm makes a saccade to that tile for foveal scrutiny. For the foveal selection, knowledge of the spatial positions of the Gabors is used. The zero-lag correlation is computed between the vector containing the 24 peak responses for the target and that containing the 24 envelope responses of the tile at the exact same locations. The foveal stage accepts the tile as being the target candidate if it is above a threshold. For the time being, we have selected the

threshold manually (for all our trials below we set it to 0.7), however by including a learning stage or a feedback mechanism, the algorithm could adjust the threshold.

Figures 7.4 and 7.5 show examples of correct detection rates as a function of the target contrast obtained for Models 1 and 2, respectively (note that we co-varied the contrast of the target and noise, giving a weight of c to the target and $(1 - c)$ to the noise). Three sets of 100 trials were run for each condition. In comparison, a random observer would have an expected correct detection rate of about 2%. The nonfoveal correct rate (gray bars) refers to the percent correct when the highest response of the nonfoveal stage is the tile containing the target. The foveal correct rate (black bars) corresponds to the percent correct at the final fixation. In most of the cases the foveal rate is higher than the nonfoveal; hence indicating that more than one fixation was performed to find the target.

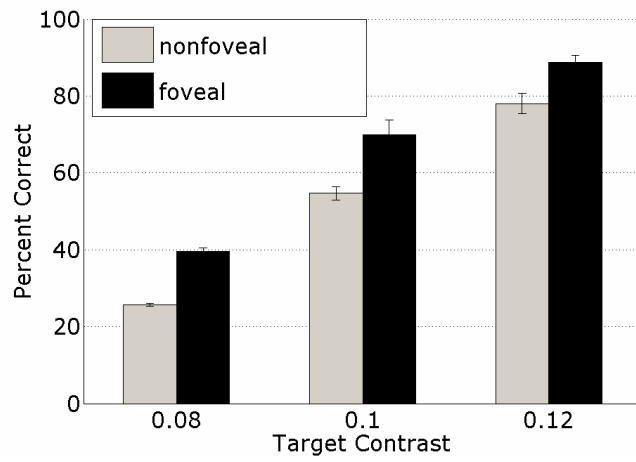


(a)

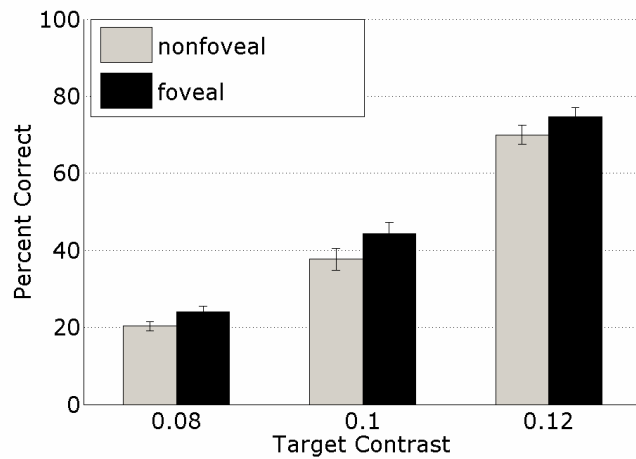


(b)

Figure. 7.4: Examples of correct detection rates, as a function of the target contrast, obtained for Model 1 for search for the (a) triangle and (b) star.



(a)



(b)

Figure. 7.5: Examples of correct detection rates, as a function of the target contrast, obtained for Model 2 for search for the (a) triangle and (b) star.

7.3 Conclusions

In this chapter, I have provided some insights into how some of our experimental findings may be used in automated visual search. I have tested some of these ideas in simple search tasks and have shown how it could be extended to object search. I plan to further develop the algorithm by including foveation. Furthermore, we could perform the nonfoveal operation on a larger patch size of the stimulus rather than the same size patch as for the foveal stage. This would be further in line with the performing operations on a much larger area. We can also combine all this with perhaps a multiscale representation of the target based on eccentricity, statistics on luminance & contrast, distributions of saccade lengths, and so on.

