IS THE HUMAN VISUAL SYSTEM OPTIMISED FOR ENCODING THE STATISTICAL INFORMATION OF NATURAL SCENES?

by

Carlos Alejandro Párraga

A thesis submitted in to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Science, Department of Experimental Psychology.

April 2003

73,760 words

ABSTRACT

Since all properties (anatomy, physiology and sensory processing mechanisms) of the human visual system are the product of evolution and postnatal development, it is reasonable to suppose that they have been configured to make the most effective use of the characteristics of the environment, according to the lifestyles of the individuals who exploit them. This optimisation has been shown most convincingly for fish and insects, but there are many information theory based studies, which point to a similar optimisation in humans and other primates. Here we aim to demonstrate this tenet experimentally.

In the first two studies, we conducted psychophysical experiments that required human observers to distinguish between achromatic pictures of slightly different faces or objects (i.e. shape discrimination). The stimuli were digitally modified to increase their deviation from the second-order statistics of natural images, making them increasingly "unnatural" in this respect. Performance was best when the stimuli had statistics similar to those found in the natural visual environment. We also explored performance when viewing stimuli monocularly, using foveal and peripheral vision. Our results show that performance in peripheral vision was best for slightly "blurred" morph sequences and that stimulus resizing (M-scaling) did not fully compensate for the deficiencies of peripheral vision. In a third study, a simple multi-resolution cortical model of the discrimination processes was shown to be capable of predicting the previous psychophysical results. A fourth study was carried out to explore the spatio-chromatic information content of natural scenes. We found that a particular subset of scenes (those containing red fruit on a background of green or yellow leaves) have properties that match those of the psychophysically measured achromatic and red-green chromatic contrast sensitivity funtions but not the blue-yellow chromatic contrast sensitivity functions, implying that the red-green opponent system is particularly well suited to the fruit/leaf discrimination task.

ACKNOWLEDGMENTS

Para Patxi, la luz de mis ojos

As with any scientific work created over a long period, this is the sum of many individual contributions, put together. They are the contributions of all those who studied the subject of vision before me and wrote about their work, those who taught me what I know, those who discussed these issues with me in an informal basis and the loved ones who supported my work at home and abroad.

Many people guided my voyage throughout this fascinating discipline of visual perception. Among those, I would like to express my gratitude, in first place, to my supervisor Prof. Tomasz Trościanko. Thank you Tom for your enthusiasm and energy, for keeping beside and supporting me along all these years. Thank you for teaching me but most importantly, thank you for always being there when I needed you.

The second person I would like to thank is Dr David Tolhurst of Cambridge University, who contributed with his enormous understanding and insightful opinions to the main ideas and discussions behind this work. He also spent hours of his precious time going through my manuscripts, graphs and calculations and making sure all was intelligible, organised and neat. Thank you Dave for your generosity.

I would also wish to thank all the other members of the Bristol University Vision Group in the Department of Experimental Psychology: Dr Iain Gilchrist, Dr Nick Scott-Samuel, Dr Chris Benton and Dr Christof Koerner who contributed with ideas, comments, and scientific literature to this work and Prof Innes Cuthill from the School of Biological Sciences in Bristol University, who recommended some exciting

references on adaptation and optimality. This acknowledgement extends to all staff and postgraduate students that shared discussion sessions, presentations and seminars where this work was debated. I would also like to thank Dr Phil Benson who created one of the morph sequences used in this work and the staff of the Bristol University Botanical Gardens for letting me work among their precious tropical plants.

As almost anything in life, this work would have been impossible without generous funding. In my case, the funds came from the Department of Experimental Psychology of Bristol University who bore the cost of my tuition fees, the Medical Research Council and the Biological and Biotechnical Research Council who provided the funds to employ me full-time throughout the whole period of my work.

The people who contributed to this work with love and affection are many. I would like to show my appreciation to all of them, but most importantly…

To my family, *mis amores*, Claudia and Patxi, for understanding me, supporting me and placing laughter in my face every time I arrived home with my head still attached to my office PC. You were the real engine behind this work.

To my family and friends back in Argentina. *A mamá, papá, Ceci y Clara. Los quiero mucho.* You will never guess how much I missed you.

To my friends and neighbours Rossana and Natalio and to Juan, Helen, Barbara, Aviv, Tony, Rohan, Hans and many others who contributed so much with their warmth and affection. I love you all. You are part of this work too.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree.

Any views expressed in the dissertation are those of the author and in no way represent those of the University of Bristol.

The dissertation has not been presented to any other University for examination either in the United Kingdom or overseas.

SIGNED: .....

DATE:.....

TABLE OF CONTENTS

Section *Page*

LIST OF TABLES AND ILLUSTRATIVE MATERIAL

Tables

Table Number *Page*

Figures

Figure Number *Page*

ABBREVIATIONS AND ACRONYMS

2AFC	Two-Alternative Forced Choice
Chrom	Chrominance
CIE	Commission Internationale de l'Eclairage
CRT	Cathode-Ray Tube
CSF	Contrast Sensitivity Function
EO	Electro-Optical
EV	Exposure Value
GoF	Goodness-of-Fit
HVS	Human Visual System
ICA	Independent Components Analysis
IR	Infra Red
IT	Integration Time
ITU	International Telecommunications Union
JPEG	Joint Photographic Experts Group
LCD	Liquid Crystal Display
LED	Light-Emitting Diode
LGN	Lateral Geniculate Nucleus
LMS	Long, Middle, Short
Lum	Luminance
LUT	Look-Up Table
ND	Neutral Density
PC	Personal Computer
PCA	Principal Components Analysis
RF	Receptive Field
RGB	Red, Green and Blue
RMS	Root Mean Square
SE	Standard Error
SF	Spatial Frequency
SR	Spectroradiometer
STD	Standard Deviation
TIFF	Tagged-Image File Format
UV	Ultra Violet

GLOSSARY

Morph sequence: indexed sequence of 41 digitised monochrome pictures of similar size and grey level depth produced artificially from two original (reference) pictures using a morphing technique. This technique allows small changes in shape, texture and contrast from one image to the next. The two original pictures are the fist and last in the sequence.

Reference image: one of the non-morphed (first or last) images of a morph sequence.

Test image: one of the morphed (intermediate) pictures of the morph sequence.

Experimental trial: single run of the morph discrimination experiment. Three pictures from the same morph sequence are presented sequentially on the screen and the observer has to decide which one is the "odd one out" (2AFC paradigm).

Experimental series: collection of 180-200 trials performed on a morph sequence with the same characteristics (same Fourier amplitude slope). They allow for the morph discrimination threshold to be calculated.

Morph discrimination threshold: value calculated from a given experimental series by fitting the psychometric function with the integral of a normal distribution.

Experiment: consists of seven complete experimental series, each one corresponding to a different Fourier amplitude slope. The seven corresponding morph discrimination thresholds are usually plotted to form a typical U-shaped graph.

Experimental conditions: refers to experiments performed using the four different morph sequences (man-to-woman, woman-to-man, car-to-bull and bull-to-car).

Chrominance: refers to the normalised push-pull signal given by a combination of two chromatic mechanisms (L vs. M in the case of *RG_chrom* and S vs. Lum the case of *BY_chrom*). It is different from "chromaticity".

Chapter 1.

C h a p t e r 1

INTRODUCTION

Background to this work

I have called this principle, by which each slight variation, if useful, is preserved, by the term of Natural Selection. Charles Darwin, On the Origin of Species (1859) ch. 3

1.1. Overview

This work focuses on the relationship between the statistical properties of natural scenes and the early stages of processing by the human visual system, with special attention to the adjustment (optimisation) of the second to the properties of the first. Chapter 1 provides a summarised review of the existing literature in the field, starting with the properties of the environment, the physiology of the HVS and the advantages of optimisation. Chapter 2 is centred on the spatial properties of foveal vision and its capacity to discriminate small changes in achromatic natural (and unnatural) scenes. The methodology here is based on psychophysical experiments and data analysis. Chapter 3 extends this analysis to peripheral vision under similar circumstances and using similar methodology. In Chapter 4, we try to explain the previous experimental results by means of a (very simple) computer model, based on plausible assumptions about the human visual system's early processing stages. In Chapter 5 we tackle issues of spatial colour vision and its relationship with the visual environment, focusing on the properties of coloured visual scenes that may have been crucial in determining the characteristics of colour vision in humans. Our starting point in that chapter is the

gathering of a database of calibrated natural scenes, to be analysed by computational methods.

This work borrows from several disciplines its research methods and techniques. Given this, it may be it beneficial for the reader to have a summary of the corresponding methods at the beginning of each individual chapter, instead of the classical approach of writing a single comprehensive "Methods Chapter", which may risk being too obscure and disconnected from the rest.

1.2. Prologue: environment, detectors and optimality

The hypothesis that the HVS (human visual system) is optimised to encoding natural scenes has become a tenet in vision research. But what does *optimisation* mean? In most cases, this term refers to the mathematical problem of optimising the assets of a function. Finding the solution for this problem is today a complex field with diverse applications in physics, engineering, computer science, cybernetics, economics, manufacturing and biology. In many cases, solutions are straightforward and familiar, and concern maximising (or minimising) the value of certain variable such us units of production, weight, temperature, etc. In biology, the solution may involve maximising fitness, or number of offsprings or caloric intake. This maximised (or minimised) quantity is usually a function of several variables subject to a set of constraints. Not all optimisation problems are clearly soluble or admit unique solutions and the characterisation of the function in question, its relevant variables, constraints etc. is usually a complex problem itself.

The mindless process of natural selection is capable of producing more intricate and exquisite designs than those produced by human skills. This process statistically

compares the small, inherited variants with respect to their effects on reproduction and allows new variants with higher-than-average fitness to supplant old variants (which were themselves fitter than those they supplanted) (Seger and Stubblefield 1996). A succession of such incremental substitutions can quickly improve the design, and eventually transform it radically (Dawkins 1986). *Genetic algorithm computing* (a method of computationally solving complex and otherwise difficult optimisation problems) is a burgeoning discipline in Computer Science that exploits the power of this process (Holland 1992).

The problem of "optimality" is, thus, central to all biological sciences. Within evolutionary ecology there is a substantial tradition treating natural selection as an optimising agent that produces maximal adeptness to the environment (Maynard Smith 1978; Richardson 1994). This approach (called "adaptationism") has been challenged from a variety of sources (Amundson 1994; Gould and Lewontin 1979), but remains the mainstream of evolutionary biology (Parker and Maynard Smith 1990; Seger and Stubblefield 1996).

Although it is natural selection, rather than the organism, that acts as "optimisation agent" (organisms do not "calculate" the optimal size to grow to or optimal calorific intake), the common approach to understanding the process is by taking an engineering view of the problem, imagining how to build a biological system given the environmental constraints. '*The optimisation approach to adaptation is not based on the assumption that organisms are "optimal" in any global or metaphysical sense, and in fact it often reveals that they are far less than "perfectly" adapted*' (Seger and Stubblefield 1996).

All biological organisms must live in, and interact with, the natural environment; and in many cases, the minute details of this interaction may determine whether an organism

will survive long enough to reproduce or not. The nature of this interaction varies depending on the physical properties of the environment and the characteristics of the other organisms that share the same surrounding space. This may take place in many ways, being commonly the case that organisms develop specialised parts of their bodies, some to detect physical perturbations and some to react to them (Morgan 1965). Many properties of the physical environment are likely to be perturbed (e.g. mechanical, electrical, magnetic, chemical perturbations, etc.) and may give rise to these specialised "sensors". However, developing sensors, although potentially beneficial, comes at some evolutionary cost, given that they consume energy, make the genetic coding more complex, are more likely to be exposed and damaged and, in many cases, need other specialised organs to function. Therefore, it makes sense to develop sensors that gather the most relevant type of information for the organism, are somehow protected from harmful effects of the environment, and require the least amount of supporting "hardware" to function.

Having a well-designed sensory system may provide an organism with evolutionary advantages over their competitors, helping it to find food, find the right mate or escape predators easier and quicker. One can imagine the properties of such a mechanism to be:

a) Cost effective. The biological advantages it supplies should outweigh the energetic cost of running it (Attwell and Laughlin 2001).

b) Reasonably protected from harmful effects of the environment (mechanical interactions, radiation, energy peaks of the physical perturbation it senses, etc.).

c) Reasonably noise-free. The signal-to-noise ratio should be large enough so that the information supplied by the sensor is useful.

d) Able to efficiently collect the most relevant information and filter out irrelevant data.

In other words, the sensory system has to be closely related to the environmental signal. The study of the properties of the sensory system (as a group of subsystems that acquire, communicate and represent information) has benefited greatly from the formulation of the theory of information (Shannon and Weaver 1949). By using the theory of information as an efficiency principle, (to assess the efficiency of the neural representation, assuming that efficient representation of information has some evolutionary advantages) formulated as an *optimisation problem*, one can produce *design principles* and predict neural processing (Atick 1992). In this optimisation problem, one must also consider the relevant activities (tasks) that the organism must perform in order for the sensory system to be biologically advantageous. Sensory systems tuned to the same physical property of the environment may have different characteristics according to the task that the organism wants to perform (e.g. it might be advantageous for a predator to have its sensors tuned to detect movement, etc.). There are many examples of ecological adaptation of the sensory systems to the visual task and the properties of the natural environment (Chiao *et al.* 2000b; Fasick and Robinson 2000; Lythgoe 1984; Lythgoe and Partridge 1989, 1991).

Sensory systems tend to be named according to the physical property that they sense, for example, the systems that are tuned to detect mechanical waves of air pressure (sound) are called *auditory senses*, the ones that are tuned to detect chemical perturbations are called *chemical senses*, etc. The sensory system that is tuned to detect

perturbations in the visible part of the electromagnetic radiation spectrum is called "visual system". To understand some of the properties of the visual system we first need to understand some of the properties of light.

1.3. Electromagnetic radiation detectors

1.3.1. Physical properties of light

The classical view of electromagnetic radiation as a perturbation (wave) that travels through space is illustrated in Figure 1.1.

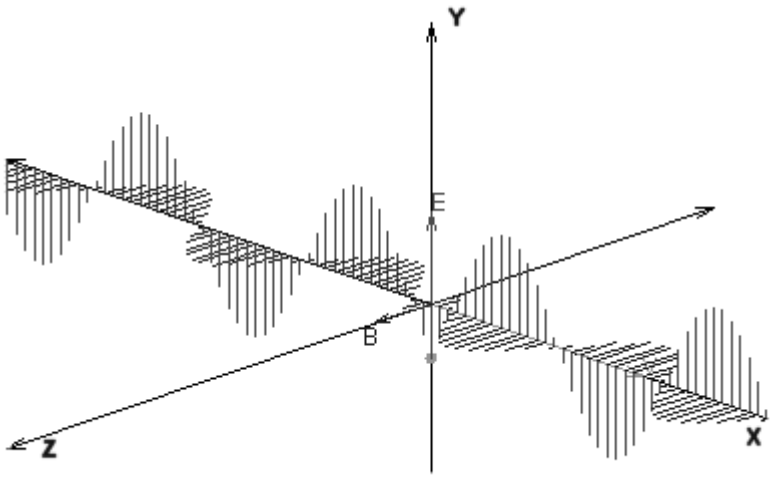

Figure 1.1: Classical view of an electromagnetic wave travelling through space.

This perturbation is generated by an electrically-charged particle that oscillates along the *y*-axis. Vector **E** (plotted on the *y*-axis) represents the electric field, vector **B** (plotted on the z-axis) is the magnetic field and the wave propagates along the *x*-axis. The frequency of the particle's oscillation determines the frequency of the propagating wave, and its energy. The direction of this oscillation in space determines the plane of polarization of the wave. The speed of the propagation is constant (the speed of light),

thus the wavelength of the oscillation is inversely related to its frequency, given that *speed = wavelength · frequency*.

This classical view is well suited for explaining the production of low frequency electromagnetic radiation. To explain the production of higher frequency electromagnetic waves we need to add quantum physics hypotheses, but the picture of the travelling wave of radiation holds in most cases. In the quantum view, an interaction between high frequency radiation and small objects or particles is represented like a hail of particles (called photons), each with energy proportional to the radiation's frequency. Electromagnetic radiation is classified according to its frequency/wavelength as shown in Figure 1.2:

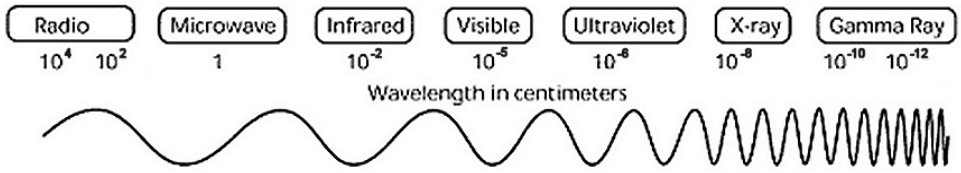

Figure 1.2: Classification of electromagnetic radiation according to its wavelength.

Organisms that possess any kind of visual system are sensitive only to a very small range of electromagnetic wavelengths (between 300-700 nm approx.). This is despite the fact that the electromagnetic energy reflected from most vegetation peaks outside these limits (in the infra-red or IR). For example, receptors found in the human visual system (HVS) are only sensitive to wavelengths between 400 and 700 nm consequently, this range is called the "visible spectrum" and this radiation is called "visible light". The biological reasons as to why this is the case are related to the materials from which all biological organisms are built. IR radiation produces a process

called *thermal isomerization*[1] of the photopigments in photoreceptors, which generates noise. Even low temperatures (like the body's temperature of 37°C) generate enough IR noise at 700 nm (Aho *et al.* 1988; Barlow 1957, 1988). Short-wavelength radiation, like the ultra-violet (UV) radiation, severely damages proteins and nucleic acids in all living cells, owing to a process called *photo-oxidation* (as a result of the interaction between oxygen and light). This constraint may not affect birds and insects (which have have a shorter lifespan than mammals and can thus tolerate the damaging effects of UV). Bennett and Cuthill (Bennett and Cuthill 1994) provided a review of the evidence of UV vision in birds and its function. There are other (physical) constraints to the use of UV-light by an animal. The degree of scattering of the light passing through a medium of small particles (that is, relatively small compared to the wavelength of the light, e.g. dust) is proportional to the inverse of the fourth power of the wavelength. Because of this, UV-light will be much more scattered than light of longer wavelengths, resulting in distant objects being indistinct in the UV. Imperfections in the optical media of the eye may also increase this UV-light scattering. These two limits (UV and IR) apparently determine the outer boundaries of the visible range that an animal can maximally cover. The advantages that an animal will have from having an UV or IR receptor have to be balanced against its ability to keep low the costs of repairing or compensating for damages (Neumeyer 1991).

1.3.2. Light and the visual environment

The properties of electromagnetic radiation from which relevant information about the world can be extracted are:

[1] An isomer is a chemical species with the same number and types of atoms as another chemical species, but possessing different properties. "Thermal isomerization" refers to this process occurring by effects of heat.

a) Amplitude of the wave or number of photons per unit of time (also called *radiance*).

b) Frequency of the wave or energy per photon.

c) Plane of polarization.

d) Direction of travel of the wave.

Natural light is often produced by a large number of charged particles that oscillate at different amplitudes, frequencies and directions in space (thus, it contains a mixture of wavelengths or photons with different energy, and different planes of polarization).

The Sun is an example of a natural light source that radiates across the whole of the electromagnetic spectrum (although its intensity varies with the wavelength considered). Most of its radiation is absorbed by the Earth's atmosphere before it reaches our visual environment. This absorption is not even across the spectrum, and some wavelengths are absorbed more than others. The interaction of light from the Sun and the Earth's atmosphere also produces other effects such as *refraction*, *diffraction* and *phase shift*. *Refraction* occurs when radiation goes from a medium to another and changes its direction of travel. It happens differently for different wavelengths. *Diffraction* occurs when a small object (the same order of magnitude in size as the wavelength of the radiation) is in the way of the electromagnetic wave. The radiation emerging from different sides of the object interfere so that the resultant energy patterns are not uniform. This is a consequence of the wave-like nature of light and determines the geometry of the smallest holes that light can enter without losing its information properties (primarily, it is the direction of the incident light that is lost during this process). The smallest hole that light can enter without suffering from this

phenomenon is called the "diffraction limit". The efficiency with which the atmosphere can scatter light and the direction in which the incident light is redistributed are dependent on all these effects.

The interaction of sunlight with air molecules in the atmosphere produces scattering of the radiation with lower wavelengths, which determines the wavelength distribution of the light coming from the sky. Scattered light reaching us from the sky contains more energy in the lower end of the visible spectrum than light coming directly from the sun. This in turn affects the composition of the light that illuminates objects, which is a mixture of direct light from the sun and indirect light from the sky. As a result, objects in the shade are illuminated by light with higher proportion of short wavelentghs. Shadows produced by direct light from the sun are "sharp" (the energy transition is steep) because all the electromagnetic waves come from a single, long-distance source. Shadows produced by indirect illumination from the sky (i.e. light coming from different directions) have "softer" edges, where the energy transition is smooth.

Small particles (e.g. dirt, cloud droplets and small ice crystals which are some 50 times larger than the light wavelengths) floating in the atmosphere scatter light of all wavelengths equally, thus producing an increase in the amount of light coming from all directions (diffuse light) and decreasing the amount of directional energy present. The overall effect is a decrease in the "contrast"[2] of the scene (e.g. fog).

Most of the light that reaches our eyes has not only interacted with the atmosphere but has also interacted with objects by means of reflection. Reflection alters some of the properties of light in various ways. It mainly changes the direction in which the light

[2] See Section 4.2.1 for a more detailed explanation of contrast in complex images.

travels, but it also attenuates the amplitude and alters the frequency composition and the plane of polarization of the light. Reflection produces light scattering, a process where the light is reflected with different degrees of likelihood in different directions. Some incident light is reflected at the interface (specular reflection – follows simple geometrical rules). The rest of the light enters the material and interacts with embedded particles. Part is absorbed and part re-emerges as reflected light (body reflection). The interface reflection is likely to be concentrated on one direction while body reflection emerges with almost equal likelihood in all directions. A reflection where most of the light is reflected by the interface is called "specular" and a reflection where most of the light is reflected by the body is called "diffuse" or "Lambertian". Most real surfaces show some degree of both of these types of reflection but the ratio varies greatly.

Objects do not reflect the same amount of light in every wavelength. They usually reflect (or inversely, absorb) more light at some wavelengths than at others. The spectral reflectance function of a surface is the property that describes the fraction of the light reflected as a function of wavelength. This gives objects their different "colours".

Since the HVS does not detect changes in the plane of polarization, we will not be concerned with this property here.

1.3.3. Biological light detectors and the physical properties of the environment

Almost every living thing (from animals and plants to single-celled organisms) is sensitive to light. It is possible to follow the evolution of photoreceptors, from simple cells towards more complex, specialised receptors (adept at detecting movement) through to human eyes. Primitive receptors may be distributed over the skin, but true

image-forming eyes started when receptors were arranged in groups, inside a depression (pit) (Lythgoe 1991). Environmental pressure can be seen as the driving force of this process; for example, primitive pits enhanced directional sensitivity and protected the receptors from the surrounding glare, although they made the receptors more susceptible to being blocked by foreign particles. Transparent membranes (which later evolved into crude lenses) were developed to keep these particles away from the pit (Cronly-Dillon 1991; Gregory 1998). The water from the sea-like environment was replaced by a fluid (aqueous humour). The crude membrane-lenses evolved from light-intensity-increasing devices towards image-forming devices, which focused the light on the (ever more complex) photoreceptors, etc.. (Gregory 1998).

There is a wealth of exciting literature on the evolution of the eye (Cronly-Dillon 1991; Gregory 1991; Land 1991), visual processing (Horridge 1991) and visual behaviour (Lythgoe 1991) involving several topics and disciplines, showing the clear physical and physiological constraints that shaped this aspect of evolution. For example, it is possible to show that the laws of optics govern the evolution of the eye by demonstrating the relationship between the dimensions of eyes, the geometry of receptors and the diffraction limit of light (Horridge 1991). The compromise met by the eye is to produce optics with the least possible blurring, and optimise the size of the receptors to match the optical resolution of the finest detail by the lens. This evolution is governed by the laws of optics while the evolution of the visual centres, which analyse the visual information, can follow along independently later. This is perhaps the reason why very sophisticated and complex eyes (e.g. like those found in medusae and worms) are not necessarily accompanied by a nervous system that can achieve much visual processing (Horridge 1991). While these sophisticated eyes may provide an evolutionary advantage for best detecting a moving shadow (which could

be an approaching predator or possible mate), for example, for the development of more sophisticated visual centres, a more complicated *visual task* is needed. We will come back to the subject of what are the possible visual tasks that may have driven the evolution of the HVS at the end of this chapter (Section 1.9) after exploring some of its properties.

1.3.4. The human eye

Figure 1.3 shows a cross section of the human eye (De Valois and De Valois 1990).

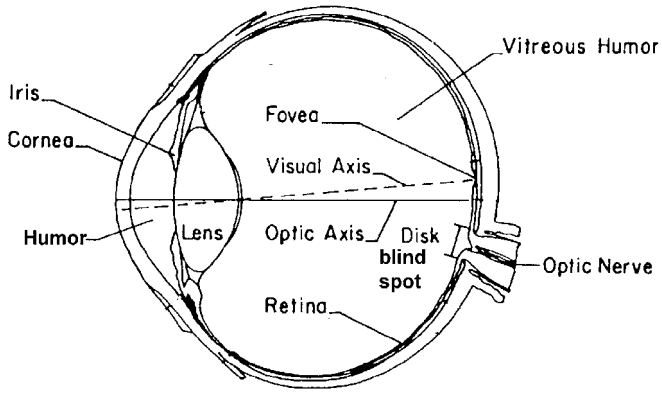

Figure 1.3: Schematic cross-section of the eye, (from De Valois and De Valois 1990).

Light enters the eye and passes through its optics (cornea and lens) before reaching the retina. Here is where photoreceptors (called *cones* and *rods*, according to their photosensitivity) convert the light into neural activity. The most common photoreceptor found in the retina is the rod (there are about 10^8 rods in the human retina). Rods are sensitive to low (scotopic) illumination levels and become saturated when exposed to moderate levels of light. Cones are less common ($6 \cdot 10^6$ cones in each retina) but are responsible for vision at higher (photopic) illumination intensity levels. The distribution of these photoreceptors is not even over the retina. Rods are

maximally concentrated at about 20° from the optic axis. Although they are found throughout the whole retina, cones are particularly concentrated over a very small area, where the visual axis meets the retina, called the *fovea*.

Figure 1.4 shows the spatial density distribution of rods and cones in the retina. The plot shows some interesting features:

a) The density of cones falls almost vertically near the limits of the fovea.

b) There is an absence of photoreceptors at about 17-19° from the visual axis, in the nasal side of the retina (temporal visual field). This area (circular in shape), which corresponds to where the ganglion cell axons (see below) exit the retina, is called the blind spot.

c) There is a big reduction in the density of rods in the fovea, and there is even a rod-free area.

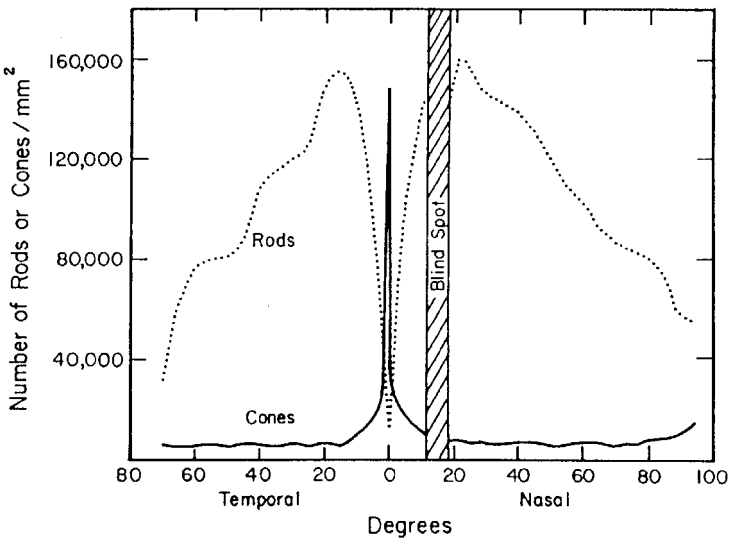

Figure 1.4: Distribution of rods and cones as a function of retinal eccentricity from the visual axis. (From De Valois and De Valois 1990).

The size of the fovea (which plays a disproportionately large role in visual perception) is about 1.5 mm in diameter or about 5.6° of the visual field (Wandell 1995). Cones are highly packed in this area, reaching densities of 160,000 cones/mm^2. The central part of the fovea is completely rod-free (size 0.5 mm, about 1.7°). The highest concentration of cones happens at the centre of the fovea, in an area that is also capillary-free, called the *foveola* (size 0.3 mm, about 1°) (Wandell 1995).

Although rods are much more numerous (and smaller) than cones and sample the retinal image more finely, their outputs are pooled into fewer neurons enhancing their sensitivity, but decreasing their fine level detail information. As was mentioned before, the interaction between light and small molecules (such as photopigments) can be modelled as a "hail" of particles. When this hail is sparse, the output of the photoreceptors suffers from noise (called "quantum noise") attributable to the discrete and probabilistic nature of this physical phenomenon. Spatial (and temporal) summation makes sense at scotopic illumination levels because it reduces the quantum noise inherent of very low illumination (De Valois and De Valois 1990).

Rods become saturated quickly (Norman and Werblin 1974) when illumination levels rise, so they are believed to be useful for nocturnal vision only and to play little role in everyday photopic vision. Figure 1.5 shows a scheme of the intensity-response functions of both rods and cones at different levels of adaptation (represented by broken lines A, B, C, etc..). For example, at adaptation level A, rods' sensitivity is higher and this will induce larger responses than those at adaptation level B, where rods are almost saturated. At higher (C) light levels, rods become completely saturated. Cones on the other hand, can adapt for a few seconds to higher light levels beyond C up to intensities that would bleach out all the photopigment molecules. When a cone

adapts to a light level (e.g. C) lights D and E would be indistinguishably bright (blocking the cones' synaptic release), but if the cone is allowed to adapt to level E, it will now respond to light around this level. This does not happen for rods, which cannot compensate for their compressed intensity response (limited dynamic range) (De Valois and De Valois 1990). Because of this, and because rods are absent from the central fovea (where most of the visual input comes from), our study will concentrate on the visual information supplied exclusively by cones.

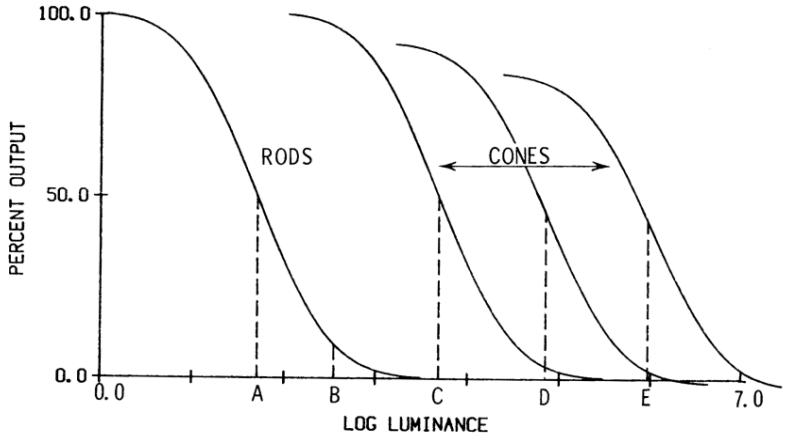

Figure 1.5: Scheme of the rod and cone intensity-level response for different adaptation levels (shown as A, B, C, D and E on the plot). Rods do not compensate for changes in the illumination levels and therefore do not "adapt" in the same way as cones do. Because of this, they show a compressed intensity-level function. (From De Valois and De Valois 1990).

1.3.5. Some characteristics of the photopic photoreceptors (cones)

Over 90% of the light entering the eye is not captured by any photoreceptor (Baylor 1987). Most of the light would just be absorbed by the layers of neural elements (and the arteries, veins and capillaries that nourish them) or simply enter the photoreceptors and exit out on the other side, before being absorbed by a black pigment layer called the *pigment epithelium*, behind the retina. The other 10% will be absorbed by one of the

photo-pigment molecules in the outer segment of the receptor (see Figure 1.6). The photo-pigment molecules absorb photons and produce chemical reactions that change the electric state of the cell, leading to changes in electrical potential at its synaptic ending (Morgan 1965). The likelihood that a photon is absorbed by a given photoreceptor depends on both the energy of the photon (or the light wavelength) and the type of photoreceptor involved. Once a photon is absorbed by a photo-pigment, the probability that the transduction succeeds is about 0.7 (Goldsmith 1991).

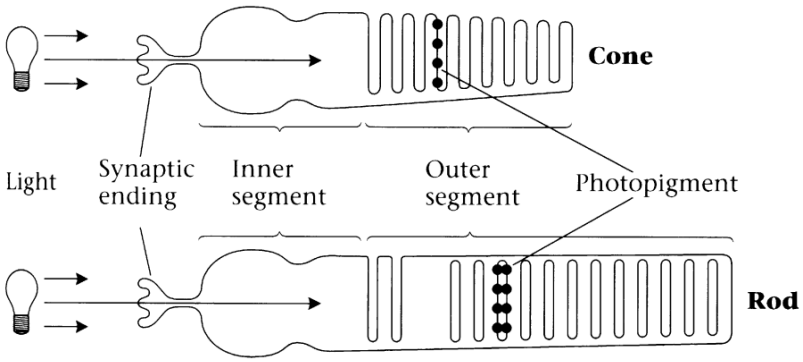

Figure 1.6: Scheme of mammalian rod and cone photoreceptors. (From Wandell 1995).

Photo-pigments are classified according to their photon absorption preferences. The absorption spectrum of a photo-pigment describes the relative probability that a photon of specific energy will be absorbed. According to this, there are many classes of photoreceptors, each with its characteristic absorption spectrum, which peaks at the most favoured wavelength.

The nature of photon absorption by photo-pigments present in the receptors means that once a photon has been absorbed, all information about the energy of this photon (or light wavelength) is lost. The receptor can only report to the central nervous system that a photon has been absorbed. This means that a moderate flux of photons near the

most favoured wavelength will produce the same rate of absorptions as a high flux at a less favoured wavelength (this is known as the *principle of univariance*) (Rushton 1964).

Three types of cone have been identified in the human retina. We name them according to their spectral response functions (L is the long-wavelength sensitive cone, M is the medium wavelength sensitive cone and S is the short-wavelength sensitive cone). Their physical size varies accordingly to foveal eccentricity (1 to 4 μ diameter in the fovea, 4 to 10 μ in the periphery) and they are not evenly distributed either. There are about 14 times more combined L and M cones than S cones in the retina (Wandell 1995). This is consistent with the strong blurring of the short-wavelength component of light due to the axial chromatic aberration of the lens. For example, the average number of S-cones in the fovea is 6-7 cones/degree of visual field (De Monasterio *et al.* 1985; Williams *et al.* 1981), much less than the average number of combined L- and M-cones of 120 cones/degree (Wandell 1995). The ratio of the number of L to M cones varies from one human observer to another. Table 1 shows some of the L/M ratios obtained and the different assumptions behind each measurement. There are very few S cones in the centre of the fovea, which, for this reason, is called S-cone blind spot (Williams *et al.* 1989; Williams *et al.* 1981). Their maximal incidence is within an annulus of 1° eccentricity (Ahnelt *et al.* 1987).

Method	Assumptions	Study	Mean ratio	Range of L/M cones ratios
Photopic luminosity function	Compares the relative numbers of L and M required to fit an individual or a standard flicker photometric function.	(De Vries 1947)		0.59 - 4.0
		(Vos and Walraven 1971)	2.0	
		(Walraven 1974)	2.0	
		(Smith and Pokorny 1975)	1.6	
		(Knau *et al.* 2001)		0.9 - 3.4
Spectral acuity	Compares visual acuities for heterochromatic gratings	(Brindley 1954)	<< 1.0	
		(Smith and Pokorny 1975)	<< 1.0	
Weber fractions	Compares the Weber fractions for Stiles's red and green mechanisms.	(Vos and Walraven 1971)	1.64	
Spectral sensitivity functions	Compares the relative heights of the spectral sensitivity functions of the cone primaries.	(Walraven 1974)	2.0	
Point source detection	Relates the detection steepness of the psychometric functions to the cone numbers contributing to detection.	(Cicerone and Nerger 1989)	2.07	1.46 - 2.36
		(Vimal *et al.* 1989)	2.8	1.6 - 4.0
		(Cicerone *et al.* 1994)	2.0	1.3 - 2.5
Human ERG	Compares the amplitudes of the isolated L- and M-cone ERG responses.	(Shapley and Brodie 1993)	>> 1.0	
Adaptive optics and retinal densitometry	Images of the arrangement of the L, M and S cone receptors in living human eyes.	(Roorda and Williams 1999)		1.15 - 3.79

Table 1: Comparisons between the mean values of L:M cone ratios produced by different workers, methods and assumptions. The table was based on the list compiled by A. Stockman and L.T. Sharpe at the Color and Vision Database: http://cvrl.ioo.ucl.ac.uk.

1.3.6. Determining the spectral sensitivity of the receptors

The spectral sensitivities of human cone photoreceptors were only measured after Rushton (Rushton 1962) developed a non-invasive technique called *retinal densitometry*[3] to study retinal photopigments in the 1960s. Retinal densitometry can determine the absorption spectrum for the L and M cones across the central part of the visible spectrum, but this technique loses accuracy in the short and long wavelength range. Another technique, called *microspectrophotometry* is based on isolating a single cone (after removing the retina from the eye) and illuminating it with monochromatic[4] light of known radiance. The transmitted light is measured and the difference between the incident and transmitted light is attributed to photopigment absorption. By repeating for many wavelengths, it is possible to determine the whole absorption spectrum for each type of cone. These measurements revealed that the absorption spectrum of the three L, M, and S pigments are surprisingly invariant across many species of Old World monkeys (Bowmaker and Mollon 1983; MacNichol *et al.* 1983) with maximum values at about 430, 535 and 565 nm. In humans, pigments for the L and M cones are similar to those of Old World monkeys (Bowmaker and Dartnall 1980; Dartnall *et al.* 1983), but S cones seem to have a pigment absorbing at shorter wavelengths with a maximum at about 420 nm (Bowmaker 1991).

[3] This technique allows the study of the absorption spectrum of one of the cone types in the fovea. For example, to study the M cones it is necessary to find a colour anomalous subject who has only M cones in the fovea (S cones are normally absent from the central fovea, so a person who is missing L cones will have only M cones in the fovea). The measurement consists of shining a monochromatic (single wavelength) light, of known radiance, onto the fovea and measuring the amount of light reflected back out. The radiance of the reflected light will be less than that of the incident light due to absorption by the photopigment. It is therefore possible to compute the relative absorption for that wavelength. By repeating the procedure for a large number of wavelengths, one can determine the absorption spectrum for the M cones. To find the absorption spectrum for the L cones, the procedure is repeated with a colour anomalous person who is missing the M cones.

[4] Light with energy concentrated on a single wavelength.

It was not until quite recently that human cone photo-pigments were available in quantities large enough (e.g. by reproducing each cone pigment using DNA cloning techniques (Merbs and Nathans 1992)) to define their properties with higher degree of certainty than those provided by retinal densitometry and microspectrophotometry. Under these new measurements, maximum absorption values were found to be 426 nm for the S pigment, 530 nm for the M pigment, and 557[5] nm for the L pigment. Before these techniques, their spectral properties were obtained indirectly, from psychophysical colour and brightness matching experiments and physiological measurements of cone photocurrents (Alpern and Pugh 1974). The first (psychophysical experiments) require the observer to match a monochromatic test light with combinations of three primary lights on a bipartite white screen and obtain a plot of the primary light intensity as a function of test light wavelength. To be able to obtain colour-matching functions, a few theoretical conditions must be met (Wandell 1995):

a) Linearity[6]: if any combination of lights $\mathbf{e}(e_1, e_2, e_3)$ matches light **t** and another combination $\mathbf{e}'(e'_1, e'_2, e'_3)$ matches light **t'**; then **t**+**t'** matches **e**+**e'** (Grassmann's additivity law) (Grassmann 1854). The colour-matching experiment linearly maps the physical stimulus to the primary intensities. Because of this property, it is possible to relate input to output with a linear matrix transformation: **e**=**C t**.

[5] There are actually two polymorphic variants of the L pigment in humans, the second being closer to the M pigment, with a maximum absorption value at 552 nm Bowmaker J.K. 1991. *Visual pigments, oil droplets and photoreceptors*. in The perception of colour. Vision and visual disfunction. J. R. Cronly-Dillon. 1991. The Macmillian press Ltd., London. 108-127, Merbs S.L. and Nathans J. 1992. *Absorption spectra of human cone photopigments*. Nature. 356, 433-435..

[6] A function that obeys superposition (*f(a)+f(b)=f(a+b)*) and homogeneity (*f(c.a)=c.f(a)*) is called a *linear* function.

b) The three primary lights should be visually independent, meaning that no additive mixture of two of the primary lights should be a visual match to the third.

c) "Negative" lights can be represented by shining one of the primaries on the test field (e.g. $e_1 + e_2 = e_3$ is the same as $e_1 = e_3 - e_2$)

There are many combinations of primary lights (all related by linear matrix transformations) that can produce colour-matching functions. In 1931, these various combinations of primaries were unified by the CIE (Commission Internationale d'Eclairage – an international standards organization) to produce the $\bar{x}(\lambda), \bar{y}(\lambda)$ and $\bar{z}(\lambda)$ standard colour matching functions (see Figure 1.7). There, the *x*-axis represents the wavelength of the test light and the *y*-axis represents the amounts of the primary lights (*tristimulus* values) that would be required to match it.

These functions are appropriate for many applications and have become an industrial standard, especially because they are non-negative and also because $\bar{y}(\lambda)$ is coincident with another CIE standard curve, the *spectral luminous efficiency* curve or V(λ), which is a set of weighting factors to convert radiometric values of monochromatic light into photometric units. The main drawback is that no set of physically realisable primary lights can yield these colour-matching functions. Inaccuracies in the short-wave region of these early estimates have been improved by Judd (Judd 1951). The CIE colour matching functions can objectively represent any spectral power distribution in terms of three values called X, Y and Z.

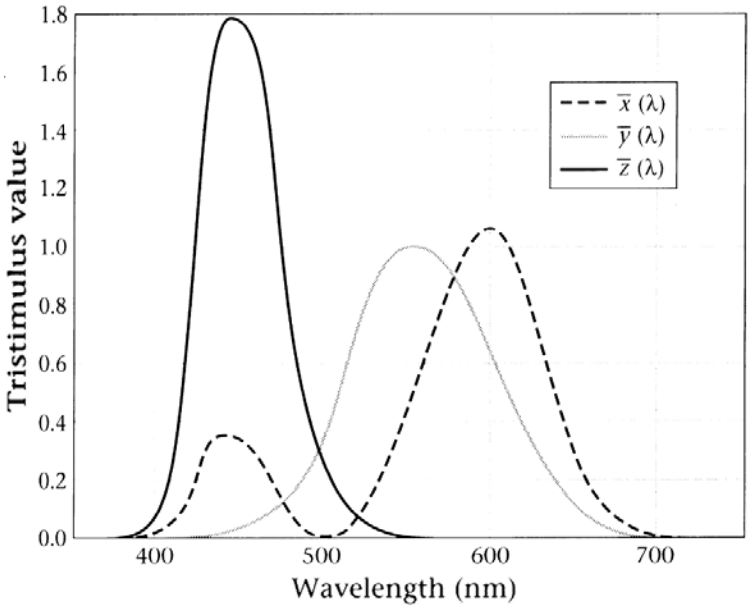

Figure 1.7: The CIE (1931) XYZ standard colour-matching functions, called $\bar{x}(\lambda), \bar{y}(\lambda)$ and $\bar{z}(\lambda)$. The *x*-axis represents the wavelength of the test light and the *y*-axis represents the amounts of the primary lights (*tristimulus* values) that would be required to match it.

It is possible to translate the tristimulus values from a set of primary lights (such as those defined by the CIE functions of Figure 1.7) to another by means of a linear matrix transformation. This also means that we can find a matrix transformation to convert the XYZ tristimulus values into the equivalent LMS (cone absorption curves) tristimulus values. This can be done by determining the nine coefficients of transformation in the **T** matrix:

$$\begin{bmatrix} S \\ L \\ M \end{bmatrix} = \begin{bmatrix} t_1 & t_2 & t_3 \\ t_4 & t_5 & t_6 \\ t_7 & t_8 & t_9 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

Equation 1.1

Where L, M and S are the tristimulus values (or absorption values) of the three classes of cones and X, Y and Z are the CIE (1930) tristimulus values. To solve Equation 1.1

we need to make certain assumptions (Travis 1991b) like, for example to suppose that *dichromacy*[7] is a reduced form of normal colour vision (*Helmholtz-König* hypothesis). A set of L, M and S cone fundamentals (called the Helmholtz-König cone fundamentals) can be deduced from this hypothesis by using co-punctal points (i.e. points that represent colours confused by the different types of dichromats). Other sets of cone fundamentals can be derived from other assumptions such as different colour matching functions, co-punctal points or different normalization constants to determine the relative heights of the functions (Travis 1991b).

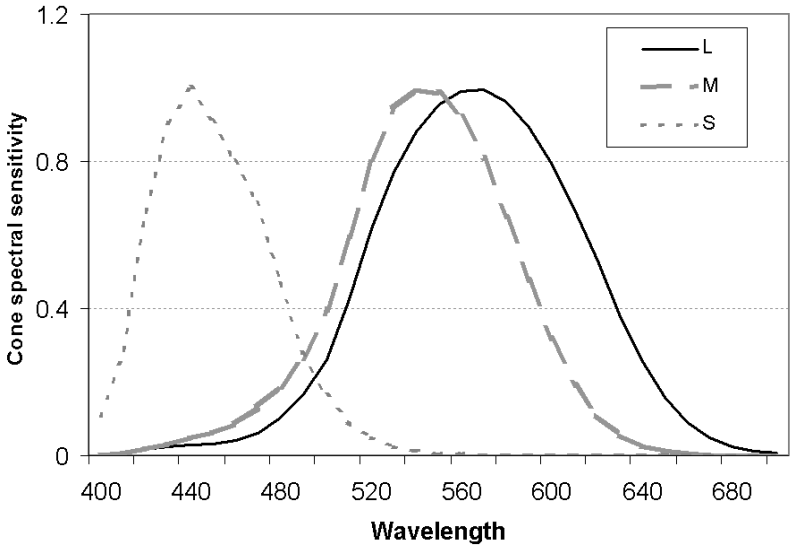

Figure 1.8: Smith and Pokorny's LMS (for long-, medium- and short-wavelength sensitive) cone sensitivity functions.

Smith and Pokorny (Smith and Pokorny 1972, 1975) were able to psychophysically derive the cone spectral sensitivity functions (see Figure 1.8) by measuring the contributions from one type of cone at a time. They used red-green dichromat subjects (people lacking either the L sensitive receptor or the M sensitive receptor) and the

[7] Dichromats are people who lack one of the three classes of cone photopigments. The remaining two classes are believed to be no different from those of people with normal colour vision.

hypothesis of *reduction dichromacy* (where only the central part of the fovea, which lacks short-wavelength receptors, is used). They showed that spectral sensitivities of these observers are similar in shape to the corresponding L or M visual pigment absorption spectra (Smith and Pokorny 1972). S-cone sensitivity here peaks at a different value from that measured for the S-photreceptor (420 nm), reflecting the filtering effect of the yellow lens (Xu *et al.* 1997).

The responses (photocurrents) of monkey cones to monochromatic light of different wavelengths were physiologically measured by Baylor *et al* (Baylor *et al.* 1987). They found cones with three different spectral response functions, with sensitivities varying over a factor of 10^6. These physiological measurements confirmed the psychophysical results obtained previously. After correcting for absorptions by the lens and inert pigments in the eye, it was also possible to derive from these three spectral response functions a system matrix for the colour matching experiments.

1.4. Post-receptoral and cortical mechanisms

Most of the neurophysiological vision research that concerns the present work was done on cats and macaques. Although the overall structure of these visual systems is thought to be similar to that of humans, there are some differences (e.g. cats have poor colour vision).

The output from retinal receptors does not travel directly to the brain. First, it is processed by an array of neurons in the retina, where it is transformed from simple measures of the number of photons to more complex spatio-chromatic information. Retinal neurons develop embryologically from the same cell lines that give rise to the brain, and because of that, the retina (which is part of the central nervous system) can

act on the visual information and transform it. This transformation is produced by the need for transmitting information through the optic nerve (Atick 1992). The optic nerve (see Figure 1.9) connects the retinal cells with the striate cortex (see below) and has two serious constraints:

a) Has to be flexible (so that we can move our eyes).

b) Like the rest of the body, is built of limited materials (has a transmission limit set by its biology).

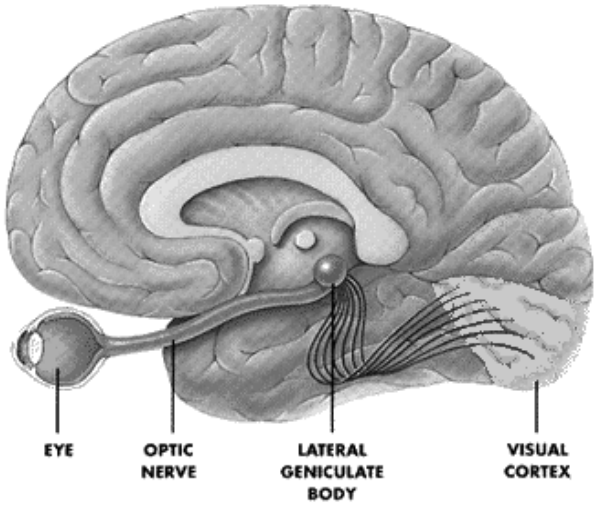

Figure 1.9: Scheme of the visual system showing the relative positions of the eye, LGN and visual cortex. Image publicly available from webvision: http://webvision.med.utah.edu).

Given that there are about 10^8 receptors in the retina and only about 10^6 optic nerve fibres, and the ambient light levels determine a receptor's dynamic range of about 10^6 (from moonless night to bright sunny day), the visual system has to somehow "compress" the amount of visual information so that its resources are maximised (Atick 1992; De Valois and De Valois 1990).

1.4.1. Spatial and spectral antagonism

The structure of post-receptoral retinal organisation contains layers (levels) of neurons. Each level does some kind of processing upon the visual information and passes its output to the next level. Neurons in the retina can be classified in two basic types of cells, those that conduct information towards the next level and those that conduct information laterally, within the same level. Figure 1.10 shows a scheme of the post-receptoral retinal organisation, with bipolar, horizontal, amacrine and ganglion cells. Bipolar cells are synaptically connected to photoreceptors, and can be divided in two classes: ON-bipolars and OFF-bipolars. ON-bipolars are excited (i.e. produce changes on their membrane potentials) when light falls over its subserving receptors and are inhibited (their membrane potentials are attenuated) when light falls over the surrounding receptors. OFF-bipolars have opposite behaviour. Horizontal cells are responsible for lateral connections within the retina. The number of receptors that a horizontal may contact varies between six (in the fovea) to 40 (in the periphery) (Boycott and Kolb 1973).

It has been proposed (Gouras 1991) that this array of antagonistic connections determines two complementary mechanisms, subserving each area of the visual space in parallel and responding in a push-pull way to increments or decrements of light. These two mechanisms (which signal increments or decrements of light intensity) provide a larger dynamic range and use less metabolic energy than a single sensor.

Most visual neurons (including receptors) receive antagonistic or opposing signals indirectly from neighbouring receptors through the retina. For any given neuron, the antagonistic influence tends to be strongest in the area surrounding the photoreceptors subserving it (Kuffler 1953). This phenomenon enhances a border's contrast formed

by gradients of light. For example, to obtain achromatic contrast the brain must somehow compare the outputs of a ON-bipolar to that of a neighbouring OFF-bipolar. To obtain chromatic contrast it must compare the output of cells that get excited by some wavelengths (e.g. long wavelengths) to that of cells that get excited by another wavelengths (e.g. medium-short wavelengths). In the fovea, cone bipolars are subserved by a single cone, whereas in the periphery, this system breaks down and bipolars are subserved by several cones. This arrangement reduces spatial resolution but does not appear to alter colour discrimination (Gouras 1991).

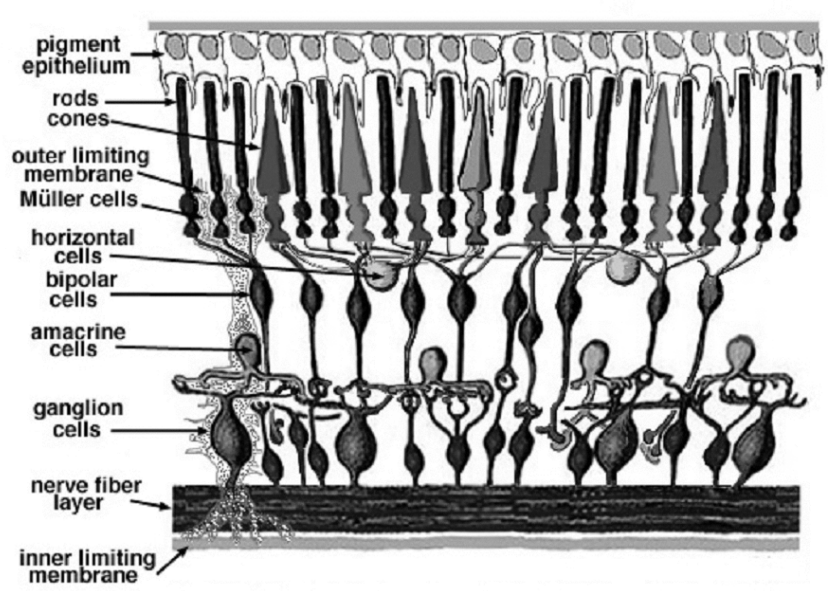

Figure 1.10: Scheme of the post-receptoral retinal connections, including bipolar, horizontal, amacrine and ganglion cells. (Kolb, Fernandez and Nelson, image publicly available from webvision: http://webvision.med.utah.edu).

Bipolar cells transmit visual information into the next layer, formed by amacrine and ganglion cells. Like the horizontal cells, amacrines make up for the lateral connections in the second layer. The axons of the ganglion cells provide the retinal output to the cortex (via a neuronal system called the *Lateral Geniculate Nucleus* (LGN)).

1.4.2. The M and P visual pathways and the Lateral Geniculate Nucleus

In addition to the mechanisms described above, there is evidence of a parallel system of ganglion cells throughout the retina. For example, the ratio of bipolar and ganglion cells to cones in the fovea is greater than two (Krebs and Krebs 1989; Schien 1988), and there are differences in the spectral sensitivity, conduction times (Gouras 1968) etc., of some ganglion cells. The parallel ganglion cell system (subserved by a separate set of bipolar cells that receive input from both L and M cones) does not show spectral opponency (Gouras 1991). Both types of cones feed the direct and antagonistic inputs to ganglion cells; therefore, the system is excited by a broad band of spectral wavelengths. Signals reach these spectrally broadband ganglion cells faster than chromatically opponent cells (Gouras 1968). The contribution of S-cones seems to be only to chromatic contrast (i.e. they do not contribute to achromatic contrast) (De Valois *et al.* 1966; Derrington *et al.* 1984; Derrington and Lennie 1984; Ingling and Martinez 1983). The connections between S cones and bipolars resemble that of L and M cones and bipolars in more peripheral areas of the retina. They receive input from S-cones and opponent signals from L and/or M cones.

Figure 1.9 shows a scheme of the connections between the retinal neurons, the LGN and the visual cortex. The LGN is the main route by which the retina communicates with the cerebral cortex in primates and it is separated in layers. Retinal axons synapse into specific layers of the LGN according to their function. For example, the output from fast (or *transient*) spectrally broadband ganglion cells synapses onto the *magnocellular* layers and all other axons (from slower or *sustained* ganglion cells) synapse in the *parvocellular* layers (Cleland *et al.* 1971; De Valois and De Valois 1990; Kulikowski and Tolhurst 1972; Tolhurst 1973).

Magnocellular and parvocellular layers segregate into two functionally different visual systems, which work in parallel in each local area of the retinal image (Derrington *et al.* 1984; Derrington and Lennie 1984; Gouras 1971; Wiesel and Hubel 1966). Magnocellular layers are more developed in nocturnal animals (Hassler 1966) and less fovealised (Chacko 1948) than parvocellular layers. The magnocellular system is crucial in the perception of movement (Schiller *et al.* 1989). It has a high sensitivity for achromatic contrast (Shapley *et al.* 1981) and lower spatial resolution than parvocellular cells (Derrington and Lennie 1984). There is also psychophysical evidence for presence of these sustained and transient mechanisms in human vision (Kulikowski and Tolhurst 1972; Tolhurst 1973).

There are four distinct types of LGN cells, subserving the L, M and S retinal receptors (in ON and OFF-centre arrangements) and the achromatic L+M system. The cells that subserve the L and M cones have extremely small receptive field[8] (RF) centres, especially in the fovea and surrounding areas. They usually subserve one L or M cone in the centre of their receptive field and receive antagonistic input from a larger area (including the surround of their receptive fields). The LGN cells that subserve the S cone mechanism have larger receptive field centres and only contribute to blue-yellow chromatic contrast. The spectrally broadband cells that subserve both the L and M cones (L+M) show weak or no cone opponency and must only contribute to achromatic contrast (Derrington *et al.* 1984; Gouras 1991). There is also evidence for a third thin layer of primate LGN cells below the magnocellular and parvocellular layers, called the *koniocellular* layer (Hendry and Reid 2000; White *et al.* 2001; Xu *et al.* 2002; Xu *et al.* 2001). This consists of smaller cells, which share some characteristics with other LGN cells, although there is evidence that their visual responses are modulated by

tactile and auditory stimuli (Irvin *et al.* 1986; Norton and Casagrande 1982; Norton *et al.* 1988), suggesting that their role is more to do with modulating rather than participating directly in spatial vision.

More recently, there has been evidence for some S cone input to other neural pathways such as the motion-related middle temporal (MT) area of the visual cortex (Seidemann *et al.* 1999). The consensus now seems to be that the S input to MT comes from a circuit originating in the koniocellular pathway (Hendry and Calkins 1998; Hendry and Reid 2000; Martin *et al.* 1997). A review on the physiology of the visual pathways associated with blue/yellow colour vision and links between the S cone and other visual channels is provided by Calkins (Calkins 2001).

Retinal connections account for only 10% of the LGN synapses. Nearly 60% of the LGN synapses are feedback signals from the cortex, and the remaining are connections from other parts of the brain, (Wandell 1995). At present, there is no clear indication of the role that these feedback signals play (they are assumed to somehow modulate the visual input from the retina).

1.4.3. Cortical Neurophysiology

In primates, the major part of the signal from the retina and the LGN arrives at a single area within the occipital lobe of the cortex called *area V1* or *primary visual cortex*. This area contains about $1.5 \cdot 10^8$ neurons, about 100 times more than the LGN or the number of optic nerve fibres. Although there are more than 20 other cortical areas that receive strong visual input, area V1 has been the subject of most of the study and it is by far the best-known cortical visual area (Wandell 1995).

[8] See a definition of receptive field later in Section 1.5.1.

V1 is a layered structure. Anatomists have identified six main layers (numbered 1 to 6, layer 1 being the most nearest to the surface) based on differences in the relative connections of neurons, axons and synapses. Layer 4 has been subdivided into several parts (labelled A, B and C) according to its interconnections with other areas of the brain. Magnocellular and parvocellular neurons of the LGN are connected to area 4C. Magnocellular neurons send their signals to the upper part of area 4C (labelled 4Cα) and Parvocellular neurons connect to the lower half (called 4Cβ) (De Valois and De Valois 1990; Wandell 1995). Layer 6 sends a large output back to the LGN (Lund *et al.* 1975).

Signals from the two eyes are initially segregated as they arrive in the cortex (layer 4C) and their organisation has been described early as forming "ocular dominance columns" (Hubel and Wiesel 1977).

Receptive fields in area V1 are qualitatively different from receptive fields in the LGN. While most LGN neurons have circularly symmetric receptive fields, most V1 neurons tend to respond better to stimuli moving in one direction than in the opposite (direction selectivity) (Hubel and Wiesel 1968). Simple cells' receptive fields differ from those of the retinal ganglion cells and LGN in the presence of excitatory and inhibitory regions that are not concentric and in their preference for stimuli of certain orientation (orientation selectivity) (Hubel and Wiesel 1959). These orientation selective neurons are found, in macaque, in layers 2 and 3, and are quite rare within layer 4C. The orientation selectivity of cortical cells has been explained in terms of summing the outputs from receptive fields of LGN cells that were in different (but aligned) spatial locations (Hubel and Wiesel 1962) as shown in Figure 1.11. The preferred orientation of neurones varies in an orderly way with the neuron's position in the cortical sheet

(De Valois and De Valois 1990; Hubel and Livingstone 1987; Hubel and Wiesel 1977). There are areas in the layers 2 and 3 of the visual cortex that lack orientation selectivity. They coincide with higher concentrations of an enzyme called *cytochrome oxidase* (Hubel and Livingstone 1987).

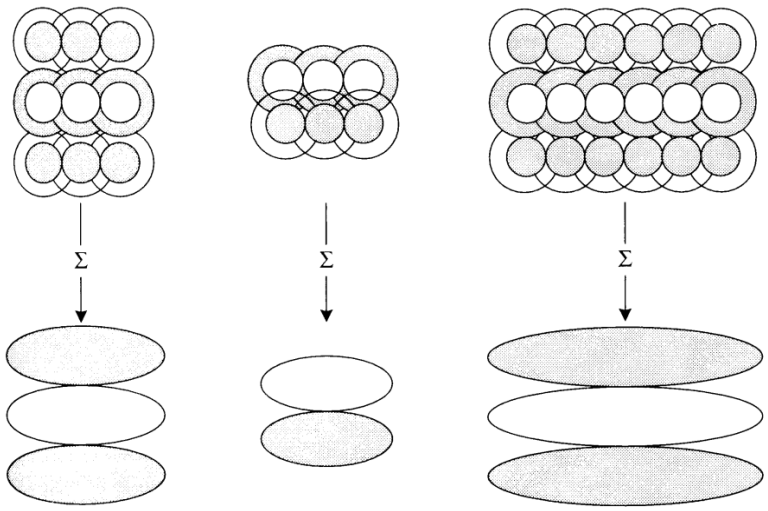

Figure 1.11: Cortical orientation-selective RFs can be created by the addition of responses from circular (non-oriented) RF such as those found in the retina and the LGN. (From Wandell 1995).

Lesions produced in the magnocellular and parvocellular pathways of the monkey have shown that performance deteriorates differently for different tasks (Merigan *et al.* 1991a; Merigan *et al.* 1991b; Schiller *et al.* 1989). These experiments demonstrate that the parvocellular pathway is concerned with a variety of tasks, such as colour discrimination and pattern detection. When magnocellular cells in the LGN are destroyed, visual performance is mostly unaffected, apart from tasks requiring the ability to code high temporal frequency flicker or movement information (Merigan *et al.* 1991a).

In summary, the fast (transient) magnocellular pathway is concerned with coarse achromatic (low-SF) visual information and movement information. The slow

(sustained) parvocellular pathway deals with fine achromatic (high-SF) and chromatic (low-SF) visual information. It responds weakly to movement and carries stereoscopic information (Gouras 1991). Given the characteristics of the magnocellular system, it is unlikely that its goal is to optimally encode the spatio-chromatic properties of the natural world. Instead, it seems more concerned with other (and equally important) survival priorities related to the detection of movement. Since the purpose of the present work is to investigate the possible optimisation of the HVS to encode the spatial and chromatic statistical information of natural scenes and not movement, we will concentrate on the properties of the parvocellular visual pathway.

1.5. Visual information coding by the parvocellular pathway

Contrast is a very important parameter for assessing vision. High contrast stimuli are usually presented in clinics (black letters on white background), but sinusoidal gratings are generally used to measure the ability of the eye and visual system to encode optical information at a wide range of spatial scales. There are several reasons why gratings became a popular type of stimuli:

a) Gratings can be made and calculated precisely in any frequency and contrast.

b) Gratings can be extended to any size, and any stimuli can be transformed into a set of component gratings and back (by a simple 2D-Fourier transform).

c) Because of the linear nature of the majority of cells in the retina and some 50% of the cortex (De Valois and De Valois 1990), it has been

argued that knowledge of grating sensitivity will give clues as to how we "see" any part of a scene (see Section 1.7, below).

The contrast of a grating is defined by:

$$C = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}}$$

Where L_{max} and L_{min} represent the maximum and minimum values of luminance and C (called the modulation, *Rayleigh* or *Michelson* contrast) ranges between 0 and 1. Since the luminance of a grating varies in a sinusoidal manner, we can alter the contrast of a grating without changing its average luminance. A very common psychophysical measure of optical and neural performance is obtained by asking observers to decide what contrast is necessary to render a grating just visible (Schade 1956). A plot of (contrast) sensitivity versus spatial frequency of the stimulus grating is called the spatial *contrast sensitivity function* (CSF).

1.5.1. The receptive field and gratings

The visual receptive field of a neuron is defined as the retinal area in which light influences the neuron's responses. It depends on the properties not only of the neuron in question, but also on the whole of the prior visual pathway, from the optics of the eye to the neuron itself.

Retinal ganglion cells responses have proven to be reasonably linear by satisfying the principle of superimposition (i.e. the resultant response of a neuron to a large stimulus is equal to the response of the centre plus the surrounding areas) (Enroth-Cugell and Pinto 1970; Enroth-Cugell and Robson 1966). These linear neurons have been named

X cells and their defining characteristic is the presence of a grating position on its receptive field which elicits null response from the neuron (Enroth-Cugell and Robson 1966). There is also evidence of the existence of non-linear ganglion cells (much less common, called *Y cells*) that do not share the same characteristics (Y cells would elicit a response to a grating at any phase) (Enroth-Cugell and Robson 1966).

Because the response of X cells to contrast patterns is linear, it is possible to map the stimulus contrast onto the change in firing rate of the neuron. Figure 1.12 shows the two-dimensional receptive field of a typical linear retinal ganglion neuron (e.g. an ON-centre). It was measured by shining small points of light on a cat's neuron receptive field and measuring its neural response (Kuffler 1953). Light areas of the figure denote retinal locations where light excites the neuron, dark areas show where light inhibits the neuron and grey areas represent retinal locations where light has little or no influence in the neuron's response. This centre-surround mechanism can be modelled as the difference between two Gaussian functions (Enroth-Cugell and Robson 1966; Rodieck 1965).

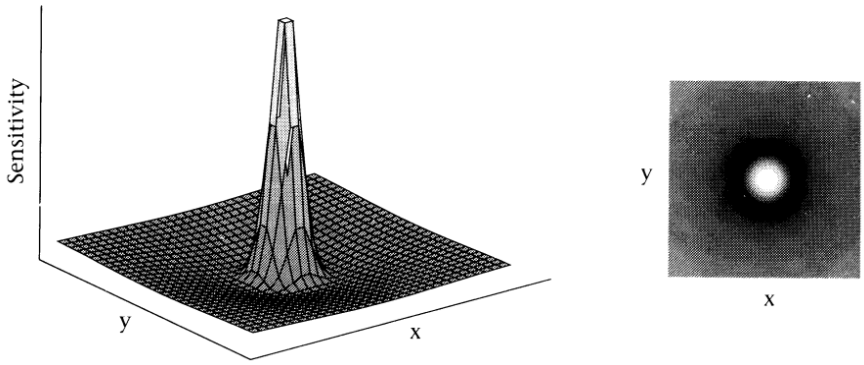

Figure 1.12: Two-dimensional scheme of a ON-centre, OFF-surround, retinal ganglion cell receptive field. Light areas of the figure show where light excites the neuron, dark areas show where light inhibits the neuron and grey areas represent retinal locations where light elicits little neuron response. (From Wandell 1995).

As mentioned before, an advantage of using gratings is that any 2-dimensional stimulus can be analysed in terms of its 2-dimensional Fourier components (sinusoidal and cosinusoidal contrast gratings). The neuron that is described by the receptive field in Figure 1.12 will be more sensitive to certain stimuli than others. For example, a sinusoidal pattern of light that illuminates the excitatory part and darkens the surround would elicit a maximum response. Gratings with other SFs would either stimulate the surroundings to produce antagonistic output (thus reducing the cell's total response) or weaken the effect of the centre. Neurons like these and others found in the visual cortex act in fact as *local* Fourier analysers, reacting more strongly to some SFs than to others. Figure 1.13 shows the contrast sensitivity function (CSF) of a neuron in the parvocellular layers of a monkey LGN (retinal ganglion cell receptive fields are indistinguishable from the corresponding LGN receptive fields). It was obtained by drifting grating patterns across the retina at a velocity such that each point on the retina saw 5.2 cycles of the pattern each second (Derrington and Lennie 1984). As a rule, low spatial frequencies (low-SF) represent broad areas in the image and high spatial frequencies (high-SF) contain information about fine detail in the image.

The first qualitative receptive-field map of neurons in the cat's visual cortex was produced by Hubel and Wiesel (Hubel and Wiesel 1959) using low-resolution stimuli (fairly large spots of light). Maps that are more detailed were obtained using oriented bars (Movshon *et al.* 1978b) and smaller spots of light (Jones and Palmer 1978b). Cortical receptive fields vary in the number of sub-regions they contain, their optimal *spatial phase* (relative position of the periodically recurring sequence of the grating) and elongation.

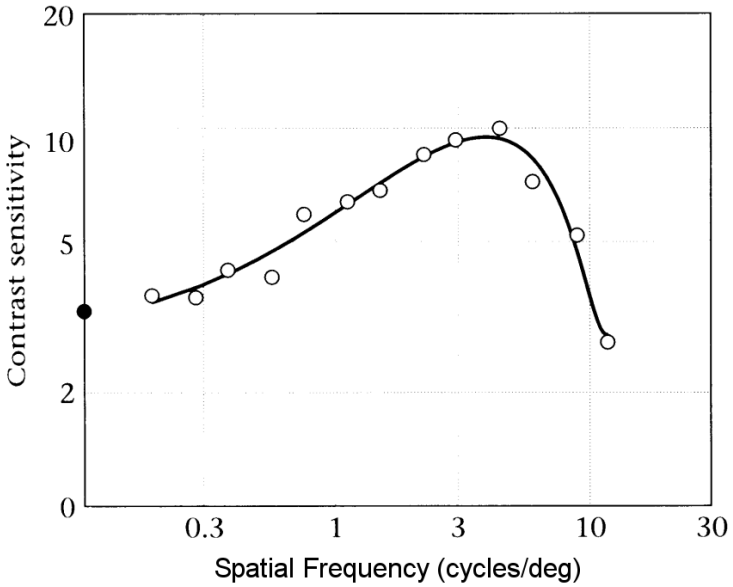

Figure 1.13: Contrast sensitivity measurements of a neuron in the parvocellular layers of the monkey LGN. (From Derrington and Lennie 1984).

1.5.2. Gabor models of receptive fields

Hubel and Wiesel (Hubel and Wiesel 1959, 1962) classified cortical neurons in area V1 according to the linearity of their responses into *simple* and *complex* cells. According to this classification, the optimal stimuli for simple (linear) cells could be predicted from the pattern of excitatory and inhibitory sub regions in the cell's receptive field. Simple cells do satisfy the homogeneity and superimposition conditions for linearity (while complex cells do not), therefore, their receptive fields can be measured using linear methods. Although the criteria used by Hubel and Wiesel for classifiying cortical simple and complex cells in terms of their linearity is functionally identical to those used by Enroth-Cugell and Robson for classifying X and Y cells, it should not follow that there is a parallel arrangement of linear and non-linear ganglion cells feeding into linear and non-linear cortical neurons.

The one-dimensional structure of simple-cells receptive fields (obtained using orientation bars) can be described by a one-dimensional function called "Gabor function" (introduced by the communications engineer Dennis Gabor in 1946) (Gabor 1946). This function consists of a localised sinusoidal variation (which is in fact the product of a sinusoidal function and a Gaussian function) that minimises the uncertainty in position and spatial frequency, and therefore provides a highly efficient encoding of arbitrary stimuli. The Gabor function can be generalised into 2-dimensional space by simply extending the Gaussian spatial envelope in the second dimension (see Figure 1.14).

From Figure 1.12 and Figure 1.14 it is possible to see that one of the main features of retinal and LGN receptor fields is their radial symmetry, while the cortical receptor fields are characterised by orientation-specific structure.

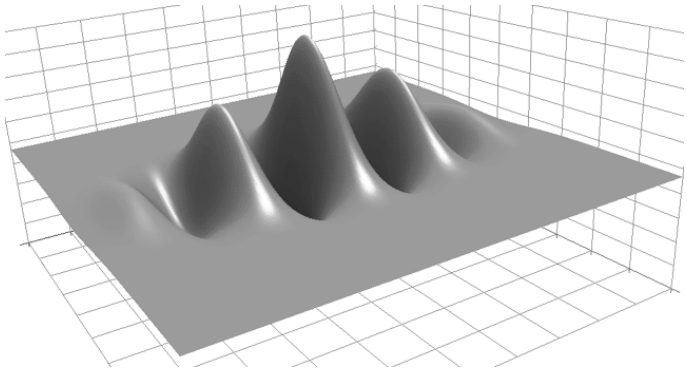

Figure 1.14: Representation of a Gabor function in 3-dimesional space. The Gaussian envelope can be tuned to give function the characteristic shape of simple-cells receptive fields.

Figure 1.15 (De Valois *et al.* 1982a) shows a plot of the CSF for six cortical neurons (both simple and complex) in area V1 of monkey. Their tunings is narrower than those of retinal ganglion cells (around 1.5 octaves) and they seem to vary in their preferred SF.

Simple cells with Gabor-like receptive fields are to a first approximation linear, i.e. the shape of the receptive field predicts some aspects of responses to gratings (Jones and Palmer 1978a, b; Movshon *et al.* 1978b; Ringach 2002). However, there are systematic mismatches in predictions (Deangelis *et al.* 1993; Tadmor and Tolhurst 1989) that may imply important non-linearities.

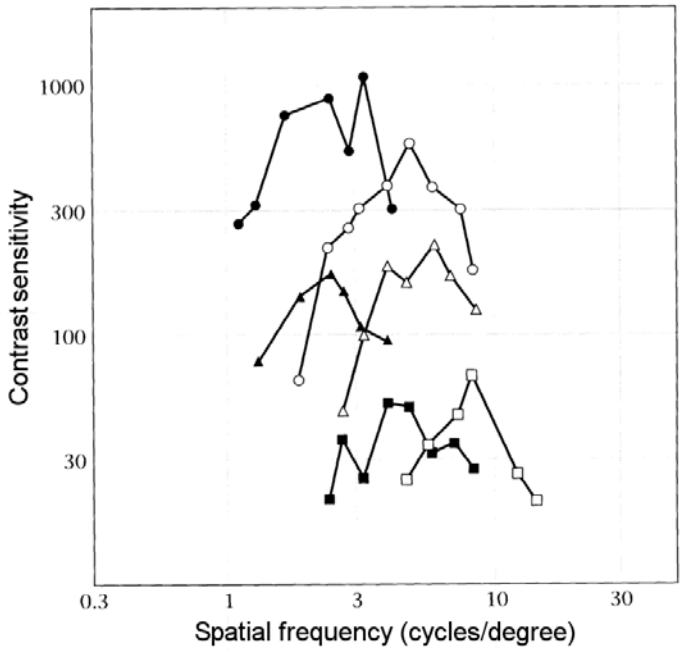

Figure 1.15: Contrast sensitivity of sex neurons in area V1 of monkey. (From De Valois *et al.* 1982a).

Gabor described a family of transforms that are capable of representing one-dimensional signals. These were later extended (Kulikowski *et al.* 1982) to transforms (also called wavelet transforms) with bandwidths increasing proportionally to SF and to two dimensions (Daugman 1985; Watson 1983). This has strong implications for modelling the processing of visual information in the cortex, since Gabor transforms can represent images with arrays of basis functions that are localised in both image and

SF space which (some authors claim) capture the intrinsic properties of the visual environment, such as scale invariance (see Section 1.8.6) (Field 1987, 1989, 1990).

1.5.3. The CSF as an envelope of numerous band-pass mechanisms

The properties of the CSF of individual ganglion neurons (Enroth-Cugell and Robson 1966) have been found to match behavioural CSF of the cat (Pasternak and Merigan 1981) for a variety of mean background intensities. This means that the information present at the output of individual neurons is available when the observer makes judgements on the presence of contrast patterns. However, single cortical neurons have narrower SF tuning curves than the overall CSF, suggesting (De Valois and De Valois 1990) that the later might represent the envelope of many narrowly tuned cortical mechanisms (or channels). Under this view, a "channel" is composed of all those cells with receptive fields that are identical in every respect except retinal location. Each of these channels is responsive to only one fraction of the total range of SFs present in the CSF and they work simultaneously (in parallel). Evidence for SF-channels comes from electrophysiological measures of striate cortex cells of macaque (De Valois *et al.* 1982a) and cat (Movshon *et al.* 1978a) as well as psychophysical adaptation experiments (Blakemore and Campbell 1969) and SF masking (Campbell and Robson 1968; Legge and Foley 1980). In these adaptation experiments, one can selectively modify the sensitivity of an observer to a certain spatial pattern (by adapting the observer to that pattern) without modifying the sensitivity to other spatial patterns (Blakemore and Campbell 1969). Figure 1.16 shows the spatial frequency adaptation effect (following adaptation to a low-SF) from one of such experiments (De Valois 1977b).

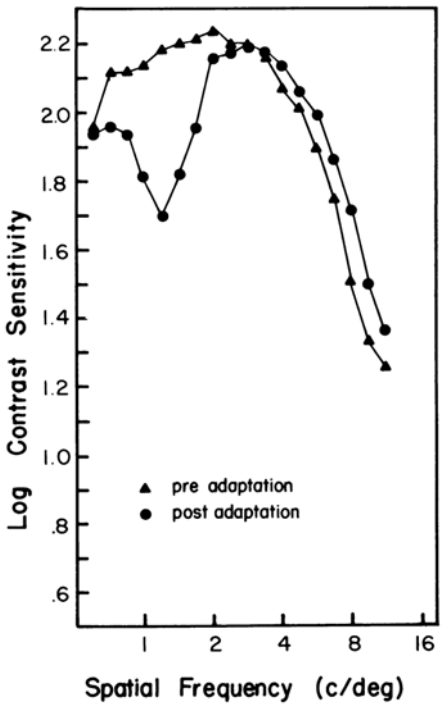

Figure 1.16: Spatial frequency adaptation effect, after adaptation to a low-SF. (From De Valois and De Valois 1990).

Careful measurements have revealed that SF adaptation may increase contrast sensitivity in other parts of the spectrum. This may indicate that SF channels are not truly independent and may have an inhibitory effect on each other (De Valois 1977b; De Valois and Tootell 1983; Tolhurst and Barfield 1978). Masking studies, where the detectability of a pattern is measured along and in the presence of another (masking) pattern have also provided psychophysical evidence for the presence of SF selective channels (Legge and Foley 1980).

Striate cortex recordings (De Valois *et al.* 1982a; Movshon *et al.* 1978a) have shown that cells, which are maximally sensitive to a certain spatial frequency, vary greatly in their SF-bandwidth tuning (and receptive field profiles). There is also activation of cells in response to increments in the contrast of gratings that do not match the cell's optimal SF tuning. This heterogeneity and variability in the properties of neurons may lead us

to review the idea of "rigid" channels and favour the image of channels being composed of whatever cells are contributing to the transmission of information with power at a particular SF (De Valois and De Valois 1990). However flexible this definition may be, SF channels have been successful in predicting the detectability of complex luminance (Blakemore and Campbell 1969; Campbell and Robson 1968; Graham *et al.* 1978) visual patterns and perceptual changes with adaptation (Blakemore and Campbell 1969). There is also evidence for the presence of multiple SF channels that are sensitive to pure colour patterns (Bradley *et al.* 1985; De Valois and Switkes 1983). The same logic applied to suprathreshold masking can be applied to produce a similar paradigm called *subthreshold summation*. Here the responses of two or more subthreshold stimuli can be added to produce a stimulus that would be detected by neurons within a channel. For this to happen, a channel needs to show linear summation within its SF-band. Gratings of closely related SF have been found (Sachs *et al.* 1971) to produce subthreshold summation (which does not happen with gratings of very different SF).

1.5.4. SF-channel bandwidth

SF-channels are orientation selective (adaptation to a horizontal grating has no effect on the detection of a vertical grating) (De Valois and De Valois 1990). Since the first neurons in the visual pathway to show narrow orientation tuning are those in the cortex (Hubel and Wiesel 1959, 1962), it is reasonable to accept that the psychophysically measured channels are cortical in origin. There is also considerable interocular transfer of the SF-adaptation effects, which implies that cortical neurons are involved (Blakemore and Campbell 1969) since the cortex is the first place where interocular summation occurs.

The bandwidth of the SF-channels has been estimated using different psychophysical (Blakemore and Campbell 1969; Legge and Foley 1980; Sachs *et al.* 1971) and physiological techniques (De Valois *et al.* 1982a; Movshon *et al.* 1978a). There is a degree of variation between these results, but most of the studies come to a rough agreement at the bandwidth (or the width at half-height of the Fourier peak) of SF-channels being 1 to 1.5 octaves in both cats and macaque (De Valois and De Valois 1990).

1.5.5. Spatial frequency interactions

Although most of the analysis of the visual system in terms of SF-processing is based on a linear systems approach (which is mathematically simpler), there are non-linearities in its responses that ought to be considered. For example, a given increment in the stimuli would produce a different resulting behaviour depending if it is above or below threshold. The non-linear response of cortical cells is evident in the different *apparent* contrast for stimuli with the same physical contrast and different SF at suprathreshold levels (Georgeson and Sullivan 1975) and the occurrence of half-wave rectification (Albrecht and De Valois 1981).

However, the non-linearities that concern us most are those related with interactions between different SF-channels: if the system's responses were truly linear, the various SF-channels would operate independently of each other. The first suggestion that this may not be the case came from (suprathreshold) adaptation experiments by Tolhurst (Tolhurst 1972). He compared the loss in contrast sensitivity resulting from adaptation to sinusoidal ratings and to square-wave gratings (which contain a larger variety of sinusoidal Fourier components) that had their fundamental frequencies equated. Adaptation to the square grating produced a smaller loss in contrast sensitivity

compared to the sinusoidal grating. Tolhurst suggested that this difference might be due to an *inhibitory* effect produced by the response of the other channels (not directly stimulated by the fundamental frequency) to sinusoidal Fourier components present in the square-wave function. These findings have been backed by physiological evidence of inhibitory effect of stimulus outside the receptive field of cortical cells (Blakemore and Tobin 1972; De Valois and Tootell 1983). A model of the mutual inhibition of striate cortex cells (effectively normalizing their responses with respect to stimulus contrast) has been presented by Heeger (Heeger 1992).

In summary, it is not right to think of the visual system as being truly linear, however, in certain circumstances (such as in threshold or just-suprathreshold contrast) its behaviour approximates that of a linear system.

1.5.6. Colour Opponency and the CSF

Psychophysical measurements of the sensitivity of human subjects to red-green-isoluminant (526 nm and 602 nm) and monochromatic (526 nm) gratings were made by Mullen in 1985 (Mullen 1985). These results are shown in Figure 1.17. Figure 1.18 shows a similar graph for blue-yellow (470 nm and 577 nm) isoluminant gratings. Both Figures display a band-pass (with respect to spatial frequencies) achromatic CSF and a low-pass chromatic CSF. The human colour and luminance CSFs have several characteristic features:

a) The presence of a low-SF decline in the achromatic CSF. This may be necessary for the visual system to filter out the low-SF luminance information, which results primarily from variations in the intensity of illumination (Mullen and Kingdom 1991).

b) The absence of a low-SF decline in the chromatic CSF. This may be related to the usefulness of detecting broad areas of homogeneous colour across variations in luminance (e.g. shadows) in natural images (Mollon 1989). An object of given chemical composition will have uniform "colour" but local luminance and shadows from its combined texture and shape.

c) Colour contrast *resolution* (related to sensitivity at the highest SF) is lower than luminance contrast resolution.

The absence of a metric for colour contrast (in the same way as there is a metric for luminance contrast) makes the comparison between these two curves difficult (e.g. a different choice of colour pairs would cause the whole of the chromatic CSFs to change). This problem can be overcome by specifying stimulus contrast by means of its physiological effectiveness (Mullen and Kingdom 1991), for example, using cone responses to luminance and colour contrast as a common measure to compare colour and luminance contrast sensitivity (Mullen 1985).

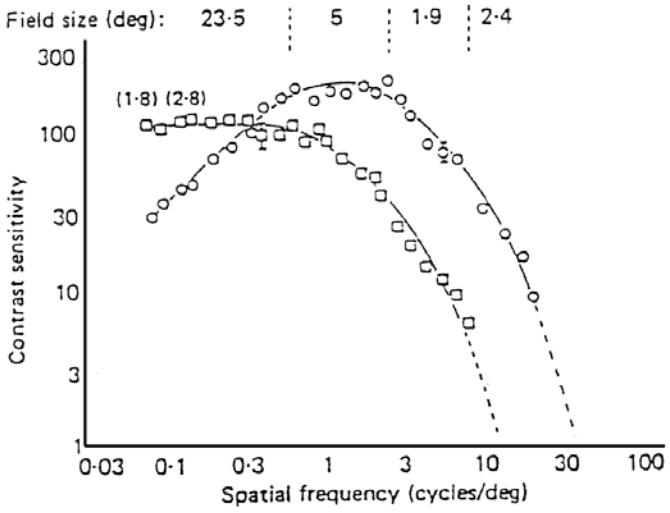

Figure 1.17: Contrast sensitivity functions (CSFs) for the red-green (squares) opponent chromatic system and the luminance (circles) system. (From Mullen 1985).

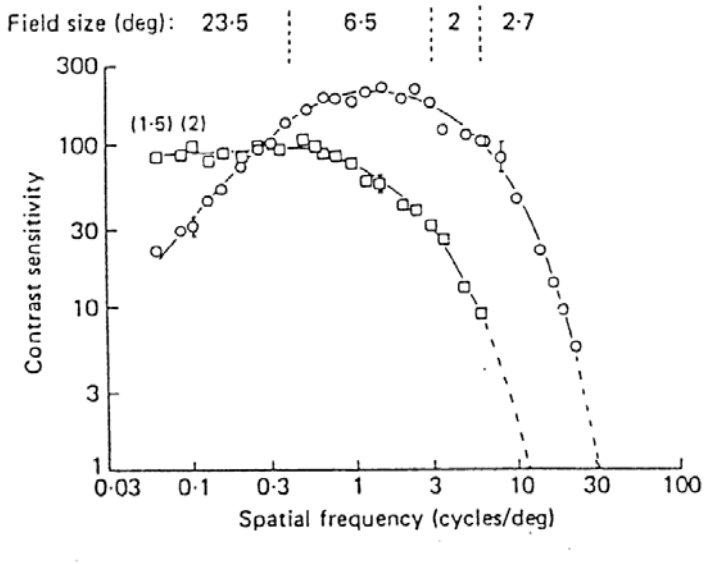

Figure 1.18: CSF for the blue-yellow (squares) opponent chromatic system and the luminance (circles) system. (From Mullen 1985).

One can also assume that these systems adapt to the dynamic range of normal stimulation and compare them in a relative manner (e.g. by normalising the areas below the curves, etc.).

The CSF clearly shows some of the limits to the performance of colour vision (e.g. in terms of spatial resolution). These limitations may have their origin in the difficulties of representing both chromatic and achromatic information in the parvocellular cells at the level of the early-post receptoral mechanisms. For example, the need to combine both spectral and spatial opponency in such cells leads to a system that is capable of carrying two or more types of information simultaneously (this is referred as *multiplexing*). There is evidence that multiplexing is done by opponent units, that act as a single mechanism carrying both luminance and chromatic information (Derrington *et al.* 1984; Ingling and Martinez Uriegas 1983; Ingling and Martinez 1983; Thorell *et al.* 1984). Demultiplexing mechanisms (to separate the luminance and colour signal in the

cortex) have been proposed by several workers (D'Zimura and Lennie 1986; Ingling and Martinez 1983; Martinez-Uriegas 1991; Mullen and Kingdom 1991).

Receptoral and post-receptoral mechanisms play an important role in setting the maximum achromatic and chromatic resolution. As seen before, chromatic contrast depends on the differences in the spectral composition of light while achromatic contrast depends on differences in the spatial distribution of energy of light. These two distributions can change independently of each other across borders (e.g. dappled shadows in the forest). At the *receptoral level* (see Figure 1.19), achromatic contrast can make use of the maximum resolution of the photoreceptor mosaic while chromatic contrast has to be obtained after a comparison between at least two analogous (localised) groups of cones in neighbouring areas of the visual space (Gouras 1991). This is because of the principle of univariance (see Section 1.3) (Rushton 1964), which states that a single cone alone cannot distinguish the wavelength of the light falling upon it. Chromatic response requires differencing of the outputs of at least two different types of cone photoreceptors and this requires doubling the summation area. At the *post-receptoral level* (see Figure 1.19b), this chromatic resolution may be further reduced by the need to obtain a differentiating signal from both centre and surround receptive fields. Another likely explanation as to why chromatic contrast sensitivity is restricted to larger areas (low spatial frequencies) is that signals from a receptor's output *difference* are likely to be smaller than from a receptor's output *addition* (especially given the close spectral overlap of the L and M receptors). Integration of the chromatic response across a larger area may provide a way of strengthening the weaker chromatic signal (Mullen and Kingdom 1991).

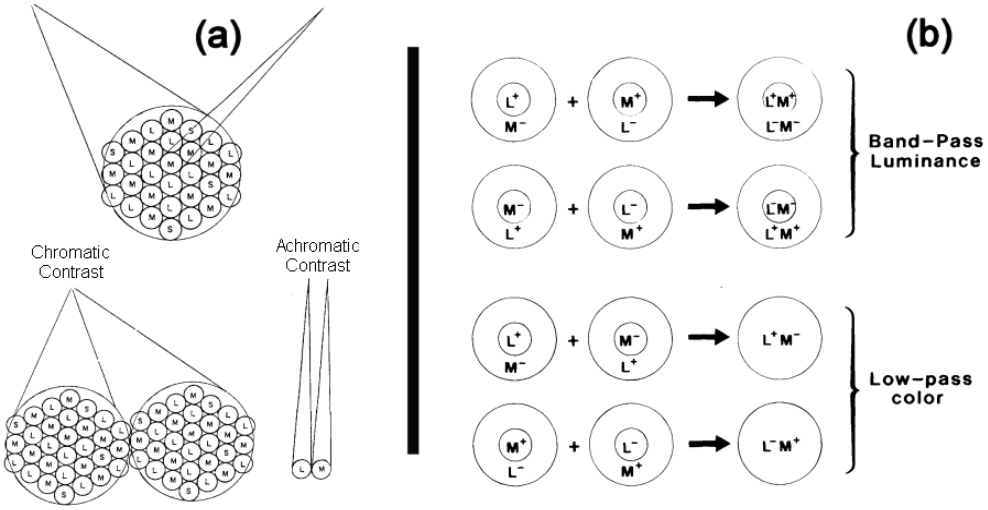

Figure 1.19: (a) Receptoral and (b) post-receptoral chromatic mechanisms. Section (a) shows how achromatic contrast can make use of the ultimately resolution of the photoreceptor mosaic while chromatic contrast occurs after a comparison between two analogous (localised) groups of cones in neighbouring areas of the visual space. Section (b) shows how chromatic resolution may be further reduced by the need to obtain a differentiating signal from both centre and surround receptive fields. (a) obtained from Gouras 1991 and (b) from Mullen and Kingdom 1991.

Physiological measures of the output of neurons in the parvocellular layer of the LGN (De Valois *et al.* 1966; Derrington *et al.* 1984; Wiesel and Hubel 1966) of macaque show cells whose output clearly depends on the wavelength of light (i.e. they are involved in colour processing). Derrington *et al* (Derrington *et al.* 1984) distinguished between two groups of these cells. One of these groups receives opposed, but not equally balanced inputs from L and M cones and they were labelled L-M cells. The other group of cells receives input from S cones almost equally opposed to a combined input from L+M cones. These were labelled S-(L+M). A second study (Derrington and Lennie 1984) reports relatively high sensitivity for achromatic contrast and higher temporal frequency tuning in magnocellular units than in parvocellular units of the LGN. These measurements and others helped to establish the physiological basis for a definition of colour contrast. Once the cone contributions to the post-receptoral mechanisms are

known, then the effectiveness of colour modulation in activating that mechanism can be established (Mullen and Kingdom 1991).

There is not general agreement on the relative cone weightings required to model the L-M opponent response. Many such types of weighting (based on different criteria) have been used, e.g. L-M (Tansley and Boynton 1976), 0.96L-1.28M (Guth *et al.* 1980; Ingling and Martinez Uriegas 1983), L-2M (Eskew and Boynton 1987; Frome *et al.* 1981), 0.8L-M (Stroemeyer *et al.* 1983). Other views (Stroemeyer *et al.* 1985) suggest that once the effects of visual adaptation within the cones (Weber's law) are taken into account, the contribution of the L and M receptors is nearly equal. Since there is more than one opponent mechanism, it is necessary to account for the interactions between all of these before a method becomes generalised (Mullen and Kingdom 1991).

There have been many studies on the influence of the relative densities of L and M cones in the retina (which varies from one individual to another (Dartnall *et al.* 1983)) on the human CSF (Ingling and Tsou 1988; Kremers *et al.* 2000; Miyahara *et al.* 1998). These findings suggest that the sensitivity of the luminance channel is directly related to the relative densities of the L and the M cones and that the red-green chromatic channel introduces a gain adjustment to compensate for differences in L and M cone signal strength. Gunther and Dobkins (Gunther and Dobkins 2002) hypothesise that the relative ratio of L versus M cones in the eye should influence red-green chromatic contrast sensitivity in the same way it influences spectral sensitivity (also called "luminous efficiency function" or $V(\lambda)$). They found a correlation between variations in the chromatic CSF across subjects and the relative number of L- versus M- cones inputting to the L-M mechanism.

There have been reports (based on electrophysiological measures) of a shift away from colour opponent axes in V1 (Lennie *et al.* 1990). Chromatically sensitive cortical neurons have also been found to have larger receptive fields and slower temporal responses than LGN neurons (Ingling and Martinez Uriegas 1983; Kulikowski and Walsh 1993).

1.5.7. Visual adaptation

In a typical day, the range of absolute intensities that the visual system may experience exceeds six orders of magnitude (De Valois and De Valois 1990). However, the range of contrasts that we encode in a typical image, from the least contrast that we can detect to 100% contrast is no more than 2 orders of magnitude (Tadmor and Tolhurst 2000). This means that the visual pathways need to have some non-linearities build in to adapt their dynamic range to such changes. Indeed, their neuronal and behavioural responses change as a function of the mean background intensity (this is called *visual adaptation*). The system behaves non-linearly at a global level, but locally, it can be approximated to a linear system, by measuring with respect to local contrast, instead of absolute intensity (Wandell 1995) (see Figure 1.5).

An example of the effects of visual adaptation is the dependence of the CSF of a neuron on the mean intensity level of the background. At low background intensities, the CSF shows little band-pass behaviour (poor response to high-SF) showing that there is little effect of its inhibitory surround. At higher background intensities, the shape of the function changes, making the reduced sensitivity to low-SF more significant (Enroth-Cugell and Robson 1966).

Another clear example of the visual adaptation effects on the response of the visual pathways is the change of the threshold sensitivity as a function of the background intensity. If we plot the threshold intensity of an incremental test flash required to elicit a criterion peak-firing rate on a retinal ganglion cell as a function of the background intensity, it follows a linear function (in *log-log* coordinates) with slope close to 1 (Enroth-Cugell *et al.* 1977). This type of relationship (very common and first discovered from measurements of human behaviour) is called *Weber's law*[9].

1.6. Peripheral vision

Up to this point, we have described some of the properties of the visual system where most of its hardware and computational resources are concentrated, i.e. the fovea (although the fovea occupies a tiny fraction of the retina, its processing occupies 25% of V1). Vision outside the fovea does not share the quality of foveal vision. For example, the task of extracting meaning out of written sentences becomes significantly difficult as soon as we move outside the fovea (Latham and Whitaker 1996). This difference in the quality of form vision inside and outside the fovea may originate from a combination of both lack of neural resources (i.e. image undersampling) and under-processing of the visual information (Kelly 1984; Rovamo and Virsu 1979; Rovamo *et al.* 1978; Virsu and Rovamo 1979).