CHAPTER 8

Conclusions and Future Work

8.1 Conclusions

I have presented a new and effective technique to investigate saccadic targeting and target selection during visual search. This framework has enabled me to provide insight into observers' strategies in visual search tasks. Results obtained for various experimental conditions clearly point to visual guidance in saccadic target selection, a much debated issue in previous studies of active visual search. I have shown that observers use target attributes such as spatial frequency and orientation in saccadic programming and in selecting the target candidate upon fixation. Moreover, I have illustrated that observers exhibit uncertainties and offsets in selectivity for such features. Under uncertain stimulus conditions, I demonstrated the use of invariant features of the target as search strategy. An orientation bias was present for these conditions. I also showed that for targets containing multiple orientation characteristics, observers use multiple bandpass strategy.

8.2 Future Work

Some avenues of studies that I envision could be pursued in the future are as follows:

1. *Observer selectivity along other feature dimensions.* A natural expansion of my work would be to adapt our novel experimental technique to the study of other attributes such as but not limited to color, motion, size, and stereoscopic depth. The grand goal is to determine the weighting of each feature in saccadic targeting and target selection, which could then be used in automated search algorithms

2. *Visual search for object-like targets.* Following a similar logic as in Chapter 6, visual search could be examined for object-like targets constructed from multiple summations and mosaicings of Gabors of a variety of spatial frequencies and orientations. Such targets have the advantage of being highly specific in frequency and orientation characteristics, and are much more controllable than the direct use of objects. An interesting question would be what aspects of these targets are used during search. One may find that certain frequency components might be weighed more than others are or that some spatial characteristics of the target may emerge.

3. *Target Prevalence.* So far, in all the studies that I have been doing, the target has always been present in every trial. Thus, other search configurations could be examined; for example, when different probabilities of appearance of a target during trials are used. The goal would be to investigate differences in search strategies and behavior. In an interesting study, inspired from baggage inspections at airports, Wolfe (2005) showed that rare items were often missed in visual search tasks; one may adapt the present framework to perform similar experiments, perhaps using Gabor targets, to examine the nature of

possible deteriorations in search behavior, reflected in the average noise images, when target prevalence is lowered.

4. *Interaction between visual perception and eye movements.* Various studies have shown that eye movements can affect visual perception and performance in certain tasks. Hence, it would be of great interest to scrutinize how eye movements enhance visual performance; for instance, task performance could be compared in unconstrained versus limited eye movement conditions.
5. *Cognitive Processes.* The understanding of eye movement behaviors in naturalistic tasks could provide insight into cognitive processes involved. Earlier experiments of Yarbus (1967) have nicely demonstrated that cognitive processes affect scanning patterns; for instance, recordings of eye movements when an observer viewed the same image but was given different instructions produced differing gaze patterns. It would be quite exciting to link eye movement behavior to cognitive processes. Such discoveries could have a great impact in applications for novel visually directed human-machine interfaces; e.g., for unmanned aerial vehicles, remote robotics especially in hostile environments, assistance to patients with limited mobility, and security/surveillance systems.

Bibliography

- Abbey, C. K., & Eckstein, M. P. (2002). Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments. *Journal of Vision*, 2, 1, 66–78.
- Abbott, A.L (1991). Selective fixation control for machine vision: A survey. *IEEE International Conference on System, Man and Cybernetics*, 1–6.
- Agarwal, S., Awan, A., & Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26, 11, 1475–1490.
- Ahumada, A.J., Jr. (1996). Perceptual classification images from vernier acuity masked by noise. *Perception*, 26, 18.
- Ahumada, A.J., Jr., & Beard, B.L. (1999). Classification images for detection. *IOVS 40 (4, ARVO Suppl.)*, S572. <http://vision.arc.nasa.gov/personnel/al/talks/99arvo/outline.htm> .
- Ahumada, A.J., Jr. (2002). Classification image weights and internal noise level estimation. *Journal of Vision*, 2, 1, 121–131.
- Ahumada, A.J., Jr., & Krebs, W. K. (2000). Signal detection in fixed pattern chromatic noise. *IOVS 41 (4, ARVO Suppl.)*, S713. <http://vision.arc.nasa.gov/personnel/al/talks/00arvo/mask/outline.htm>.
- Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The oblique effect in man and animals. *Psychological Bulletin*, 78, 4, 266–278.
- Beard, B.L. & Ahumada, A.J., Jr. (1998). A technique to extract relevant image features for visual tasks. *SPIE Proc. on Human Vision and Electronic Imaging III*, 3299, 79–85.
- Beard, B.L., & Ahumada, A. J., Jr. (1999). Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. *Journal of the Optical Society of America A*, 16, 755–763.
- Berkley, M.A., Kitterle, F., & Watkins D.W. (1975). Grating visibility as a function of orientation and retinal eccentricity. *Vision Research*, 15, 2, 239–244.