As we shift the stimuli from the fovea into the periphery (see Figure 1.20), the overall CSF decreases and both, the CSF peak and the high-SF falloff move towards lower SFs. Low-SF attenuations become less apparent in the case of luminance-CSFs

[9] Weber's law states that the change in stimulus necessary to elicit a given response in a system is proportional to the initial state of the system. It is generally enunciated as $\Delta I/I$=constant. This relationship becomes a line with slope equal to one when plotted in *log-log* co-ordinates. Although very common, in most biological systems this law has proven to be only a rough approximation.

(Rovamo *et al.* 1978). The precise explanation of these effects is not fully known, since results largely depend on the size of the stimulus grating. For example, if the peripheral patterns are made increasingly large, (to compensate for the presumed retinal lack of ganglion cells), the peak CSF shifts to lower frequencies with eccentricity, while the contrast sensitivity to lower SF increases in absolute terms (Kelly 1984; Rovamo *et al.* 1978). This later effect could be a consequence of probability summation (see below) since the extent of the stimuli increases with eccentricity.

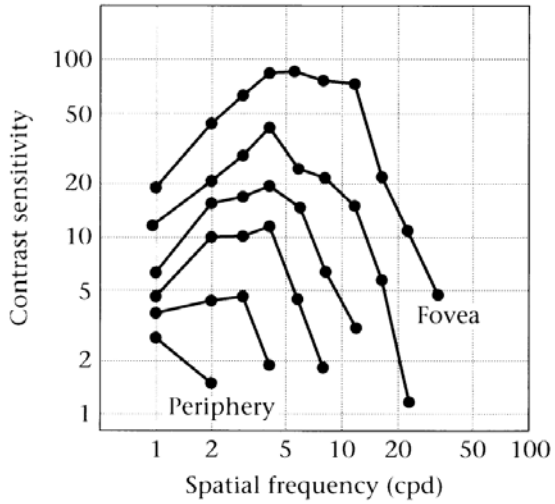

Figure 1.20: Variation of the CSF with retinal eccentricity. Curves correspond to 0, 1.5, 4, 7.5, 14 and 30 deg. From Rovamo *et al.* 1978.

Evidence (Movshon *et al.* 1978a; Robson and Graham 1981) suggests that every small region of the visual space is analysed by several channels tuned to a variety of SFs. Sensitivity to high-SF tends to drop out in peripheral channels, but the entire range of SF is detected in the fovea itself. However, the loss of spatial frequency sensitivity (acuity) may be only one of the contributing causes to the degradation of form vision in the periphery. Hyperacuity tasks (e.g. to discriminate between targets at different spatial *positions*) have revealed some qualitative differences between the processing of

foveal and peripheral stimuli. These may be related to the lack of receptors and ganglion cells (undersampling) (Levi and Klein 1986), the uncalibrated disarray of cortical connections (Hess and Field 1993), loss of sensors at some phase angles (Bennett and Banks 1987) and to the topological map of connections from the retina to the cortex (Levi *et al.* 1985; Westheimer 1982).

The degree of fall in acuity with eccentricity cannot be explained from the spacing between photoreceptors (which determined the minimum resolvable angle) (Drasdo 1977) or the quality of the eye's optics (Jennings and Charman 1981). It was then hypothesised that this should be related to the amount of neural convergence between the photoreceptors and ganglion cells (or ganglion cell density) (Brindley and Lewin 1968). In its turn, ganglion density (*D* -number of cells per solid degree in the retina) was believed proportional to the cortical representation in the striate area. To relate changes in the quality of vision with the amount of cortical representation of the corresponding retinal space, a compensating factor (called *cortical magnification factor* or M-factor) was introduced (Daniel and Whitteridge 1961). The M-factor represents the linear distance in mm of the cortical projection on the striate area that corresponds to 1 deg of visual space. In order to estimate the M-factor psychophysically, a few assumptions must be made:

a) Ganglion cell density (*D*) is proportional to the square of the M-factor.

b) Visual acuity is directly proportional to the M-factor

c) Cortical measurements of cortical striate projections made in the macaque monkey cortex are also representative of cell projections in humans.

There is not a complete agreement on the validity of these assumptions. Many researchers claim that there is a relative expansion of the foveal projection at the level of the striate cortex, bigger than that predicted from the ganglion cell density (Azzopardi and Cowey 1993; Dow *et al.* 1981; Van Essen *et al.* 1984).

Cowey and Rolls (Cowey and Rolls 1974) found a correlation between the reciprocal value of M and visual acuity (minimum angle of resolution). They later estimated, based on this correlation, the central value of the M-factor from measurements made within 1.5-35 deg. Tolhurst and Ling (Tolhurst and Ling 1988) reviewed the evidence relating to the organisation of the visual cortex to determine whether this is organised in a similar way to the macaque monkey. They concluded that it is reasonable to accept that both species should have similar dependencies of the magnification factor upon eccentricity. Human magnification factors were estimated to be 1.6 times greater than those of the macaque (Brindley and Lewin 1968; Dobelle *et al.* 1979). Following the evidence from a number of studies, (Daniel and Whitteridge 1961; Dow *et al.* 1981; Hubel and Wiesel 1974; Tootell *et al.* 1982; Van Essen *et al.* 1984) Tolhurst and Ling (Tolhurst and Ling 1988) estimated the foveal value of M to be between 20-25 mm/degree. Their calculations of M-factor values are based on a modification of the equation provided by Tootell *et al* (Tootell *et al.* 1982) adapted for the central 10° of the human visual field. This estimate accounts for a fall of M by a factor of 12-18 over 10° eccentricity.

Rovamo and Virsu (Rovamo and Virsu 1979) obtained a set of four equations (one for each visual field half-meridian) to compute the value of M. Their computations were based on successive approximations so that a simple set of equations would give values

of D in agreement with the total number of retinal ganglion cells and the density of the centremost cones.

Since performance in visual tasks is generally worse in the periphery, an increase in the size of the stimuli given by the M-factor should produce results similar to those found using foveal vision. The use of the M-factor (or M-scaling) has been successful in compensating for the fall of performance with eccentricity in a number of visual tasks (e.g. simple visual acuity) (Rovamo and Virsu 1979; Rovamo *et al.* 1978) but it has been less able to account for the fall in performance in others (e.g. vernier acuity or stereo-acuity). Figure 1.21 illustrates the fact that hyperacuity declines with eccentricity more steeply than visual acuity and a single M factor cannot explain all thresholds. Estimates of the decrease in performance for grating resolution tasks show a fall by a factor of 4-6 (Cowey and Rolls 1974; Rovamo and Virsu 1979; Rovamo *et al.* 1978). Vernier acuity falls by a factor of 10-16 (Westheimer 1982) and stereo acuity falls by a factor of 13 (Fendick and Westheimer 1983) in 10 degrees of eccentricity. The M scaling factor adopted in this thesis is consistent with the later values.

The deterioration of contrast coding in the periphery can be explained by the changes in the CSF, but the greater extent of deterioration of positional coding is more difficult to explain. Bennett and Banks (Bennett and Banks 1987) have suggested that *'only phase shifts that alter contrast should be discriminated in the periphery'*, thus relating shift changes with changes in the *local contrast*. This relationship between simple local contrast and phase coding was also investigated by Bradcock ((Badcock 1984a, b), who concluded that observers may not discriminate relative phase directly, but instead they may find differences in the local changes of contrast in the stimuli.

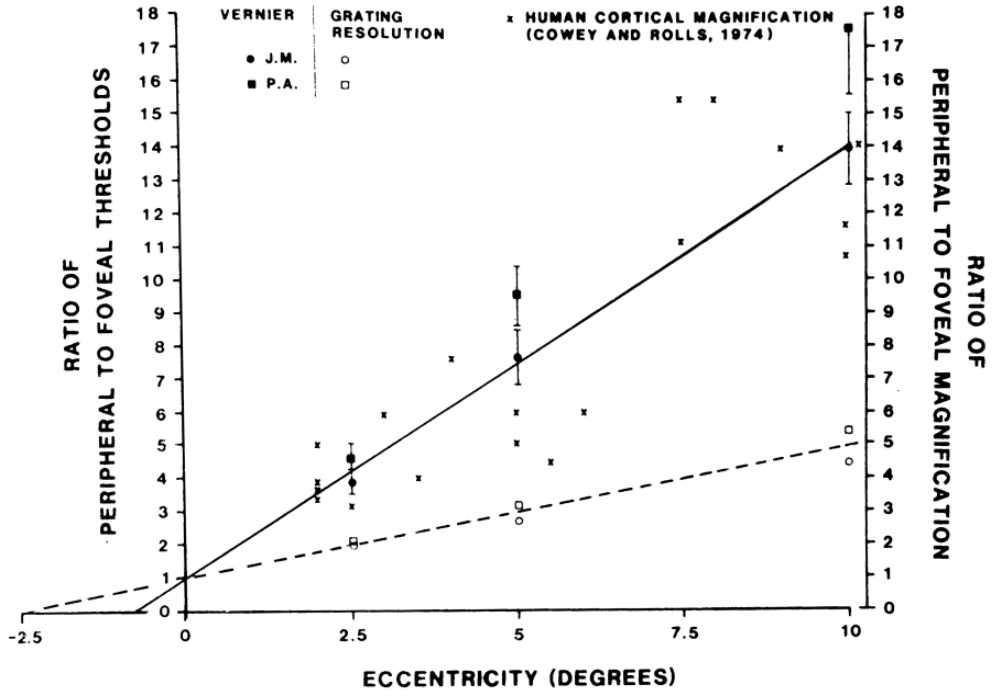

Figure 1.21: Ratio of peripheral to foveal thresholds for optimal vernier acuity (filled symbols) and for grating acuity (empty symbols) for two subjects. Units are normalised to show the increase of task difficulty as a proportion of the foveal task difficulty (lines meet the *y*-axis at y=1). Values of estimates of human cortical magnification from Cowley and Rolls (Cowey and Rolls 1974) are included (x-symbols). (Taken from Levi *et al.* 1985).

Reading performance becomes quickly impaired with eccentricity. Latham and Whitaker (Latham and Whitaker 1996) compared word recognition and reading rates for both unrelated words and meaningful sentences viewed foveal and peripherally. They concluded that these could be equated by an increase in letter size only in the case of unrelated words, implying that periphery is qualitatively inferior to the fovea at extracting meaning from sentences. Subjects reading with peripheral vision also experience an effect called "visual crowding" that affects reading when letter separation is kept constant (Toet and Levi 1992). The effect of destructive interaction between adjacent contours (called contour interaction) was studied by Hess *et al* (Hess

et al. 2000). They concluded that results in central vision could be explained by the physics of the stimuli, which is not the case for peripheral vision.

The two parallel systems (transient and sustained) that exist in foveal vision also exist (with different distributions) in the periphery (Harwerth and Levi 1978). In the same ways as in central vision, these systems are believed to correspond to the parvocellular and magnocellular neurons in the LGN, linked to the retinal and cortical visual areas in the brain. Data from primates (Daniel and Whitteridge 1961; De Monasterio and Gouras 1975) and cat (Kaplan and Shapley 1986) suggests that the distributions of neurons that mediate these two systems vary across the visual field, with a predominant increase of the (sustained) chromatic system towards the fovea (De Monasterio and Gouras 1975).

The role of SF-channels in peripheral vision has been intensively investigated for letter identification (Chung *et al.* 2002; Chung *et al.* 1998; Solomon and Pelli 1994) although less research has been conducted with complex stimuli (such as natural images discrimination) (Peli and Geri 2001).

1.7. Modelling contrast discrimination

Much of the basic concepts, ideas and models on how the HVS represents visual information are derived from what we know about detection and discrimination of spatial patterns. This is because threshold performance offers a good chance of isolating parts of complex systems (i.e. only the most sensitive mechanisms of these systems are active during threshold performance). Following this rationale, it ought to be possible to predict the HVS's sensitivity to any spatial pattern from sensitivity

measures of a relative small and restricted number of patterns (e.g. contrast sensitivity function).

1.7.1. Neural representations

Early computational models (Marr 1982; Schade 1956) of vision were based on the idea that there is a subset of visual neurons whose responses (taken as a collective) capture the properties of the visual stimuli. This representation, which condenses the effects of the many HVS spatial components, was called the *neural image* (Robson 1980). According to these models, different neuron populations may represent different types of information (e.g. some neurons may represent some coarse aspects of the stimuli and others include some level of detail, as well as colour, motion, depth, etc.) depending on the different transformations that occur to the signal along the visual pathways.

One of the most common ways to calculate a neural representation of a given stimuli is by using a shift-invariant linear mapping or *convolution* (Gonzalez and Woods 1992). This is equivalent to evaluating the (linear) responses of neurons whose identical receptive fields are uniformly spaced, covering the whole region of the stimuli. The fact that retina is not uniform (e.g. the highest concentration of receptors is in the fovea) means that this kind of analysis is restricted to the relatively small central portion of the visual field and places where the distribution of receptors (and cortical areas devoted to them) are similar. Since we move our eyes to direct our gaze (and the cortical visual power related to it) to places of visual interest, foveal vision remains our main source of information. However useful, this analogy between convolutions (shift invariant calculations) and neural receptive fields must be taken with caution since it refers to behavioural measurements, not real neural receptive fields (Wandell 1995) and the

neural representation (or neural image) is a computational tool (not the same as having a "picture" of the visual stimuli inside the head).

1.7.2. Linespread functions and vector-length calculations

To perform the shift-invariant linear mapping of the input stimuli, we need to decide the shape of the convolution operator or *linespread function*. These operators are traditionally based on models of receptive fields of retinal ganglion cells (Schade 1956) (e.g. difference of Gaussians (Enroth-Cugell and Robson 1966; Rodieck 1965)), which are maximally sensitive to certain SFs and have contrast sensitivity functions that resemble those of retinal ganglion cells -as shown in Figure 1.13 and Figure 1.15. After the convolution operator has been defined, the amplitude of the neural image (at detection threshold) has to be decided based on the responses of individual operators (neurons). Since detectability has to be related to the pooled responses across the neural image, not the response of a single neuron *n*, a possible solution is to add all the squared responses of the receptors and to form a squared neural image (Graham *et al.* 1978):

$$(neural\_image)^2 = \sum_i n_i^2$$

Equation 1.2

The squaring of the neural image has the advantage of removing the sign of the individual neuron responses (which are unimportant, since any deviation, positive or negative, from the mean firing rate of real neurons should indicate a detection). The criterion for deciding the exponent in the previous equation, which determines the stimulus visibility, is extremely important and varies from author to author. In Section 4.2.4 we present a slightly different approach.

A different convolution kernel will produce a different shift-invariant linear transformation of the original image (and a different neural representation). As we have seen before, there is strong evidence of the presence of SF-channels, which are maximally sensitive to a small range (or band) of SFs and interact with one another. A complete model of contrast *detection* should consider this and produce several neural images; each corresponding to neurons with different SF-bandwidths, and the decision process should be based on a pooling of all their outputs (*multi-resolution* image). In the case of contrast *discrimination*, the decision should be based upon information collected from two input images, the test and reference images. Psychophysical evidence on pattern adaptation (Blakemore and Campbell 1969) and detection/discrimination of square-wave gratings (Campbell and Robson 1968) support this idea of a multiple channel representation. In the later study, Campbell and Robson demonstrated that the different harmonics of a square-wave grating are encoded by different SF-channels.

1.7.3. Masking and facilitation effects: the dipper function

Several workers (Campbell and Kulikowski 1966; Legge and Foley 1980; Nachmias and Sansbury 1974; Tolhurst and Barfield 1978) measured the contrast needed to detect a sinusoidal target grating of fixed SF as a function of the contrast of another (masking pattern) superimposed to it. Figure 1.22 shows a plot of such results from Legge and Foley (Legge and Foley 1980). When the masking pattern or *pedestal* has low contrast and similar SF as the target pattern, it facilitates the detection and the opposite occurs when the masking pattern is of high contrast.

The two effects (masking and facilitation) determine the "dipper" shape of the curve. The masking effect is reduced when the SF of the mask and test stimuli differ by a factor greatly than 3 (De Valois 1977a, b). These pedestal effects can be incorporated

into a contrast discrimination model (see Section 4.2), along with more sophisticated assumptions such as convolution kernels that are maximally sensitive to certain orientations, neural noise (Gorea and Sagi 2001), and increasingly complex decision making rules (Ahumada *et al.* 1998; Ahumada *et al.* 1995; Daly 1992; Foley and Legge 1981; Watson 1983; Watson and Solomon 1997).

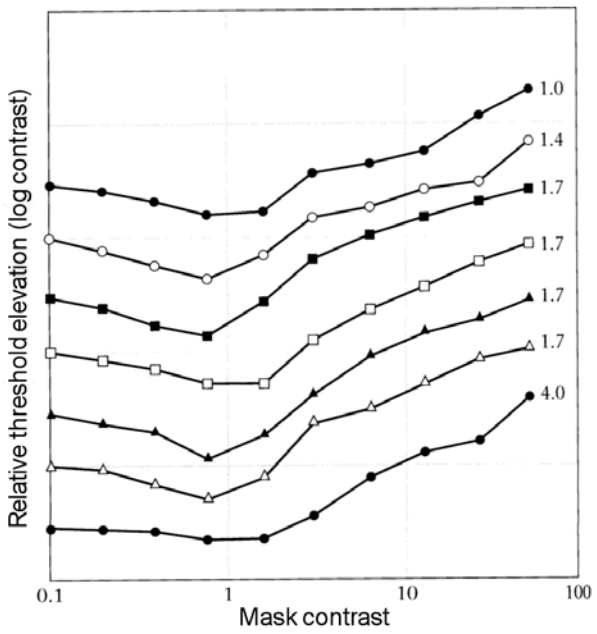

Figure 1.22: Contrast needed to detect a 2 cycles/deg grating as a function of the contrast of a second, superimposed grating (masking pattern). The different lines represent the SF of the masking grating. (From Legge and Foley 1980).

1.7.4. Model's predictions of psychophysical data

Since it is more difficult to compare the performance of these *multi-resolution* models (i.e. models that rely on a collection of component neural images, each representing a SF narrow-band "channel" and orientation) with psychophysical data, most of their predictions come from computer simulations. However, several authors (Menendez and Peli 1995; Rohaly *et al.* 1997; Watson 1993b) have published reviews of popular

vision models for detection, recognition and image difference metrics, and their ability to predict psychophysically measured data.

Brady and Field (Brady and Field 1995) were able to predict *contrast constancy* results (i.e. the ability of observers to perceive objects as maintaining a constant contrast independently of size or distance) by using a model where channels' SF bandwidths increase with SF and peak sensitivity is equal across SF.

Peli (Peli 1996) used a definition of local band-limited contrast (see Section 4.2.1) that agrees with *perceived* contrast (Peli 1990) as the basis for a model of contrast discrimination. Simulations of distant views were produced using the CSF of observers who were later asked to distinguish the simulations from the originals at increasingly longer distances. The model predicted the distance at which the simulation started to appear indistinguishable from the original image.

Watson and Solomon (Watson and Solomon 1995) fitted a contrast model to psychophysical measures (thresholds for a Gabor patch masked by gratings of various orientations or combinations of orientations) produced by Foley (Foley 1994). Their model used Gabor filters with octave bandwidths at eight different orientations as operators. Excitatory and inhibitory nonlinearities were modelled as power functions with exponents of 2.4 and 2. The decision was based on a *Minkowski* pooling (also called *generalised vector magnitude* -see Equation 4.10 later in Chapter 4) of the outputs with an exponent of 4.

Tadmor and Tolhurst (Tadmor and Tolhurst 1994; Tolhurst and Tadmor 1997) modelled the discrimination of changes in the Fourier amplitude slope of natural and

synthetic images using a *local* contrast discrimination model, which compares the contrast within restricted SF bands using circularly symmetric kernel operators.

Rohaly *et al* (Rohaly *et al.* 1997) compared the performance of three models that predict image discriminability (the visibility of the difference between pairs of images). These were (a) a multiple channel model where the effects of within-channel masking were implemented; (b) a single-channel contrast sensitivity model and (c) a simple arithmetic difference metric between images. They also explored the effects of altering the exponent of the Minkowski summation (see Equation 4.10) used to combine the responses of channels/receptors and the inclusion of arbitrary contrast gain factors. They found that all models would produce reasonable predictions (the best exponent for the Minkowski sum was 4); especially after the addition of contrast gain factors.

1.8. Naturalistic stimuli

Although much has been learnt about the visual system from the use of simple spots of light, bars or gratings as stimuli in vision research, these do not represent the type of visual environment that shaped the properties of the visual system throughout its evolution and postnatal development. Such stimuli (called *natural images*) have particular and more complex properties. For example, natural images have very low ranges of contrast, compared to physiological or psychophysical artificial stimuli. This means that it might not be possible to predict the responses of simple cells to natural stimuli based on the responses to artificial stimuli (Brady and Field 2000; Tadmor and Tolhurst 2000). Besides this, for much that the traditional research methods tell us much about the responses of simple cells, they say nothing about the possible reasons why these cells have evolved to behave in the way they do.

1.8.1. Natural images and ecological tasks

If we assume that the primary role of any visual system is to improve the chances of survival for an organism by aiding it to navigate, find food or suitable mates, avoid predators, etc. then it makes sense to study vision and the visual environment in the presence of such ecologically-relevant tasks. Under this view, it is fundamental to identify properties of the natural environment that are relevant for these tasks (e.g. finding food in the forest). Some workers put forward the notion that colour vision in primates has evolved to aid them in finding food (Dominy and Lucas 2001; Mollon 1989, 1991; Sumner and Mollon 2000a, b). If this is the case, then it is important to study the colour statistical properties of images from the foraging environment where primates find their food, which may be different from the rest of the visual scenes in many respects. For example, tropical jungle environments are characterised by high tree canopies. Here, part of the direct light coming from the sun is filtered by the leaves before reaching the ground, some light is subject to multiple reflections and some shines though gaps in the vegetation producing a characteristic "dappling". The light passing though the leaves or reflected from the leaves is likely to be spectrally biased towards the green regions of the visual spectrum, while light coming from the sky is likely to be bluish. This bias towards blue regions is stronger in cloudy days. The overall effect of these interactions is to produce strong bluish-greenish shadows alternating with large luminance variations in the illumination, up in the canopy and on the ground. A visual system optimised for foraging in these conditions may benefit from removing the effect of the dappling and compensating for changes in the colour of light. Other activities such as identifying members of your own species or detecting rapid changes in the local distribution of luminance (as may occur when there is a

sudden movement) are also ecologically-relevant visual tasks, which should be taken into consideration when studying the natural environment.

Another view emphasizes that the HVS seems to be adequate at performing a variety of tasks and its main advantage seems to be its flexibility in adapting to many requirements. This has prompted many researchers (see below) to study some of the most common properties of natural scenes, based on information theory (Shannon and Weaver 1949) assumptions that apply indiscriminately to all images, regardless of the environmental conditions or particular visual task. In this view, it is the stimulus-response specificity of single visual neurons what makes them efficient at coding features in all natural scenes. Although "efficiency" has yet to be defined, this proposal has prevailed in the mainstream vision community. "Optimality" is measured by means of some kind of informational-processing goal such as reducing the "wastage" in visual channels or reducing the signal-to-noise ratio in neurons, etc.

1.8.2. Redundancy reduction

The simplest kinds of statistics to analyse on digital images are based on local brightness intensity (pixel grey levels) and are called *first order* statistics. They refer to means, variance and probability distributions of brightness. A more complex way of analysing an image includes studying the inter-relations between each pixel and its neighbours (e.g. how does a pixel value depends on that of its neighbours?). These relationships are called *second order* statistics.

Natural images contain complex statistical regularities (in terms of redundancy, power spectra, variance, phase spectra, etc.) that provide a very different input from that of bars, spots or gratings. The idea that the statistics of sensory stimuli we receive from

the environment are important for perception and cognition started to gain momentum at the end of the XIX century (see Barlow (Barlow 2001) for a review) and achieved extra force after the theory of communication was introduced by Shannon and Weaver in 1949 (Shannon and Weaver 1949). From there, concepts like *channel capacity*, *information* and *redundancy* started to influence vision scientists. Many theories of efficient coding have been put forward since.

One of the most influential theories of visual efficient coding is centred on a principle called *redundancy reduction* (Attneave 1954; Barlow 1961). According to information theory, *redundancy* is what wastes channel capacity (it is the difference between the entropy of the messages actually transmitted and the maximum entropy that the channel could actually transmit) (Barlow 2001). This theory argues that the removal of redundant statistical structure in the sensory input is the main aim of the sensory neurons. The concept of redundancy is related to the *predictability* of certain structures in the natural visual input. For example, in natural scenes it is very often the case that once we know the luminance of one point we can predict the luminance of most of the nearby points (i.e. they are highly correlated). Psychophysical measurements of redundancy in monochromatic images were obtained by Kersten (Kersten 1987). He asked observers to reconstruct partially corrupted images using their innate knowledge of the structure of natural scenes and estimated that about 46% to 74% of the information present in his natural scenes was redundant. Redundancy removal is different from selective coding, where some information is deliberately discarded by the HVS for a reason (e.g. it is biologically unimportant). When redundancy is reduced, full reconstruction of the original image is possible, thus, no information is lost. Barlow (Barlow 2001) later reviewed these assumptions and concluded that redundancy is not something completely useless which can be safely stripped from the visual

environment. In his view, it is more important for an animal to identify what is redundant in its sensory messages and use it to gather important clues about the structure and statistical properties of the environment.

Since the ratio of photoreceptors to optic nerve fibres is high, it has been suggested that retinal coding represents a redundancy reduction that compresses the information into a channel of limited capacity (Atick and Redlich 1992; Srinivasan *et al.* 1982). However, the channel capacity *increases* in the cortex, this meaning that redundancy must also increase at these stages (since information cannot be created). Barlow has argued that sensory coding "*should convert hidden redundancy into a manifest, explicit immediately recognisable form, rather than reduce or eliminate it*" (Barlow 2001).

Redundancy reduction is still important, although other ideas have been incorporated to account for the fact that the brain is not just a communications device. For example, there might be cases where a simple compression is not the best way to exploit statistical structure and some knowledge of the environmental properties may help to match the characteristics of the detector to these properties (and thus reduce the signal-to-noise ratio of the process) (Barlow 2001).

1.8.3. Second order redundancy

The information content in natural scenes (statistical regularities that distinguish them from random noise), either static or dynamic (movie sequences) (Dong and Attick 1995) has been the subject of various studies. Many of these studies have concentrated on the variability contained in the signal as a function of the spatial frequency (also called second order correlation or *power spectra*) of monochrome natural images (Burton and Moorhead 1987; Carlson 1978; Field 1987; Párraga *et al.* 1998a; Tolhurst *et al.*

1992). The analysis of the properties of the Fourier domain led to the discovery of *second order* redundancy.

The most commonly used quantitative measure of second order-redundancy (e.g. the dependency of a pixel intensity to that of its neighbours) is the autocorrelation function (Field 1987). This function measures the average of the product of the intensity at two positions as a function of their separation. Typically, autocorrelation falls with relative distance. Correlation measures assume that the statistics of natural images are *translation invariant*. This means that the presence of strong features that occur typically at certain positions (such as horizon and sky usually appearing at the top) would upset this measure (Ruderman 1997). The assumption of translation invariance allows us to decorrelate the images by transforming them into the frequency (Fourier) domain.

Television engineers (Kretzmer 1952) discovered empirically that the *power* spectrum of TV images approximately follows a simple relationship: Fourier power = $1/f^2$; where f is the spatial frequency (in cycles/degree of viewing or cycles/picture) and all orientation dependencies are ignored (averaged) in the analysis. The power spectrum is related to the autocorrelation function through the Fourier transform (Gonzalez and Woods 1992). The same equation can be expressed in terms of the Fourier *amplitude* spectrum (which is the square root of power).

$$\textit{Fourier amplitude} = 1/f^{\alpha}$$

Equation 1.3

The exponent (α) of f in Equation 1.3 was measured for a small sample of achromatic natural scenes by Field (Field 1987) and found to be around 1 (2 in the case of power spectrum). Figure 1.23 shows a cylindrical plot (Fourier *amplitude* is represented on the

vertical axis, SF is measured by the perpendicular distance to this axis and *orientation* is determined by the angle θ) the of the Fourier amplitude of one of such images.

More exhaustive measures (using larger databases of calibrated achromatic images) have shown that the average exponent α is closer to 1.2 (Párraga *et al.* 1998a; Ruderman 1995; Tolhurst *et al.* 1992) and that there are substantial variations among individual images (Langer 2000; Van Der Schaaf and Van Hateren 1996). An exponent α equal to 1 means that each octave in spatial frequency *f* contains the same amount of Fourier power. A smaller value of α means that there is proportionally more power in the short distance details of the image, etc. This proximity of the α to unitary value is related to a property of natural scenes called "scale invariance" (the statistical structure of the image is independent of the pixel size).

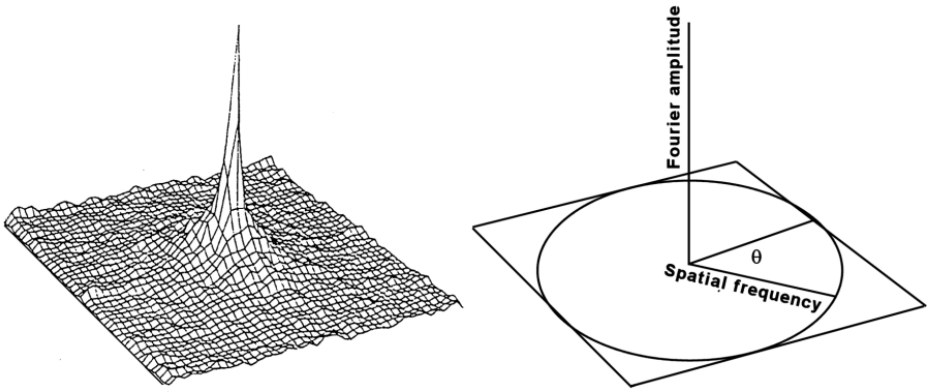

Figure 1.23: Plot of the two dimensional amplitude spectrum of a typical natural image. The scheme on the right illustrates the meanings of the axis. The centre of the plot represents 0 spatial frequency (mean amplitude of the image). For the sake of clarity, a smaller grid of 32x32 points is shown. (From Field 1987).

Figure 1.24 shows Fourier amplitude plots corresponding to four natural scenes. Each point was obtained by averaging across orientations (the same as averaging along the SF-constant circles in Figure 1.23). The values are plotted on a *log-log* co-ordinate axis,

which is best to show the $1/f^{\alpha}$ dependency. Lines show the corresponding fitted $1/f^{\alpha}$ functions (notice how the measured amplitude slope depends of the large concentration of high-SF points near the right side).

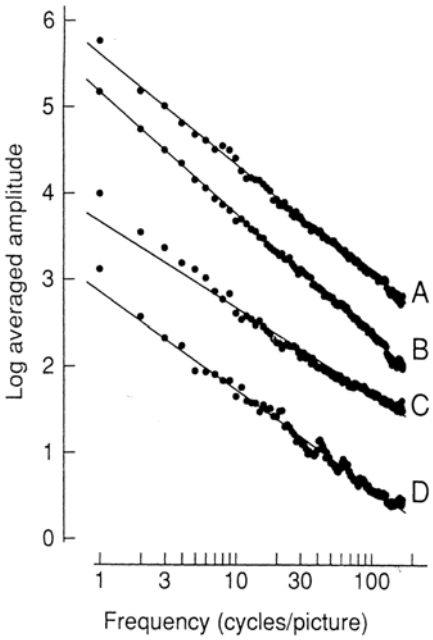

Figure 1.24: Double-logarithmic plots of Fourier amplitude spectra of four different natural scenes (averaged across orientations). (From Tolhurst *et al.* 1992).

The environmental causes of the power law (or $1/f^{\alpha}$ law) in natural scenes have been the subject of much speculation. Some authors argue that they are a product of scale invariance in the natural visual world (Field 1987, 1990). Others argued that this property is because of the presence of edges in images (since edges themselves obey a power law spectral distribution) (Carlson 1978). Ruderman (Ruderman 1997) argued that scale invariance of natural scenes comes from the superimposition of statistically independent objects (surfaces in 2-dimensions) with power-law distributions of sizes. Balboa and Grzywacz refuted this hypothesis arguing that a power-law distribution of sizes in natural images is not constrained to produce scale-invariant images (Balboa *et*

al. 2001). The premise behind these analyses is that the HVS should be optimised or "tuned" to extract information from the natural visual environment, thus exploiting any statistical regularity.

1.8.4. The "ecological approach"

The idea that efficiency in the information representation by the brain potentially has evolutionary advantages can be used as a design principle to predict neural processing. *This approach can be called 'ecological' since it attempts to predict neural processing from physical properties of the stimulus environment* (Atick 1992).

Does the visual system take advantage of the structure present in natural scenes? Several authors (Atick 1992; Barlow 2001; Field 1994; Laughlin 1981; Srinivasan *et al.* 1982; Van Hateren 1992b) have argued that it is possible to take account of signal redundancy and produce a neural encoding where it is minimised, thus improving the efficiency. Srinivasan *et al* (Srinivasan *et al.* 1982) proposed that the size of retinal receptive fields could be estimated from the amount of subtractive inhibition necessary to cancel out the correlation in natural scenes. They compared the predicted inhibitory receptive fields surround to that measured from the compound eye of the fly and both agreed. Further research (Van Hateren 1992b) considered the presence of white photoreceptor noise and changes in the system to account for variations in the signal-to-noise ratios at lower luminance levels. Attick (Atick 1992) has reviewed the uses of information theory as the basis for an approach to neural computing. He examined cases where this approach has been successful in predicting coding strategies (mammalian retina and fly's LCM -Large Monopolar Cells).

Psychophysical measures of the discriminability of natural images were made by Knill *et al* (Knill *et al.* 1990). They imposed Fourier amplitude spectra consistent with a power law to *synthetic* images and measured observers' performance in discriminating between images of slightly different slopes. Other workers (Tadmor and Tolhurst 1994) considered these slope-modified synthetic images to be inadequate as natural visual stimuli. Slope-modified synthetic images produced lower discrimination thresholds than *real* natural images. They were also perceived to be at best quality when filtered with less steeper amplitude slope (α= 0.5) than that corresponding to natural images (α=0.8-1.5). The reason for this may be the lack of a complete randomness (possible anisotropy) in the distribution of energy with spatial orientation that may exist in real natural images as opposed to synthetic images. Tolhurst and Tadmor (Tolhurst and Tadmor 1997) proposed a SF-band limited model of contrast, which explains their measured thresholds for detecting changes in the slope of natural scenes. They argued that observers might perform this task by detecting changes in contrast within a single SF narrow-band. Further research (Párraga and Tolhurst 2000) concluded that observers do not usually detect changes in contrast within a single narrow band; rather, they do compare contrast between at least two bands.

Human horizontal retinal cells have also been studied in an ecological context. Balboa and Grzywacz (Balboa and Grzywacz 2000a, b) proposed that the role of early lateral inhibition is to deal with noise without missing relevant clues from the visual world, especially, the occlusion boundaries between objects.

Geisler *et al* (Geisler *et al.* 2001) examined the statistical properties of contours in natural images arguing that these statistics may have driven the contour grouping

mechanisms that allow the brain to link local edge elements into global contours. They found that contour detection performance might be predicted from these statistics.

Colour vision scientists have also benefited from the redundancy reduction and optimisation hypothesis (Buchsbaum and Gottschalk 1983). Burton and Moorhead (Burton and Moorhead 1987) analysed a set of nineteen natural scenes in terms of the L, M, and S relative cone absorption and found high levels of correlation between the signals from the red-and green-sensitive cones. They pointed out that this correlation could be removed by a linear mapping into three new channels, two carrying chromatic information and one carrying luminance information, as suggested by Buchsbaum and Gottschalk (Buchsbaum and Gottschalk 1983). Ruderman *et al* (Ruderman *et al.* 1998) examined the statistical properties of a *hyperspectral*[10] image dataset of foliage and concluded that the principal component axes of the dataset lay along directions (in terms of the L, M and S cone responses) very similar to the chromatically opponent mechanisms discussed in previous sections. However, it is worth noting that the L-M opponent channel encoded very little variance. Multiscaling properties were also found in hyperspectral images (Turiel *et al.* 2000).

Clement and Moorhead (Clement and Moorhead 2000) pointed out that there exists an analogy between the redundancy introduced by the spectral overlap of the three (L, M and S) cone receptors (later removed by the chromatically-opponent mechanisms) and the spatial sampling of the SF-channels (also overlapping in SF-space). They investigated what aspects of the natural environment are captured by decorrelating the SF-channels and found filters whose properties may explain the subjective appearance of some visual stimuli.

Nascimento *et al* (Nascimento *et al.* 2002) studied the reflectance properties of 30 natural scenes and concluded that spatial L and M cone-excitation ratios were remarkably invariant across different illuminants. They suggested that this may represent a reliable property of the visual environment and may form the foundation for visual *colour constancy*[11].

Other attempts to relate the spatio-chromatic properties of the visual environment and those of the HVS was made by Párraga *et al* (Párraga *et al.* 1998a). They analysed the achromatic and L-M chromatic properties of a data set of 29 natural scenes in terms of the distribution of Fourier amplitude along the SF spectrum and reported that this distribution did not match the (psychophysically measured) SF filtering characteristics (Mullen 1985) of both luminance and L-M chromatic channels in the parvocellular visual pathway. To model the luminance and L-M chromatic channels, they employed four different sets of weightings for the combinations of L and M receptors outputs based on different criteria (Buchsbaum and Gottschalk 1983; Ingling and Tsou 1988). Although the ecological approach predicted a higher content of low-SF energy in the chromatic image representations than in the achromatic representations, they reported no difference in the spatial properties of the scenes between these signals. This failure to find a match between the spatio-chromatic properties of the HVS and those of natural scenes may be caused by analysing images without considering any specific task for the visual system. If primate colour vision has evolved for specific tasks such as finding red fruit in the forest, it would be critical to analyse spatio-chromatic properties of natural images in view of such task (i.e. images of red fruit and green leaves).

---

[10] Each pixel is calculated from many (usually 31) spectrally narrowband filtered pixels which sample the same region of the image.

[11] Colour constancy is a fundamental cortical mechanism that compensates for spectral and spatial changes in the illumination, keeping the colour of objects more or less constant (Hurlbert A. 1999).

1.8.5. Higher order statistics

Second (and higher-order) redundancies in the code means that the responses of one neuron can be predicted from the responses of other neurons. If these correlations are among groups of three or more neurons, they constitute higher-order redundancy. Higher order redundancy is directly related to the phase spectra of natural scenes. Information redundancy was measured for a large ensemble of natural scenes of (only) 64 greylevels by Petrov and Zhaoping (Petrov and Zhaoping 2003). They found that 2-pixel correlations produce about 50% of the total redundancy while 3-pixel correlations have only a marginal effect (4%).

The scale invariance of the Fourier amplitude (or power) spectrum is the simplest regularity that natural images possess. Although very convenient, the analysis of the Fourier amplitude spectra provides by no means a complete characterisation of natural scenes. The influence of the phase spectra on natural images was reported by Piotrowski and Campbell (Piotrowski and Campbell 1982). Tadmor and Tolhurst (Tadmor and Tolhurst 1993) investigated the effects of swapping the amplitude and phase spectra of a natural image with those of another. They found that '*although the average amplitude spectra of different natural images may be similar in their overall form, a realistic description of the amplitude spectra must also incorporate the particular way in which the energy is distributed across different orientations'*.

To distinguish a natural scene from the next we also need to characterise their *phase* spectra. The phase structure of natural scenes has proven quite difficult to characterise. A number of authors (Ruderman 1995; Thomson 1999a, b) have investigated their phase (also called higher-order statistics) structure and found some invariant properties.

1.8.6. Natural images and cortical processing

Much of the research on natural images is devoted to derive a set of linear basis functions that remove image redundancy and that are similar to the receptive fields found in the primary visual cortex (Field 1987; Hancock *et al.* 1992; Olshausen and Field 1996; Shouval *et al.* 1997; Wachtler *et al.* 2001). To resemble cortical visual cells, these basis functions (filters) have to be SF band-pass, spatially oriented, localised and symmetric.

An important attempt to relate the statistical structure of the environment to the coding (derived the narrowly tuned channel mechanisms) of the visual system was made by Field (Field 1987). He considered the statistics of six achromatic scenes from the natural environment and compared various coding schemes on how they represent information in such natural scenes. Field based his analysis on a model of the behaviour of cortical cells derived from principles from information theory (Shannon and Weaver 1949) and Gabor's (Gabor 1946) theory of communication. Field labels as "sensors" individual Gabor functions located at a point within the scene and representing a single hypothetical cortex cell. He organised these sensors into "channels" (spatial arrays of sensors tuned to a common orientation and SF). He concluded that such a collection of channels with SF-bandwidths constant in octaves and orientation bandwidths constant in degrees produces an even distribution of the information (from the set of six scenes) across the array. Figure 1.25 shows the relations between the size of a channel in the frequency domain and the size and spacing in the space domain.

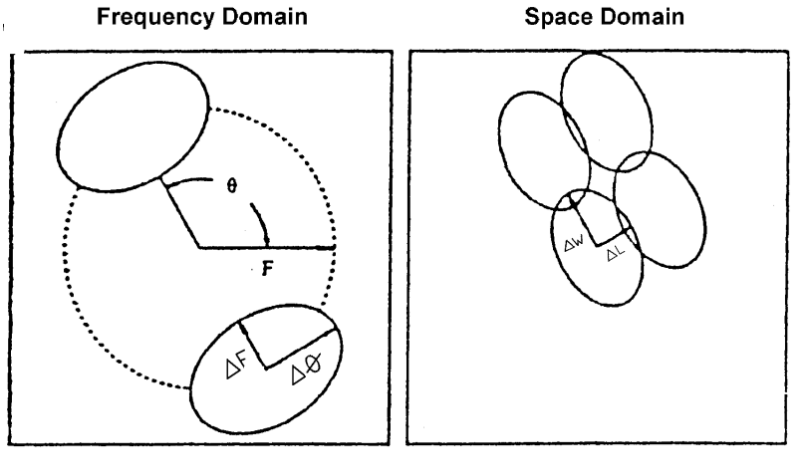

Figure 1.25: Fields channels represented in the frequency domain and in the space domain. A channel with a bandwidth ΔF in the frequency domain consists of an array of sensors with a width ΔW in the space domain. (Taken from Field 1987).

Field found the optimal SF-bandwidth to be in the range 0.5 to 1.5 octaves. In addition, the optimal constant ratio between SF and orientation bandwidth (measured in the two-dimensional Fourier space) was found to lie between 0.5 and 1.0. This is despite the fact that Field's model of evenly distributed channels over a rigid sensor grid does not reflect the variability in the spatial frequency tuning of different cortical cells (De Valois *et al.* 1982a).

Hancock *et al* (Hancock *et al.* 1992) investigated the distribution of variance in fifteen natural images using principal components analysis (PCA). PCA is a form of factor analysis that conveys the most information from the image using a limited number of linear descriptors. The very first components are the linear descriptors that convey most of the image information. In the dataset that they analysed, Hancock *et al* found a bias towards vertical and (to lesser extent) horizontal orientations. This was confirmed by other research (Van Der Schaaf and Van Hateren 1996). Coppola *et al* (Coppola *et al.* 1998) analysed a larger image dataset (150 pictures, 3 different types of visual

environment: outdoor, indoor and "natural") and found that there are differences in the distribution of oriented contours between the different types of scenes.

Another commonly used technique to remove correlations between pairs of neurons (second order correlations) is called independent component analysis (ICA) (Bell and Sejnowski 1995). This technique has the advantage of reducing higher order dependencies (although it increases the first-order redundancy, which may lead to an increase of overall redundancy). It also provides a good account of the receptive fields properties of mammalian visual cortex (Bell and Sejnowski 1997; Fyfe and Baddeley 1995; Van Hateren and Ruderman 1998; Van Hateren and Van Der Schaaf 1998; Wachtler *et al.* 2001).

Field's efficiency criterion was re-examined to take into account the *sparseness* (i.e. only a small subset of is active to respond to each stimulus, thus reducing the number of action potentials at any time) of the set of cortical cells (Field 1994). Learning rules that produce sparse-coding for natural images were introduced later by Olshausen and Field (Olshausen and Field 1997). This algorithm is related to ICA (Bell and Sejnowski 1997) and produced 'receptive fields' that are similar to those of simple cells (Van Hateren and Van Der Schaaf 1998).

Field and Brady (Field and Brady 1997) identified physical blur (with the consequent loss of high-SF energy) and the variability in the *density of structure* as a major source of variability of the Fourier power spectra of natural scenes. They noticed that white noise (which does not follow a power law) appears perceptually to be dominated by high frequency structure, while noise artificially filtered to follow a power law seems to have structure distributed across all spatial scales (Brady and Field 1995). This agrees with the idea that the responses of cortical cells are independent of spatial scale (Field 1987)

and shows that the same limited dynamic range is present in the cell's responses across spatial scale (Field and Brady 1997).

Baddeley *et al* (Baddeley *et al.* 1997) found an exponential distribution of instantaneous firing rates of neurons in primary and inferior temporal visual cortices of cats and monkeys when visually stimulated with videos of natural scenes. This shows that neurons maximise their information-carrying capacity (with a fixed, long-term-average firing rate), consistently with an optimal coding.

1.8.7. General stimuli versus specific stimuli

The ecological approach to vision research combined with a widespread availability of naturalistic stimuli databases (Chiao *et al.* 2000a; Párraga *et al.* 1998a; Ruderman *et al.* 1998) has lead to a great progress in our understanding of the relationship between the properties of the HVS and the environment by treating the it as a simple communications device (i.e. information-theory based studies) (Attneave 1954; Barlow 1961). Many properties of the HVS can be explained by this approach; these include *compression* (Atick and Redlich 1992; Srinivasan *et al.* 1982) of the visual information before transmission through the optic nerve, *match* between the range of contrast found in the natural world and neuronal response ranges (Laughlin 1981; Tadmor and Tolhurst 2000), etc. In these cases, the criterion for efficiency was usually borrowed from signal processing theory (redundancy, entropy, signal-to-noise ratio, etc.). In other cases, the efficiency criteria may include more complex items, such as *even distribution* (Field 1987) of Fourier energy across a set of neuronal receptors maximally sensitive to given bandwidths and orientations or maximum *reduction* (Olshausen and Field 1996, 1997) of the number of active neurons at any moment (sparseness), etc. However successful these efficiency criteria may be, an analysis of the HVS and its

optimisation to the properties of the natural environment needs to be based in a reliable definition of what it is to be *efficient* (see Section1.2). For example, it is unclear what the significance of a match between the properties of the HVS and the second-order statistics of natural scenes may be, since nobody knows the advantage of detecting deviations from this average "natural" structure. Moreover, the distribution of energy in the Fourier spectrum for natural scenes has been reported to be nearly the same across images that look completely different. It is necessary at this point to introduce the idea of a *task-related* efficiency criterion. Under this view, a system is optimised when it performs an ecologically-relevant task in an efficient manner. For example, one could (psychophysically) measure visual performance in ecologically-relevant visual tasks that involve stimuli obtained from photographs of natural scenes and later repeat the task with statistics that deviate more or less from those of truly natural scenes. Best performance (optimality) in the task when the stimuli have natural rather than "unnatural" statistics will suggest that the visual system does work best with natural scenes, and therefore, is more efficient at coding the features found in natural scenes.

Examples of how vision scientists are starting to consider naturalistic stimuli in terms of their ecological significance are the studies of colour spectral sensitivities of human photoreceptors matching to the characteristics of visual input. The wavelength distribution of most natural visual input extends over the whole spectrum (Stiles *et al.* 1977), thus making it possible to represent most of its variance with only three photoreceptors (Maloney 1986). The problem with primate photoreceptors is that their sensitivity is not evenly spread across the spectrum, making them not optimal, as is the case of some animals (Wright and Bowmaker 2001). This inefficient sampling of the colour spectra was confirmed by Nagle and Osorio (Nagle and Osorio 1993), who

found that the tuning of human photoreceptors might minimize (instead of maximising) red-green chromatic signals in natural conditions. Mollon (Mollon 1991) proposed a solution for this problem, considering that the L and M photoreceptors evolved for the purpose of aiding primates to distinguish reddish fruit against a background of leaves. Analysis of the coloration of specific (and ecologically-relevant) visual stimuli supported Mollon's ideas (Osorio and Vorobyev 1996; Regan *et al.* 1998). Other authors (Dominy and Lucas 2001) support the idea that the L-M opponent signal is optimal for detecting the youngest and most edible leaves from plants.