- Blakemore, C., Carpenter, R.H.S., & Georgeson, M.A. (1970). Lateral inhibition between orientation detectors in the human visual system. *Nature*, 228, 37–39.
- Blakemore, C. & Campbell, F. W. (1969). Adaptation to spatial stimuli. *Journal of Physiology*, 200, 11–13.
- Boltz, R.L., Harwerth, R.S., & Smith, E.L. III (1979). Orientation anisotropy of visual stimuli in rhesus monkey: a behavioral study. *Science*, 205, 4405, 511–513.
- Bosch, H., Milanese, R., & Labbi, A. (1998). Object segmentation by attention induced oscillations. *Proc. IEEE Int. Joint Conf. Neural Networks*, 2, 1167–1171.
- Bouet, R., & Knoblauch, K. (2004). Perceptual classification of chromatic modulation. *Visual Neuroscience*, 21, 3, 283–290.
- Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Brunelli, R., & Poggio, T. (1997). Template matching: Matched spatial filters and beyond. *Pattern Recognition*, 30, 5, 751–768.
- Bundesen, C. (1996). Formal models of visual attention: A tutorial review. In A. Kramer, G. H. Cole, and G. D. Logan (Eds.), *Converging operations in the study of visual selective attention*. 1–44.
- Campbell, F.W., Cooper, G.F., & Enroth-Cugell, C. (1969). The spatial selectivity of visual cells of the cat. *Journal of Physiology*, 203, 223–235.
- Campbell, F.W. & Robson, J.G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197, 551–566.
- Campbell, F.W., Kulikowski, J.J., & Levinson, J. (1966). The effect of orientation on the visual resolution of gratings. *Journal of Physiology*, 187, 2, 427–436.
- Carrasco, M., McLean, T.L., Katz, S.M., & Frieder, K.S. (1998). Feature asymmetries in visual search: Effects of display duration, target eccentricity, orientation and spatial frequency. *Vision Research*, 38, 3, 347–374.
- Coppola, D.M., White, L.E., Fitzpatrick, D., & Purves, D. (1998). Unequal representation of cardinal and oblique contours in ferret visual cortex. *Proceedings of the National Academy of Sciences USA*, 95, 5, 2621–2623.
- De Valois, R.L., Albrecht, D.G. & Thorell, L.G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22, 545–559.
- Eckstein, M. (1998). The lower efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science*, 9, 111–118.

- Eckstein, M.P., & Ahumada, A.J. Jr. (2002). Classification images: A tool to analyze visual strategies. *Journal of Vision*, 2, 1.
- Eckstein, M. P., Beutter, B. R., & Stone, L. S. (2001). Quantifying the performance limits of human saccadic targeting during visual search. *Perception*, 30, 1389–1401.
- Eckstein, M.P., Shimozaki, S.S., & Abbey, C.K. (2002). The footprints of visual attention in the Posner cueing paradigm revealed by classification images. *Journal of Vision*, 2, 1, 25–45.
- Eckstein, M.P., Beutter, B.R., Pham, B.T., Shimozaki, S.S., & Stone, L.S. (2007). Similar neural representations of the target for saccades and perception during search. *Journal of Neuroscience*, 27, 6, 1266–1270.
- Enroth-Cugell, C. & Robson J.G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, 187, 3, 517–552.
- Ferster, D., Chung, S. & Wheat, H. (1996). Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature*, 380, 249-252.
- Field, D.J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4, 2379–2394.
- Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, 37, 617–631.
- Findlay, J. M. (2004). Eye scanning & visual search. In J. M. Henderson and F. Ferreira (Eds.). *The interface of language, vision, and action: Eye movements and the visual world*. 135–159. New York: Psychology Press.
- Findlay J.M., & Gilchrist I.D. (1998). Eye guidance during visual search. In Underwood, G. (Ed.), *Eye guidance in reading and scene perception*, Elsevier, 297–314.
- Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The Psychology of Looking and Seeing*. Oxford University Press.
- Foster, D.H., & Ward, P.A. (1991). Asymmetries in oriented-line detection indicate two orthogonal filters in early vision. *Proceedings of the Royal Society B (London)*, 243, 1306, 75–81.
- Furmanski, C.S., & Engel, S.A. (2000). An oblique effect in human primary visual cortex. *Nature Neuroscience*, 3, 6, 535-536.