The differences between the general stimuli approach, which considers the brain as a communications device and the specific stimuli approach, which tends to view the brain as a *decision* device become apparent after we consider the visual tasks that drive the evolution of the HVS.

1.9. What are the visual tasks that drive the evolution of the HVS?

In our view, optimality is related to performing an ecologically-relevant visual *task* in the most efficient way. This leads us to the following question: what are the visual tasks that drive the evolution of the HVS? The answer is not a simple one, but it can be summarised in three categories as follows:

1.9.1. Vision and attention

It is known from studies of the relationship between attention and perception (inattentional blindness (Mack *et al.* 1992; Moore and Egeth 1997; Newby and Rock 1998) and change blindness (Becken and Cervone 1983; Littman and Becklen 1976; Simons and Chabris 1999)) that without the observer's attention, most features of the

visual environment are not consciously perceived[12] at all. For example, Simons and Chabris (Simons and Chabris 1999) using naturalistic stimuli, showed that about half of the observers failed to notice a highly salient and unexpected event when engaged in a monitoring task on a display screen, implying that only the features of the visual environment that receive our attention are encoded, processed and retained in the memory. These failures to notice events outside the region of attentional focus point out to the fact that, although all the information in the visual field is potentially available for attentive processing, only a small region is processed *at any time*. The environment acts in fact as an information storage device that our eyes use to search and retrieve what it is useful, saving the visual system the burden of having to process all the information contained in the visual field, allowing it to concentrate its power in a much smaller region. However, there are instances when meaningful stimuli (such as smiley faces or the observer's own name) are detected, suggesting that although not consciously perceiving visual features, the observer can perceive their meaning (Rubin and Hua 1998). To act like this, processing only a fraction of the information imaged on the retina, our visual system needs to assume that somehow the general properties of the visual environment do not change dramatically over a short period. Apparently, this is the general case, since changes in illumination are generally slow compared with the time it takes to move our body, head or perform a saccade. In summary, inattentional blindness and change blindness seem to indicate that the task of the visual system is *not* to process information from the whole of the visual field, but only its relevant features.

[12] Although failing to perceive an object, observers can still be influenced on their performance by the object, thus it is necessary to distinguish between 'conscious' perception and 'implicit' perception. In this case, the term refers to the conscious experience of an object's presence or an event.

1.9.2. Vision and action

However useful, concentrating all the HVS power on a small region of the visual field may prove a disadvantage if a predator suddenly appears from outside this attentional and processing focus area. This is supported by the fact that, although almost all visual functions decline with eccentricity, those related with monitoring change (flicker and movement sensitivity) actually improve in the periphery (Baker and Braddick 1985). It is as though the major role of peripheral vision is to provide the information required for subsequently orienting the movements of the head and the eyes so that the fovea is aligned with the region of interest. The orientation of the eyes is done by three different systems called the *saccadic*, *pursuit* and *vergence* systems. The saccadic system is in charge of rotating the eye, bringing the target on to the fovea. The pursuit system allows the eyes to follow a target in motion and the vergence system allows maintaining both eyes on a target at different depths (Findlay and Gilchrist 2003).

Another fundamental task for a visual system is to provide navigational information to the brain (which is later passed to the muscles of the body) for many different activities such as grasping, walking, climbing, etc. Spatial-memory formation experiments (Ludvig *et al.* 2003) in freely moving New World monkeys show that they are able to efficiently generate short- and long-term spatial memories (i.e. they remember the positions of baited and non-baited food-ports) of a fixed environment. There is also evidence of two separate attentional systems mediating stimuli in the far and near (within reach) space (Rizzolatti *et al.* 1983) in macaque. In humans, evidence of separate neural systems concerned with the perception of stimuli in near and far space (and the response to them) was produced by Cowey *et al* (Cowey *et al.* 1994). This supports the idea that spatial awareness (in monkeys) is a combination of several

perceptuo-motor systems, each with a different neural representation of space and concerned with the movements of eyes, head, arms and body, suggesting that they evolved for different tasks (grasping, walking, etc.).

1.9.3. Detection, discrimination and recognition

Since all organisms have to gather enough energy to be able to live and reproduce, an early ecologically-relevant task for the HVS is to efficiently detect and discriminate the best sources of food (foraging). It is also important to recognise members of their own species or group and distinguish friends from foes. There is also a need for discriminating the best possible mates to reproduce with. Such tasks can be optimally performed by concentrating the processing power of the HVS on certain attributes of the visual environment and ignoring others. For example, there is evidence that line-drawing depictions of scenes (edge figures) can lead to similar performance in object naming as full-colour photographs (Biederman and Ju 1988). Computer scientists, whose algorithms are mostly based on heuristic formulations tailored to solve specific problem needs, have exploited the enlarged role of edges in object and pattern detection (Campbell *et al.* 1997; Clark *et al.* 2000; Marr 1982) and recognition (Gonzalez and Woods 1992). Other characteristics of visual scenes that may be extremely useful in a foraging task are, for example, shadow distributions that may reveal the shape of objects and their relative positions. On the other hand, the HVS may be interested in removing the effects of some characteristics of the visual environment that may interfere with the foraging task such as daily changes in the spectral distribution or the overall illumination levels, the "dappling" effect of shadows, etc.

1.10. Specific questions addressed by this dissertation

The relationship between the "general stimuli" approach and the "specific stimuli" approach to human vision in the presence of a visual discrimination/detection task has not been explored in depth. This lack of understanding has motivated our research, which explores the possible optimisation of the HVS to the spatial and chromatic properties of the natural environment, where it is supposed to have evolved. In particular, the questions addressed in the next chapters are the following:

a) Is the foveal achromatic vision in humans optimised for performing ecologically-relevant visual discrimination tasks in the presence of second-order statistics corresponding to natural scenes?

b) What are the differences in terms of performance and optimisation between foveal and peripheral achromatic vision for the same ecologically-relevant discrimination task?

c) Can the psychophysical performance of the HVS for a visual discrimination task in both, periphery and fovea, be replicated by a simple multi-resolution model of local contrast discrimination involving only "low level" visual processes?

d) Is the (physiological) imbalance found between the spatial properties of the chromatic and achromatic mechanisms of the parvocellular pathway reflected in the (statistical) spatial and chromatic properties of natural scenes or is it relevant to some specific task?

1.11. Published work

Parts of this dissertation (Párraga *et al.* 1998c, d; Párraga *et al.* 1999; Párraga *et al.* 2000a, 2002; Párraga *et al.* 2000b) have been published/presented in scientific journals and conferences.

Chapter 2.

Chapter 2

THE PROPERTIES OF SPATIAL INFORMATION IN NATURAL IMAGES AND SPATIAL PROCESSING IN FOVEAL VISION

Foveal spatial vision: Discriminating small changes in natural images

2.1. Overview

The idea that the comprehensive properties of the HVS are not accidents but emerge from a close match between evolution (and postnatal experience) and the environment where this visual system must function, is a fundamental principle of visual science (Atick 1992; Barlow 1961; Field 1987; Laughlin 1983; Marr 1982; Srinivasan *et al.* 1982; Van Hateren 1992a). There is evidence for this in insects (Laughlin 1981; Ocarroll *et al.* 1996; Weckstrom and Laughlin 1995), fish (Lythgoe 1991), and marine mammals (Fasick and Robinson 2000).

Evidence of visual optimisation in mammals comes more from theoretical arguments (Atick and Redlich 1992; Hancock *et al.* 1992; Olshausen and Field 1996; Van Hateren and Ruderman 1998) than from direct observations (Dan *et al.* 1996; Lauritzen *et al.* 1999; Vinje and Gallant 2000). These arguments (based on Shannon's information theory (Shannon and Weaver 1949), a review of some of them was presented in the introductory chapter), are presented as follows (Atick 1992):

a) *Efficiency* (in information theory terms) in the representation of information has evolutionary advantages.

b) We can reformulate the above principle as: much of the early neural processing is *optimised* to exploit the visual environmental structure (i.e. to build efficient representations of the sensory stimuli).

c) A model designed to maximally exploit the structure of the visual environment (within certain biological constraints) can *predict* neural processing.

Therefore, we can use information theory to assess the efficiency of neural information representation. Many authors (Field 1987; Hancock *et al.* 1992; Olshausen and Field 1996; Shouval *et al.* 1997; Wachtler *et al.* 2001) have used this *design principle* to produce models of the neural representation of visual information that are efficient in exploiting statistical structures of the environment ("ecological" approach). For example, Olshausen and Field (Olshausen and Field 1996) have proposed that simple-cell receptive fields properties "emerge" by training a neural network to use a coding strategy that maximises *sparseness* (i.e. finding the smallest number of descriptors can represent any given image from the natural images dataset). By doing this, they implicitly assume that neural processing is optimised to exploit natural image redundancy (the design principle (c)). Nevertheless, to be able to claim that the design principle (c) is valid, we need to prove (b) that the HVS can really see things better when the statistics of the visual environment correspond to those of real natural scenes. This seems like a question with an obvious answer, since we have assumed the efficiency principle (a) to be valid (see introductory chapter (Maynard Smith 1978; Richardson 1994)), but it is not so. The evidence of an evolutionary advantage of efficient representation of the sensory information does not actually prove that the HVS is optimised to process the natural statistics of the visual environment.

Attempts to prove this optimisation were done by Knill *et al* (Knill *et al.* 1990) and Tadmor and Tolhurst (Tadmor and Tolhurst 1994). The former measured the ability of human observers to discriminate changes in the Fourier slope (α) of synthetic scenes and reported that the highest sensitivity was around α values of 1.4-1.8. They suggested that the HVS may be tuned to these (rather "unnatural") statistics. The later did the same measure for *natural and synthetic* scenes. They found that thresholds for the discrimination of changes in α were higher (lower sensitivity) when the image statistics were close to natural statistics and lower when they were not. This is analogous for a blur discrimination task. They interpreted these results as a proof that the HVS is optimised for discriminating natural statistics, since such a system is more tolerant to image-distortions, (such as changes in pupil diameter, accommodative errors, variance in the environmental statistics, etc.) when the image is in focus or has "best quality". However interesting as these results may be, it is not clear whether being able to optimally discriminate between changes in the slope of natural scenes constitutes an evolutionary advantage.

In the present chapter, we will attempt to prove that the HVS is optimised for scenes with natural statistics. To do it we expect to demonstrate experimentally that a human can perform an evolutionary relevant task better in the presence of natural statistics than in the presence of unnatural ones.

Our approach here is a simple one: to psychophysically measure the performance of observers doing a naturalistic visual task in different conditions (with different image datasets) where the statistics of the stimuli are gradually modified to be more or less "natural". If the visual system is "optimised" somehow to see best when the image statistics are "natural", then performance for such naturalistic task should be optimally

higher. To produce our experimental dataset we used a *morphing* technique described in the following chapters.

2.1.1. Why use a morphing technique?

From an evolutionary point of view, it makes sense to be able to discriminate as quickly as possible between familiar or unfamiliar faces or between known and unknown (or possibly threatening) animals. Following this idea, we have developed a naturalistic, form-discrimination task, which we believe is a likely exemplar of this kind of activity. The task requires an observer to distinguish between slightly different monochrome pictures where the component objects differ slightly in shape, position, texture and brightness (see Figure 2.4). At its core, this requires discrimination of small changes in local contrast of the stimuli images. We decided to use a morphing technique (as opposed, for example, to a superimposition of two images to different degrees) because it produces a set of stimuli where each one of the component pictures is an image of a plausible object, sharing the Fourier natural statistics of the original ones. Merging of pictures or of their spectra (Tolhurst and Tadmor 2000) produces intermediate images that can arguably be considered "unnatural" or unrepresentative of the visual environment, since they contain "impossible" (superimposed) objects. Figure 2.1 shows a double-logarithmic plot of the Fourier amplitude for the two reference pictures of the morph sequence car-to-bull (car and bull) and an exemplary intermediate one.

Each point in the figure represents the average Fourier amplitude across all orientations for a given SF. All three plots share remarkably similar second-order statistics (linear plots with very similar slopes on a *log-log* axis, as discussed in the introductory chapter), which indicates that the morphing process does not produce unnatural stimuli in the statistical

sense. Besides, morphing allows good experimental control because it provides a way to introduce very small changes in a sequence (less than 0.1%) in linear steps.

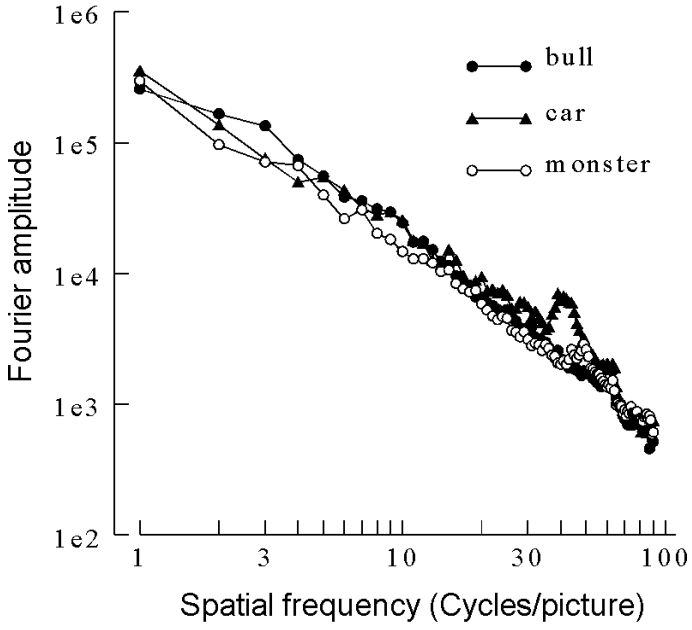

Figure 2.1: Logarithmic plot of Fourier amplitude for three pictures of the sequence car-to-bull. Both reference pictures (car and bull) were compared with an intermediate one (monster). The bump around 40-50 cycles/pic corresponds to the car's radiator.

In this experiment, our observer must discriminate visually between pictures of very similar faces or objects. We wanted to know whether the ability of the observer to perform this task would be degraded if the spatial contrast of the pictures is made unnatural in some way. Figure 2.2 shows an example of the kind of modification to the Fourier spectrum of the pictures that was made in order to make them more or less "unnatural". We decrease (Figure 2.2a) or increase (Figure 2.2b) the slopes of the amplitude spectra of all the pictures within a sequence to alter their second-order statistics.

If performance is optimal when the second-order statistics correspond to those of natural scenes, we may be able to conclude that those statistics have been important in shaping the function of the HVS.

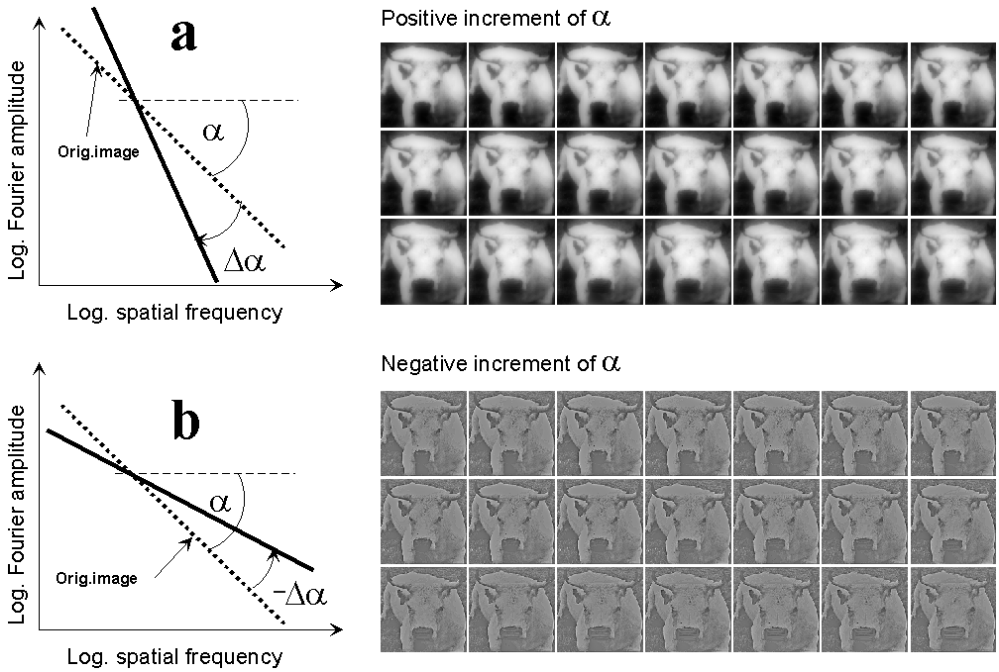

Figure 2.2: Effects of changes to the slope of the amplitude spectrum on part of the car-to-bull sequence. Part a shows the effects of a positive increment of the slope and part b shows the effects of a negative increment. The positive increment produces a picture that resembles a "blurring" and the negative increment resembles a "whitening".

The differences in overall contrast between the two sets of images in Figure 2.2 reflect the limits imposed by the dynamic range of the digital imaging device and the quality of the paper printouts. There are also limits imposed by the dynamic range of the monitor in which the images were presented. The effects of these are discussed below.

There are many ways of altering the natural statistics of digitised images. Another common way of producing "unnatural" visual images is to alter the Fourier *phase* spectrum (Párraga *et al.* 1998b; Tadmor and Tolhurst 1993; Thomson 2001) by filtering it or randomising it. A reason not to apply any filtering or randomisation to the phase spectra of our morph sequences is that *morphing itself modifies the phase spectra* of natural images, since it alters local values of brightness in the picture. By asking observers to perform a visual

discrimination task in this morph/phase domain, while gradually modifying the amplitude spectra, we can manipulate these two dimensions independently of each other.

2.1.2. Effects of eye movements in scene perception

For an observer to be able to perform this task, she must move her eyes and find the region of the image where contrast changes are more salient. This points to an issue arising from eye movements: there is a strong sensitivity loss during rapid jumps of eye position or *saccades*. Saccades can be very quick (up to 500° per second) but our eyes remain relatively still between saccades(Rayner 1998). Typical mean fixation duration times are, for visual search tasks, 275 msec and for scene perception tasks, 330 msec, with mean saccade sizes of 3° for the first and 4° for the second (Rayner 1998). The loss of sensitivity (effect of neural, rather than optical factors) during a saccade is called *saccadic suppression* (Rayner 1998) and is related to the motion (magno) pathway since it occurs mainly for low-SF achromatic patterns (Volkman *et al.* 1978) and not for coloured edges (Burr *et al.* 1994). It has been advocated (Biederman *et al.* 1982; Loftus and Mackworth 1978) that the gist (or essential meaning) of a scene can be extracted very early, sometimes from a single brief exposure, corresponding to the first couple of fixations. There also evidence that our eyes are quickly drawn to informative regions of the scene (Antes 1974; Mackworth and Morandi 1967). "Higher" cortical processes (like memory) are increasingly involved in recognition tasks with increasing number of fixations (Christianson *et al.* 1991; Loftus 1972) and there is evidence that cognitive processing activities are suspended *during* a saccade, although this has not been proved for scene perception (Rayner 1998). To safely ignore any possible effect of saccadic suppression we decided to restrict saccades to the minimum that would allow our task to be performed. These restrictions were spatial, by using relatively small images (2.57º of subtended visual

angle) and temporal, by presenting the stimuli for a short period of time (500 msec). However, the presentation time was long enough for the stimuli to be considered static (Braddick 1973).

2.2. Methods

2.2.1. Visual stimulus creation

In the first experiment, human observers were required to discriminate between pictures of slightly different faces or objects. This was done by producing two sequences from monochrome digitised pictures (128x128 pixels, 256 grey levels), each consisting of 41 pictures of similar size and grey level depth. In the first sequence, two faces (one of a man and one of a woman) were slowly blended (or morphed) together producing a series of plausible intermediate pictures of slightly different faces[13] (Benson 1994; Tolhurst *et al.* 1998). The term "morphing" refers to the creation of a series of intermediate images between a start image and an end image. The simplest way of morphing is called "cross-fading" which is the direct transformation of each pixel of the start image into the corresponding pixel of the end image. During the production of more complex morph sequences, one can direct the spatial transition by linking particular areas of each image with "key points" or "key lines" and by accelerating or delaying the transformation of specified areas. Figure 2.3 shows an example of such "key lines" and "key points".

In the first sequence (created by Dr Phil Benson and called "man-to-woman" here), the shape, contrast and texture vary 2.5% from each picture to the next. The second sequence, (created by the author of this work and shown in Figure 2.4(a)) consists of a morph between an animal (bull) and an object (Morris Minor car). To avoid distortions

generated by the morphing algorithm, the background was digitally removed from both initial pictures. An effort was made to match the salient features of the front of the car (lamps, radiator, the sides of the windscreen, etc.) to the salient features of the bull's face (eyes, nose, horns, etc.). Some implausible objects that were not recognisable either as animate or inanimate were produced for mixtures near 50% car and 50% bull. This sequence is thereafter called "car-to-bull".

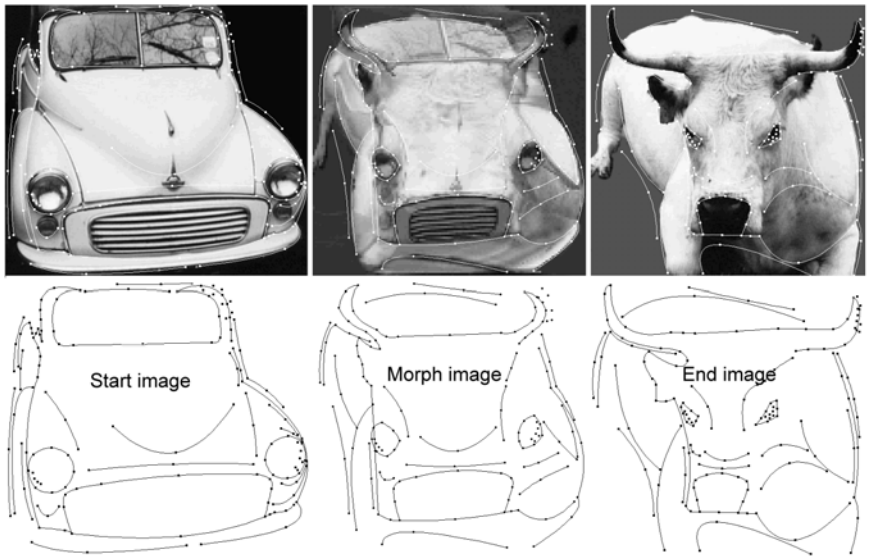

Figure 2.3: Example of the morphing technique: the morphing sequence is produced by determining "key points" and if necessary, joining them with "key lines" in the s*tart* Image. These will move to specific areas in the e*nd* Image. It is necessary to adjust the corresponding key points/lines in the *End* Image by dragging the point to the appropriate position.

Given the greater difference between both original images, observers were able to discriminate between much smaller percentages of change than in the man-to-woman sequence, and so smaller (0.5%) morph steps were created. This is one of the key differences between both morph sequences: spatial changes between the car and the bull

[13] The man-to-woman morph sequence was kindly supplied by Dr Phil Benson.

images are larger and more salient than between the two human faces, therefore the minimum morph step was adjusted to represent this.

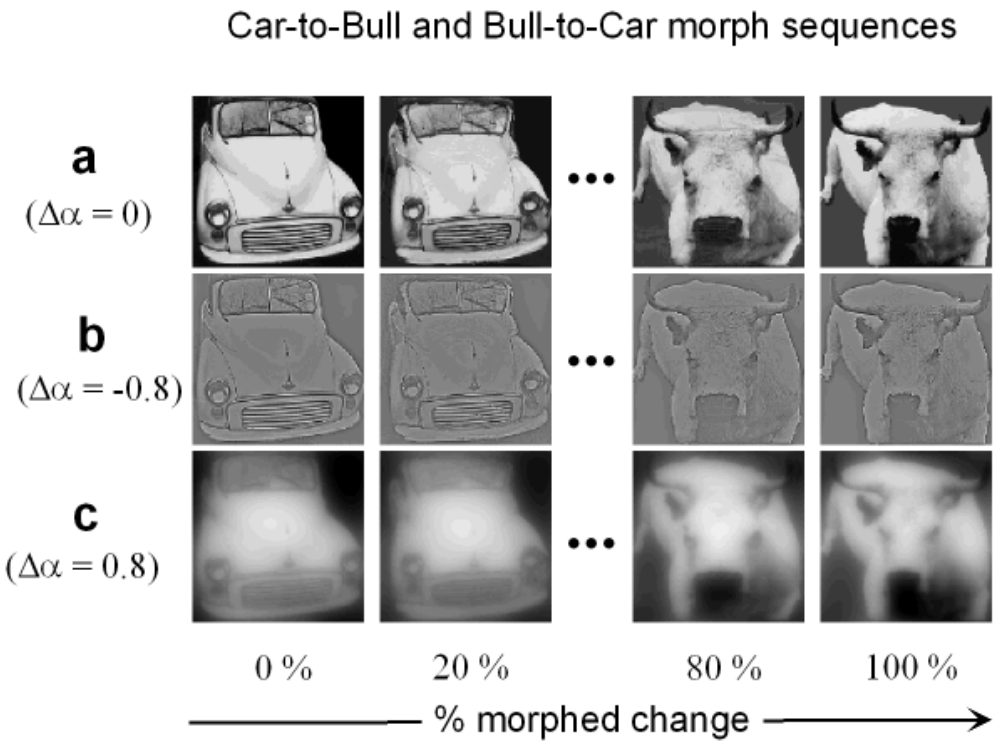

Figure 2.4: Examples of pictures from two of the four morph sequences used in our psychophysical experiments. (a) Original car-to-bull morph sequence in which all pictures have approximately natural statistics. The bull-to-car sequence runs from right to left. (b) Same pictures processed to decrease ('whiten') the slopes of the amplitude spectra. (C) Same for an increase ('blur') of the slopes. Slope offset ($\Delta\alpha$) for each sequence is shown on the left.

"Higher" cortical processes (memory, face features integration (Thompson 1980), etc.) mediate object discrimination tasks such as the one proposed here (Wandell 1995). The middle parts of our morph sequences (near 50% morph change) contain some objects that, although not implausible, are certainly "unfamiliar" to the observer. To reduce the role of higher cortical processing and keep our morph discrimination closer to the ecologically-relevant task of local contrast discrimination, we decided to use the images near the reference pictures of both morph sequences. This means that our subjects were

asked to discriminate, for example, between a "familiar" bull and a slightly fatter version of the same bull and not between a "monster" (see middle picture in Figure 2.3) and a slightly different "monster".

We believe the image components of the morph sequences used in this research are representative of natural world in general. Their amplitude spectra slope is between the range reported by several other workers (Field 1987; Párraga *et al.* 1998a; Tolhurst *et al.* 1992) and their phase statistics were not randomised or altered in a way that would make them statistically different from the original pictures.

Since each sequence can be used in two ways (either the first or the last picture can be the reference picture), in practice we used four different sequences in our experiments. They were: "car-to-bull", "bull-to-car", "man-to-woman" and "woman-to-man".

The digital modifications of the original car-to-bull pictures and morphs were done using commercial algorithms (Paint Shop Pro v5.0 from JascSoftware and Morph Artist v1.0 from Gryphon) in a PC running under Windows 95 operating system.

From both initial morph sequences, further sets were made where the spectral slopes of the component pictures were increased or decreased from the natural value. This was made by taking the Fourier transform of each image, removing the DC (mean amplitude) and multiplying its amplitude spectrum by a filter of the form:

$$Weight(f) \propto f^{-\Delta\alpha}$$

where f is spatial frequency and $\Delta\alpha$ determines the increase or decrease in the slope of the amplitude spectrum. Negative values of $\Delta\alpha$ make the slope of the amplitude spectrum shallower (whitened pictures), therefore transforming the picture as shown in Figure

2.4(b) and positive values make the slope steeper (blurry[14] pictures) as in Figure 2.4(C). The new images were constructed by inverse-Fourier transformation. After the inverse-Fourier transformation, some pixel values become larger than the available range (0-255 greylevels) and had to be scaled to fit within the display limit (127 grey levels on each side of the mean display luminance). This was especially true for large negative values of $\Delta\alpha$ (-0.8 and -1.2). This operation (the scaling of the signal to fit the available dynamic range of the digital image) determines the "pivot point" in the rotation of the Fourier amplitude slope plot (see Figure 2.2).

For example, an extremely "whitened" picture would contain much more energy near sharp edges between objects than in any other areas. This would compress the range of greylevels used by most of the image (and thus, decrease the contrast) and produce bright pixels near salient borders. The effect of these contrast variations on our results will be analysed below. Given that normalising (adjusting the dynamic range of the image to fit that of the display) each component image of the morph sequence would produce variations in contrast within the same morph sequence, we used one normalisation factor for each of the sequences of images at any given spectral slope. Different normalisation factors were used for sequences with different values of $\Delta\alpha$. To avoid edge effects in the Fourier transform and in our morph discrimination experiment, the square edges of each picture were smoothed with a Gaussian roll off mask (SD= 15 pixels). All manipulations were made using a purpose-written algorithm in Turbo Pascal under DOS operating system.

[14] Although this transformation makes the pictures look "blurred", it is not technically a blurring operation since the energy present in the low-SF side of the spectrum is increased.

2.2.2. Experiments

The pictures were presented on a 19' Sony Trinitron monitor driven by a Cambridge Research Systems VSG 2/4 Graphics Card. The screen area was 1024 x 768 pixels. The frame rate was 80 Hz.

The display was calibrated to compensate for luminance non-linearities. This is necessary because in CRT displays, such as the one used in this experiment, the transfer function is non-linear, (i.e. the luminance measured on the screen is not linearly related to the driving voltage). The characteristic transfer function for CRTs has the form:

$$L = k(V - V_0)^{\gamma}$$

Equation 2.1

Where L is luminance, V is the applied voltage; V_0 is the brightness level and γ is about 2.5 for CRT monitors. The process of compensating for these display non-linearities (also called "gamma correction") consists of making a series of measurements across the full luminance range and fitting a function using a standard minimization technique. An inverse transfer function is then derived and applied to compensate for the non-linearities (Carpenter and Robson 1999; Travis 1991b).

Each picture was resized (multiplied by 2) to measure 9 x 9 cm when presented on the centre of the display and subtended 2.57° square to the eyes of an observer situated 2 m away. This ensured that the whole picture was within foveal range (diameter of the fovea = 5.2° (Wandell 1995)). Each of the original picture's pixels subtended an angle of 1.2 arc min. The range of SFs involved in this discrimination task could be estimated from the geometry of the experimental set up (between 0.38 and 24 cycles $\cdot deg^{-1}$, which is within the visible range, although higher spatial frequencies were present at oblique orientations).

The part of the screen that was not occupied by the stimulus had a fixed luminance of 85 cd/m^2, the midpoint of the luminance range of the pictures. The whole monitor subtended 10.28 x 7.43º to the observer.

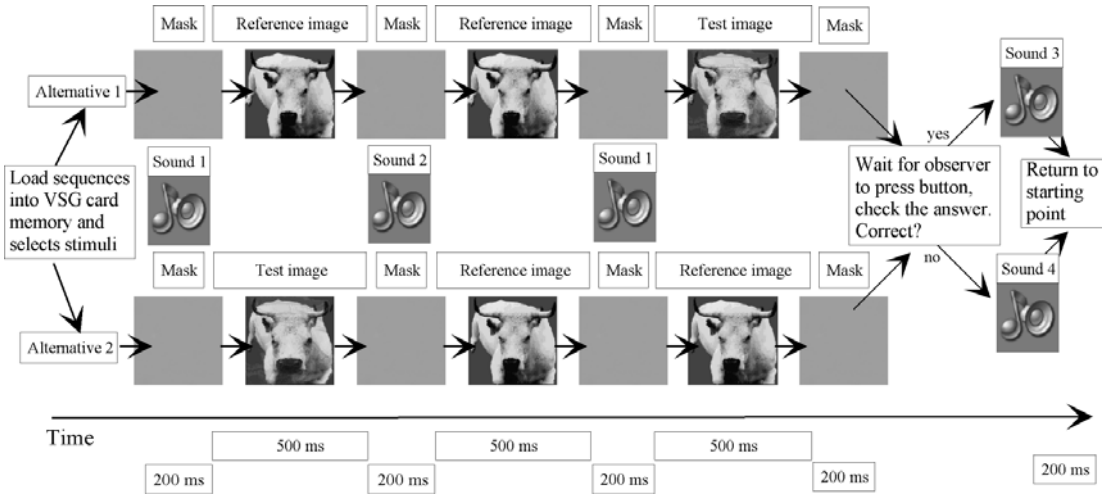

Figure 2.5: Scheme of the experimental protocol. The computer algorithm choses randomly between two alternatives (names as Alternative 1 and Alternative 2). In Alternative 1, the test image is the last one to be presented and in Alternative 2 is the first. The second picture presented is always a reference image. The bottom of the figure shows the timing of the events, which include the production of different sounds to tell the observer whether the answer was correct, etc.

A modified two-alternative forced choice (2AFC) paradigm (Tadmor and Tolhurst 1994) was used to measure the observer's thresholds for discriminating among the morphed pictures in the sequence as follows. One of the original (non-morphed) pictures was presented on every trial and was called the *reference*. Three pictures were presented sequentially. Each one was presented for 500 msec with intervals of 200 msec between them (to reduce the number of saccades and prevent masking). The second presentation always contained a copy of the reference; one of the other two presentations (chosen randomly by the computer) also contained an identical copy of the reference, while the remaining presentation contained a morphed picture (called the *test* picture). The observer had to press the left/right mouse buttons to tell the controlling computer if the morphed

picture was the first/last one. Auditory feedback was given to the observer to tell her whether the choice was correct. The discrimination task was made harder or easier using a conventional staircase technique, in order to find how much morphing was required to just allow discrimination. If the observer correctly identified the presentation containing the morphed picture five times, then a new morphed picture was chosen that should make the discrimination task harder. If the observer made one or more errors in a sequence of five trials, then a potentially easier morphed picture was chosen for the subsequent trials. This continued until 2-3 reversals occurred (normally after about 180 - 200 responses). Two independent staircases were run for each slope. One of the staircases starts with an "easy" pair of pictures and becomes gradually more "difficult". The other starts with a "difficult" pair and goes in the opposite direction. Eventually the staircases meet at the threshold point. A third mouse button allowed observers to repeat the last trial, but they were instructed to use this only in exceptional cases. A complete experiment included seven morph sequences: the original sequence ($\Delta\alpha = 0$) and six spectral-slope modified sequences ($\Delta\alpha$ = -1.2, -0.8, -0.4, 0.4, 0.8, 1.2). Stimuli from the seven sets were interleaved randomly.

Observers sat in a darkened room and looked at the pictures binocularly (they were allowed to fixate freely and to take breaks to avoid tiredness). Free fixation was required to allow observers perform the task, since it involved searching for regions of the scene where local changes in contrast have occurred. The presentation time allowed approximately two fixations to be done in each presentation (the mean fixation duration for a typical visual search task is about 275 msec (Rayner 1998)). A typical experiment (consisting of 7 experimental series, each corresponding to a different Fourier amplitude slope and involving about 180-200 trials) was divided into two sessions that lasted about 45 minutes each. The controlling computer was a 300Mhz Pentium II PC running under

Windows 95 in DOS mode. Both the algorithm running the experiment and the one used to analyse the results were adapted from custom algorithms originally written by Dr D.J. Tolhurst in Pascal programming language.

2.2.3. Data analysis

The results from each single experimental series were plotted as psychometric functions and fitted with the integral of a normal distribution, which was constrained to fall within the range 50% (the guess rate in a 2AFC) to 98% (allowing for a 2% "finger error"). "Finger errors" or "observer lapses" are mistakes committed in repetitive tasks by observers who want to press one button and accidentally press another (see Appendix A for a more complete discussion on the effects of these errors). The slope (β) and position (ε) of the best-fitting cumulative normal were sought with a SIMPLEX routine, which maximised *log*-likelihood (Press *et al.* 1986). The χ^2 values of the fits, estimated as $-2 \cdot log$ (likelihood ratio) were usually lower than the number of degrees of freedom, indicating an acceptable goodness of fit. Threshold was considered to be the magnitude of ε that would allow the observer to correctly identify the interval containing the morphed stimulus on 74% of trials. It was possible to estimate the standard error (SE) of the measured threshold as:

$$SE = \sqrt{-1 \Big/ \frac{\partial^2 L(\hat{\varepsilon}, \hat{\beta})}{\partial \varepsilon^2}}$$

Equation 2.2

where $L(\hat{\varepsilon}, \hat{\beta})$ is the *log*-likelihood function calculated for $\hat{\varepsilon}$ and $\hat{\beta}$, the values of the two parameters which produced the best fit (Edwards 1972). In fact, simulation of such staircases shows that the distribution of $\hat{\varepsilon}$ is slightly skewed, so that Equation 2.2 slightly

underestimates[15] the standard error (see Appendix A for an analysis on how the errors were obtained).

Once the psychometric functions were fitted and the errors calculated, a diagram showing the dependence of the discrimination threshold (measured as a percentage of the morph sequence where 0% was the reference and 100% was the other extreme of the morph) with the slope offset ($\Delta\alpha$) was produced.

2.3. Results

The experiments were performed on two observers (TT and CAP) who were psychophysically experienced and knew about the details of the experiment but the main results were replicated on a naïve observer (IG) who was psychophysically experienced but did not know about the details of the experiment at the time. TT and CAP were presented with four types of reference stimulus: pictures of a bull, a car, a man's face and a woman's face. The bull and the car formed one sequence (called bull-to-car), the man's face and the woman's face another (called man-to-woman). IG was presented only with the man-to-woman sequence. The observer had to identify test pictures that were different from the reference in the direction of the other end of the morph series. Thus, the car will become slightly more like the bull, the bull more like the car, the man's face like the woman's face and vice versa.

Figure 2.6 and Figure 2.7 show the results obtained for observers TT and CAP. The discrimination thresholds in both figures are expressed as a percentage movement through the morph series. The plotted points show the discrimination thresholds and the error bars show the standard errors. The abscissas show the deviation from the "normal"

[15] D.J. Tolhurst and M. Chirimuuta, personal communication.

value of α, the spectral slope (as discussed in Section 2.2.1). A positive amplitude slope offset value ($\Delta\alpha$) means that the slope was steeper from the normal (the scene was "blurred" -Figure 2.2a) and a negative value means that the slope was shallower (the scene was "whitened" - Figure 2.2b). Zero offset means that the picture had "natural" second-order statistics.

We predict that, if natural scenes are optimally encoded, the thresholds at zero amplitude slope offset will be lower than at the other values. As we can see from the figures, the eight experimental sets do show thresholds that are generally lower between amplitude slope offsets ($\Delta\alpha$) of – 0.4 and +0.4. The thresholds rise more or less symmetrically from this lowest point.

The numbers on the ordinate axis differ, reflecting the differences between the morph scenes. For example, thresholds for the man-to-woman and woman-to-man series normally range from 10% to 30% whereas thresholds for the bull-to-car are lower (1% to 5%). This reflects the fact that the car and bull pictures were more dissimilar to each other than the man and woman pictures. Thus, the car-to-bull sequence had to be formed of pictures differing by a smaller percentage from the next than those forming the man-to-woman sequence. There are other justifications for the differences in discrimination thresholds obtained using one sequence or the other. One possible reason is that the car-to-bull sequence was produced by a commercially available software package whereas man-to-woman was produced using a custom-written algorithm. The other possible reason has to do with man-to-woman consisting of human faces, which might be encoded differently by the HVS.

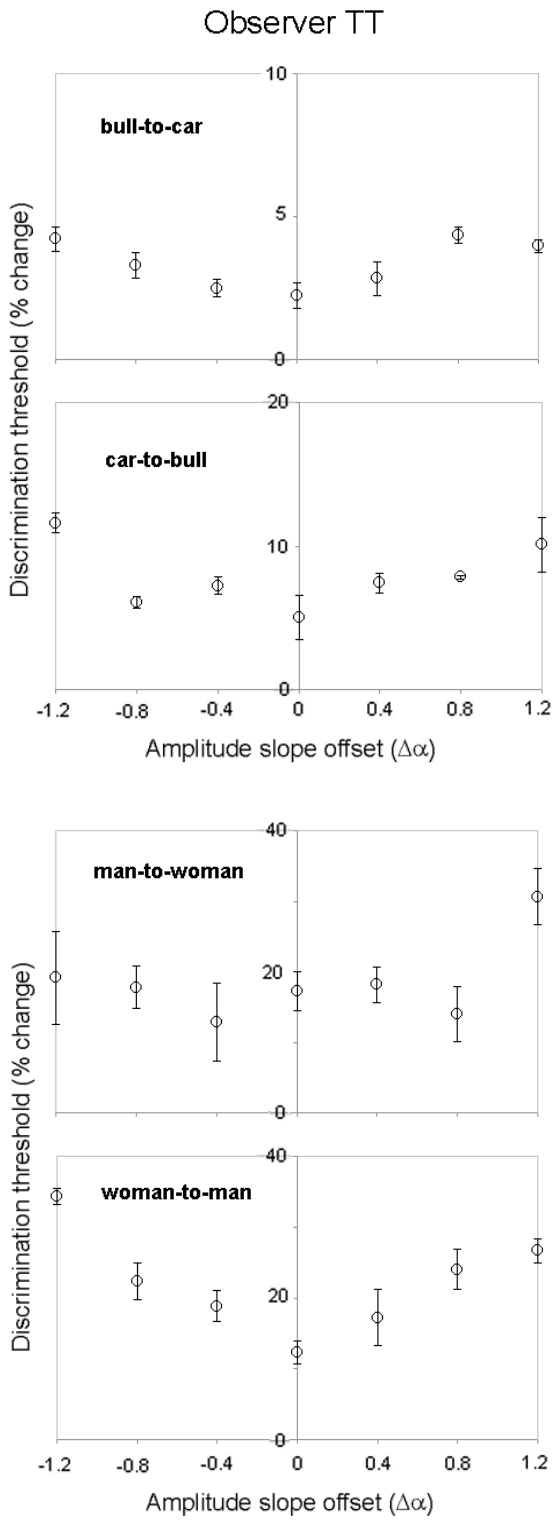

Figure 2.6: Experimental results obtained for observer TT. The plots show the discrimination thresholds for all four morph sequences versus the deviation ($\Delta\alpha$) from the natural value of α. Threshold is expressed as a percentage change through the morph continuum (0% represents one original picture and 100% represents the other).

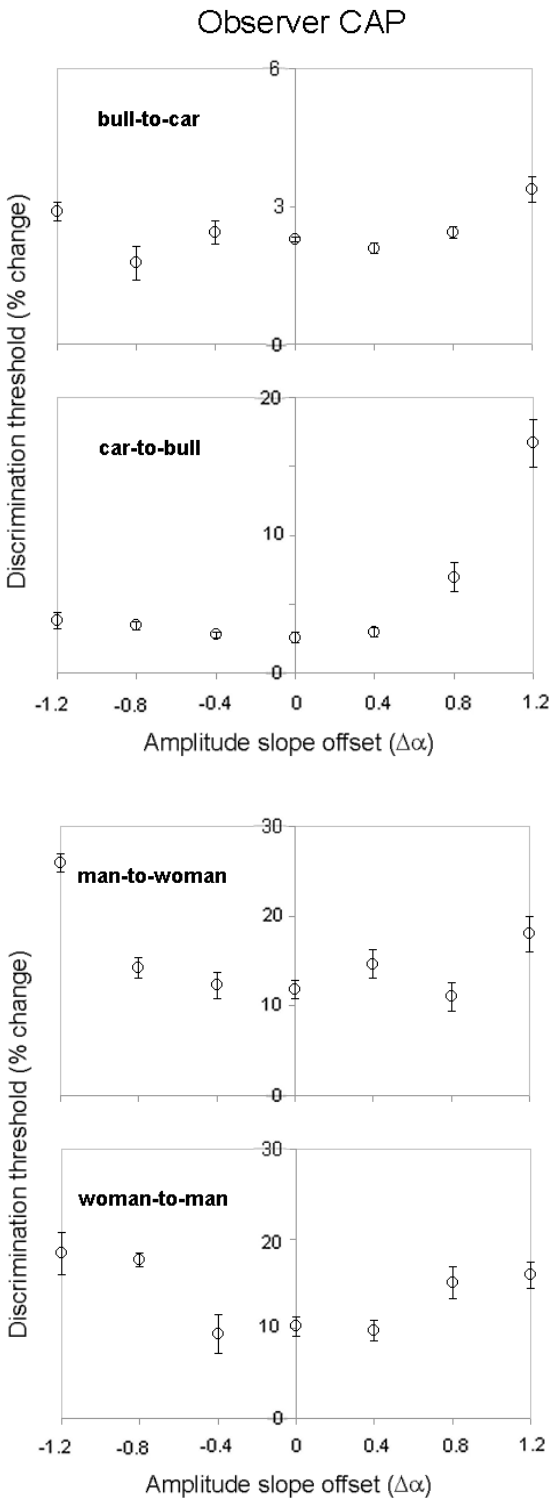

Figure 2.7: Experimental results obtained for observer CAP. Axes are similar to those for observer TT.

Figure 2.8 shows similar results for observer IG, who was naïve about the experimental hypothesis but was an experienced psychophysical observer.

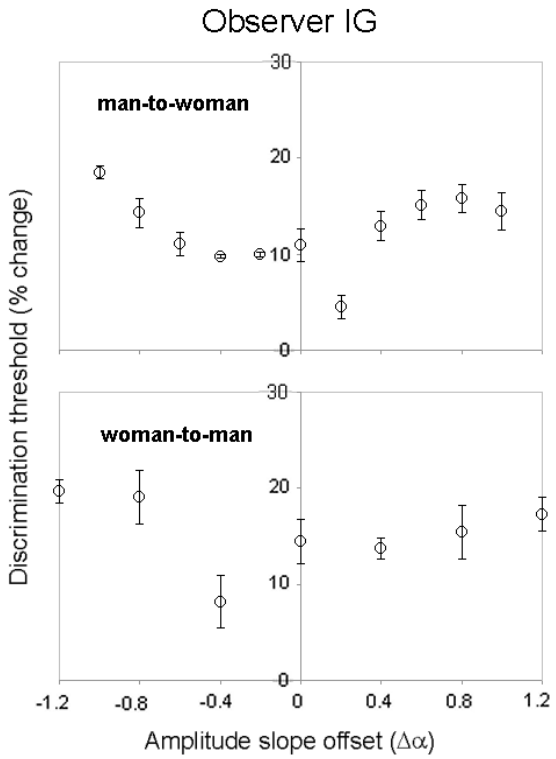

Figure 2.8: Experimental results for observer IG (psychophysically experienced but naïve about this particular experiment). Axis are similar as the above.

To show more formally that thresholds are lowest near zero amplitude offsets, we fitted two linear equations to each dataset and looked at where they intersected. The mean intersection point for pairs of lines was at slope offset ($\Delta\alpha$) of –0.071 (n= 10, SD= 0.425). We also fitted second-order polynomials (Figure 2.9 and Figure 2.10). The mean slope offset to the minima of the fitted polynomials was $\Delta\alpha$ = -0.016 (n = 10, SD=0.165).

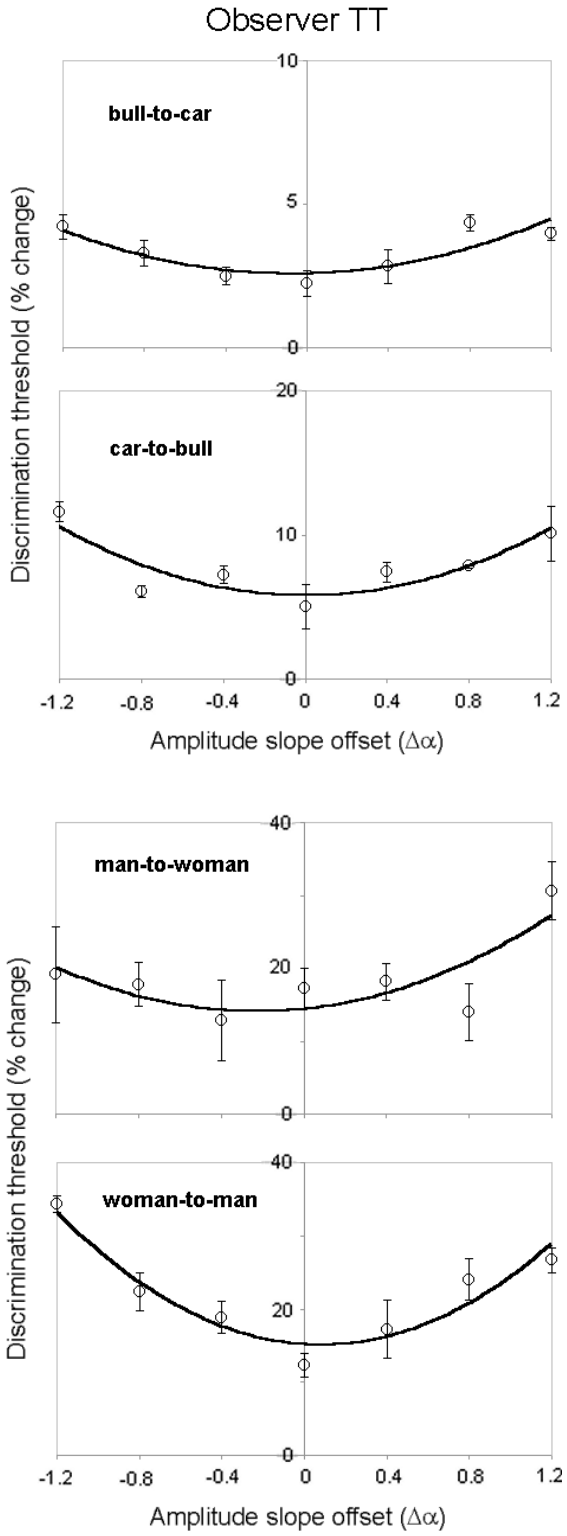

Figure 2.9: Polynomial fittings to results for observer TT.