- Geisler, W. S., Perry, J. S., & Najemnik, J. (2006). Visual search: The role of peripheral information measured using gaze-contingent displays. *Journal of Vision*, 6, 9, 858–873.
- Geisler, W. S. (2003). Ideal observer analysis. In L. M. Chalupa & J. S. Werner (Eds.), *Visual Neurosciences*. Cambridge, MA: MIT Press.
- Geisler, W. S., & Chou, K.-L. (1995). Separation of low-level and high-level factors in complex tasks: Visual search. *Psychological Review*, 102, 356–378.
- Geisler, W. S., & Perry, J. S. (1998). A real-time foveated multiresolution system for low-bandwidth video communication. *Proceedings of SPIE*, 3299, 294–305.
- Gieffing, G.J., Janssen, H., & Mallot, H. (1992). Saccadic Object Recognition with an Active Vision System. *Proc. 10th Eur. Conf. Art. Intell.*, 803–805.
- Gold, J.M., Murray, R.F., Bennett, P.J., & Sekuler, A.B. (2000). Deriving behavioral receptive fields for visually completed contours. *Current Biology*, 10, 11, 663–666.
- Gold, J.M., Sekuler, A.B., & Bennett, P.J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, 28, 167–207.
- Graham, N. & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channel models. *Vision Research*, 11, 251–259.
- Gosselin, F., & Schyns, P. G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychological Science*, 14, 505–509.
- Hansen, B.C. & Essock, E.A. (2004). A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes. *Journal of Vision*, 4, 12, 1044–1060.
- Hansen, T., & Gegenfurtner, K. R. (2005). Classification images for chromatic signal detection. *Journal of the Optical Society of America (A)*, Feature issue on Visual coding in memory of Russell L. De Valois, 22, 10, 2018–2089.
- Henderson, J.M., Brockmole, J.R., Castelano, M.S., & Mack, M.L. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain*. Oxford: Elsevier.
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, 50, 243–271.

- Hoffman, J. E. (1998). Visual attention and eye movements. In H. Pashler (Ed.), *Attention*, San Diego: Psychology Press, 119–153.
- Hooge, I. T. C., & Erkelens, C. J. (1999). Peripheral vision and oculomotor control during visual search. *Vision Research*, 39, 8, 1567–1575.
- Hubel D.H. & T.N. Wiesel (1968). Receptive fields and functional architecture of monkey striate cortex, *Journal of Physiology*, 195, 215–244.
- Hubel, D. & Wiesel, T. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160, 106–154.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489–1506.
- Jacob, R.J.K. (1995). Eye tracking in advanced interface design. In W. Barfield and T.A. Furness (Eds.), *Virtual Environments and Advanced Interface Design*, Oxford University Press, New York, 258–288.
- Kim, M.-S., & Cave, K. R. (1999b). Grouping effects on spatial attention in visual search. *Journal of General Psychology*, 126, 326–352.
- Kinchla, R. A. (1992). Attention. *Annual Review of Psychology*, 43, 711–742
- Klarquist, W.N., & Bovik, A.C. (1998). FOVEA: A foveated, multi-fixation, vergent active stereo system for dynamic three-dimensional scene recovery. *IEEE Transactions on Robotics and Automation*, 14, 5, 755–770.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4, 219–227.
- Kontsevich, L.L., & Tyler, C.W. (2004). What makes Mona Lisa smile? *Vision Research*, 44, 13, 1493–1498.
- Kuffler, S.W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16, 37–68.
- Li, B., Peterson, M.R., & Freeman, R.D. (2003). Oblique effect: a neural basis in the visual cortex. *Journal of Neurophysiology*, 90, 1, 204–217.
- Liversedge, S.P., & Findlay, J.M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Science*, 4, 6–14.
- Mansfield, R.J.W. (1974). Neural basis of orientation perception in primate vision. *Science*, 186, 4169, 1133–1135.