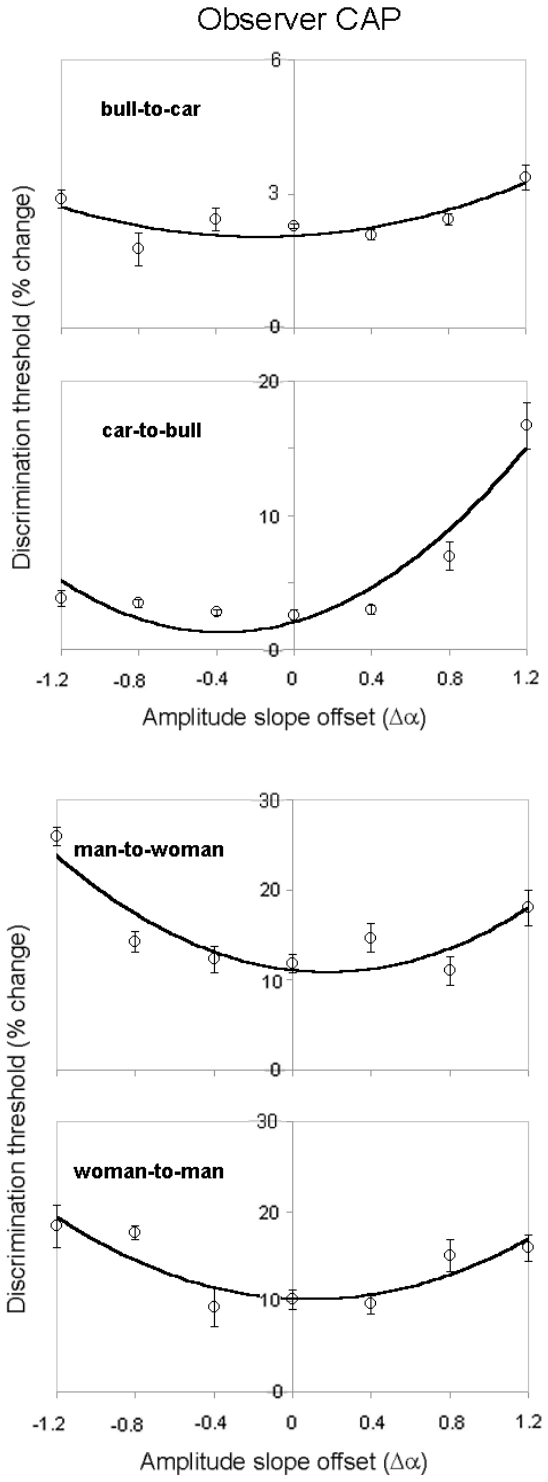

Figure 2.10: Polynomial fittings to results for observer CAP.

Table 2 shows the mean slope offset of the minima for all polynomials and χ^2 for all observers and conditions. The χ^2 of these fits were compared to the χ^2 resulting from the

fitting of a pair of lines. Our results show that the average χ^2 for the pair of lines was 8 times higher than the average χ^2 for the polynomials. A more complete statistical analysis of these results and others is included in the next chapter (Section 3.4.2).

Condition	Subject	Minimum $\Delta\alpha$	χ^2
Man-to-woman	IG	-0.136	2.727
Woman-to-man	IG	0.106	6.292
Bull-to-car	CAP	0.002	12.826
Car-to-bull	CAP	-0.282	34.096
Man-to-woman	CAP	0.173	18.811
Woman-to-man	CAP	0.204	10.342
Bull-to-car	TT	-0.096	15.157
Car-to-bull	TT	-0.044	24.354
Man-to-woman	TT	-0.204	5.145
Woman-to-man	TT	0.116	5.768

Table 2: Mean slope offset of the minima and calculated χ^2 for all observers and conditions.

2.4. Discussion

It is noticeable in Figure 2.2 that the change in the slopes of the amplitude spectra also changes the apparent contrast of the pictures, thus it is possible that our results (optimisation of performance for discriminating morphed images with natural slope) are a trivial result of differences in contrast. To rule out this possibility we performed a control experiment in which we reduced the contrast of the original pictures without any change in the amplitude slope and investigated the effects on the observer's performance for doing the same discrimination task.

In this new experiment, observers had to reduce the contrast of a picture with zero offset ($\Delta\alpha$= 0 or natural statistics) until it appeared to match the contrast of the pictures with

other slope offsets ($\Delta\alpha \neq 0$). For example, in a single trial tha observer is presented with two pictures on the screen, one is the original, unmodified image of the woman's face and the other is a Fourier-modified versions of it ($\Delta\alpha \neq 0$). The subject has then to press the mouse's buttons to change the contrast of the second image so that it matches the contrast of the first. All other conditions were the same as in the normal morph discrimination threshold measurements. The contrast corresponding to the original, unmodified image ($\Delta\alpha = 0$) was considered equal to 1 and all reductions were measured as a fraction of this. The results (see Figure 2.11a) show that the (relative) subjective contrast of pictures of the woman's face is greater for the natural slope than for either steeper or shallower slopes. From the same figure we can see that the contrast of the zero offset picture had to be reduced more to match the more "blurred" or "whitened" pictures.

Figure 2.11b shows the normalised root mean square (RMS) amplitude of the same pictures. RMS is related to the total area under the Fourier power spectra by Rayleigh's theorem:

$$\int |g(x)|^2 = \int |G(f)|^2$$

Equation 2.3

This translates into Parseval's theorem for discrete functions:

$$RMS = \frac{1}{N} \sum_N |g(x)|^2 = \sum_N |G(f)|^2$$

Equation 2.4

In the previous equations, *G(f)* is the Fourier transform of the function *g(x)*. RMS values were normalised so that unity represents the contrast of the image without any

modification (($\Delta\alpha$= 0 or natural statistics) and other contrasts are indicated as a fraction of this. From Figure 2.11 (a and b) we can see that subjective contrast (a) is not closely related to physical power (Troscianko *et al.* 2000a, b) (b). Both figures are plotted on the same scale.

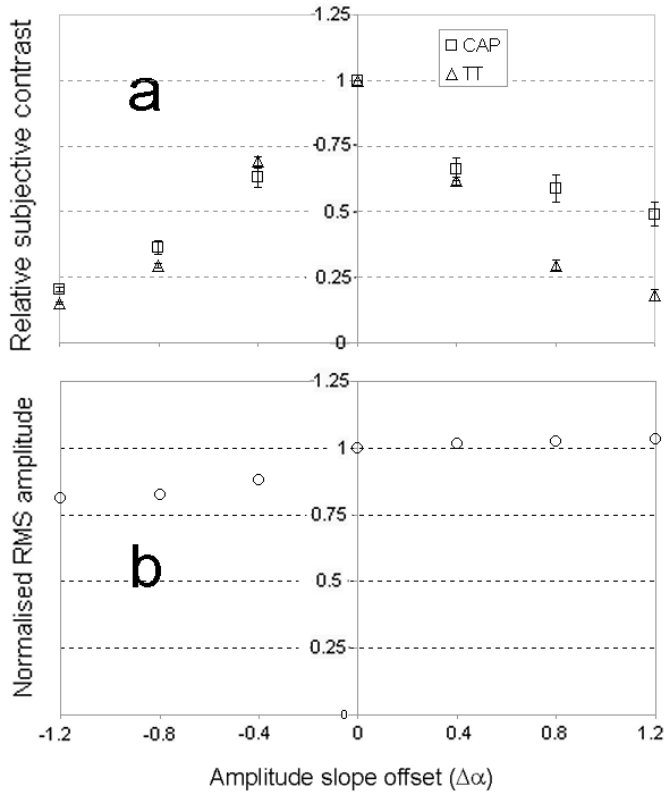

Figure 2.11: Contrast variations across images. (a)The subjects had to reduce the contrast of the "natural" picture of the woman's face to match the "blurred" and "whitened" ones. (b) Normalised RMS of the same pictures.

The reporting of a higher relative contrast by observers is consistent with claims made by other workers. For example, Brady and Field (Brady and Field 1995; Field 1987) proposed that natural statistics (1/*f* power spectra) might allow the visual system to optimally encode the spatial structure of natural scenes, because the responses of cortical cells are the same regardless of the spatial scale (limited dynamic range). This even response across

different spatial scales (i.e. an optimised distribution of neural activity) corresponds to focused natural images. In their view: "…*an image which is "in focus" will have structure (e.g., edges) which has roughly the same magnitude across scale. That is, the loss of high frequency energy in some images is due to the reduction of the number of regions that contain structure rather than the amplitude of that structure. An "in focus" image will have structure (e.g., edges) across scale that have roughly equal magnitude but may vary in the area covered by structure"* (Field and Brady 1997, page 3367). Differences in "perceived blur" as judged by human observers were related to measures of image structure (the "rectified contrast spectrum" or "RCS") (Field and Brady 1997).

The effects of reducing the contrast of the original pictures without any change in the amplitude spectra were investigated by measuring thresholds for discriminating between morphed pictures (with $\Delta\alpha = 0$) at a variety of contrast attenuations in a control experiment. In this new experiment, one of the previous discrimination threshold measurements (woman-to-man series ($\Delta\alpha = 0$)) was repeated at increasingly lower contrast levels. At first, all conditions, including contrast, were similar as before, but later the experiment was repeated (same sequence without any changes in the Fourier spectrum) at contrasts values of 56%, 32% and 20% (for observer CAP) and 51%, 21% and 15% (for observer TT) relative to the original sequence. Figure 2.12 shows the discrimination thresholds for two observers (TT and CAP) as a function of the sequence's relative contrast. Here, a relative image contrast of 1 means that the image contrast remains the same and a relative contrast of 0.5 means a 50% contrast reduction. From the figure, it is possible to see that a modest contrast reduction does not affect the discrimination task. The thresholds rise slightly when the contrast of the morph sequence is attenuated by 75% or more (0.25% relative image contrast in Figure 2.12). This is equivalent to the

subjective contrast of a much "whitened" ($\Delta\alpha$= -1.20) or "blurred" ($\Delta\alpha$= 1.20) image as seen in Figure 2.11a.

The results of the control experiment suggest that the optimisation of the performance for discriminating morphed images with a natural spectral slope is not an artifact caused by differences in contrast.

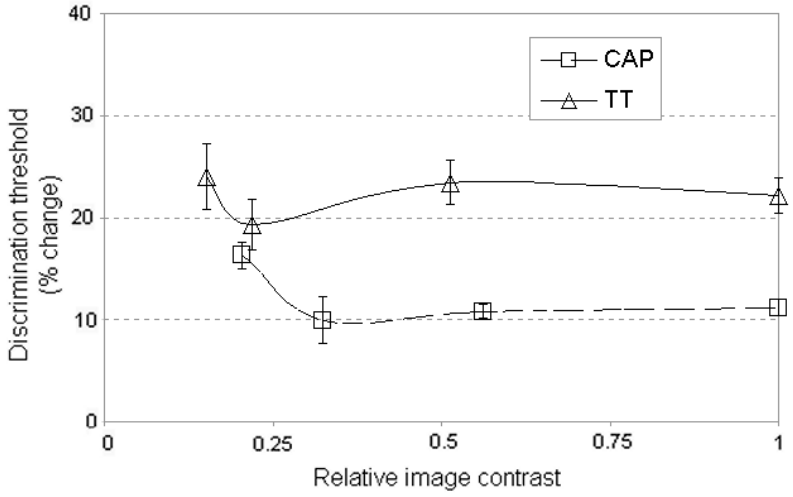

Figure 2.12: Contrast attenuation. Experimental results for two observers who measured their discrimination thresholds at a variety of contrast attenuations (woman-to man series).

One reason why we changed the second-order statistics (or power spectra) to make our images unnatural is that this was especially straightforward. There is however, a more important reason: many natural scenes have been found to have similar spectral slopes (Burton and Moorhead 1987; Field 1987; Tolhurst *et al.* 1992), thus one can imagine this to be a driving force in the shaping of the properties of the HVS.

Other manipulations would make an image unnatural, as for example, changes in the phase spectra. Phase spectra are said (Piotrowski and Campbell 1982) to be more crucial to the appearance of pictures than the power spectra, since it reflects third-order (or

higher) image statistics (Thomson and Foster 1997). According to this view, a manipulation of the third-order statistics might make our morph sequences even harder to distinguish than the changes in the power spectra reported here. However, Tadmor and Tolhurst (Tadmor and Tolhurst 1993) have pointed out that the global amplitude spectrum is also essential for specifying the particular content of natural images (see Section 1.8.5). Our results also support this view, given that changes in the amplitude spectra of natural scenes increased the difficulty in performing our morph discrimination task. This does not contradict the evidence (Piotrowski and Campbell 1982) showing that changes in the phase spectra produce more dramatic effects on the appearance of natural scenes than changes in the amplitude spectra. There is a case for saying that both, the amplitude and the phase spectrum may be important for reliable specification of such images.

2.5. Conclusions

a) We argue that morph discrimination as a visual task is both evolutionary relevant and relatively easy to control by incremental steps.

b) Our results for this task show that the HVS performs optimally when the second-order statistics of the images are natural. This optimisation is shown in the form of morph discrimination functions that are U-shaped and can be fitted by a second order polynomial, which has a minimum value corresponding to images with "natural" second-order statistics.

c) The fact that observers report a higher subjective contrast for the natural images than for the ones with unnatural slopes, is consistent with the proposal that the amplitude spectra of natural scenes might be most

appropriately sampled by neurons with natural bandwidths (Brady 1997; Brady and Field 1995; Burton and Moorhead 1987; Field 1987; Field and Brady 1997).

Chapter 3.

Chapter 3

THE PROPERTIES OF SPATIAL INFORMATION IN NATURAL IMAGES AND SPATIAL PROCESSING IN PERIPHERAL VISION

Peripheral spatial vision: discriminating small changes in natural images

3.1. Overview

The majority of the research on the HVS is dedicated to the central 1% of the visual field. This is understandable since central vision is extremely important for everyday life, but it means that the characteristics of peripheral vision and its mechanisms are poorly understood in comparison with central vision.

In general terms, the HVS is far more deficient when processing information in its periphery than it is in its central part and this deficiency increases with eccentricity. Early work (Aubert and Förster 1857; Wertheim 1894) established that visual acuity falls inversely with eccentricity of target location from the centre of the visual field, and that iso-acuity contours (contours of similar two-point resolution), are horizontally-elongated ovals displaced by increasing amounts into the temporal visual field with increasing eccentricity. The conceptual framework for relating these topological features of visual resolution to the anatomy and physiology of the retina was developed by Helmholtz (Von Helmholtz 1911).

The fact that the size of the receptive fields tends to increase with the distance from the fovea was noted by Hubel and Weisel (Hubel and Wiesel 1960, 1974). They suggested that

'*differences between central and peripheral visual acuity in man may well be related to variations in receptive-field central size'*. Contrast sensitivity measures (Kelly 1984; Robson and Graham 1981; Rovamo and Virsu 1979; Rovamo *et al.* 1978), show that our sensitivity to high-SFs also decreases with increasing retinal eccentricity but a re-scaling approach was proposed to compensate for this. Under this point of view, if the size of targets is expressed in terms of the *cortical dimensions* (or amount of cortical representation corresponding to the retinal space) rather than visual space dimensions, visual performance is invariant across the visual field. In practice, visual targets were magnified in the periphery in order to compensate for the reduced size of the receptive fields. The argument for re-scaling is based on both the increasing size of the receptive fields (Rovamo and Virsu 1979) and the decreasing amount of cortical processing devoted to the visual field (Brindley and Lewin 1968; Daniel and Whitteridge 1961; Van Essen *et al.* 1984) with increasing eccentricity. This approach has been relatively successful (Rovamo and Virsu 1979; Rovamo *et al.* 1978; Virsu *et al.* 1987; Virsu and Rovamo 1979); especially because it is possible to produce similar contrast sensitivity functions at any part of the visual field, providing that the stimulus is re-scaled (see Figure 3.1).

Several studies revealed that there might be a wide choice of threshold gradients, many of which are substantially steeper or flatter than the proposed inverse cortical magnification (Klein and Levi 1987; Pointer 1986). For example, gradients for grating acuity (Klein and Levi 1987) are steeper than for unreferenced motion tasks (Levi *et al.* 1984) and flatter than for letter acuity (Virsu *et al.* 1987), vernier acuity (Levi *et al.* 1985) or relative motion tasks (Levi *et al.* 1984). In general, when tasks require local analysis, performance in the periphery is worse than when the task require some global analysis.

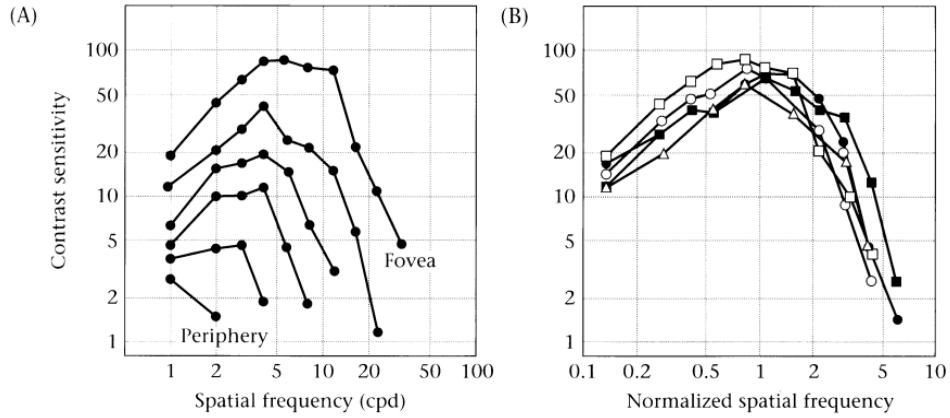

Figure 3.1: Part a: variation of the CSF with eccentricity. Part b: same measures with the test stimuli scaled in size and spatial frequency to compensate for the cortical area devoted to processing. Curves correspond to 0, 1.5, 4, 7.5, 14 and 30 deg. From Rovamo *et al*, 1978 (Rovamo *et al.* 1978).

The large variance between observers' performance, the variations across tasks and the general agreement about the decrease of neural representation with eccentricity in the periphery, does not make it possible to attribute the decline in performance to any simple anatomical structure (Wandell 1995). Tolhurst and Ling (Tolhurst and Ling 1988) suggest that *'there may indeed be more than one useful scaling procedure, each reflecting a different mode of visual processing and a different aspect of the organisation of the visual system'.*

3.1.1. What the morph discrimination experiment may tell us about peripheral vision?

In the previous chapter, we argued that the human visual system is optimised to encode the second-order statistics of the visual environment. We measured the human discrimination thresholds for small spatial changes in stimuli with natural and unnatural Fourier statistics, using a 'morphing' technique, assuming that the subject can direct his/her gaze at the region of interest so as to project this into the fovea. We know that peripheral vision differs from foveal vision in the representation of spatial information,

but it is less clear how this would affect our ability to perform naturalistic tasks (such as the one considered in our experiment). We believe that repeating the same morph experiments for peripheral vision may improve our understanding of the visual processing occurring in the periphery. For example, if peripheral vision is optimised to encode the second-order statistics, as is the case of foveal vision, the minima of the U-shaped discrimination function should coincide with natural stimuli. If we assume that the human CSF is the envelope of a certain number of SF-channels, and these "signal" to the brain when they detect a change in the morph sequence, then "blurring" or "whitening" the signal may reduce the number of signalling channels. For example, a flat discrimination threshold vs. $\Delta\alpha$ plot (as opposed as the "U-shaped" ones found in the previous chapter) may indicate that those channels that are involved in the discrimination task remain unaffected by changes in the distribution of energy. Since the "blurring" and whitening were extreme, this is unlikely. However, if the number of SF-channels involved in the task decreases, one might expect that changes in the distribution of Fourier energy may upset the response of the system in a greater manner, thus provoking larger discrimination thresholds (and less flat "U-shapes"). In summary, a measure of the flatness (or "tuning") of the discrimination threshold vs. $\Delta\alpha$ function may relate to the number of SF-channels involved in the task (a broad tuning indicate more channels and fine tuning indicates less channels). If the same number of SF-channels is active, then the U-shape of the discrimination function should remain the same.

Regarding the absolute values of the discrimination thresholds, we expect that, since observers' performance on the previous experiment is based on *local* discrimination of contrast variations in the images, thresholds should be increase faster with eccentricity than thresholds for normal CSF measures (which requires no *local* analysis). We expect this

decrease on observers' performance across eccentricity to be closer to that experienced for vernier acuity tasks (Levi *et al.* 1985; Westheimer 1982).

The effect of stimulus re-scaling in the periphery for naturalistic tasks is also not very clear. Rovamo *et al* (Rovamo *et al.* 1997) investigated the effects of scaling on observers' abilities to perceive small amounts of geometric change in human faces in peripheral vision. They claim that performance could be made equal across the visual field using the appropriate enlargement. In their experiments, a decrease in performance at 1.73-2.45° of eccentricity could be compensated by a magnification of twice the foveal size of the stimuli. This agrees with previously measured values for vernier acuity, orientation discrimination, etc.

Melmoth *et al* (Melmoth *et al.* 2000) investigated the effects of contrast and size scaling on face perception in foveal and peripheral vision. They claim that failures of spatial scaling may simply reflect the need to compensate for task-specific reductions of efficiency (i.e. reduction of the signal-to-noise ratio) which can be done by increasing the stimulus contrast. Other authors (Thorpe *et al.* 2001) have investigated the performance for the detection of animals in natural images using far peripheral vision. They reported linear decreases in accuracy with eccentricity up to about 70.5°, where detection was still possible (60.5% correct answers). They have shown that even high-level visual tasks can be performed using the poor spatial information provided by the peripheral retina.

The compensatory effects of re-scaling are linked with "simple" cortical processes and anatomical structure. Obtaining the same discrimination thresholds for foveal stimuli and M-scaled peripherally-viewed stimuli may indicate a correlation between this task and the anatomical structure of the cortex.

For these reasons, we decided to extend our study to peripheral viewing.

3.2. Methods

We treated this set of experiments as an extension of the ones in the previous chapter. Subjects were recruited and trained to fixate so that the targets were peripheral to those in the previous chapter. All conditions were similar unless stated otherwise. Observers' acuity was tested using a letter chart to confirm that it fell within the normal range. Their "better" eye was selected as being the eye that they normally used for pointing, shooting or looking through a telescope, for example.

3.2.1. Visual stimuli and experiments

The same sequences of visual stimuli as described in section 2.2.1 were used. These were car-to-bull (cb), bull-to-car (bc), man-to-woman (mw) and woman-to-man (wm) with the same values of $\Delta\alpha$. They were displayed in a similar fashion; using the same equipment and the same blurring mask to avoid edge-interference.

A morph sequence of natural scenes was presented to observers in modified 2AFC experiments. They were asked to discriminate between reference (original) images and a slightly morphed version of these. Observers were asked to fixate either freely (for foveal viewing) or upon a red light-emitting diode (LED) at 3 or 6 deg along the horizontal axis to the side of the screen. Images were viewed on the temporal visual hemi-field (i.e. retinal images were projected on the *nasal* hemi-field) in all cases. Discrimination thresholds were obtained in the same way as before. The statistics of each morph sequence were manipulated by controlling the falloff of Fourier amplitude with SF (α), and thresholds

for morph sequences with different α values were measured. In this case there were four different experiments depending on the eccentricity and size of the target:

a) *Experiment 1:* viewing distance: 2 m. Eccentricity: 0 deg (i.e. foveal viewing). Target size: 9 x 9 cm (i.e. it subtended about 2.6 deg square), monocular viewing. Number of subjects: 5.

b) *Experiment 2:* similar to experiment 1 except that subjects were instructed to fixate on a target that was 3º off-centre to the temporal side of the visual field (eccentricity: 3º. Temporal). Number of subjects: 5.

c) *Experiment 3:* similar to experiment 2 except that subjects were instructed to fixate on a target that was 6º off-centre (eccentricity: 6º. Temporal). Number of subjects: 5.

d) *Experiment 4*: similar to experiment 1 except that subjects were asked to fixate freely and pictures were reduced in size by a factor of 0.369 (target size: 3.3 x 3.3 cm). Number of subjects: 2 (see below for a justification of the scaling factor).

e) *CSF measurement:* this was done using the same methods as in the other experiments. Vertical sinusoidal luminance gratings were presented instead of pictures. The "reference" picture in this case was a blank (midgrey) mask and the observer was instructed to press the button indicating in which of the two intervals the visible grating was. The contrast of the visible grating was thus reduced/increased (following two interleaved staircases) until the threshold was reached. Psychometric functions were fitted. A similar Gaussian roll-off was applied to minimise edge effects in all experiments.

The CSF measurements were repeated for foveal, 3°, 6° and foveal-small conditions.

In all cases, the stimuli were viewed monocularly so that the temporal retina could be assessed independently of the nasal retina. We ensured that in all cases, no part of the stimuli reached near the 10º-15° eccentricity area, due to interference of the blind spot.

A phenomenon was observed in 1804 by Troxler (Troxler 1804) who noticed that objects in peripheral vision tended to fade in visibility if a steady gaze was held. This is a property of a system which codes dynamic events for the sake of efficiency. The more eccentric the object in peripheral vision, the quicker it fades when fixation is constant. This effect is not noticeable in foveal vision due to regular small eye movements, which stimulate new retinal cells and prevent fading of the object (Carpenter 1988). Observers were therefore encouraged to occasionally stop between trials and move their eyes around the fixation point to reduce the 'Troxler effect'. After the experiment, psychometric data was screened for the presence of "secondary peaks" which may represent "glimpses" of the peripherally viewed image. Experiments were repeated in such (unusual) cases.

To avoid the effects of 'binocular rivalry' subjects were instructed to take breaks when needed and to press the mouse's middle button to repeat a trial when rivalry had occurred. Binocular rivalry is a product of monocular viewing, which entails the input from the covered eye becoming temporarily more dominant than the uncovered eye. The frequency of this phenomenon increases as the uncovered eye becomes tired.

As before, both the algorithm running the experiment and the algorithm used to analyse the results were adapted from custom algorithms originally written by Dr D.J. Tolhurst in Pascal programming language.

3.2.2. Cortical magnification factor and M-factor

The original aim of experiment 4 was to compare the results obtained for central (foveal) vision with those of peripheral vision using a compensatory magnification factor. Preliminary results showed that, when the peripheral image was magnified, subjects were often concentrating on changes occurring near the edge of the picture that was closer to the central part of the visual field. To avoid these artifacts and to encourage subjects to pay attention to changes occurring in the whole picture instead of its borders, we decided to compare the results obtained for 6 º peripheral viewing (see experiment 3) with those obtained for central viewing of images *reduced* by a similar factor instead (see experiment 4).

The M-factor values were calculated from to the following equation:

$$M = (1 + 0.29E + 0.000012E^3)^{-1} M_{fov}$$

Equation 3.1

where M is the cortical magnification factor (in mm/deg) and E is visual eccentricity (in degrees). It was made in agreement to values of ganglion cell density (D) (Rovamo and Virsu 1979; Rovamo *et al.* 1978). Estimations of cortical representation of the fovea (M_{fov}) by the same authors give a value of 7.99 mm/deg (Rovamo and Virsu 1979). The cortical representation at 6º (M_{6deg}) eccentricity (according to Equation 3.1) is 2.91 mm/deg. The scaling factor for the 6º eccentric retinal position used in this study was calculated by dividing the foveal value of M (M_{fov}) by the eccentric value (M_{6deg}).

$$C = \frac{M_{fov}}{M_{6deg}} = 2.7$$

Equation 3.2

This number represents the ratio of cortical projections between to the two retinal areas (corresponding to the fovea and 6°). Since there is not a general agreement, the values of *M* obtained here may not coincide with those used by other workers (see Drasdo (Drasdo 1991) for a comprehensive review of the literature). Even if the average value of *M* in humans can be established with reasonable precision, it could at best coincide with only one of the diverse values that have been reported for different visual tasks (see discussion at the end of the chapter). There have been reports (Rovamo *et al.* 1997) of an equivalent decrease in performance in tasks such as detection of geometric image distortions on facial features, vernier acuity, orientation discrimination and curvature detection and discrimination. In the present study, we decided to use Rovamo and Virsu's measures because of the physical limitations of the screen and graphics card resolution.

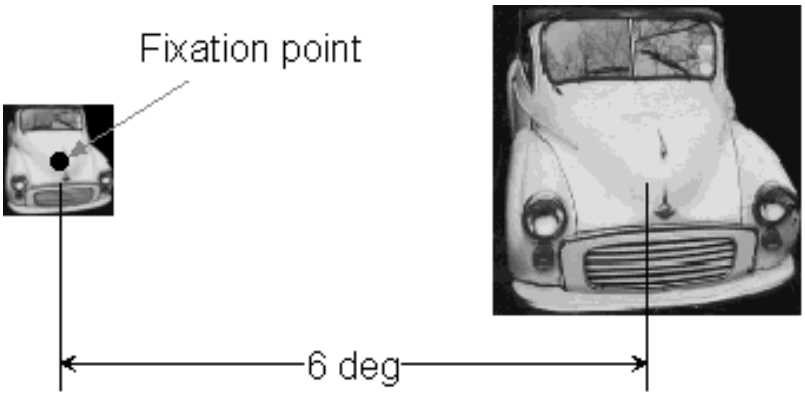

Figure 3.2: Relative sizes of the images present in experiments 3 (picture of the car on the right) and 4 (smaller version of the same picture, on the left). The small picture was presented foveally and the large one at 6º eccentricity. The fixation point was present only in experiment 3 and consisted of a red LED light.

Since, in practice, we are reducing the size of the foveal picture instead of magnifying the size of the eccentric one, foveal pictures presented in experiment 4 were (in height and width) multiplied by $1/C = 0.37$.

The reduction in size of the stimuli was achieved by using the customary algorithms provided by the Graphics Card manufacturers (Cambridge Research Systems, UK). This did not affect the subjective quality of the stimuli (see Figure 3.2).

3.2.3. Data analysis

The results were analysed as described in section 2.2.3. Psychometric functions were fitted to each experimental series in each of the four experiments and the relationship between discrimination threshold (% morph) and $\Delta\alpha$ was plotted. In order to evaluate how this relationship may differ in the periphery from the one that characterises foveal data, second-order polynomials were fitted to the data plots. The position of the minima in the four types of experiment was measured. The second derivative of the second-order polynomials was calculated to see if there were any qualitative differences between the different functions. The value of the second derivative shows the rate of change of the slope of the polynomial, which is related to the "shallowness" of the curve. The smaller the value of the second derivative, the more "shallow" or "flat" the curve fitting is. As discussed before, the shallowness of the curve is also a measure of the tuning of the visual system to the image statistics: broad tuning (i.e. more SF-channels are "active") is related to a small value of the second derivative and fine tuning (i.e. less channels are "active") will be related to a high value of the second derivative. Qualitative differences in tuning were measured.

3.3. Results

3.3.1. Overview

The results refer to the four experiments described above; each one of them was performed with four *conditions*, corresponding to the morph sequences (named mw, wm, bc and cb). There was also the measurement of the CSF, which was done with four conditions, corresponding to the eccentricities and sizes of each of the morph stimuli.

The experiments were performed on seven observers (all postgraduate students or members of staff at Bristol University). All observers but CAP were naïve to the purposes of the experiment and two of them (CAP and BO) were psychophysically experienced. Because of availability and time constraints, not all experiments were performed on all subjects, but two subjects (KB and CAP) completed all four experiments and the CSF measurement in all conditions.

Table 3 shows the details of all experiments. The first column shows the initials of each of the subjects and the first row shows the corresponding experiment. The second row details each of the conditions. A cross represents a completed experiment.

Given the amount of data collected and the available space, we chose to show the complete results from only two subjects (KB and CAP). They are the ones that completed all experiments in all conditions and are representative of the rest of the measurements. However, all statistics were calculated using the whole collection of results.

	Experiment 1 Foveal				**Experiment 2** 3 deg				**Experiment 3** 6 deg				**Experiment 4** Foveal - small				CSF measurement			
	mw	wm	bc	cb	mw	wm	bc	cb	mw	wm	bc	cb	mw	wm	bc	cb	foveal	3 deg	6 deg	small
KB	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
CAP	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
BO	x	x	x	x	x	x	x	x			x	x				x	x	x	x	
TW	x	x	x	x	x	x	x	x			x	x					x	x	x	
JB	x	x	x	x	x	x	x	x	x	x	x	x					x	x	x	
AG	x	x		x	x	x		x									x	x		
IM	x	x			x	x														

Table 3: Summary of all the subjects, experiments and conditions. The first column shows the initials of the subjects, the first row shows which experiment and the second row shows the conditions. A cross means that the given experiment was completed.

Figure 3.3 shows the results for experiments 1 and 2 for subject KB for all four morph sequences. Similar morph sequences are placed side by side to make comparisons easier. The abscissas show the variation ($\Delta\alpha$) in the slope of the amplitude spectrum and the ordinates the percentage of change in the morph sequence. The *y*-axis labels were chosen to keep similar scales in adjacent plots. Continuous curves represent the second-order polynomials that best fit all the data. Different y-axes were chosen to best show the data, given that the ranges of discrimination thresholds vary from one condition to another (e.g. it was necessary to use a much smaller percent step to discriminate the car from the bull than the man from the woman).

The results for experiments 3 and 4 (observer KB) are shown in Figure 3.4. A second curve (made of broken lines) shows the same fit considering only part of the data points (see below). Results for the other main subject (CAP), considering the same experiments

and conditions, are shown in Figure 3.5 (experiments 1 and 2) and Figure 3.6 (experiments 3 and 4).

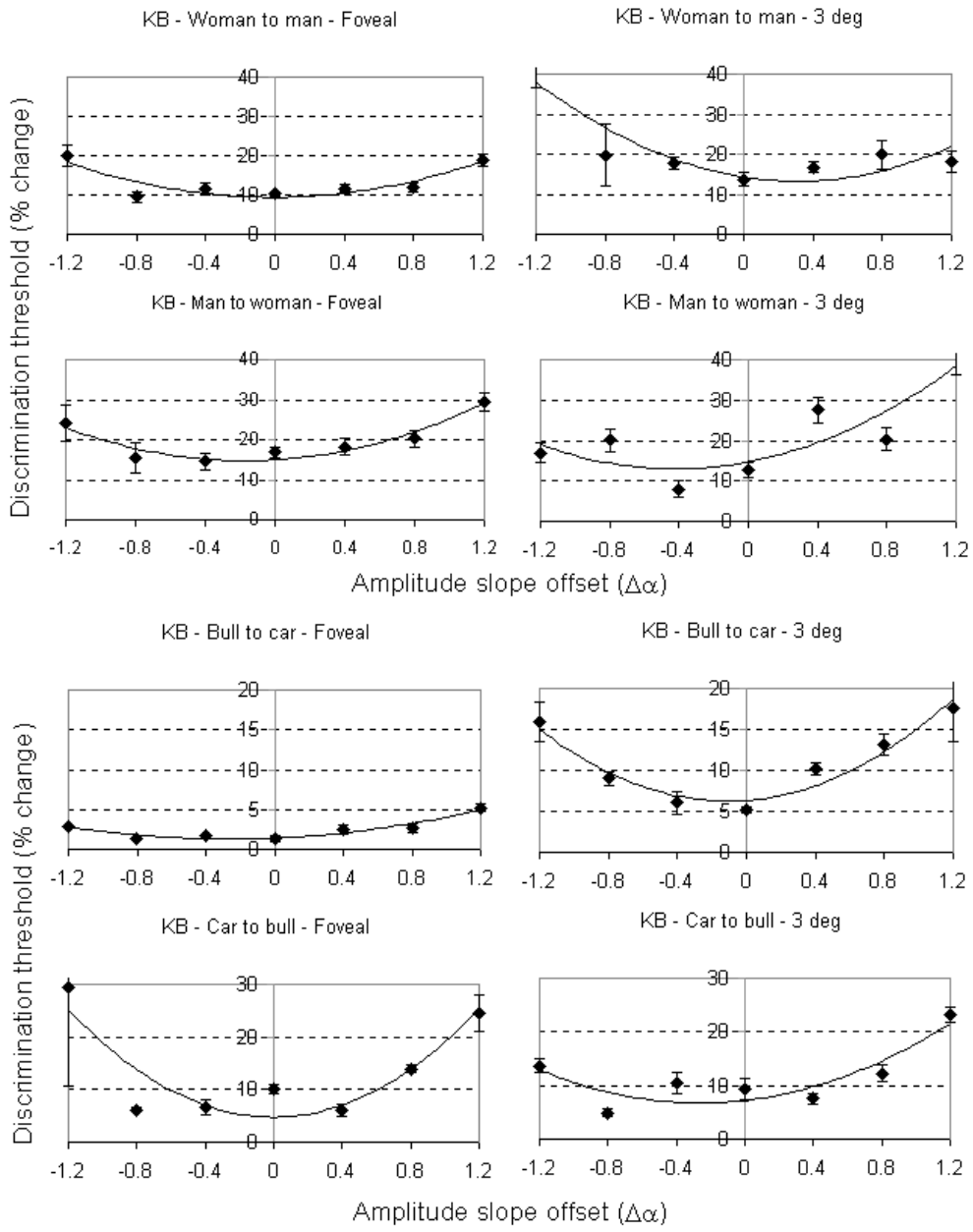

Figure 3.3: Results for experiments 1 (left side) and 2 (right side) for subject KB in all four conditions. The abscissas show the variation ($\Delta\alpha$) in the slope of the amplitude spectrum and the ordinates the percentage of change in the morph sequence. The curve represents the second-order polynomial that best fits the data.

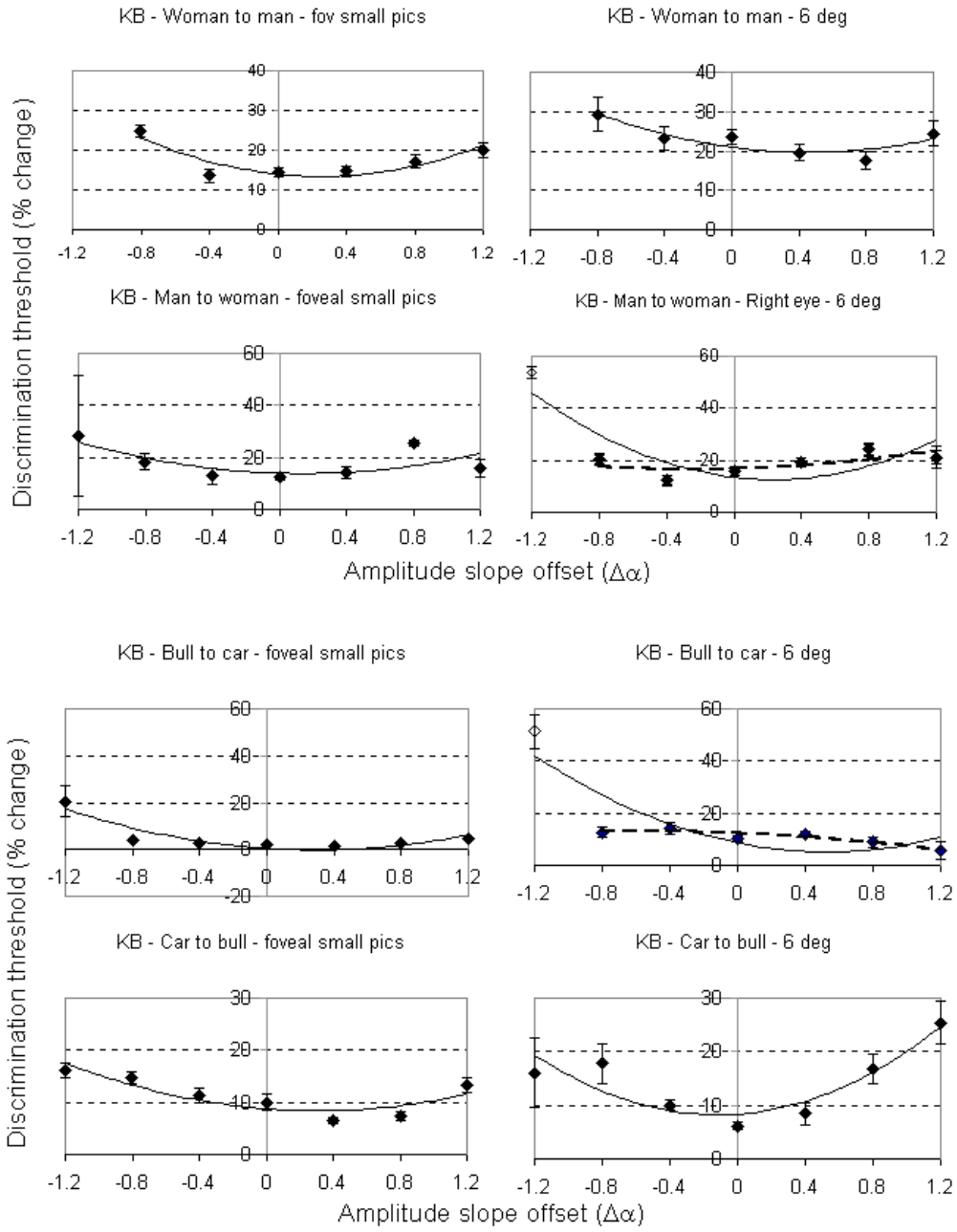

Figure 3.4: Results for experiments 3 (left side) and 4 (right side) for subject KB in all four conditions. The continuous curve represents the second-order polynomial that best fits the data. The other curve, made of broken lines, shows the same fit with less data points.

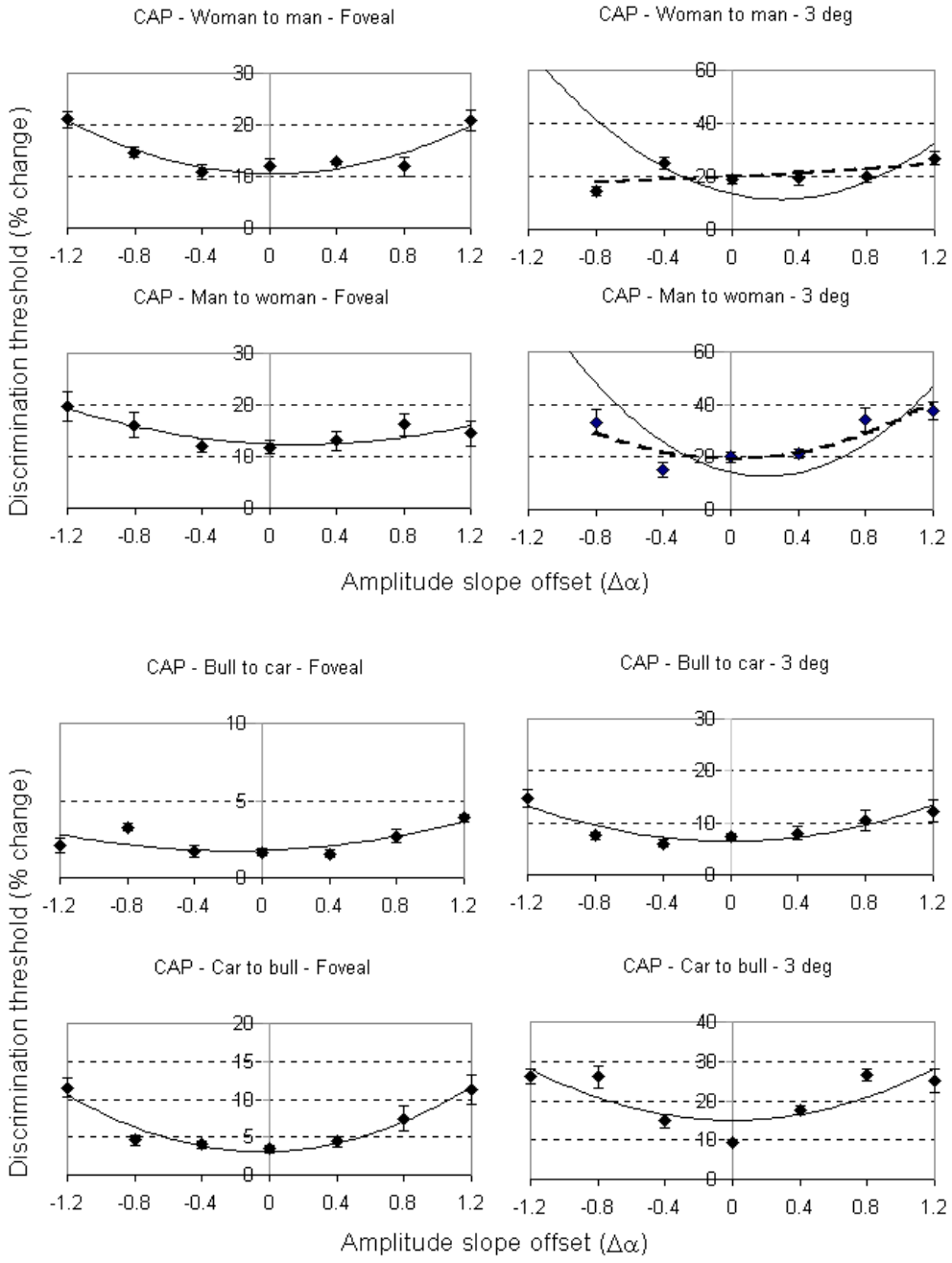

Figure 3.5: Results for experiments 1 (left side) and 2 (right side) for subject CAP in all four conditions. The continuous curve represents the second-order polynomial that best fits the data. The other curve, made of broken lines, shows the same fit with less data points.

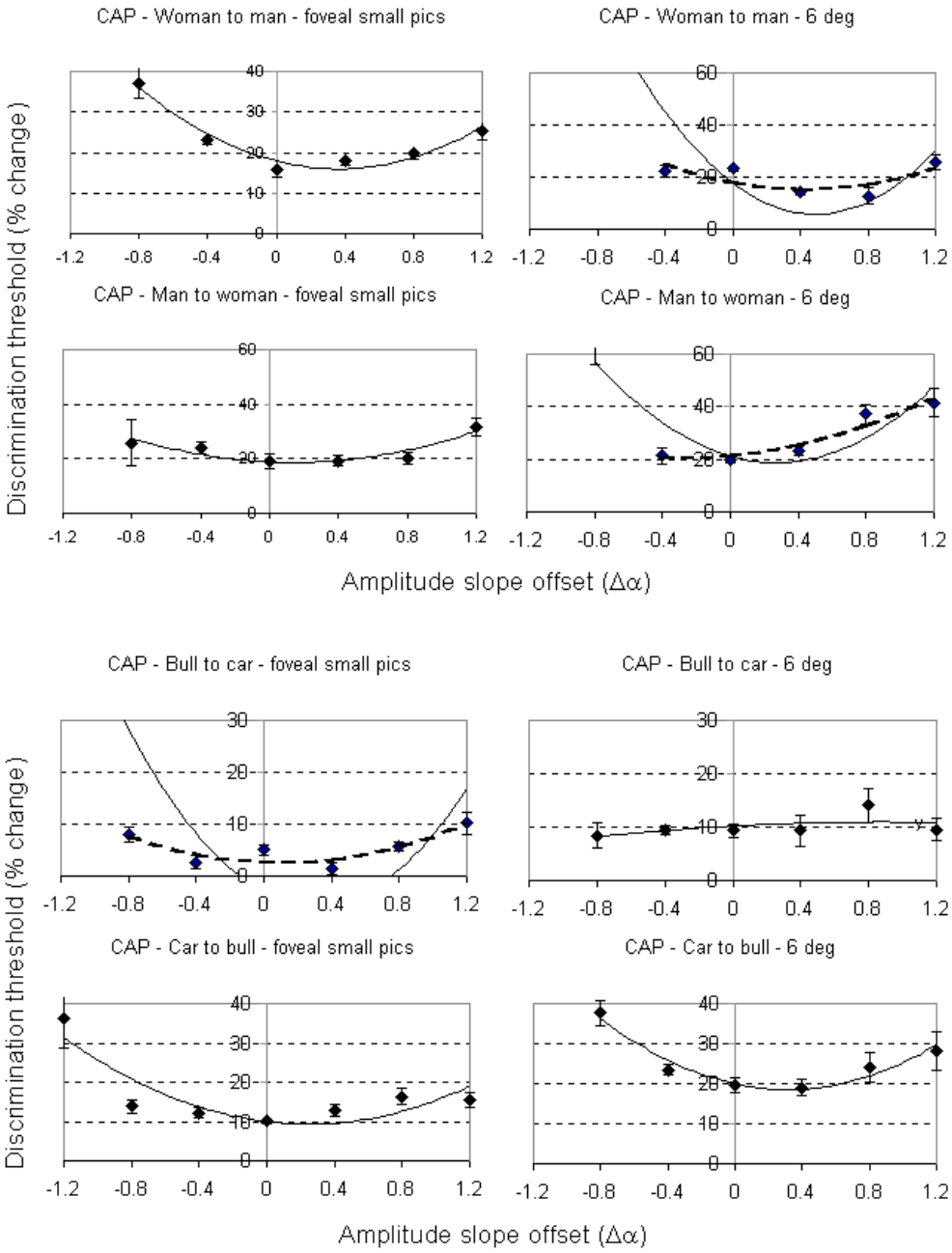

Figure 3.6: Results for experiments 3 (left side) and 4 (right side) for subject CAP in all four conditions. The continuous curve represents the second-order polynomial that best fits the data. The other curve, made of broken lines, shows the same fit with less data points.

In some cases, (e.g. experiment 3, 6° eccentricity and $\Delta\alpha$ = -1.2) the observer was unable to discriminate changes in the morph sequence even when the opposite extreme of the sequence was presented (morph change = 100%). This was the result of the extreme

whitening of the sequence, which removed all Fourier energy detectable by the observer at that eccentricity.

In general, peripheral viewing at 3°or 6° has had surprisingly little effect on discrimination thresholds. The effect of extreme "blurring" or "whitening" on the thresholds is higher than the effects of displacing the stimuli to the visual periphery. In fact, seeing unmodified "natural" stimuli at 6° eccentricity produced thresholds roughly equivalent to those of "unnatural" stimuli viewed at $\Delta\alpha=\pm 0.8$. The shape of the data remains roughly similar to the previous experiments (U-shaped), but the minima seem to be displaced towards more positive values of $\Delta\alpha$ when eccentricity increases.

There are some variations in the results between subjects. The size of the error bars reflects the fact that some discrimination thresholds were more clearly determined than others (sharper transition in the psychometric function leads to smaller error bars). Larger error bars are more common in extremely whitened images.

3.3.2. Distribution of the minima

After examining the polynomial fittings of the exemplary results above, it is possible to see how the minimum value moves towards larger positive values of $\Delta\alpha$ (i.e. images rich in low-SFs) when the eccentricity increases for all subjects. Figure 3.7 shows the average position of the minima for all the different experiments, considering *all subjects*. A Student t-test (with 95% confidence interval) shows that there is no difference between the foveal and 3° data (experiments 1 and 2), but there is a significant shift for the 6° pictures (experiments 3). The data for small-foveal results are also displaced towards the positive

$\Delta\alpha$ region. The same t-test reveals that the differences between 6° and small-foveal results are not significant.

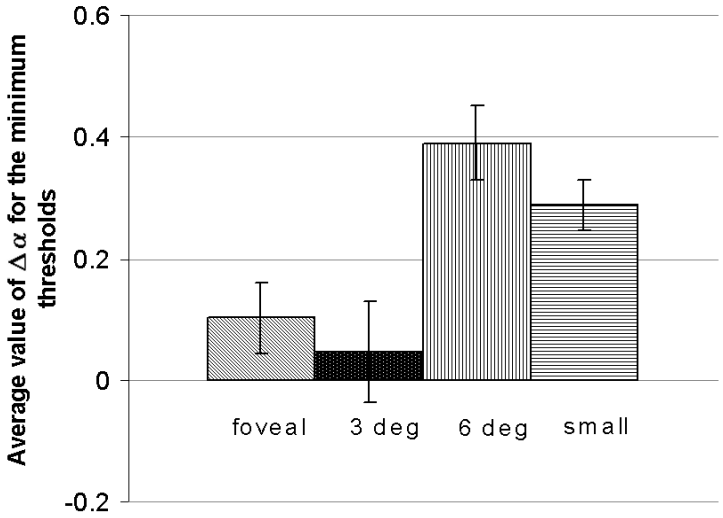

Figure 3.7: Average values of $\Delta\alpha$ corresponding to the minima of the polynomial fittings, for each of the 4 experiments. Standard errors are also shown. Data from all subjects included.

Table 4 shows the numerical values corresponding to the plot above, including *n*, the number of experiments in each case.

	Experiment 1 foveal	**Experiment 2** 3 deg	**Experiment 3** 6 deg	**Experiment 4** foveal-small
Avg	0.102	0.046	0.390	0.288
STD	0.292	0.417	0.248	0.122
n	25	25	16	9
SE	0.058	0.083	0.061	0.040592

Table 4: Average values of $\Delta\alpha$ corresponding to the minima of the polynomial fittings for all observers. Values of the standard errors (SE) and number of experiments are also provided. Data from all subjects included.

Figure 3.8(a) shows the histograms and boxplots (which relate to the statistical *dispersion* of the data) (Howell 1997) corresponding to the distribution of the minima, for two of the experiments described above (experiments 1 and 2). The plot shows similar distributions for the foveal and the 3° data. Figure 3.8(b) shows the same kind of graphs for experiments 1 and 3 (foveal and 6° data) where it is possible to appreciate the separation between the histograms (confirmed by the t-tests). However, the value corresponding to experiment 4 (foveal-small pictures) lies statistically close to that corresponding to experiment 1 (foveal pictures) thus, we cannot consider the position of the minima as different (see Figure 3.9).

If we interpret position of the minima in terms of optimality, we could argue that foveal (and near foveal vision) is optimised to discriminate morph changes in scenes with natural statistics and peripheral vision is better at discriminating changes in scenes with a steeper amplitude slope (blurred images). However, at this point we need to discuss the validity of these polynomial fits.

If data points that have very high discrimination threshold values (see Figure 3.4, Figure 3.5 and Figure 3.6) are included in the fits, they produce a strong bias on some of the second order polynomials (from where the minima values are extracted) towards the region of positive values of $\Delta\alpha$ in a systematic way. This effect is considerable in 17 out of the 75 sets of results. Very high discrimination thresholds are generally produced when the stimuli are severely whitened ($\Delta\alpha$ = -0.8 or -1.2) and vary from observer to observer. For example, an observer may produce much higher thresholds for certain extreme conditions than another observer may.

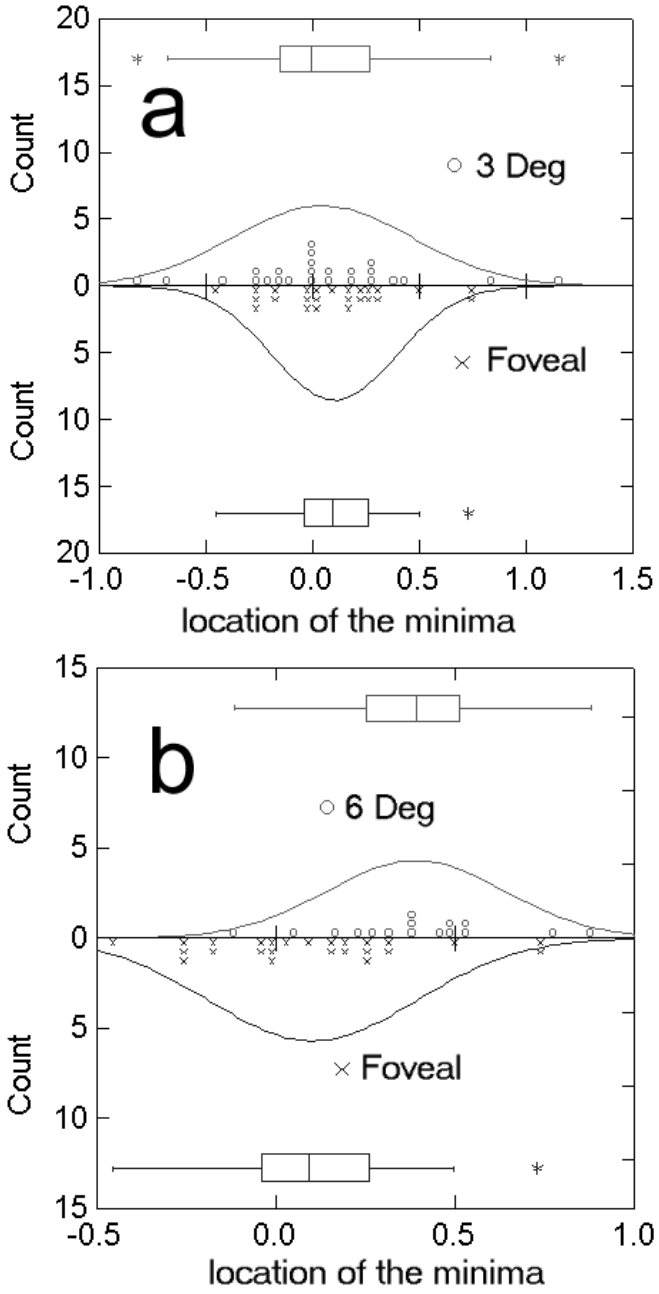

Figure 3.8: Distribution of the minima for the Foveal and 3 Deg experiments (section a) and the Foveal and 6 Deg experiments (section b). Histograms and boxplots are shown. In part a, the average difference is not significant and in part b it is. Data from all subjects included.

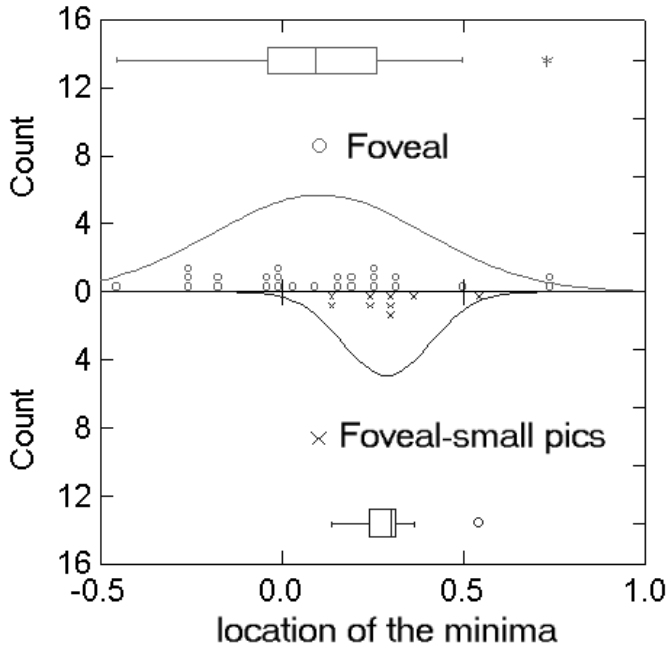

Figure 3.9: Distribution of the minima for the Foveal-small pictures experiments and 6 Deg experiments. Histograms and boxplots are shown. The difference between the average values is not significant. Data from all subjects included.

High discrimination thresholds are badly fit by the polynomials and clearly distort our results systematically towards one side of the graph. In order to estimate how big is this bias on the position of the minima, we need to fit our polynomials only to the data points that belong to the polynomial and compare with the previous results. To do this, we inspected the data and eliminated extreme cases before re-plotting. Broken lines on Figure 3.4, Figure 3.5 and Figure 3.6 show examples of polynomials fitted to the data without considering these points.

Figure 3.10 shows the average values and standard errors for the position of the minima in all four experiments. Points with high discrimination thresholds that clearly do not fit the polynomials were removed here. The effects of considering only data that clearly fit the polynomials are:

a) Dispersion of the data around its mean values increases (larger standard error bars).

b) The shift of the minima towards positive values of $\Delta\alpha$ is less marked than before.

c) Unlike previous findings, Student t-test shows no significant difference between the two first bars (foveal/near-foveal results) and the two last bars (peripheral/size-reduced).

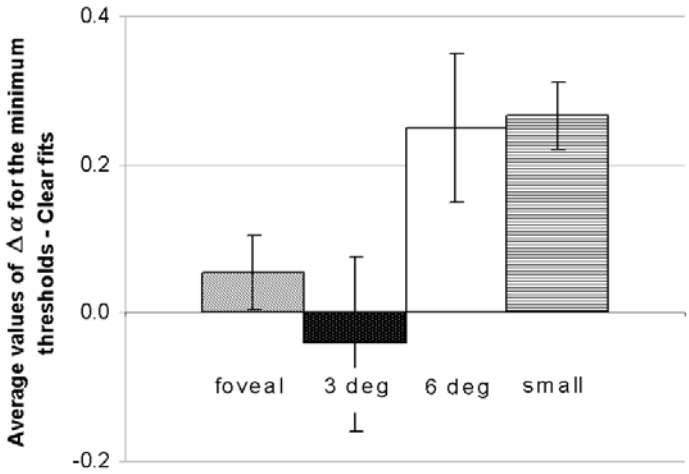

Figure 3.10: Average values of $\Delta\alpha$ corresponding to the minima of the polynomial fittings, for each of the 4 experiments. Only clear polynomial fits were considered. Standard errors are shown. Data from all subjects included.

Table 5 shows the numerical values corresponding to the plot above. In this case, data points that clearly do not belong to the polynomials were not considered.

	Experiment 1 foveal	**Experiment 2** 3 deg	**Experiment 3** 6 deg	**Experiment 4** fov-small
Avg	0.054	-0.043	0.250	0.266
STD	0.255	0.588	0.400	0.135
n	25	25	16	9
SE	0.051	0.118	0.100	0.045

Table 5: Average values of $\Delta\alpha$ corresponding to the minima of the polynomial fitting for all observers. Points that do not belong to the polynomials were not considered. Values of standard errors and number of experiments are also provided. Data from all subjects included.

The removal of questionable data points from the polynomial fits has arguably produced data that are noisier, increasing the standard errors, but has not changed the shape of the plots or contradicted the previous results. Although error bars have increased, they have not overlapped each other and there is still a clear trend towards two separated sets of results: Based on this evidence, we can conclude that, unlike foveal and near-foveal (3°) vision, peripheral vision (at 6°) is not optimised for discriminating small spatial changes in achromatic natural scenes. This lack of optimisation is also reflected by foveal vision when the number of cortical projections is reduced to simulate peripheral conditions by means of an M-scaling.

3.3.3. Discrimination thresholds values and M-scaling

The results for experiments 3 and 4 (Figure 3.4 and Figure 3.6) show that discrimination thresholds for peripheral stimuli are consistently larger than for the corresponding small foveal stimuli.

Table 6 shows the average values of discrimination thresholds for morph sequences with "natural statistics" (i.e. $\Delta\alpha = 0$) according to the eccentricity region where the stimuli was

presented (data from all subjects was included). Given that the absolute threshold values (in % morph) are different for the two sequences of morphed pictures used, they are presented separately in the table. The numbers on the table increase (this was confirmed by t-tests) when eccentricity increases.

	Eccentricity			
	Car-to-bull and bull-to-car			
	0 (foveal)	3 deg	6 deg	0 (foveal-small pics)
Avg	4.544	11.788	15.583	6.410
STD	3.204	4.468	6.553	5.100
n	11	11	10	5
	Man-to-woman and woman-to-man			
Avg	12.65	18.922	20.860	15.457
STD	3.737	6.752	5.132	2.410
n	14	14	6	4

Table 6: Average values for discrimination thresholds for morph sequences with "natural statistics" ($\Delta\alpha = 0$) in different eccentricity conditions. Standard deviations are also shown. Data from all subjects included.

From Table 6 we can infer the following:

a) The decrease in performance between foveal and 6° experiments is about a factor of 2-3.

b) The average values of the discrimination thresholds for foveal stimuli do not change dramatically with the change of stimulus size.

c) M-scaling of the stimuli did not produce the same results for the foveal-small experiment and the 6° eccentricity experiment. This is a common

feature of peripheral vs. foveal performance, where, despite the use of enlarged stimuli, foveal vision performs consistently better.

This significance of this will be discussed below.

3.4. Discussion

3.4.1. Presence of points below the discrimination threshold

As pointed out before, a single experiment consisted of seven experimental series (and seven threshold values). As eccentricities were increased or retinal size projection reduced, some of the threshold values became increasingly more difficult to measure. Values of $\Delta\alpha$ equal to –1.2 produced a very strong removal of low-SFs (whitening) that rendered the morph changes extremely difficult for peripheral vision to discriminate. Even when changes in the morph were equal to 100% (i.e. the face of the man replaced that of the woman, the car replaced the bull, etc.) observers could not perform the task. No thresholds were obtained and the corresponding threshold values (data points) were not included in this analysis (i.e. some of the graphs include only 6 data points). One could argue that if the differences between both reference pictures of the morph were larger, the observers could have discriminated them even in extreme whitening conditions. If this were so, there would have been an extra data point biasing the minima of the second-order polynomials more towards larger positive values of $\Delta\alpha$ and producing larger values of the second derivative (more narrowly tuned functions). Another argument would be that a less strong whitening (e.g. $\Delta\alpha$=-1.0) may have produced discrimination thresholds within the input contrast range of some of the SF-channels and therefore measurable (although high). Our view here is that including those points in our analysis would have produced biased results and therefore some of the second order polynomials were

calculated on six data points instead of seven. A similar effect is assumed for foveally viewed, reduced images. Other points (17 in total) could be measured (i.e. the observer was able to see something changing before reaching the 100% morph change value) but clearly do not belong to the polynomials; thus their overall effect is also to distort the U-shape of the discrimination threshold vs. $\Delta\alpha$ function. These 17 points were also removed from the analysis of the shape and position of the second-order polynomials.

3.4.2. Are the "U-shapes" fitted by polynomials?

In order to determine whether a second-order polynomial is the best equation to fit our experimental results, we performed a series of statistical analyses on the foveal morph experiment results, including both results obtained binocularly and monocularly (31 experiments, including those in the previous chapter). It is easy to see that a second-order polynomial will fit our data better than a line (since a line is an extreme case of a second order polynomial with quadratic coefficient equal to 0) and in turn a third-order polynomial will fit the data better than a second-order one for the same reasons. What we want to demonstrate here is that the improvement in the fitting in the first case (second-order polynomial over line) will be much bigger than the improvement in the second case (third-order over second-order polynomials). To estimate the *goodness-of-fit*, we used the same measure as in the previous chapter (χ^2), defined by the following equation:

$$\chi^2 = \sum_i \left(\frac{Y_i - Y}{SE} \right)^2$$

Equation 3.3

where Y and Y_i correspond to both the actual measures and the fitting and SE represents the standard error of the psychophysical measurements as described in Appendix A.

Table 7 shows the average and median values of χ^2 for 31 sets of experimental results fitted with a second-order polynomial. The same table also shows how this measure of goodness-of-fit decreases when the order of the polynomial is changed, from 1st order (line) to 2nd order and from 2nd order to 3rd order. As the table shows, there is a much greater improvement when the order is increased from 1 to 2 than when is increased from 2 to 3. This shows that a second order polynomial describes our data much better than a line and that there is definitively a dip in the shape of the data.

Average values (n=31)	Average	Median
χ^2 (second-order polynomial)	16.108	10.477
Decrease in χ^2 after increasing polynomial order from 1 to 2	45.639	26.377
Decrease in χ^2 after increasing polynomial order from 2 to 3	5.963	1.486

Table 7: Average values of χ^2 for the second order polynomial fits of our foveal data (31 experiments) and the decrease in this values when considering polynomials of 1st and 3rd order.

Our second analysis is concerned with the statistical significance of the different threshold values obtained throughout this work. For example, we may want to establish that the variations in morph discrimination thresholds obtained for different values of $\Delta\alpha$ are significant and also that there is a significant variation of these results across observers.

Table 8 shows the raw experimental data (morph discrimination thresholds) obtained for our four experiments performed foveally. All observers were included here.

	Woman-to-man								
	AG	BO	CAP	IM	JB	KB	TW	TT	IG
-1.2	8.70	35.87	21.10	50.23	28.96	20.07	34.29	27.10	19.67
-0.8	11.08	19.99	14.69	35.01	27.05	9.53	10.35	23.22	19.07
-0.4	9.92	23.23	10.79	17.70	13.15	11.49	11.39	21.25	8.18
0	8.46	6.88	12.06	20.46	10.17	10.25	10.61	11.76	14.51
0.4	6.75	20.73	12.72	22.99	9.53	11.62	8.57	14.55	13.68
0.8	8.38	11.44	11.88	27.56	9.72	12.03	11.21	19.40	15.43
1.2	11.53	23.00	20.86	32.98	13.33	18.88	16.84	27.72	17.26
	Man-to-woman								
	AG	BO	CAP	IM	JB	KB	TW	IG	TT
-1.2	24.48	100.00	19.83	98.68	35.29	24.18	19.10	18.45	18.51
-0.8	16.61	31.08	16.13	25.72	31.52	15.58	21.33	14.77	14.27
-0.4	14.81	20.83	12.04	17.36	19.09	14.55	15.69	16.53	9.78
0	13.38	17.47	11.72	13.94	18.70	16.82	12.11	21.03	10.95
0.4	11.20	19.65	13.11	18.62	17.99	18.25	11.33	14.87	12.93
0.8	11.44	30.82	16.17	22.94	25.02	20.21	11.94	11.86	15.79
1.2	10.86	51.51	14.45	34.11	28.59	29.48	17.36	19.51	14.45
	Bull-to-car								
	AG	BO	CAP	KB	TW	TT	JB		
-1.2	11.06	**72.01**	2.11	2.91	4.53	4.25	3.48		
-0.8	7.88	5.74	3.28	1.39	3.51	2.16	1.98		
-0.4	8.80	4.57	1.74	1.66	2.76	2.26	3.24		
0	10.43	0.81	1.66	1.26	1.99	1.87	3.75		
0.4	6.93	6.08	1.55	2.49	2.46	2.08	3.12		
0.8	11.39	6.85	2.69	2.63	2.90	4.58	2.74		
1.2	15.85	8.45	3.88	5.22	3.00	7.77	4.61		
	Car-to-bull								
	BO	CAP	JB	KB	TW	TT			
-1.2	74.85	11.54	8.95	29.38	19.99	9.39			
-0.8	18.79	4.63	6.54	6.10	17.21	7.67			
-0.4	9.31	4.00	4.58	6.69	12.15	5.94			
0	5.75	3.40	2.14	10.04	11.32	4.08			
0.4	10.37	4.40	7.38	6.15	9.24	2.40			
0.8	16.95	7.47	9.20	13.95	18.75	7.09			
1.2	21.34	11.22	20.18	24.56	16.31	6.61			

Table 8: Morph discrimination thresholds obtained for four foveal experiments (both monocular and binocular). All subjects included.

Two-factor ANOVAs without replication were performed on the data corresponding to each of the four experiments. The results are summarised below:

Woman-to-man

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	1604.467	6	267.4112	**11.64108**	5.24E-08	**2.294598**
Columns	1872.106	8	234.0132	**10.18718**	3.54E-08	**2.138229**
Error	1102.625	48	22.97134			
Total	4579.197	62				

Man-to-woman

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	4652.156	6	775.3594	**4.939719**	0.000664	**2.323993**
Columns	4425.095	7	632.1564	**4.02739**	0.001864	**2.237073**
Error	6592.5	42	156.9643			
Total	15669.75	55				

Bull-to-car

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	778.2379	6	129.7063	**1.232801**	0.31761	**2.420521**
Columns	967.7649	5	193.553	**1.839636**	0.135148	**2.533554**
Error	3156.38	30	105.2127			
Total	4902.383	41				

Car-to-bull

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	1798.711	6	299.7852	**3.549666**	0.008935	**2.420521**
Columns	1381.507	5	276.3014	**3.271601**	0.017826	**2.533554**
Error	2533.635	30	84.45449			
Total	5713.853	41				

The results produced by the ANOVA (Microsoft Excel 2000) show that the F statistic is higher than the critical value in all cases, except for the data obtained from the bull-to-car experiment, in which it is smaller. A close examination of the raw data (see bold lettering in Table 8) shows that a single very high threshold value of 72.01% might be producing these ANOVA results by increasing the total variance of the dataset. The removal of the experiment that produced such a extreme value derives in an immediate improvement of the F statistic (see below).

Bull-to-car (less observer BO)

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	61.16203	6	10.19367	**7.646051**	2.44E-05	**2.363748**
Columns	342.9567	6	57.15944	**42.87405**	5.95E-15	**2.363748**
Error	47.99499	36	1.333194			
Total	452.1137	48				

In summary, ANOVA results show that the morph discrimination thresholds values measured in our experiments are statistically independent both across morph sequences (changes in $\Delta\alpha$) and observers.

3.4.3. Degree of optimisation

From the data shown in Figure 3.10, it follows that vision in the periphery is not optimised to the statistics of natural scenes. In order to evaluate more quantitatively our results and relate them to the number of active SF-channels, we analyse here how the U-shapes may differ from foveal to peripheral data. To do this we calculate the second derivative of the second-order polynomial fittings of our results for experiments 1, 2, 3 and 4 (having eliminated non-fitting points, all observers included). The value of the second derivative shows the rate of change of the slope of the polynomial, which is related to the "shallowness" of the "U-shaped" curve. The smaller the value of the second derivative, the more "shallow" or "flat" the curve fitting is. The shallowness of the curve is also a measure of the tuning of the visual system to the corresponding image statistics: a broad tuning (more active SF-channels) will be related to a small value of the second derivative and a fine tuning (less active SF-channels) will be related to a high value of the second derivative.

Figure 3.11 shows the average results and standard errors for the values of the second derivative of the polynomial fittings to the experimental data. Student t-tests (95% confidence interval) show that the differences between foveal and 6º are significant. The same is true for the difference between foveal and 3º but it is not true for the difference between foveal and foveal-small second derivative values. If we interpret these second derivative values as a higher degree of tuning, then 3º and 6° eccentricity plots (experiment 3) are more sharply tuned than the rest. According to these results, peripherally viewed images stimulate a smaller number of channels than foveally viewed images, disregarding image size.

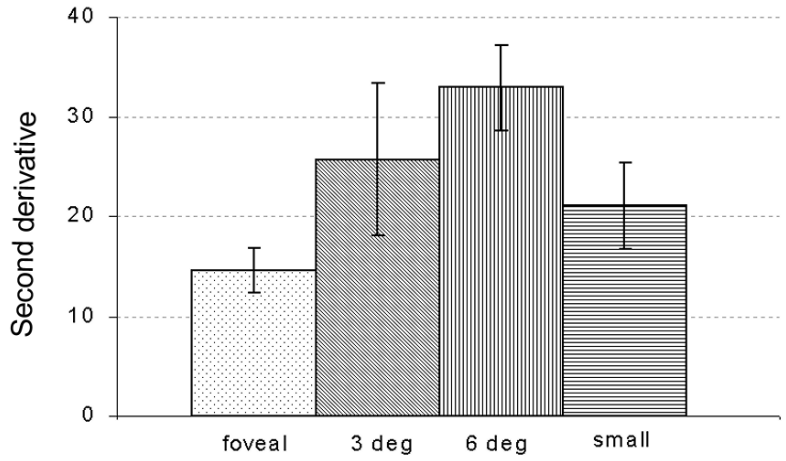

Figure 3.11: Average values of the second derivative of the polynomial fittings. Only points clearly belonging to the polynomials were considered. Standard errors are shown.

3.4.4. M-scaling: is there an effect?

One way of assessing the success of the M-scaling factor would be to plot simple CSFs for M-scaled gratings at different retinal locations. Figure 3.12 shows the CSF for observer CAP, obtained for small gratings (same size as the stimuli in experiment 4) viewed foveally and normal-sized gratings viewed at 6º eccentricity. The ordinate shows the reciprocal of

the lowest Michelson contrast needed for the observer to detect a sinusoidal grating of a specified SF. To be able to compare both plots, spatial frequency is expressed as cycles per picture width.

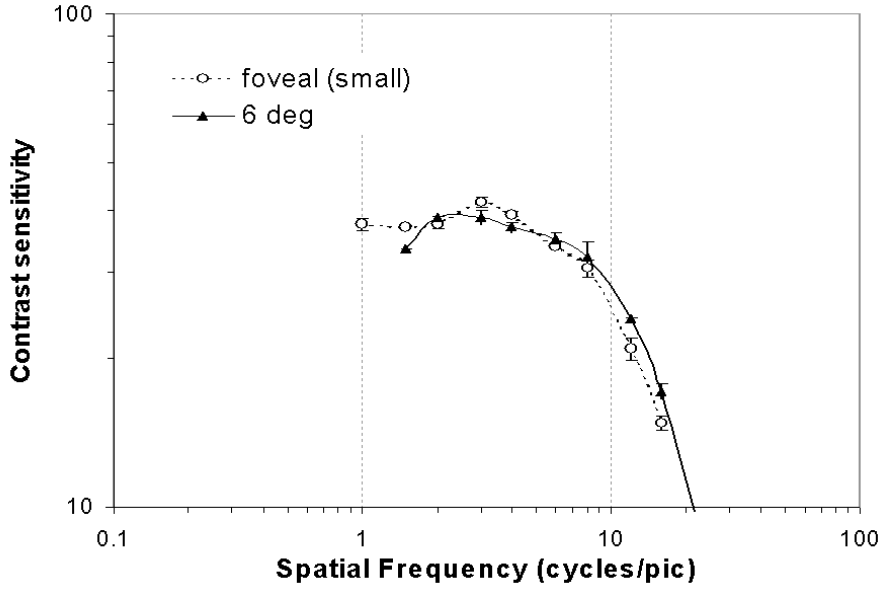

Figure 3.12: Contrast sensitivity functions for observer CAP. The *y*-axis shows the reciprocal of the lowest Michelson contrast needed for the observer to detect a sinusoidal grating of a specified SF. The SF values are plotted in cycles per picture width. Two conditions are shown: small gratings viewed foveally and large gratings viewed at 6º eccentricity. Error bars for foveal viewing are shown. The stimuli were viewed monocularly with the dominant eye.

Given that the M-scaling factor was calculated on the assumption that only differences in the neural representation of the fovea in the cortex are relevant, once this is corrected, one might expect a similar form of the CSF at all retinal positions. From Figure 3.12 we see that this is the case. Foveal CSF is coincident with peripheral CSF for the retinal positions used in this study (6º eccentricity). Figure 3.13 shows CSF plots (in cycles/degree) for the other experimental conditions considered in this study (foveal, 3° and 6°).

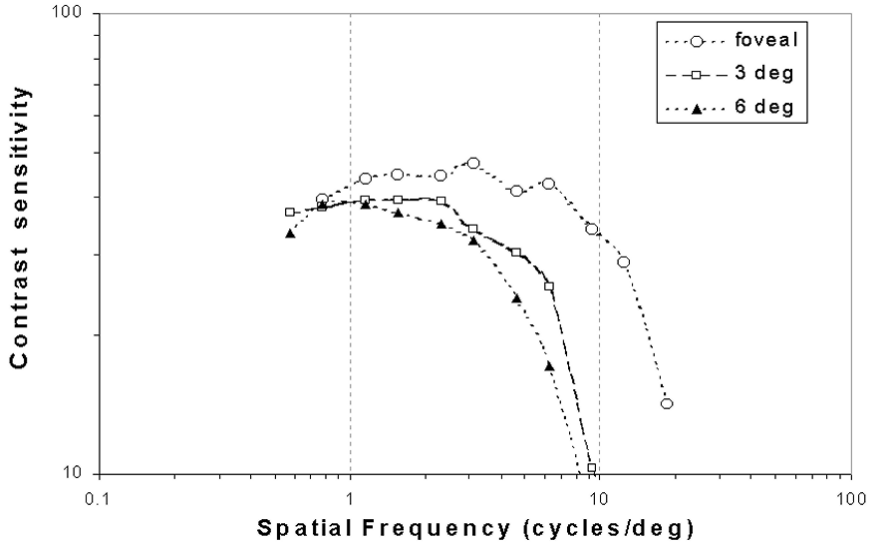

Figure 3.13: Contrast sensitivity functions (measured with normal-sized gratings) for three different eccentricities: 0deg (foveal) 3º and 6º. Observer CAP.

Figure 3.14 shows the average second-order polynomials obtained from our dataset (all observers) for all experiments. Because different morph sequences yield different discrimination thresholds, the top part of Figure 3.14 shows the average results for cb and bc sequences and the bottom part shows the same for the wm and mw sequences. To evaluate if M-scaling has compensated for the loss of visual processing in the periphery we look at the following:

a) The shape and position of the CSF measured for 6° peripheral stimuli is very close to that of foveal stimuli when the stimuli are corrected (M-scaled). This indicates that our M-scaling factor is adequate to correct for any grating acuity difference (Figure 3.12).

b) Results produced in experiments 3 and 4 (see 6º and foveal-small plots in Figure 3.14) show that the position of the minima ($\Delta\alpha$) and thus, the degree

of optimisation to natural statistics are similar for these stimuli. Foveal and near-foveal (3º) vision are optimised to discriminate changes in natural scenes while 6º eccentric vision is not. Scaling foveal stimuli to stimulate the same number of cortical receptors as 6º vision produces the same (non-optimised) results.

c) Table 6 shows that the average discrimination threshold values for M-scaled (i.e. reduced) stimuli are slightly higher than for normal stimuli viewed foveally and different from thresholds for peripherally viewed (at 6°) stimuli. This has also been confirmed by t-tests. In fact discrimination threshold values seem to depend on the eccentricity of the stimuli and not on their size (see vertical shifts of the plots in Figure 3.14).

d) The value of the second derivative (related to the "tuning" or number of channels involved in the discrimination process) seems also to depend on the eccentricity of the stimuli, regardless of the size (M-scaling does not yield the same results as peripheral viewing). Figure 3.14 shows that U-shaped polynomials become more sharply tuned as the eccentricity increases.

Items (a) and (b) indicate that the "simple" magnification factor estimated by Equation 3.1 may work well for the task, on the other hand, (c) and (d) show that M-scaling the stimuli does not seem to equate the thresholds between foveal and peripheral stimuli.

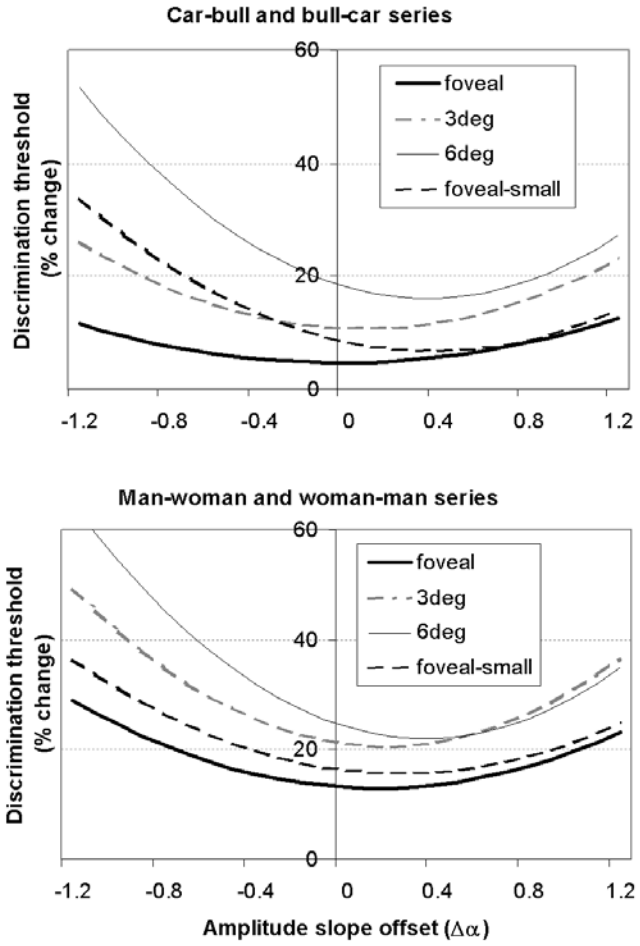

Figure 3.14: Average second-order polynomial fits obtained from the dataset (all observers).

In general, the overall effect of reducing the size of the foveal stimuli (M-scaling) is to produce discrimination threshold functions that are shifted towards the right side of the plot (minima at more positive $\Delta\alpha$ values), with slightly higher threshold values, while keeping its U-shape (similar second order derivative). A possible explanation for the shift (more positive $\Delta\alpha$ values) is the reduction in the Fourier energy available for the discrimination task. In summary what "blurring" or "whitening" the morph sequences does is to reduce/boost the energy available to some of the SF-channels to do the task. For example, if the task was done foveally comparing local changes of contrast within a medium SF-band, let's say centred around 3 cycles/degree, a reduction on the size of the

picture would produce a shift of this local Fourier energy towards higher SFs (e.g. to 11 cycles/degree). This energy may be no longer suprathreshold after the shift, thus making the task harder (increasing contrast thresholds). In the case of the "blurred" pictures (positive values of $\Delta\alpha$), the lack of high-SF frequency energy means that the task was done using the (enhanced) low-SF bands. If the SF-band chosen was centred, for example, around 1 cycle/degree, a reduction in size would shift this energy to 3.7 cycles/degree, which is well within the peak of SF-discrimination. The opposite effect would happen for "whitened images". Overall, this would explain the shift of the lowest discrimination thresholds towards the region of positive $\Delta\alpha$ values. This effect may also explain the similarity between the absolute results (threshold values) from experiment 1 (foveal stimuli) and experiment 4 (foveal-small stimuli).

The U-shape (related to the second order derivative) of the foveal-small data plot does not change substantially from the other foveal curves. This agrees with the idea that the number of SF-channels participating in the discrimination process changes from foveal to peripheral vision, regardless of the magnification factor.

A plausible explanation for shape and position of the 6º (experiment 3) data comes from modelling the discrimination process (see Figure 4.21, next chapter). An increase of the bandwidth of the SF-channels involved (e.g. from 1.5 to 1.9 octaves) in the discrimination process may also produce an increase the discrimination thresholds and a shift of the minima towards more positive values of $\Delta\alpha$, similar to the curve produced by experiment 3.

3.4.5. Comparison with other results from the literature

To compare the deterioration of observers' performance with results reported by other workers, we evaluated the increase in the discrimination threshold for unmodified (i.e. $\Delta\alpha = 0$) morph series with visual eccentricity. Figure 3.15 shows the average increase in the discrimination thresholds in normalised units (with respect to the foveal value) for both types of morph sequences used in our experiments. Averages for man-to-woman and woman-to-man series (filled circles) are plotted separately from averages for car-to-bull and bull-to-car series (open circles).

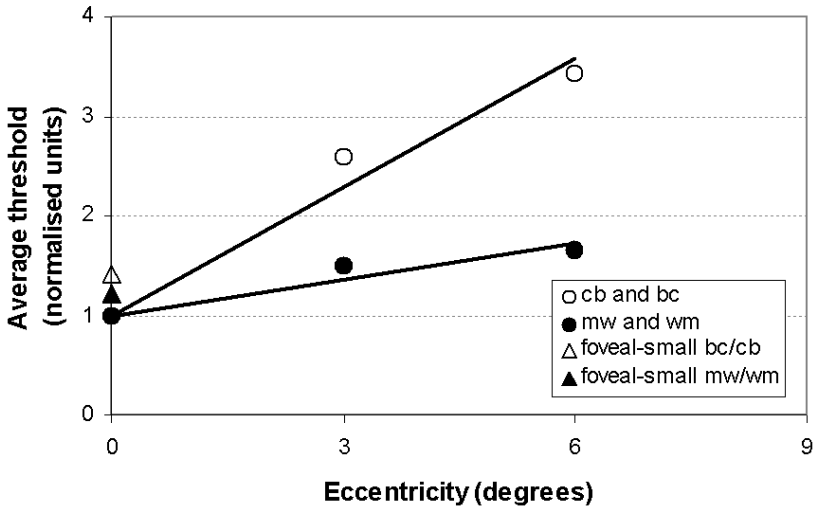

Figure 3.15: Task difficulty versus eccentricity for the two kinds of morph sequences used in our experiments: mw/wm (filled circles) and cb/bc (open circles). The triangles represent the values threshold values produced by the *reduced* foveal images: mw/wm (filled) and cb/bc (open). Units are normalised (measured with respect to the foveal value).

The average values corresponding to the foveal, M-scaled (reduced) images are also plotted as triangles (the filled triangle correspond to the mw/wm series and the open triangle to the cb/bc series). If our M-scaling had fully worked, these triangles should have the same *y*-axis values as 6° eccentric stimuli has. Average discrimination threshold values

were calculated from lowest point of the fitted second-order polynomials. All observers were included.

From Figure 3.15, we calculated the value of the slope *S*, corresponding to the deterioration of the visual performance (as a proportion of the foveal performance) with eccentricity. A popular parameter used to compare results across studies is E_2 (the eccentricity at which the task becomes twice as difficult as in the fovea). S can be estimated from the existing literature as the inverse of E_2. Our calculated value of S is equal to 0.12 deg^{-1} for the mw/wm task and equal to 0.43 deg^{-1} for the cb/bc task.

Table 9 shows some of the existing estimations of *S* for different visual tasks using different stimuli across the literature. Our own estimated values seem to lie close to those obtained for grating acuity tasks. The failure of the M-scaling implemented in this study may have its origins in the nature of the morph discrimination task being different from a grating detection task (such as the one used by Rovamo and Virsu (Rovamo and Virsu 1979) to derive their magnification factors).

In summary, M-scaling the stimuli cannot compensate for the differences in discrimination threshold results produced by experiments 3 and 4. Although our threshold measurements increase with eccentricity at a ratio not far from values measured for *grating acuity* tasks (see Table 9 above) just scaling the stimuli does not equalize foveal and 6° thresholds. It might be possible that using a different magnification factor (such as the one corresponding to measures of vernier acuity (Tolhurst and Ling 1988)) may have increased the thresholds measured in experiment 4.

M-scaling only produces changes in the U-shape of the discrimination function, and its position along the *x*-axis, which become closer for both tasks when M-scaled.

Task	Stimulus	S	
		from:	to:
Vernier acuity (Whitaker *et al.* 1992)	line	0.56	0.81
Vernier acuity (Whitaker *et al.* 1992)	two isoccentric dots	0.51	0.94
Vernier acuity (Virsu *et al.* 1987)	two dots	0.53	
Spatial interval discrimination (Whitaker *et al.* 1992)	two isoccentric dots	4.54	14.28
Curvature detection and discrimination (Whitaker *et al.* 1993)		0.44	0.70
Bisection (Whitaker *et al.* 1992)		12.5	14.28
Displacement detection (Whitaker *et al.* 1992)	referenced	0.74	0.94
Displacement detection (Whitaker *et al.* 1992)	unreferenced	0.09	0.15
Displacement detection (Whitaker *et al.* 1992)	unreferenced	0.054	0.074
Orientation (Makela *et al.* 1993)	line	0.51	
Visual acuity (Virsu *et al.* 1987)	**grating**	**0.41**	
Visual acuity (Klein and Levi 1987)	**grating**	**0.38**	
Detection of geometric image distortions (Rovamo *et al.* 1997)	pictures of human faces	0.55	
Visual acuity (Virsu *et al.* 1987)	Snellen E chart	0.58	
Visual acuity (Virsu *et al.* 1987)	Landolt C chart	1	
Crowding (Toet and Levi 1992)	Letter T	2.5	5

Table 9: Value of S (deterioration of the visual task normalised to the foveal value) obtained by different workers. Notice that our values are close to those measured for grating acuity tasks. Calculations made by Rovamo *et al* (Rovamo *et al.* 1997).

These results are not surprising, given that for some peripheral vision tasks, foveal vision has been reported to perform better than peripheral vision, regardless of M-scaling. Examples of these are tasks requiring perimetric thresholds (Drasdo and Thompson 1989; Wood *et al.* 1986) and processing of spatial relationships (Levi and Klein 1986; Levi *et al.*

1985; Paradiso and Carney 1988; Paradiso *et al.* 1989; Rentschler and Treutwein 1985; Stephenson and Braddick 1983) among others. There is also substantial amount of evidence showing that for various perceptual tasks, the threshold gradients might be steeper or flatter than the proposed inverse of the cortical magnification (Fendick and Westheimer 1983; Hampton and Kertesz 1983a, b; Klein and Levi 1987; Levi and Klein 1986; Levi *et al.* 1985; Rentschler and Treutwein 1985; Stephenson and Braddick 1983). That is, our choice of scaling parameters in the calculation of the cortical magnification factor (see Table 9 above) may not be consistent with other workers.

3.4.6. Possible effects of image type

Figure 3.15 shows a different slope S (deterioration of the visual performance as a function of visual eccentricity) for the two types of morph sequences (mw and cb). This difference is not related to the fact that the metric used to measure discrimination in each of the sequences is different (since the units in Figure 3.15 are normalised) but it might depend of another factors such as differences in the type of stimuli or different morphing technique.

Since one type of morph sequence (mw and wm) contains a picture of a human face, we cannot discard the effects of higher-level visual functions affecting the discrimination task. There is evidence that attention is not evenly distributed within static frontal views of a face. In such cases, there is a hierarchy of salience in which more attention is paid to the hair, then to the eyes, mouth, nose, etc. following roughly that order (Shepherd *et al.* 1981). These particular properties of human faces may have determined a different approach to the same task involving higher level (attention-related) processes, which may account for the different performance degradation in the two types of stimuli (see Figure

3.15). These effects might be eliminated by performing the same morph discrimination experiment on an inverted (upside down) version of the mw and wm sequence.

Another possible reason as to why the two types of morph sequence produced different values of S (deterioration of visual performance) is that they were created by a different morphing algorithm, which may have introduced differences in the way intermediate images were produced.

3.5. Conclusions

a) Our results show that peripheral vision at 6º temporal eccentricity is not optimised to discriminate local changes in scenes with natural second-order statistics in the same way as foveal vision is (shift on the position of the minima towards the "blurred" side of the U-function). This effect could be produced by an increase in the SF- channels bandwidths (see next chapter on modelling these results).

b) Peripheral vision is increasingly more "tuned" (steeper U-shapes in our morph discrimination functions) than foveal vision. This may be related to less SF-channels being used in the periphery than in the fovea with increasing bandwidths (see next chapter on modelling these results).

c) The values of the discrimination thresholds for "natural" scenes ($\Delta\alpha = 0$) do not change substantially when the foveal stimuli are reduced in size (M-scaled) to equate to those presented peripherally at 6°. M-scaling can compensate for other effects like the shift of the minima or the changes in shape (sharper tuning) of the U-function, but does not yield the same

discrimination threshold results in both experiments. For the discrimination task considered, the fovea is consistently superior in performance to peripheral vision in a way that cannot be compensated by our single scaling factor.

Chapter 4.

Chapter 4

MODELLING THE RESULTS

Can a simple computational model predict the discrimination thresholds for morphed objects in natural scenes?

4.1. Overview

In the previous chapters, we focused on the relationship between the human visual system and the second-order information present in the visual environment. There, we argued that the HVS is optimised to encode the information in the natural visual environment. Our conclusions derive from psychophysical experiments, i.e. comparing people's discrimination thresholds for small spatial changes (produced by 'morphing') in natural and unnatural (spectral slope modified) visual stimuli. In this chapter, we attempt to explain our previous results using a relatively simple computational model of the low-level discrimination process. Our general aim is to determine whether our results for this visual discrimination task can be explained without involving "higher" (and more complex) cortical processes.

There are a number of computer-based image discriminability models (Ahumada *et al.* 1998; Ahumada *et al.* 1995; Daly 1992; Watson and Solomon 1997) capable of predicting the visibility difference between a pair of images. Some of these models have been tested or compared for different tasks, such as detecting objects against natural backgrounds (Rohaly *et al.* 1997). Model algorithms can have different levels of complexity, from a simple difference between two images to (physiologically plausible) multiple spatial

frequency channel models with several types of masking (Ahumada 1987; Girod 1989). The advantages of using models that are more complex are not yet very clear (Rohaly *et al.* 1997).