- McGowan, J.W., Kowler, E., Sharma, A., & Chubb, C. (1998). Saccadic localization of random dot targets. *Vision Research*, 38, 895–909.
- McSorley E., & Findlay J.M. (2003). Saccade target selection in visual search: accuracy improves when more distractors are present. *Journal of Vision*, 3, 877–892.
- Milanese, R. (1993). Detecting salient regions in an image: from biological evidence to computer implementation. Ph.D. Dissertation No. 2647, University of Geneva, Switzerland.
- Mohan, A., Papageorgiou, C., & Poggio, T. (2001). Example-based object detection in images by components. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23, 4, 349–361.
- Morrone, M. C., Burr, D. C. & Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proceedings of the Royal Society of London, Series B*, 216, 335–54.
- Motter, B.C., & Belky, E.J. (1998). The guidance of eye movements during active visual search. *Vision Research*, 38, 1805–1815.
- Neri, P., & Heeger, D.J. (2002). *Spatiotemporal mechanisms for detecting and identifying image features in human vision. Nature Neuroscience*, 5, 8, 812–816.
- Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature*, 401, 695–698.
- Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, 40, 1227–1268.
- Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Rajashekar, U., Cormack, L. K., & Bovik, A. C. (2002). Visual search: Structure from noise. *Proceedings of the Eye Tracking Research & Applications Symposium*, New Orleans, LA, USA, March 25–27, 2002, 119–123.
- Rajashekar, U., Cormack, L.K., & Bovik, A.C. (2004). Point of gaze analysis reveals visual search strategies. *Proc. of SPIE on Human Vision and Electronic Imaging IX*, 5292, 296–206.
- Rajashekar, U., Bovik, A.C., & Cormack, L.K. (2006). Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision*, Special Issue on Finding visual features: Using stochastic stimuli, 6, 4, 379–386.

- Ramoá, A. S., Shadlen, M., Skottun, B. C. & Freeman, R. D. (1986). A comparison of inhibition in orientation and spatial frequency selectivity of cat visual cortex. *Nature*, *321*, 237–239.
- Ringach, D.L., Hawken, M.J. & Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature*, *387*, 281–84.
- Ringach, D. L. (1998). Tuning of orientation detectors in human vision. *Vision Research*, *38*, 963–972.
- Schiller, P.H., Finlay, B.L. & Volman, S.F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. III. Spatial frequency. *Journal of Neurophysiology*, *39*, 1334–1351.
- Schneider, W., & Schiffman, R. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*, 127–190.
- Scialfa, C. T., & Joffe, K. M. (1998). Response times and eye movements in feature and conjunction search as a function of target eccentricity. *Perception and Psychophysics*, *60*, 1067–1082.
- Sekuler, A.B., Gaspar, C.M., Gold, J.M., & Bennett, P.J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, *14*, 391–396.
- Shevelev, I.A., Lazareva, N.A., Novikova, B.V., Tikhomirov, A.S., & Sharaev, G.A. (1994). Double orientation tuning in cat visual cortex units. *Neuroscience*, *61*, 965–973.
- Sillito, A.M., Grieve, K.L., Jones, H.E., Cudeiro, J. & Davis, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, *378*, 492–496.
- Simoncelli, E. P. (2002). Seeing patterns in noise. *Trends in Cognitive Sciences*. *7*, 51–53.
- Snowden, R.J. (1992). Orientation bandwidth: the effect of spatial and temporal frequency. *Vision Research*, *32*, 1965–1974.
- Solomon, J.A. (2002). Noise reveals visual mechanisms of detection and discrimination. *Journal of Vision*, *2*, 1, 105–120.
- Stromeyer, C.F. & Julesz, B. (1972). Spatial frequency masking in vision: critical bands and spread of masking. *Journal of the Optical Society of America*, *64*, 1221–1232.

- Swain, M. J., & Stricker, M., editors (1991). Promising directions in active vision: report of the NSF Active Vision Workshop. Technical Report 91-27, Department of Computer Science, University of Chicago.
- Tavassoli, A., van der Linde, I., Bovik, A. C., & Cormack, L. K. (2006). Noise unveils spatial frequency and orientation selectivity during visual search. *Journal of Vision*, 6(6):450, 450a, <http://journalofvision.org/6/6/450/>, doi:10.1167/6.6.450.
- Tavassoli, A., van der Linde, I., Bovik, A.C., & Cormack, L.K. (2007a). An efficient technique for revealing visual search strategies with classification images. *Perception & Psychophysics*, 69, 1, 103-112.
- Tavassoli, A., van der Linde, I., Bovik, A. C., & Cormack, L. K. (2007b). Orientation anisotropies in visual search revealed by noise. *Journal of Vision*, 12, 11, 1-8, <http://journalofvision.org/7/12/11/>, doi:10.1167/7.12.11.
- Tavassoli, A., van der Linde, I., Bovik, A. C., & Cormack, L. K. Characterizing spatial frequency and orientation selectivity in visual search. (In preparation for *Vision Research*).
- Theeuwes, J., & Kooi, F. L. (1994). Parallel search for a conjunction of contrast polarity and shape. *Vision Research*, 34, 3013-3016.
- Tolhurst, D. J. (1972). Adaptation to square-wave gratings: inhibition between spatial frequency channels in the human visual system. *Journal of Physiology*, 226, 231-248.
- Treisman A. & Gelade G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136
- Treisman, A. & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychology Review*, 95, 1, 15-48.
- Verghese, P., & Nakayama, K. (1994). Stimulus discriminability in visual search. *Vision Research*, 34, 2453-2467.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33, 2, 113-20.
- Williams, D. E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple-conjunction search tasks. *Psychonomic Bulletin and Review*, 8, 476-488.
- Willmore, B. & Smyth, D. (2003). Methods for first-order kernel estimation: simple-cell receptive fields from responses to natural scenes. *Network*, 14, 553-577.

- Wilson, H.R., McFarlane, D.K., & Phillips, G.C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, 23, 9, 873–882.
- Wolfe, J.M. & Horowitz, T.S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, 5, 495–501.
- Wolfe, J. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, 1, 202–238.
- Wolfe, J. M. (1998). Visual Search, Chapter One. In H. Pashler (Ed.), *Attention*, San Diego: Psychology Press, 13–73.
- Wolfe, J.M., Friedman-Hill, S.R., Stewart, M I., & O'Connell, K. M. (1992). The Role of Categorization in Visual Search for Orientation. *Journal of Experimental Psychology*, 18, 1, 34–49.
- Yang, G.Z., Dempere–Marco, L., Hu, X.–P., & Rowe, A. (2002). Visual search: Psychophysical models and practical applications. *Image and Vision Computing*, 20, 291–305.
- Yarbus, A.L. (1967). *Eye movements and vision*. New York: Plenum.
- Zelinsky, G. J. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research*, 36, 2177–2187.
- Zelinsky, G. J., Rao, R. P., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements reveal the spatio–temporal dynamics of visual search. *Psychological Science*, 8, 448–453.
- Zelinsky, G.J. & Sheinberg, D.L. (1997). Eye movements during parallel–serial visual search. *J. Exp. Psychol.*, 23, 244–262.

Vita

Abtine Tavassoli was born on May 20th, 1978. He received his Diplôme d'Ingénieur title (Master of Engineering) from Polytech'Lille, Lille, France in 2001. He was an exchange student at the University of Texas at Austin in the Department of Electrical and Computer Engineering (ECE), between August 2000 and May 2001. He then joined the Master's program in the ECE department at UT Austin in 2001, under the supervision of Dr. Michael F. Becker, and received his Master of Science degree in 2002. He then joined the Ph.D. program in the ECE department at UT Austin in 2002, under the supervision of Dr. Alan C. Bovik and Dr. Lawrence K. Cormack. He is a member of the Center for Perceptual Systems and the Laboratory for Image and Video Engineering at UT Austin.

His research interests include research interests include visual attention, visual search, eye movements, scene perception, visual working memory, color vision, image processing, and computer vision.

Permanent address: 63, Quai Marcel Dassault
92210 Saint Cloud France

This dissertation was typed by the author.