One of the simplest and yet plausible ways for the HVS to evaluate differences between pictures would be to compare local spatial information within each spatial-frequency channel. According to this, we calculate differences in contrast between two images (reference and test) within a number of SF-channels designed to have the spatial-frequency bandwidth of simple cells in the visual cortex. Our model assumes that simple cells in several independent spatial-frequency bands (i.e. at each SF and location) sample the reference and test stimuli point-by-point, and that each cell then signals any local differences in the spatial structure of the two stimuli. We considered two types of interaction among these cells within a given channel (within-channel masking) using a Minkowski summation with two different exponents (Quick 1974; Watson 1987). There was also a consideration of interactions among channels (the contrast gain of a channel is modified by activity from the others), which may produce masking (Foley 1994; Teo and Heeger 1994, 1995).

By "customising" the model to include each observer's contrast sensitivity to sinusoidal gratings, we are able to replicate the forms of the relationships between discrimination threshold and spectral slope, and the ways that these differ between picture sets and observers. To keep things simple, no attempt was made to consider different spatial frequency orientations in our modelling.

4.2. Methods

Our model is based on evidence (both physiological and psychophysical – see Introductory Chapter) for the existence of multiple channels tuned to SF and it is similar in some ways to DCTune[16] (Watson 1993a) and Daly's VDP[17] (Daly 1992). Under this view, the spatial contrast sensitivity function is the envelope of many narrowly tuned SF selective channels (Blakemore and Campbell 1969). Each of these spatial frequency filters responds to the power in the region that falls within its frequency band. Here the analysis over the same restricted region is carried out in parallel by these filters (on a point-by-point basis).

4.2.1. Contrast in complex (natural) images

We calculate the contrast for every image in the morph sequence within spatial-frequency channels (bands) designed to have the spatial-frequency bandwidth of simple cells found in the visual cortex (about 1 - 1.5 octaves (De Valois *et al.* 1982a; De Valois *et al.* 1982b; Movshon *et al.* 1978a; Tolhurst and Thompson 1981)). We define contrast (Peli 1990) within a frequency-band F at a point (*x, y*) as:

$$C_F(x,y) = \frac{a_F(x,y)}{l_m(x,y)}$$

Equation 4.1

[16] The DCTune compression algorithm was developed to improve the quality of compressed pictures using a quantization matrix that is optimised according to local variations within the image. Since each image has local variations, having different quantization matrices that are adapted for individual regions of the image improves the compression. The most important local variables that DCTune handles are luminance, local spatial frequency and orientation.

[17] Daly's VDP (visual difference predictor) model was developed to measure visual differences in images. It is based on filtering techniques and involves luminance masking, detailed CSF for luminance, intra-channel contrast masking and error pooling.

In this equation, $a_F(x,y)$ is the band-passed filtered version of the image convolved with a circularly-symmetric operator, whose spatial-frequency characteristic is:

$$A_F(f) = e^{-\left(\frac{(f-F)^2}{2\sigma^2}\right)}$$

Equation 4.2

and $l_m(x,y)$ is an estimate of the local mean luminance, derived from a low-pass filtered version of the image using the circularly-symmetric operator whose spatial-frequency characteristic is:

$$L_F(f) = e^{-\left(\frac{f^2}{2\sigma^2}\right)}$$

Equation 4.3

Here *f* is spatial frequency, F is the centre frequency, σ is the standard deviation of the Gaussian-frequency tuning curves. This definition of contrast conforms to the general principle that it is some measurable luminance modulation divided by the mean luminance and captures, the *perceived* contrast of complex images (Peli 1990).

No provision was made to account for orientation tuning (our filters are circularly-symmetric), mainly because the level of complexity of such a model (and the computational power required to run it) would be much bigger. Another point to consider here is that our original images (man, woman, car and bull) did not contain any strongly oriented features, (such as horizon/sky, trees, buildings, etc.) which could bias the distributions of SF-orientations in any specific direction. In our approach, we aim to keep things as simple as possible, trying to find the basic properties of the HVS that might allow it to perform our discrimination task. The filters were centred on 2, 3, 4, 6, 8, 12, 16,

24 and 32 cycles/picture. σ in Equation 4.2 and Equation 4.3 was set to be 0.28 times F resulting in a bandwidth of 1.5 octaves. This value of σ was chosen to be reasonably close to that of human channels (Blakemore and Campbell 1969) and of neurons in the primary visual cortex of cat and monkey (De Valois *et al.* 1982a; Movshon *et al.* 1978a; Tolhurst and Thompson 1981). Given a spatial-frequency band, we can compare the contrast of the reference (first image of the morph sequence) to the contrast in the same location of any of the other images the other (test) images.

4.2.2. Detection thresholds for sinusoidal gratings

To determine whether the difference in equivalent contrast of the test and reference images in any frequency band centred on F is large enough to be detectable by the observer, we need to know how well the observer can discriminate changes in the "real" Michelson contrast of simple sinusoidal gratings. Michelson contrast is defined as:

$$C = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}}$$

Equation 4.4

Where $L_{\max}$ and $L_{\min}$ are the maximal and minimal luminances in the grating.

Rearranging the equation as:

$$C = \frac{\frac{L_{\max} - L_{\min}}{2}}{\frac{L_{\max} + L_{\min}}{2}}$$

Equation 4.5

We can see that the upper term $(L_{max} - L_{min})/2$ represents the luminance modulation whereas the lower term $(L_{max} + L_{min})/2$ is the mean luminance of the grating.

The observer's contrast discrimination functions for gratings were not measured directly. Instead, we used the template for the "dipper" function (see Figure 4.1) for contrast discrimination (Legge 1981; Legge and Foley 1980; Nachmias and Sansbury 1974). This dipper was found to be valid for a variety of observers and threshold discrimination experimental conditions (Foley and Legge 1981; Legge and Foley 1980). Figure 4.1 represents the just discriminable contrast difference between a pair of gratings plotted against the lesser contrast in *log-log* coordinates. The shape of the plot in Figure 4.1 was obtained using 2AFC techniques where observers were presented with two gratings, one of contrast C and another of contrast $C+\Delta C$ and had to decide which one was $C+\Delta C$ (Nachmias and Sansbury 1974). The function increases monotonically on the right side of the plot but on the left side (very small values of reference contrast) the function decreases. This effect is also known as *facilitation effect* (the increment threshold is actually lower than the reference), and it is confined to very low background contrasts. Although there is a general agreement about the facilitation effect, there is still some uncertainty as to whether the rising part of the discrimination follows Weber's law ($\Delta C = kC^{1.0}$) for *suprathreshold* stimuli. This is especially true for the detection of sine wave gratings in visual noise, where threshold contrast rises in proportion to the Fourier energy of the noise (Pelli 1979; Pollehn and Roehring 1970; Stromeyer and Julesz 1972). The value of the exponent that best describes the line part of Figure 4.1 has been found to be less than 1.0 in several studies (Nachmias and Sansbury 1974; Tolhurst and Barfield 1978) sometimes as low as 0.5 (Pelli 1979) and up to 0.7 (Legge 1979). However, our results (see below) show that the exact shape of this template is not critical for our modelling.

Figure 4.1 shows an example of a contrast discrimination "dipper" template function. The "line" part of the dipper function has a slope (in log-log coordinates) of about 0.7. The function was produced by averaging psychophysical measurements of 3 observers (Dr D.J. Tolhurst, personal communication). The numerical values on the axis of Figure 4.1 were normalised so that the template can be applied to results from test gratings of various SF, using the adequate contrast sensitivity value.

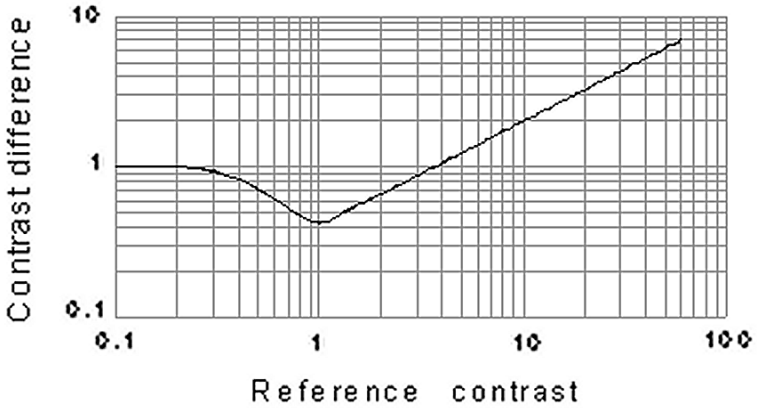

Figure 4.1: Standard contrast discrimination "dipper" template. Its overall shape was obtained for observers discriminating between gratings of slightly different Michelson contrast. The plot represents the just discriminable contrast difference between a pair of gratings plotted against the lesser contrast in *log-log* coordinates.

The positions of the template for *discriminating* between gratings of spatial frequency F (on abscissa and ordinate) were determined by measuring the observer's contrast thresholds for *detecting* sinusoidal gratings of spatial frequency F. To do this we used a 2AFC technique where the stimuli were square patches of stationary grating of the same size as the pictures used in each of the morph discrimination experiments. The same type of masking (a Gaussian roll-off –see section 2.2.2) was used to reduce edge effects. Figure 4.2 shows an exemplary plot of these detection thresholds for sinusoidal gratings in terms of their Michelson contrast. These contrast *detection* thresholds (e.g. equal to 0.006 for a

grating of 2.5 cycles/degree in Figure 4.2) were used to determine the position of the standard discrimination template (dipper) for the corresponding spatial frequency.

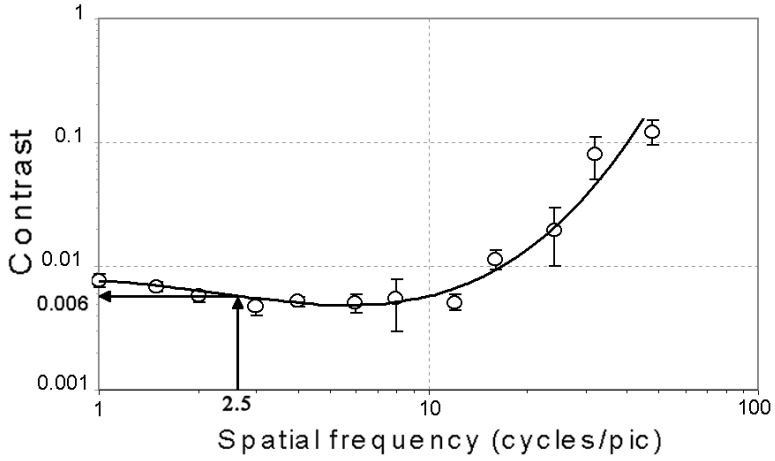

Figure 4.2: Exemplary plot of (foveal) detection thresholds for sinusoidal gratings of different SFs. The arrow shows the corresponding detection threshold for a SF of 2.5 cycles/picture.

Figure 4.3 shows how the positioning of the dipper template is done, using the value of contrast *detection* obtained psychophysically. In this figure, both absisa and ordinate of the dipper template have been multiplied (scaled) by a factor of 0.006, so that the minimum value for contrast difference occurs for a reference Michelson contrast value of 0.006 and the plot tends asymptotically to 0.006 when the reference Michelson contrast tends to 0. This corresponds to the fact that "discriminating" between two gratings, one of which has a contrast equal to 0, is equivalent to "detecting" the other. This is done for each of the central SF corresponding to the model's bands.

In summary, thresholds for *detecting* sinusoidal vertical gratings were measured for observers TT, CAP and KB. The gratings had the same size as the pictures described in section 3.2.1. They were also viewed in similar conditions (same monitor and same distance). The spatial frequencies of the gratings were 2, 3, 4, 6, 8, 12, 16, 24 and 32 cycles

per picture. Observers CAP and KB were also presented with the same set of gratings for monocular viewing, foveally and at 3º and 6º eccentricities. A special set of gratings, which was reduced in size according to the cortical magnification factor described in Section 3.2.2, was produced for measuring jnds foveally.

To be able to produce very small changes in the Michelson contrast of the sinusoidal gratings presented on the screen we used our VSG2/3 graphics card in 12-bit resolution[18] (or Pelli/Watson modification mode) (CRS 1995; Pelli and Zhang 1991). This method allows higher resolution than 8-bit on an 8-bit-per-pixel system, by adding together the two palette outputs for each colour.

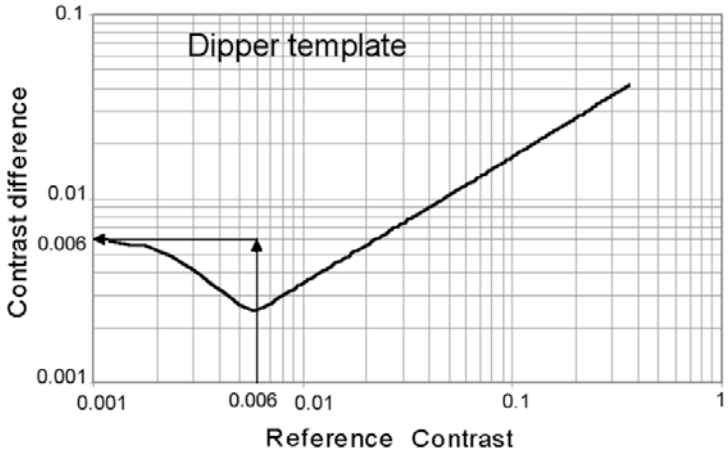

Figure 4.3: Positioning of the dipper template for 2.5 cycles/degree, using the contrast detection threshold for that SF, in the conditions of Figure 4.2.

18 As explained in VSG user's guide version 4.02: Palettes and LUT's. Cambridge Research Systems Ltd., Kent, UK. 73-75 (1995).

4.2.3. Modelling detection thresholds

The absolute value of the contrast difference at each point (x, y) between the reference picture and all the test pictures was measured for the central 68 x 68 points, corresponding to an area of the picture not modified by the masking (see Figure 4.4).

$$\Delta C_{F,j}(x,y) = \left| C_{F,j}(x,y) - C_{F,0}(x,y) \right|$$

Equation 4.6

Equation 4.6 shows this relationship where *F* is the SF band considered, and *j* is the morphed (test) image considered. *j=0* stands for the reference image.

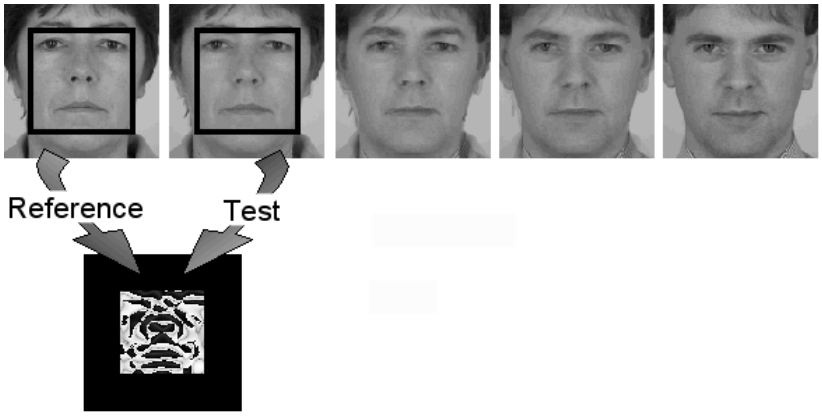

Figure 4.4: Sampling between the reference and each of the pictures of the morph series. The absolute value of the contrast difference at each point (*x, y*) between the reference picture and all the test pictures is measured for the central 68 x 68 points. This is compared to the average of those two contrasts.

The average contrast value $\overline{C}_{F,j}(x,y)$ was also computed for each pair of points.

$$\overline{C}_{F,j}(x,y) = \frac{\left| C_{F,j}(x,y) + C_{F,0}(x,y) \right|}{2}$$

Equation 4.7

Subsequently, we recorded for each value of $\overline{C}_{F,j}(x,y)$, the difference between $\Delta C_{F,j}(x,y)$ and the corresponding point on the dipper (See Figure 4.5).

$$I_{F,j}(x,y) = \left[\log\left(\Delta C_{F,j}(x,y,\overline{C})\right) - \log\left(D_F(\overline{C})\right)\right]$$

Equation 4.8

In the previous equation $I_{F,j}$ stands for the difference between our dataset and the dipper (similar to that of Figure 4.6) and D_F is the corresponding dipper value. In Figure 4.5, points above the dipper represent changes in contrast that are already *suprathreshold* and should be easily detected by the observer. Points below the dipper represent changes in contrast that are not yet detectable by the observer and points lying on the dipper represent values of contrast change just discriminable by the observer.

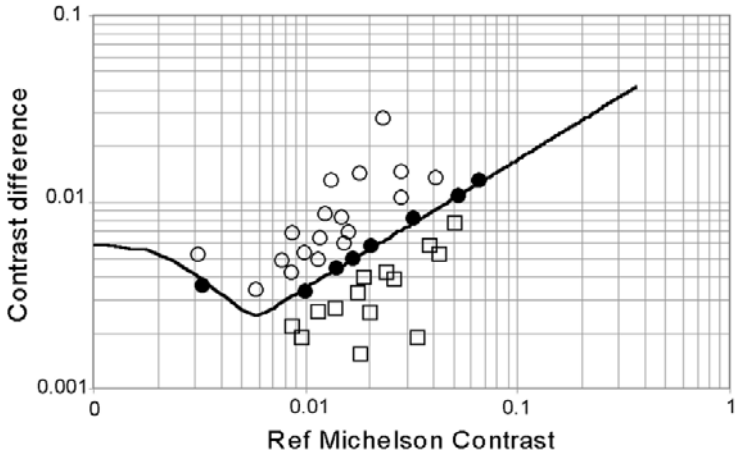

Figure 4.5: Detectability of changes in the morph sequence according to the 'dipper'. Points represent predicted values of ΔC Plotted against their average value $\overline{C}$. Hollow circles are values that are *suprathreshold* (i.e. easily detected by the observer). Filled circled are values that are just discriminable (at threshold) and squares are values that are not yet discriminable.

The average difference for all the central (68x68) points in each picture *j* and for each band *F* was calculated.

$$\bar{I}_{F,j} = \frac{\sum_{x,y} I_{F,j}(x,y)}{68 \cdot 68}$$

Equation 4.9

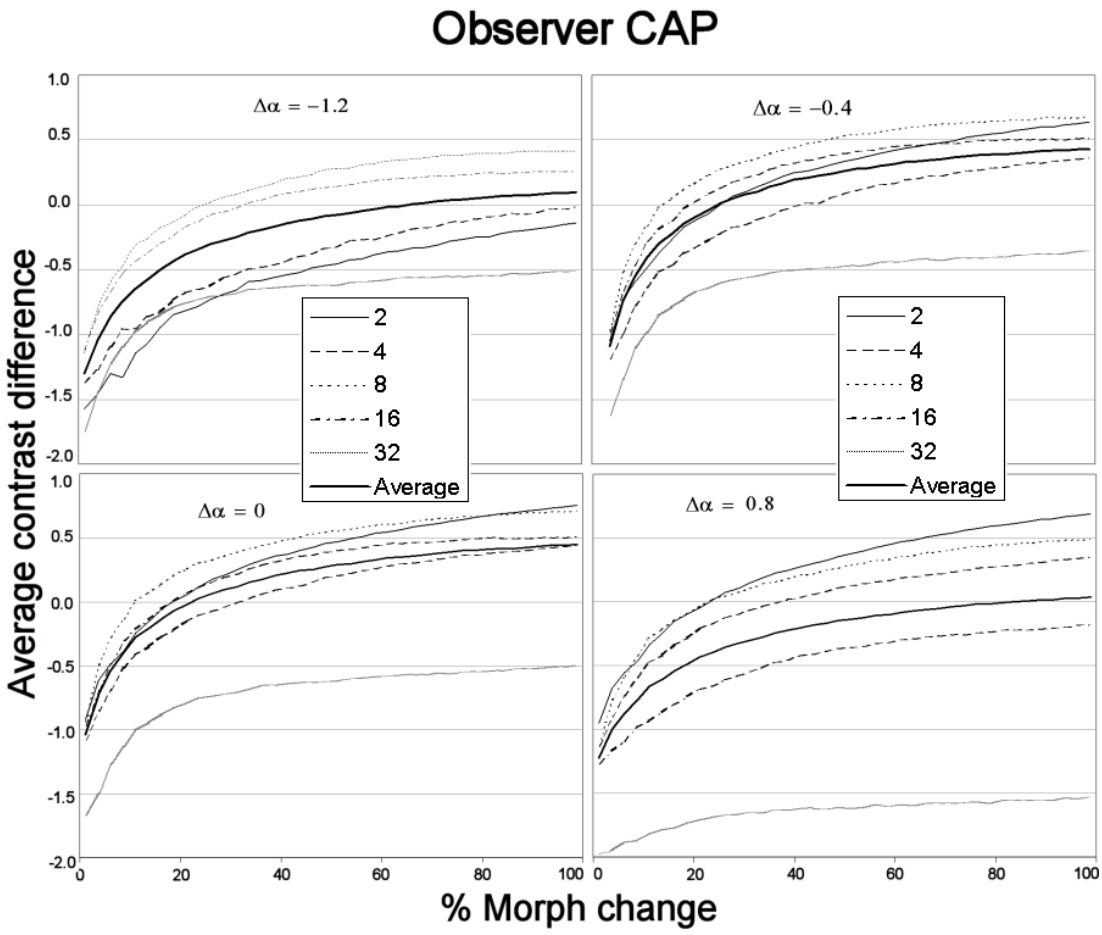

Figure 4.6: Four examples of averaged contrast difference between reference and test pictures (jnds) for each of the pictures in a sequence. The *x*-axis shows the percent of morph change in the sequence. The exemplary plots correspond to Δα equal to –1.2, -0.4, 0 and 1.2.

Plots in Figure 4.6 show the average difference $\bar{I}_{F,j}$ (just noticeable differences or *jnd* - in *log* units) between the predicted contrast difference and the dipper. Each line corresponds

to one SF band (central SF = 2, 4, 8, etc. cycles per picture). Values of jnd below zero represent contrast changes that should be invisible to the observer and values above zero represent changes in contrast that should be visible. The process was repeated for all SF bands *F* and all the values of $\Delta\alpha$ including the "unnatural" sequences ($\Delta\alpha \neq 0$).

Figure 4.7 shows the morph discrimination thresholds predicted by each of the SF channels as a function of the amplitude slope offset ($\Delta\alpha$) for the same data shown in Figure 4.6. In this exemplary case, the lowest morph discrimination threshold is predicted for $\Delta\alpha = 0$ and the channel corresponding to SF= 8 cycles/degree.

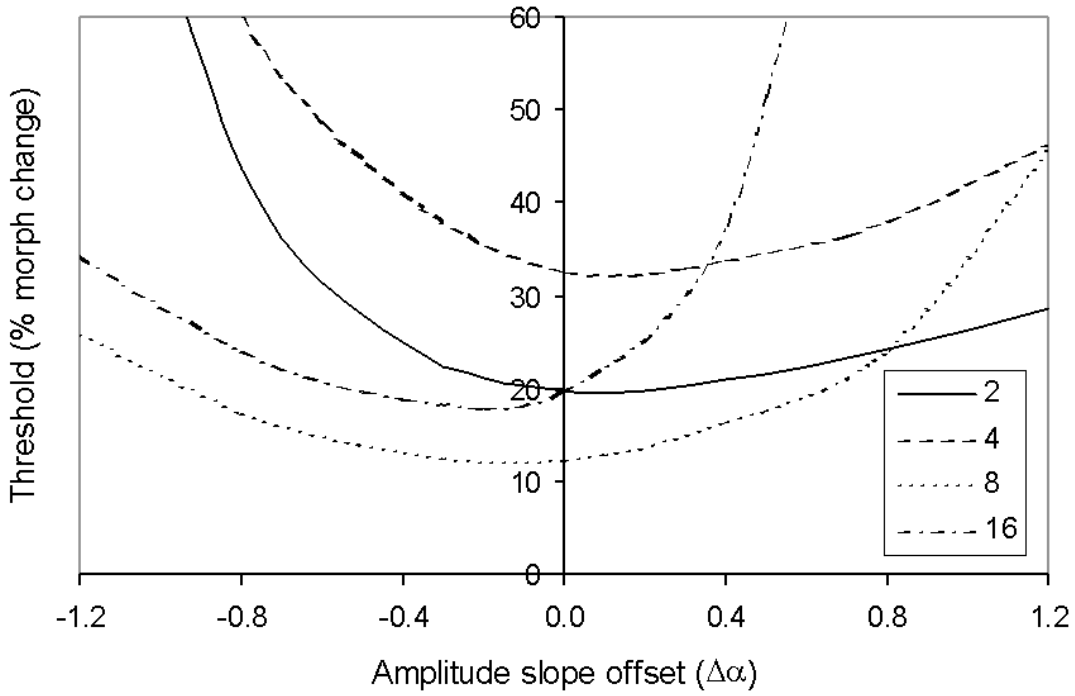

Figure 4.7: Values of morph threshold predicted by each of the SF channels as a function of the amplitude slope offset ($\Delta\alpha$). This figure was obtained by replotting the data from Figure 4.6. The values corresponding to SF= 32 cycles/deg are not shown here since they never produced a "crossing".

4.2.4. Contrast threshold and probability summation across space

Early studies (Robson and Graham 1981; Sachs *et al.* 1971) on the relationship between low-contrast stimuli and the narrowly-tuned spatial frequency channels have pointed out the need to account for non-linear interactions among channels when modelling. One of these models (called the *probability summation model*) (Sachs *et al.* 1971) consists of a large number of independent channels applied to the original stimulus. Noise (modelled as a Gaussian probabilistic sum) is then added to the stimulus to account for the variability of the subject's responses. A computationally simpler version (called the *vector magnitude model*) combines the outputs of all the channels in a non-Euclidean summation before the noise is added and the detection decision taken (Quick 1974). These models account for a range of psychophysical results and can predict detectability of periodic patterns (Graham *et al.* 1978; Kulikowski and Tolhurst 1972; Robson and Graham 1981), image discrimination (Watson and Robson 1981) and detection (Rohaly *et al.* 1997).

A second stage of our algorithm was developed to model any interactions among receptors in different SF-channels and among the channels themselves. It contemplates the possibility that all receptive fields (at each *x, y* location) interact, placing a weight on the fields that cross the dipper first and also considering the same type of rule to act across channels. We used the Minkowski sum of individual contributions with exponent β equal to 4 (approximation to probability summation) (Graham *et al.* 1978; Quick 1974; Robson and Graham 1981; Rohaly *et al.* 1997). General case:

$$Output = \left(\sum input^{\beta} \right)^{1/\beta}$$

Equation 4.10

The interaction among receptive fields is modelled as:

$$I_{F,j}(x,y) = \frac{\Delta C_{F,j}(x,y,\overline{C})}{D_F(\overline{C})}$$

Equation 4.11

which is similar to writing the following expression:

$$I_{F,j}(x,y) = \log^{-1}\left[\log\left(\Delta C_{F,j}(x,y,\overline{C})\right) - \log\left(D_F(\overline{C})\right)\right]$$

Equation 4.12

Equation 4.11 and Equation 4.12 show a change in the metric used to estimate the difference between $\Delta C_{F,j}(x,y)$ and the corresponding point on the dipper. Instead of computing this as a difference in log space, it is now computed as a ratio. The reason for this metric change is purely computational and its overall effect is to increase the weighting of points that are well over the dipper over those that are below when calculating the average difference $\bar{I}_{F,j}$ (see Equation 4.13).

$$\bar{I}_{F,j} = \sqrt[4]{\frac{\sum_{x,y}\left(I_{F,j}(x,y)\right)^4}{68 \cdot 68}}$$

Equation 4.13

If two SF-channels reach the detection criteria at the same time, then the picture is likely to be more discriminable than if only one crosses. With this in mind, we applied the same weighting laterally across channels.

$$jnd = \sqrt[4]{\sum_{F} \bar{I}_{F,j}^{\,4}}$$

Equation 4.14

Equation 4.14 shows how we estimated the just noticeable difference from one morph picture to another including probability summation across space in our calculations. Note that Equation 4.9 is a special case of Equation 4.13 where the Minkowski exponent β is equal to 1.

Figure 4.8 shows an example of how the discrimination threshold is calculated from the Minkowski lateral summation across channels. The plots show how $\bar{I}_{F,j}$ varies as the morph sequence changes in four different cases ($\Delta\alpha = -1.2, -0.4, 0, 0.8$). Thick lines show how the Minkowski summation of these values (determined by Equation 4.14) vary along the same axis. The experimentally-measured morph discrimination threshold for natural scenes ($\Delta\alpha = 0$) is introduced (see bottom-left panel in Figure 4.8) and its value determines the "*magic number*" that will represent the threshold for all other morph sequences in this version of our model.

$\bar{I}_{F,j}$ (see Equation 4.9) represents the contribution of each channel F to the detection process in every image j of the morph sequence. In our first modelling approach, we treat the various spatial frequency channels as operating independently of each other. This is despite results from research investigating masking of one spatial frequency component by another (Foley 1994). To account for these masking effects, lateral interactions are commonly added so that the contrast gain of a channel is reduced by activity from other channels (Teo and Heeger 1994; Watson and Solomon 1997). Rohaly *et al* (Rohaly *et al.* 1997) explored the effect of adding an overall contrast gain factor to three different types

of model and concluded that the least improved was the multiple spatial frequency channel model. In our case, we decided to start with the simplest case (no arbitrary parameters) and introduce arbitrary parameters later.

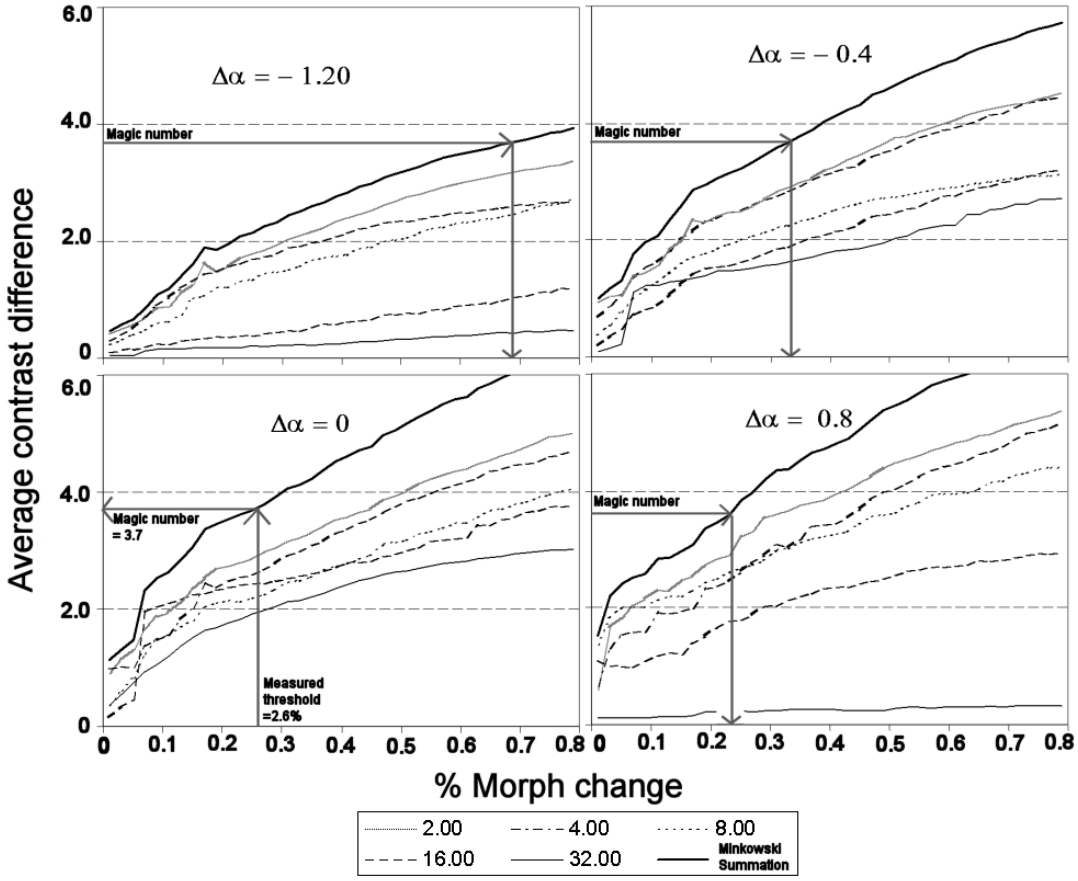

Figure 4.8: Example of how the morph discrimination threshold is determined for Rule 4. The graphs show the values of $\bar{I}_{F,j}$ along the morph sequence for each SF channel. The thick line represents the Minkowski summation of these values. In the simplest case, measured morph discrimination thesolds for natural statistics determine the "magic number" that would decide the model's predictions for all other morph series.

To simplify matters, we used two different Minkowski exponents (β) to represent models with or without receptor and channel interaction and two different jnd threshold criteria, which are explained below.

4.2.5. Modelling rules

The contribution of the SF-channels *F* to the determination of the threshold image *j* was modelled using 2 different rules (labelled after the Minkowski exponent β) Rule 1 and Rule 4, each one includes two variations *a* and *b* (depending on the free parameters used)

a) *Rule 1a:* No receptor and channel interactions. No free parameters. This could be also called "winner takes all" rule. The Minkowski exponent considered is β=1 which means that the probability that one channel will signal is independent from the probability that another channel will. The threshold value of *jnd* considered is also the simplest possible: the *first spatial frequency band* where the jnd is zero. The picture pair *j* and *j+1* where this happens is then recorded and the corresponding % morph threshold is calculated by linear interpolation. This extremely simple criterion allowed us to determine a discrimination threshold without adding any free parameters to the model.

b) *Rule 1b:* No receptor and channel interactions. Free parameter (best experimental data fit). β is still equal to 1 but a free parameter was introduced. This parameter determines the amount of vertical shift that needs to be applied equally to all dipper functions so that the best fit between the model results and the data points is achieved (considering all seven morph sequences). This is equivalent to the addition of the same constant value to all curves in Figure 4.6 so that the jnds obtained best fit the experimental data.

c) *Rule 4a:* Interaction among receptors and channels. Free parameter (fit forced through central point). β is equal to 4 (probability summation) to account for interactions between the different receptors and channels. A threshold *jnd* value was decided by arbitrarily forcing the model to fit one of the measured data points ($\Delta\alpha = 0$). In practice, this means to find what value of a threshold *jnd* that would make the model fit the central point and use this "*magic number*" to determine the discrimination threshold for all other conditions (different values of $\Delta\alpha$). See Figure 4.8 for an illustration on how this is accomplished.

d) *Rule 4b:* Interaction among receptors and channels. Free parameter (best experimental data fit). β is again equal to 4 but a threshold *jnd* value was arbitrarily decided so that the model best fits the psychophysically measured data points for the 7 sequences tested ($\Delta\alpha$ = -1.2, -0.8, -0.4, -0, 0.4, 0.8, 1.2). The best fit was decided in the same way as in Rule 1b.

Figure 4.9 shows a flow chart (from top to bottom) of the different modelling stages. There it is possible to see how every picture in the morph sequence is compared to the reference and its visibility is determined by a set of rules that range from the very simple (Rule 1a) to the more complex interactions between channels and receptors (Rule 4a and 4b).

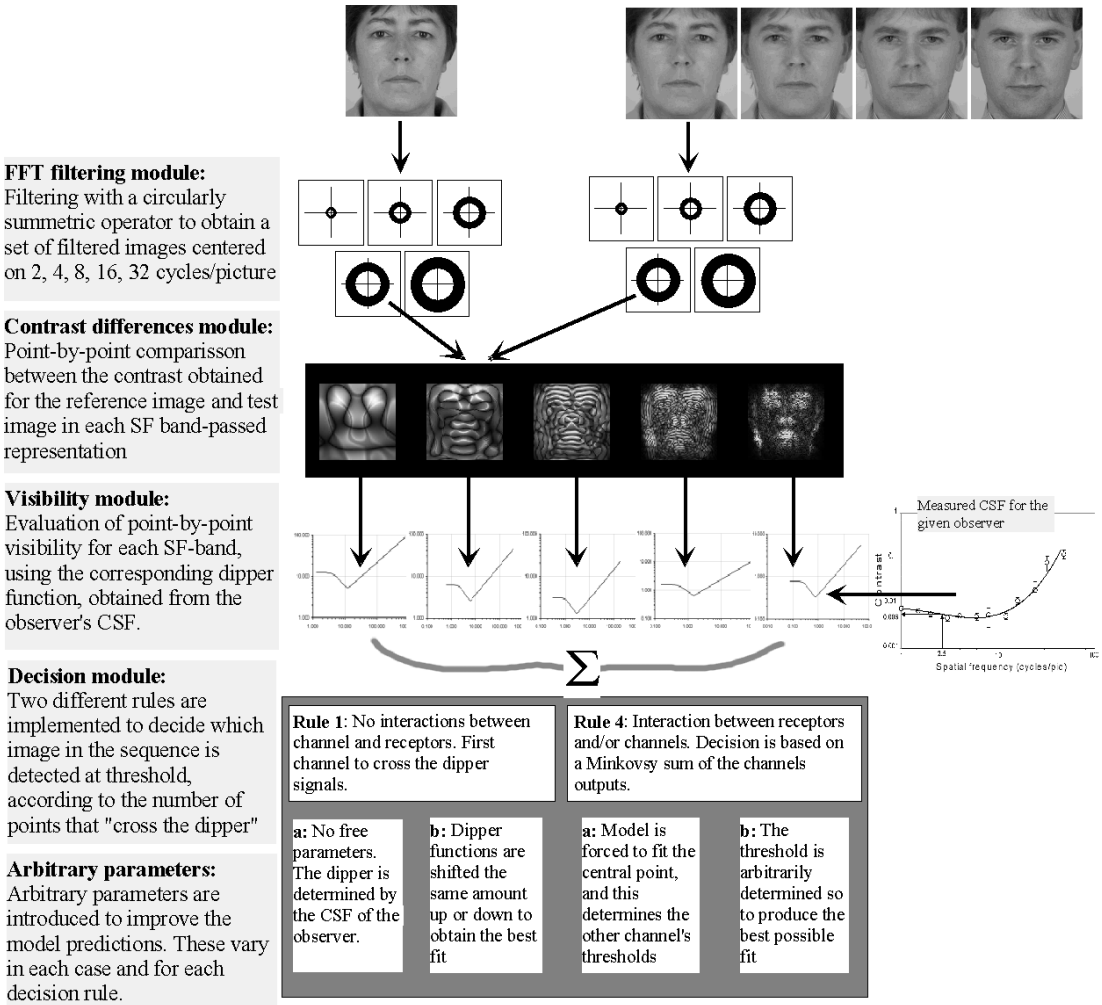

Figure 4.9: Flow chart of the model including all its rules and variants. From top to bottom it shows the analysis that is performed in every single image of the morph sequence until a set of decision rules determine which one is "visible at threshold".

To estimate the best experimental data fit in all cases, an iterative algorithm was used to find the value of the discrimination threshold (or "magic number") that minimises the relative residual sum of squares (SSE_R) between the model and the experimental results, where SSE_R is defined by:

$$SSE_R = \sum_{i=1}^{n}\left[(Y_i - \hat{Y}_i)^2 \cdot W_i\right] = \text{Relative (or weighted) residual sum of squares.}$$

Equation 4.15

With Y_i being the experimental values and $\hat{Y}_i$ the model's predictions. n is the total number of data points in one plot (n= 7). W_i (the set of weighting parameters) is determined by:

$$W_i = \frac{1}{(SE)^2} \text{; normalised so that } \sum_{i=1}^{n} W_i = n$$

SE represents the standard error of each experimental data point i as discussed in Appendix A. Under this definition, if SSE_R is equal to 0, the curve passes through every data point.

The computer algorithm used to implement our discrimination model in all versions was based on an algorithm originally written by Dr D.J. Tolhurst in Pascal programming language. Our algorith was written in Delphi 3 and runs on a 32 bit (Windows 95) platform.

4.3. Results

4.3.1. Modelling experimental results I (binocular, foveal data).

Rule 1:

Figure 4.10 shows the experimental results described in section 2.3 and the predictions corresponding to Rule 1a (solid lines) and Rule 1b (broken lines) of the model. The points represent the actual experimental results and standard errors.

From Figure 4.10 it is possible to see that the predictions of the model (using both variations of Rule 1) fit the discrimination data quite well for both observers. In general, the model predictions have the right shape for both sets of data and in many cases the

broken line goes through much of the data (in few occasions the two curves overlap). This is generally true for the two cases considered.

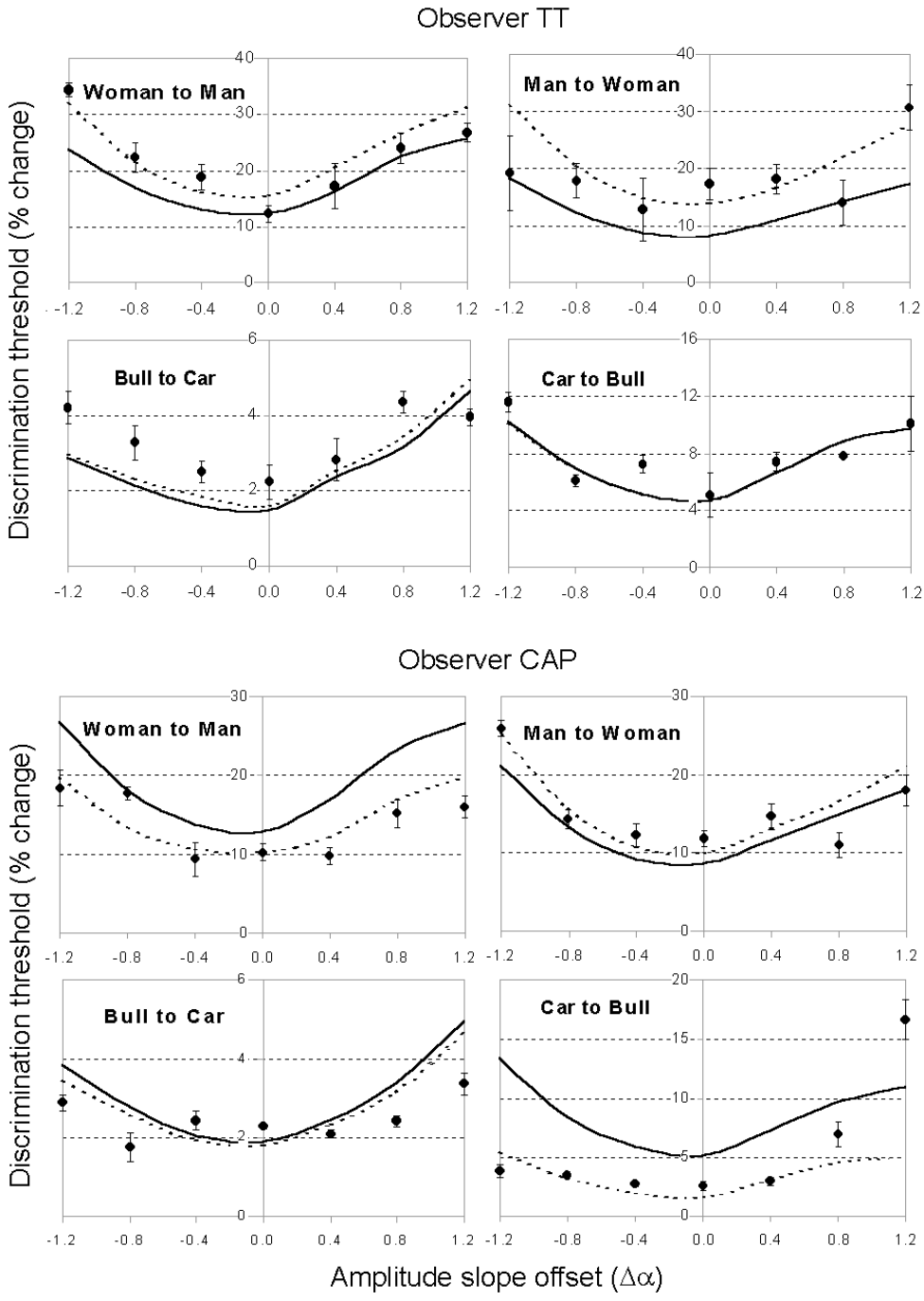

Figure 4.10: Experimental results (black dots) and model predictions using Rule 1a (solid lines) and Rule 1b (broken lines) for observers TT and CAP. Error bars are also shown. Foveal binocular viewing.

One of the observers (TT) seems to agree thoroughly with the model while the other (CAP) agrees in the general shape (U-shape) but seems to have detected some clues that the simplest version of the model did not account for, especially for the car-to-bull sequence. Adjusting the discrimination threshold for *jnds* so that it fits as much as possible data points did produce an improvement on the model's results. Notice that any shift of the model results along the *y*-axis comes accompanied by (non-linear) changes in shape in the model results' curves. In practice, these non-linearities mean that sometimes the adjustment of the model's parameters to best fit the data does not produce dramatic changes in the goodness-of-fit of the model. This is discussed later.

Rule 4:

Figure 4.11 shows the same experimental results and the model's predictions using Rule 4a (central point forced fit) and Rule 4b (best experimental data fit). Solid lines represent the model's prediction in the case of Rule 4a and broken lines show the predictions for Rule 4b for observers TT and CAP.

An inspection of Figure 4.11 shows a predictable improvement on the model's fit when we allow the whole system to adjust to best fit the dataset. The choice of the central data point ($\Delta\alpha=0$) is not always the best one for Rule 4a. In some cases, for example man-to-woman for observer TT, forcing the model to go through this point reduces considerably the quality of the prediction. In section 4.4.3 (discussion section below), we evaluate whether the addition of the hypothesis of interaction among channels and the free parameter has produced a significant improvement to our modelling.

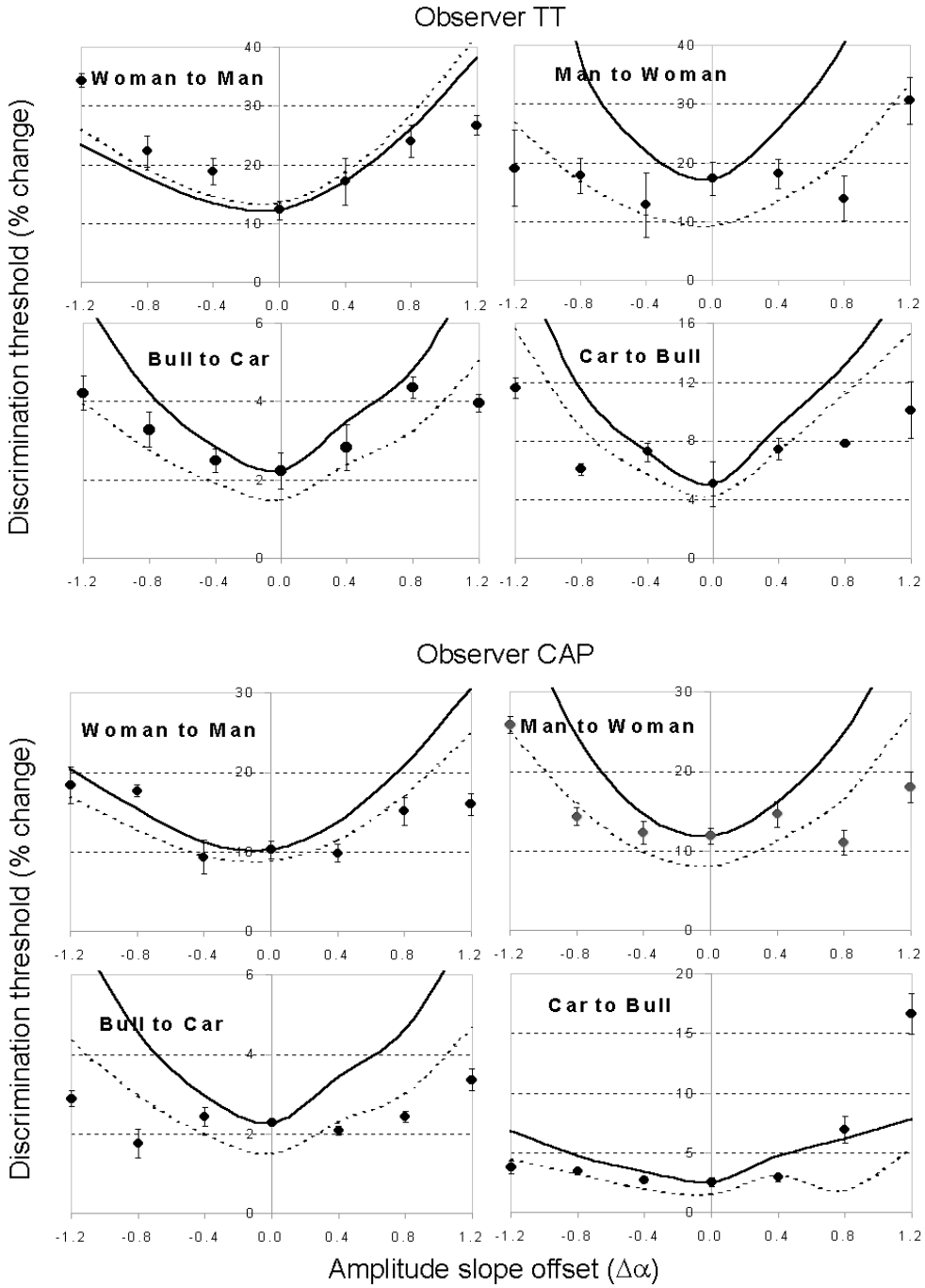

Figure 4.11: Experimental results (black dots) and model predictions using Rule 4a (solid lines) and Rule 4b (broken lines). Observers TT and CAP in foveal, binocular viewing.

4.3.2. Modelling experimental results II (monocular, foveal and peripheral data).

Using a similar criterion as in the previous section, we applied our model to some of the monocular data obtained before (see section 3.2.1) for stimuli viewed foveally and peripherally. Only two subjects (the ones who completed all experiments and conditions, KB and CAP) had their results modelled. The number of spatial frequency channels considered was also reduced to five (comprising 2, 4, 8, 16 and 32 cycles/picture) to decrease the computational load of the model.

The thresholds for detecting sinusoidal luminance gratings were determined for all conditions relevant to the experiments described in 3.2.1 (foveal, 3º, 6º and foveal-small) viewed monocularly with the dominant eye. The thresholds were estimated as described above in section 4.2.2. When model predictions were outside the range of the dataset employed (e.g. when the model did not reach threshold at 100% morph change) it was considered to be equal to the extreme of the sequence (100% morph change). This happened in only in very few cases for extremely whitened sequences ($\Delta\alpha$= -1.2).

Rule 1:

Figure 4.12 to Figure 4.15 show a comparison between the modelling results (for Rule 1a and Rule 1b) and the experimental results. As with the previous results, there is an improvement with the addition of a degree of freedom to the model (broken lines in the figures), although in many cases this improvement is only marginal. In almost all cases the model follows the data and has the correct U-shape.

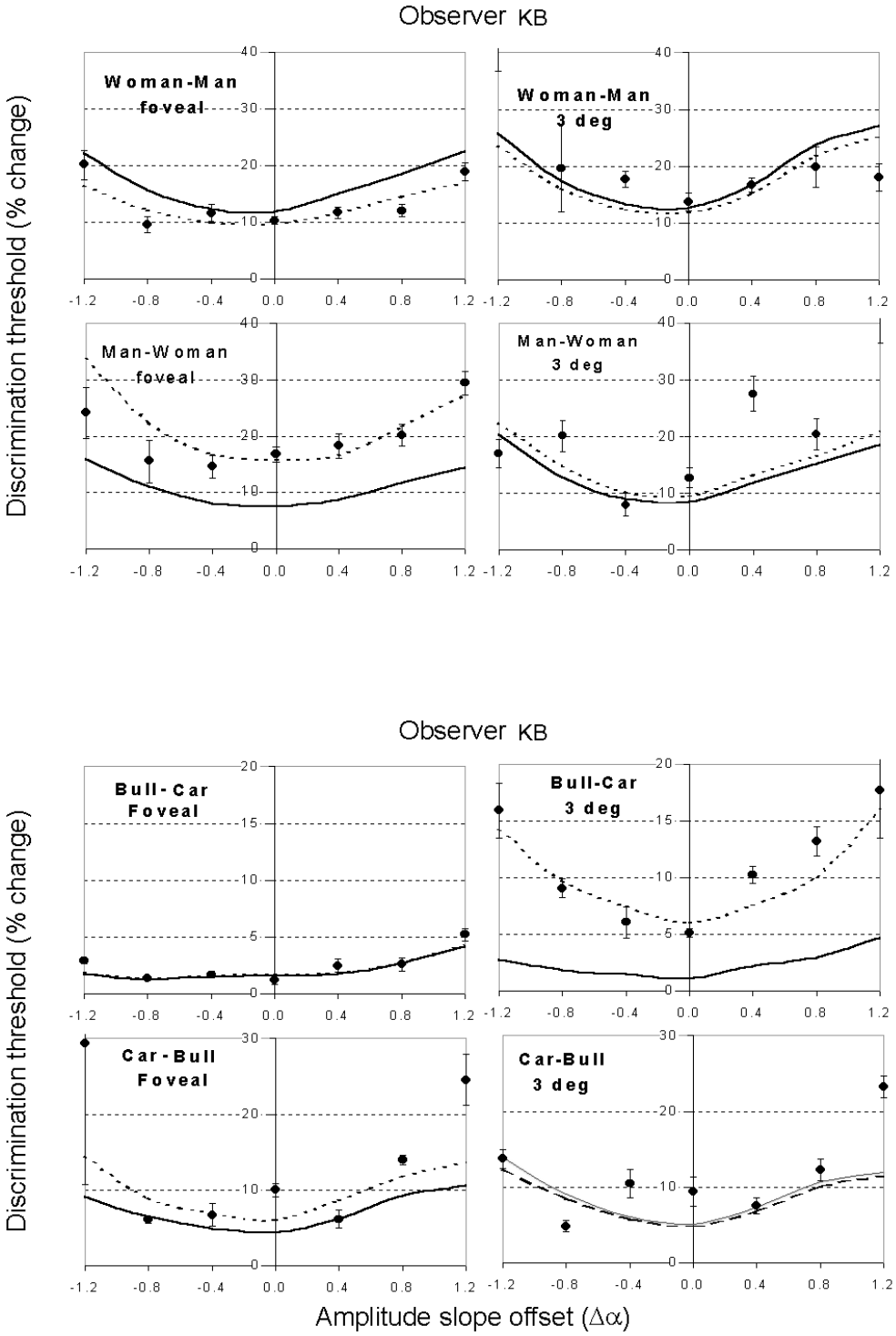

Figure 4.12: Experimental results (black dots) and model predictions using Rule 1a (solid lines) and Rule 1b (broken lines). Observer KB in monocular foveal and 3º viewing conditions.

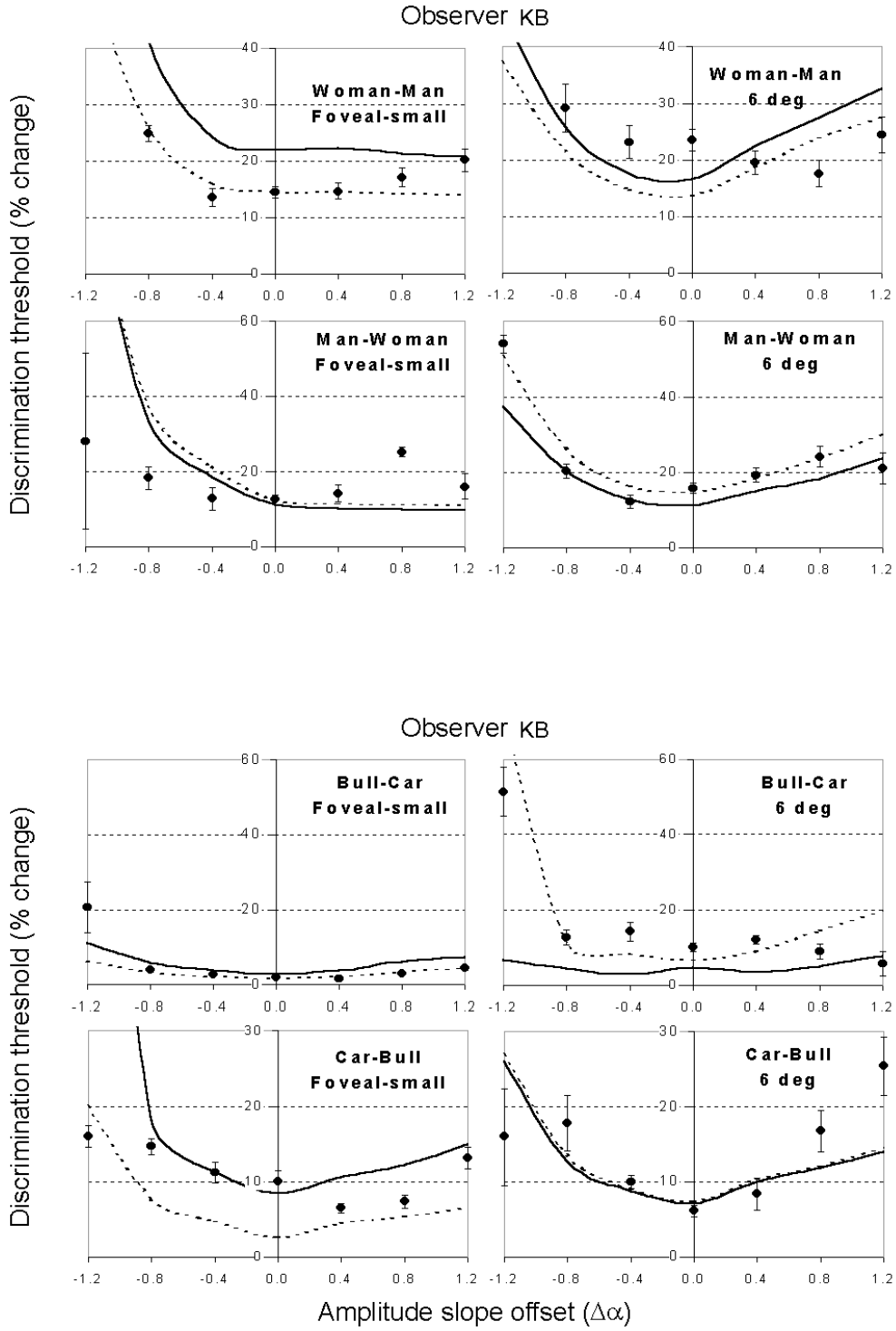

Figure 4.13: Experimental results (black dots) and model predictions using Rule 1a (solid lines) and Rule 1b (broken lines). Observer KB in foveal (small pictures) and 6º monocular viewing.

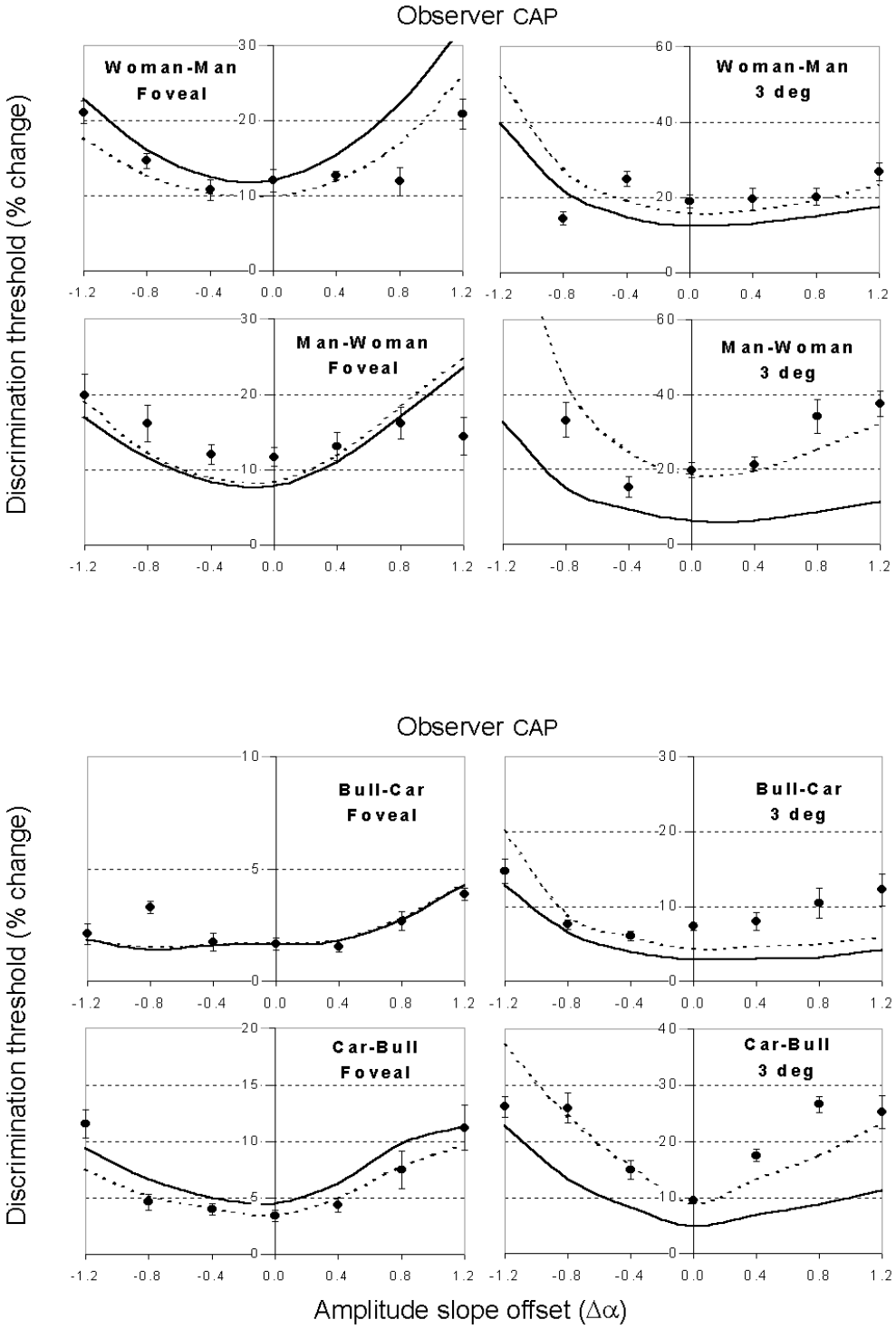

Figure 4.14: Experimental results (black dots) and model predictions using Rule 1a (solid lines) and Rule 1b (broken lines). Observer CAP in foveal and 3º monocular viewing.

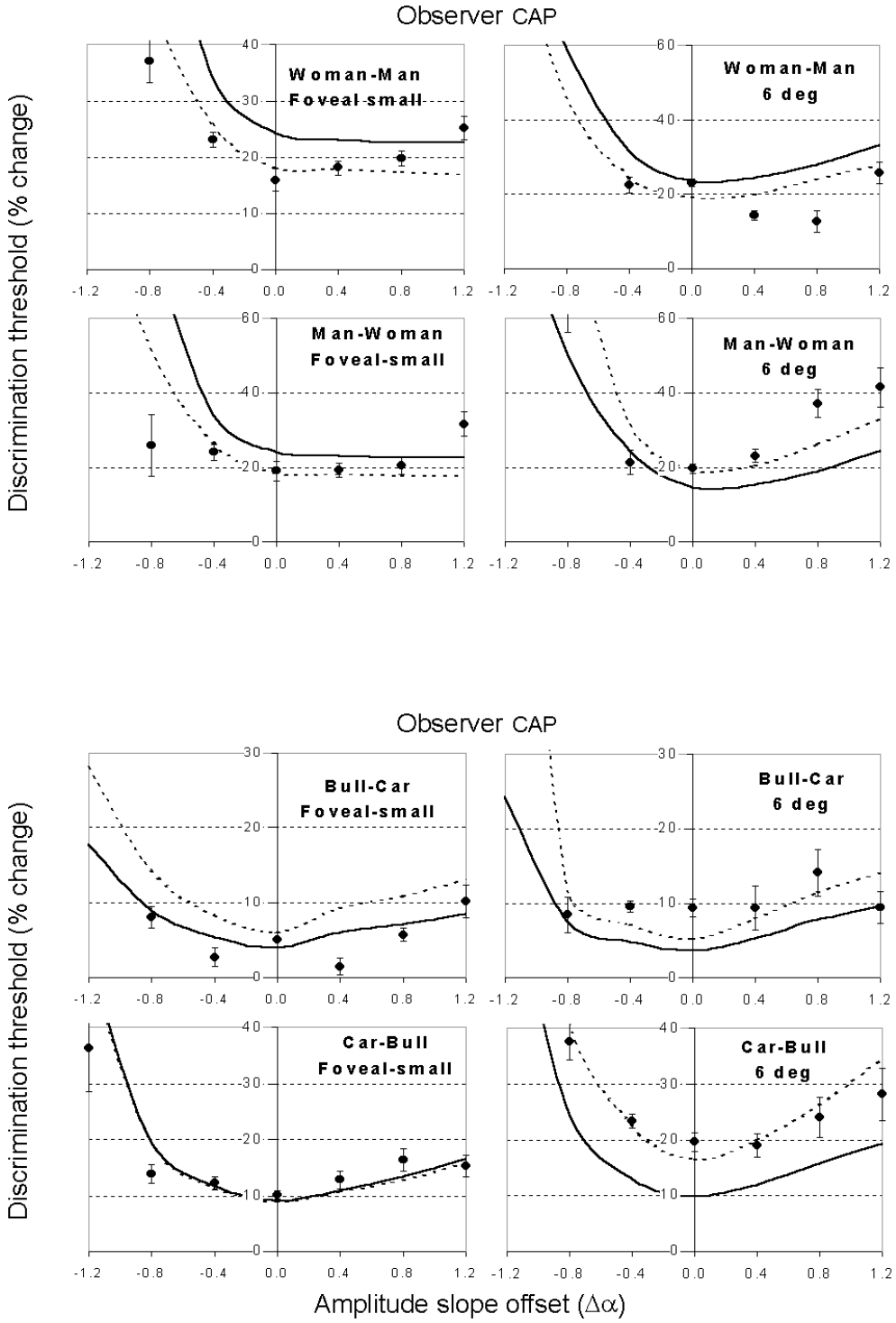

Figure 4.15: Experimental results (black dots) and model predictions using Rule 1a (solid lines) and Rule 1b (broken lines). Observer CAP in foveal (small pictures) and 6º monocular viewing.

There seems to be no qualitative difference in the modelling results between the observers. The success of Rule 1a in predicting the shape and in many cases the position of the data, is most striking, especially if one considers the simplicity of the hypothesis on which it was built.

Rule 4:

Figure 4.16 to Figure 4.19 show a comparison between the experimental results and the model's predictions using Rule 4a (solid lines) and Rule 4b (broken lines) for observers KB and CAP in all the conditions. As before, when the model did not reach threshold within the allowed range (0-100% morph change) its value was made equal to 100%, to represent a very high number. Again, there is a substantial improvement in allowing the model to adjust to the experimental dataset by means of a free parameter (Rule 4b, broken lines). Rule 4a is less compelling than we expected, especially considering that we forced it to fit the model at the central point ($\Delta\alpha = 0$). Some of the predictions of Rule 4a miss the data completely and are worse even than the simplest modelling case (Rule 1a). Again, there is no significant qualitative difference between observers and the predictions have in general the right U-shape.

There is an argument for choosing the central data point and not others as a determinant of the model's additional parameter for Rule 4a. The central point ($\Delta\alpha = 0$) generally produces the most consistent result across observers and experimental sets. Rule 4b has a truly "free" parameter whose only purpose is to minimise the relative residual sum of squares, thus providing the best experimental data fit. There is no question of an improvement in the model's fit between these two variants of rule 4, but we ask here: is there an improvement over the simplest version of the model, Rule 1a?

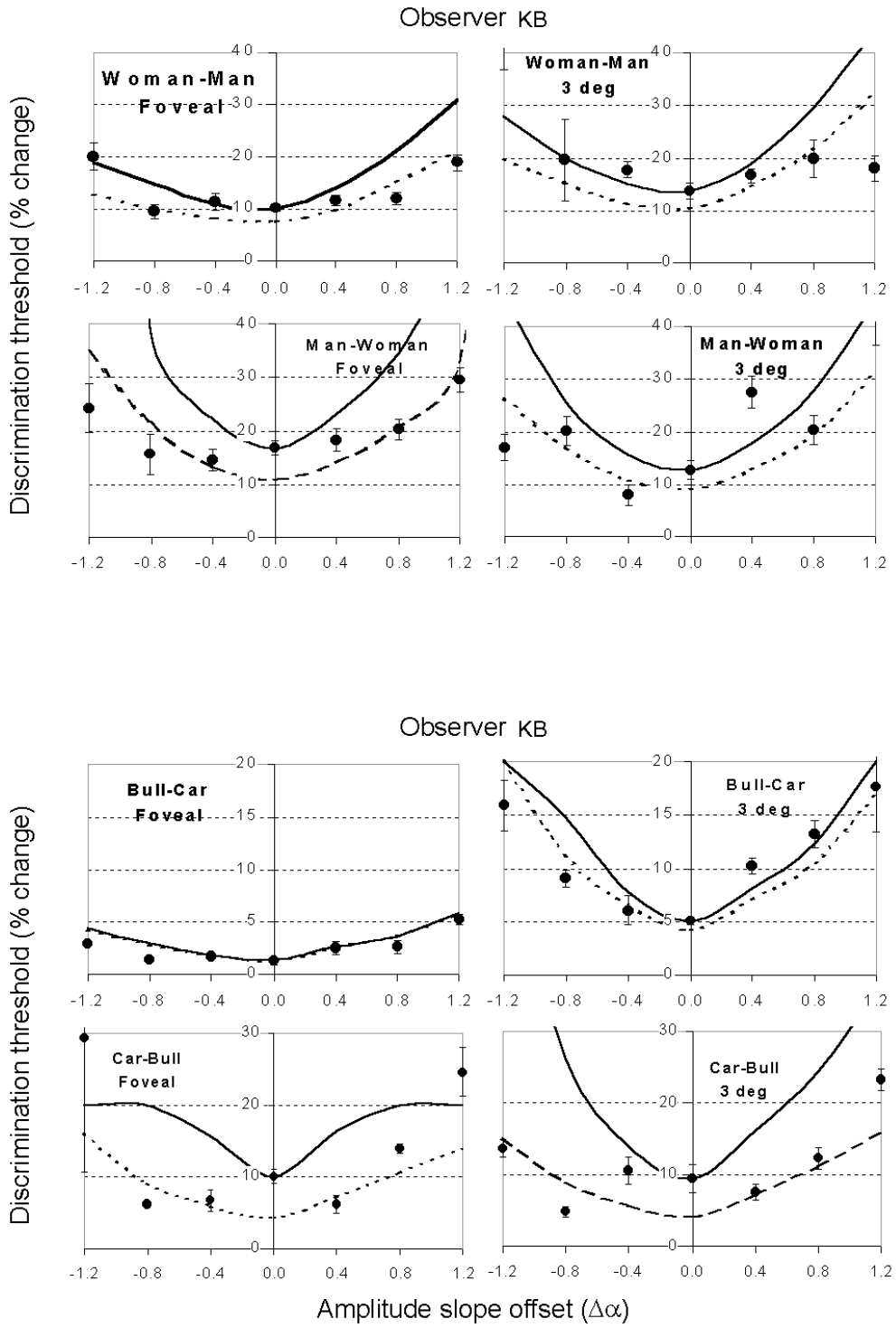

Figure 4.16: Experimental results (black dots) and model predictions using Rule 4a (solid lines) and Rule 4b (broken lines). Observer KB in monocular foveal and 3º viewing conditions.

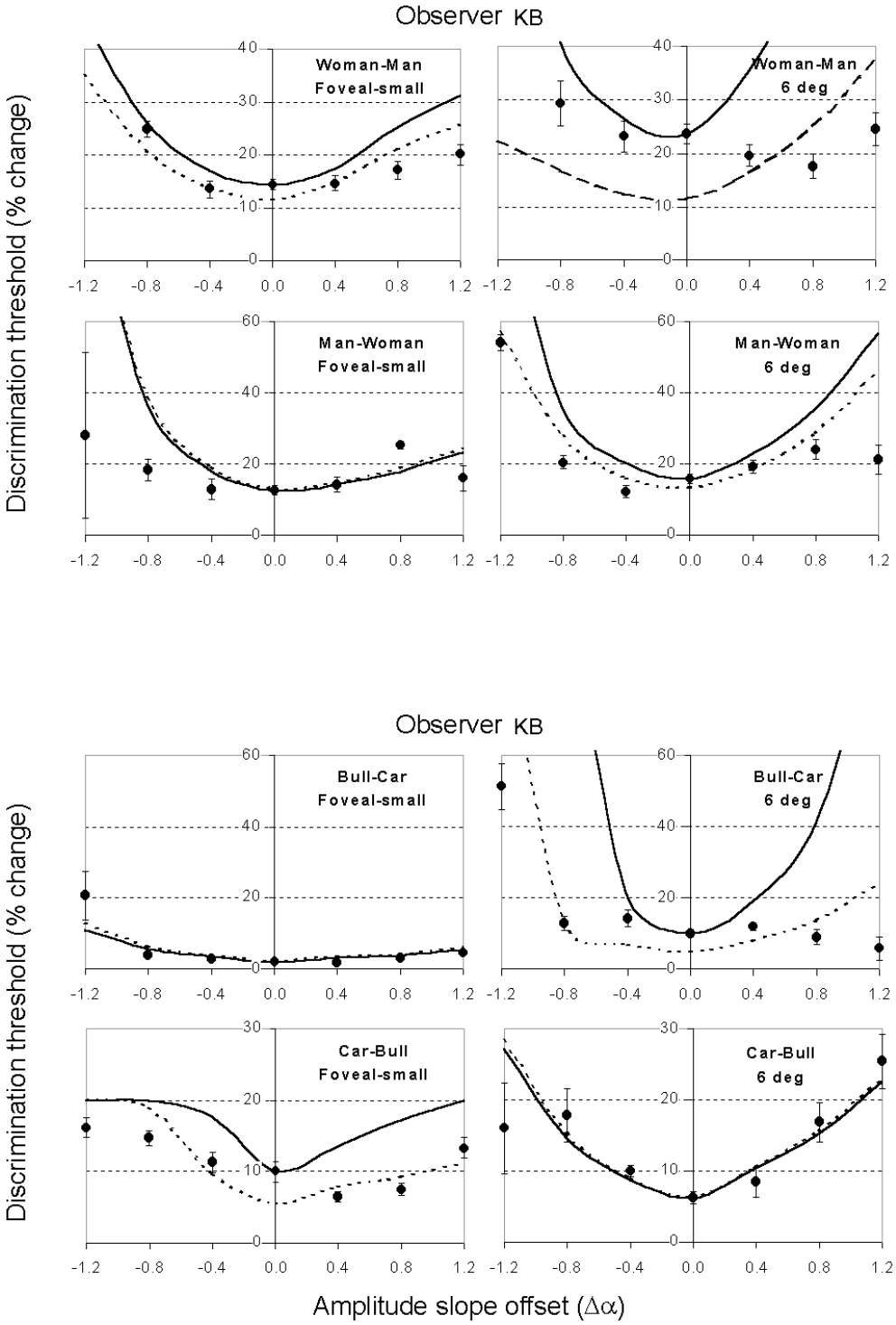

Figure 4.17: Experimental results (black dots) and model predictions using Rule 4a (solid lines) and Rule 4b (broken lines). Observer KB in foveal (small pictures) and 6º monocular viewing.

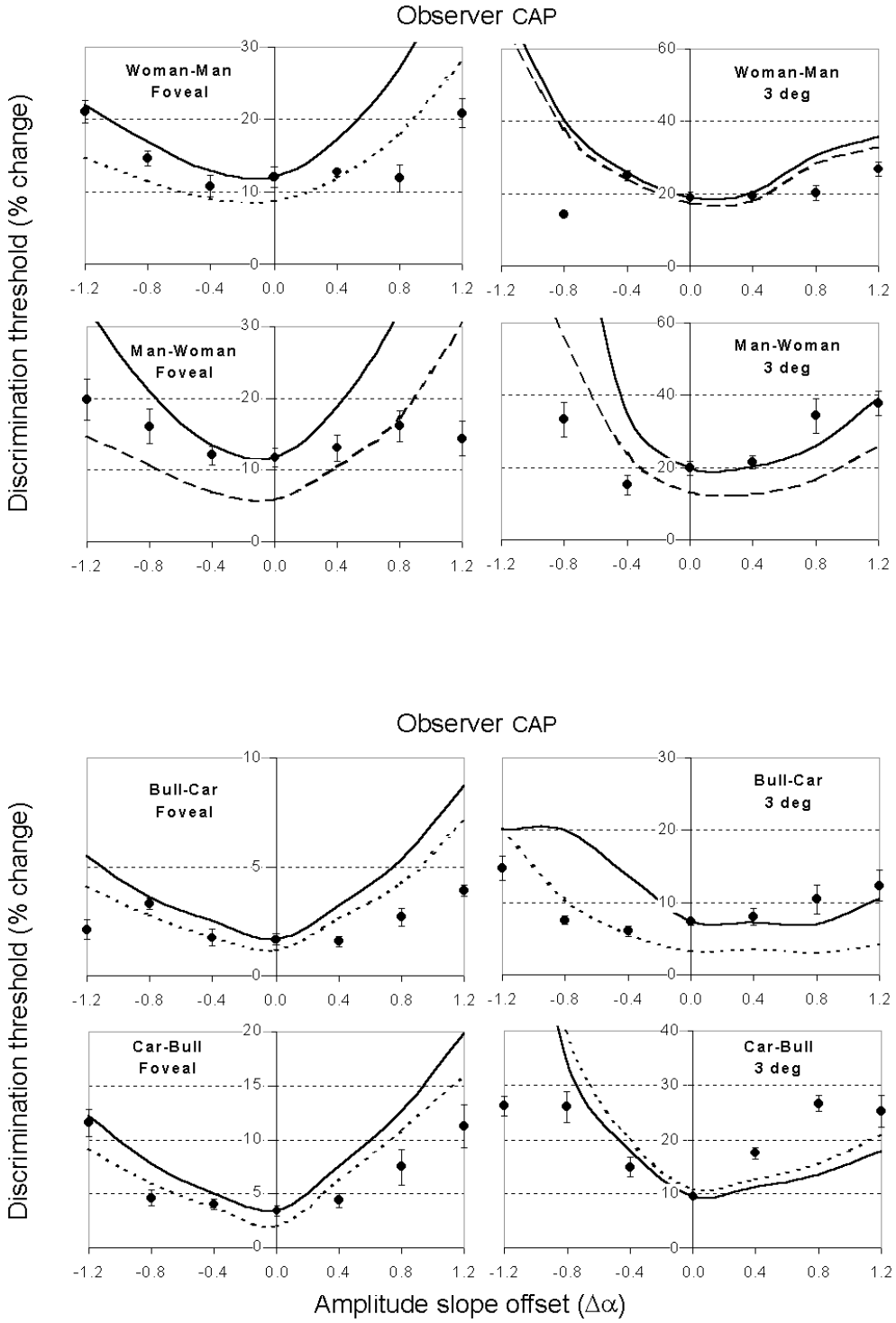

Figure 4.18: Experimental results (black dots) and model predictions using Rule 4a (solid lines) and Rule 4b (broken lines). Observer CAP in monocular foveal and 3º viewing conditions.

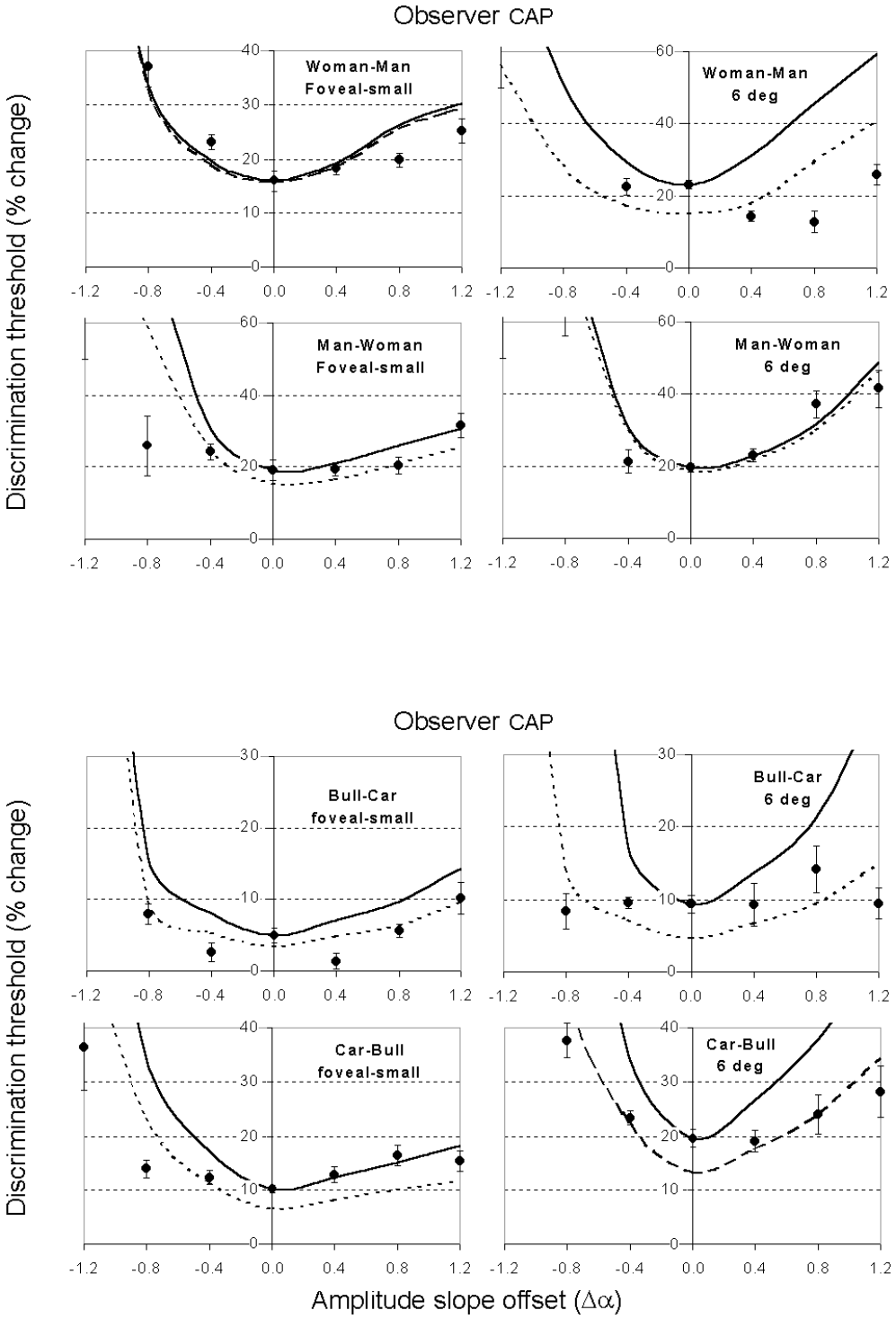

Figure 4.19: Experimental results (black dots) and model predictions using Rule 4a (solid lines) and Rule 4b (broken lines). Observer CAP in foveal (small pictures) and 6º monocular viewing.

4.4. Discussion

4.4.1. Binocular/foveal data modelling

Table 10 shows the calculated goodness-of-fit ($\hat{\sigma}$) for all our foveal modelling results (Figure 4.10 and Figure 4.11). $\hat{\sigma}$ is the standard deviation of the differences between the measured data and the curve generated by the fitted model. It gives an idea of how scattered the residuals are around the average. A perfect fit would yield a value equal to zero. It was calculated from the error variance as follows:

$$\hat{\sigma}^2 = \frac{SSE_R}{n-p} = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \cdot W_i}{n-p} = \text{Error variance.}$$

Equation 4.16

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \text{Goodness of fit (GoF).}$$

Equation 4.17

Where p is the number of free parameters in the model and n is the number of data points. SSE_R stands for Residual or Error Sum of Squares (relative) as defined in Equation 4.15. It is the sum of the squares of the differences between the actual data points and the predicted values, weighted by the standard error of the i^{th} data point W_i:

$$W_i = 1 \Big/ SE_i^2$$

normalised so that

$$\sum_{i=1}^{n} W_i = n$$

Equation 4.18

For Rule 1a, we considered p (number of free parameters) to be equal to 0. In all other cases, it was equal to 1.

	Goodness of Fit ($\hat{\sigma}$)			
	Rule 1(a)	Rule 1(b)	Rule 4 (a)	Rule 4 (b)
TT				
Woman to Man	0.0689	0.0324	0.0916	0.0895
Man to Woman	0.0764	0.0475	0.2597	0.0581
Bull to Car	0.0095	0.0098	0.0234	0.0094
Car to Bull	0.0108	0.0116	0.0579	0.0361
CAP				
Woman to Man	0.0564	0.0330	0.1233	0.0400
Man to Woman	0.0333	0.0246	0.0597	0.0462
Bull to Car	0.0058	0.0058	0.0164	0.0082
Car to Bull	0.0442	0.0135	0.0160	0.0140

Table 10: GoF for all models applied to the foveal/binocular data. Rule 1a does not contain any free parameter. Rule 4a corresponds to the central-point forced-fit variant. Rule 1b and Rule 4b adjusts the free parameter to obtain the best experimental data fit.

Table 10 shows that the standard error of the estimate is smaller for Rule 1b (average value of GoF=0.0182) in 75% of the experiments. The second best fit is (surprisingly) provided by Rule 1a, with no free parameters involved (average GoF=0.0256). The predictions obtained with Rule 4b were also very close (average GoF=0.0285) and the worst modelling rule was definitely Rule 4b (average GoF=0.0682). It is important to note that sometimes, the calculation of GoF could produce smaller values for Rule 1a than, for example, the optimised Rule 1b because of the inclusion of p (or number of free parameters) to calculate GoF *after* SSE_R has been minimised.

4.4.2. Monocular/foveal/peripheral data modelling

Table 11 shows the values of GoF considering the model's fit to the monocular data as shown in section 4.3.2 for subject KB. The table reveals that again, Rule 1b provides the best fit (50% of the experiments, average GoF= 0.417), while Rule 1a and Rule 4b come second with 25% of the experiments each. Rule 4a is definitely the worst, against our expectations.

KB	**GoF ($\hat{\sigma}$)**			
	Rule 1(a)	Rule 1(b)	Rule 4 (a)	Rule 4 (b)
KB- foveal				
Woman to Man	0.0363	0.0169	0.0557	0.0296
Man to Woman	0.0959	0.0287	0.1913	0.0513
Bull to Car	0.0068	0.0069	0.0130	0.0121
Car to Bull	0.0315	0.0322	0.1221	0.0378
KB- 3 deg				
Woman to Man	0.0427	0.0485	0.0938	0.0710
Man to Woman	0.0789	0.0769	0.1306	0.0709
Bull to Car	0.0597	0.0163	0.0242	0.0193
Car to Bull	0.0473	0.0504	0.2363	0.0421
KB- foveal-small				
Woman to Man	0.0952	0.0255	0.0503	0.0351
Man to Woman	0.1077	0.1144	0.3366	0.0776
Bull to Car	0.0174	0.0060	0.0081	0.0121
Car to Bull	0.2355	0.0497	0.0748	0.0286
KB- 6 deg				
Woman to Man	0.0692	0.0767	0.2690	0.1071
Man to Woman	0.0653	0.0393	0.1997	0.0659
Bull to Car	0.0862	0.0536	0.3882	0.0817
Car to Bull	0.0246	0.0258	0.0162	0.0158

Table 11: GoF for observer KB. All models were applied to the monocular/foveal/peripheral data. Rule 1a does not contain any free parameter. Rule 4a corresponds to the central-point forced-fit variant. Rule 1b and Rule 4b adjusts the free parameter to obtain the best experimental data fit.

Table 12 shows the equivalent data for observer CAP. It shows the same trend regarding the GoF of the model to the data as before. Rule 1b is the best in 56% of the experiments (average GoF=0.040), followed by Rule 4b (31% of the cases, average GoF= 0.06). Again, Rule 1a does a good job at predicting the results, considering the lack of free parameters (12.5% of the cases, average GoF= 0.07).

CAP	**GoF ($\hat{\sigma}$)**			
	Rule 1(a)	Rule 1(b)	Rule 4 (a)	Rule 4 (b)
CAP- foveal				
Woman to Man	0.0387	0.0207	0.0712	0.0338
Man to Woman	0.0416	0.0405	0.1461	0.0697
Bull to Car	0.0085	0.0081	0.1224	0.1147
Car to Bull	0.0154	0.0099	0.1795	0.1093
CAP- 3 deg				
Woman to Man	0.0843	0.0757	0.0276	0.0183
Man to Woman	0.2010	0.0519	0.0261	0.0172
Bull to Car	0.0362	0.0271	0.0774	0.0399
Car to Bull	0.0759	0.0361	0.1577	0.1682
CAP- foveal-small				
Woman to Man	0.0956	0.0379	0.0421	0.0407
Man to Woman	0.0891	0.0572	0.0856	0.0532
Bull to Car	0.0873	0.0854	0.0697	0.0320
Car to Bull	0.0237	0.0239	0.0826	0.0508
CAP- 6 deg				
Woman to Man	0.0849	0.0688	0.1710	0.1129
Man to Woman	0.0762	0.0461	0.0401	0.0361
Bull to Car	0.0480	0.0316	0.2662	0.0394
Car to Bull	0.0982	0.0230	0.1837	0.0450

Table 12: GoF for observer CAP. All models were applied to the monocular/foveal/peripheral data. Rule 1a does not contain any free parameter. Rule 4a corresponds to the central-point forced-fit variant. Rule 1b and Rule 4b adjusts the free parameter to obtain the best experimental data fit.

4.4.3. Does the addition of a degree of freedom and interaction among channels offer any significant improvements to the data modelling?

Our results show that, predictably, adding a free parameter improves the 'goodness of fit' of the model in both Rule 1 and Rule 4 cases. Is this improvement significant? A simple analysis (see Figure 4.20) of the average value of GoF across all observers and conditions tells us that in most cases, there is a significant improvement in the modelling when we add a free parameter (see darker bars in the same figure). In all but one case (man-to-woman), Rule 1a is as good as Rule 4b (error bars in the graph represent the standard error of the average, with n=10). Rule 4a produces always the largest values of GoF.

Figure 4.20: Average value of the standard errors of the estimate (or goodness of fit) across all observers and viewing conditions. Error bars represent the standard errors of the average. n=10.

From the same figure, it is possible to see that Rule 1 predictions are consistently better than Rule 4. This makes a strong case against the addition of interaction between channels

to our simple model. In practice it means that the simplest 'winner takes all' rule performs consistently better than the equivalent 'receptors and channels interactions' hypothesis.

4.4.4. Why is the simplest version best?

Perhaps Rule 4a is worse than Rule 1a because of inadequate choices of the modelling parameters. For example, the arbitrary value chosen for the discrimination threshold (or *magic number*) may have a determining influence on the model's poor predictions. There might be the case that choosing other values for this threshold (different to the central point of the experimental results) improves the model's results. An argument against this view is that choosing the best possible threshold value (i.e. Rule 4b) does not improve the best modelling results.

Another view suggests that the model may need to be calibrated properly. One way of doing so would be to create a pair of gratings that are *by definition* just discriminable (e.g. of contrasts C_t and $C_t + \Delta C_t$) and adjust the model so that all the points produced within a given channel are distributed equally above and below the dipper function. This is the approach taken by Rohaly *et al* (Rohaly *et al.* 1997) when they arbitrarily calibrated their models to predict contrast thresholds for grating patches in each of the SF-band pass channels. This calibration has to be done for all the SF-channels and for each of the Minkowski summation exponents considered.

4.4.5. Altering the model's assumptions

Arguably, making our model more sophisticated by calibrating each SF-channel to predict the observer's performance is likely to improve the model's predictions, but at the cost of adding another level of complexity. Our approach up to here has been very simple and has produced a good set of predictions. We believe there are other factors that may be

also considered if we want to truly understand the workings of the model, like the role of each of the assumptions based on physiological/psychophysical evidence. For example, what effect would channel bandwidths other than the physiologically measured 1.5 octaves (Blakemore and Campbell 1969; De Valois *et al.* 1982a; Movshon *et al.* 1978a; Tolhurst and Thompson 1981) have in our model? Our idea at this point is to investigate why is this simple model so robust and what are the main features of the model responsible for predicting our discrimination thresholds. We hope that exploring the consequences of changing some of the model parameters may help us draw some conclusions about the ways in which the HVS might operate.

Different filter bandwidths:

The top part of Figure 4.21 shows how the discrimination thresholds change when the spatial frequency bandwidth of the nine filters is increased or decreased in Rule 1a (0.5, 1.0, 1.5, 1.9, and 2.3 octaves). The morph sequence considered was woman-to-man and the observer was CAP (see Figure 4.10). All other values (shape of the CSF, shape of the dipper template, overall sensitivity, etc.) remain unchanged. The discrimination thresholds predicted by the model do not experience large changes for bandwidths smaller than 1.5 octaves. At this point, predictions start to diverge from measurements. The shape of the curve predicted by the model becomes more "tuned" and the lowest point shifts upwards and to the right side of the plot. These results are qualitatively similar to discrimination threshold results measured for 6º peripheral vision (Figure 3.14). This similarity may be related to changes in bandwidth of SF-channels in peripheral vision. The bottom part of the figure shows the same type of comparison in the rather extreme case when only one broadband channel was considered. Here, the model failed to produce the characteristic U-shaped curve.

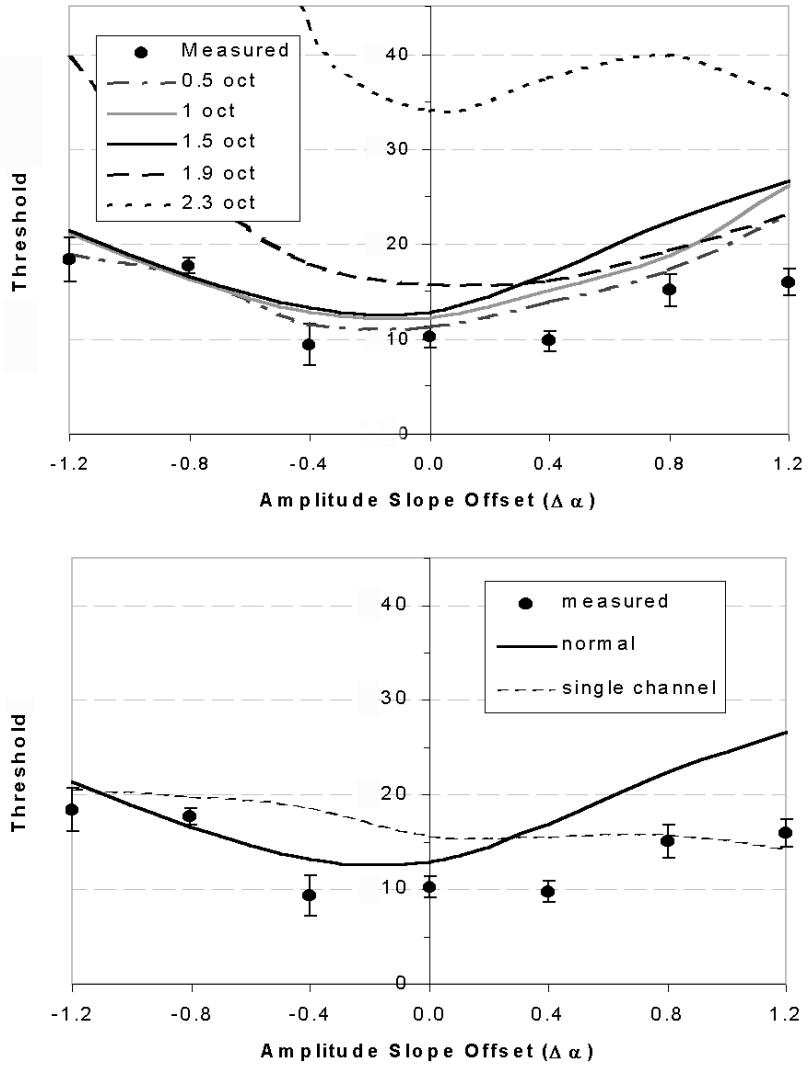

Figure 4.21: Different filter bandwidths. Top part shows the model's predictions when the filter bandwidths are modified to values different from 1.5 octaves (dark solid line). Bottom part shows the same for a single broadband channel. The morph sequence considered was woman-to-man for observer CAP. Corresponding experimental results are also shown (black dots). Rule 1a considered.

Different dipper template shapes:

The top part of Figure 4.22 shows a similar kind of plot as before, but in this case, the shape of the dipper template was modified to reflect five different shapes (named

"biological", "Weber", "no-dip" and "simple"). These are shown at the bottom part of the figure.

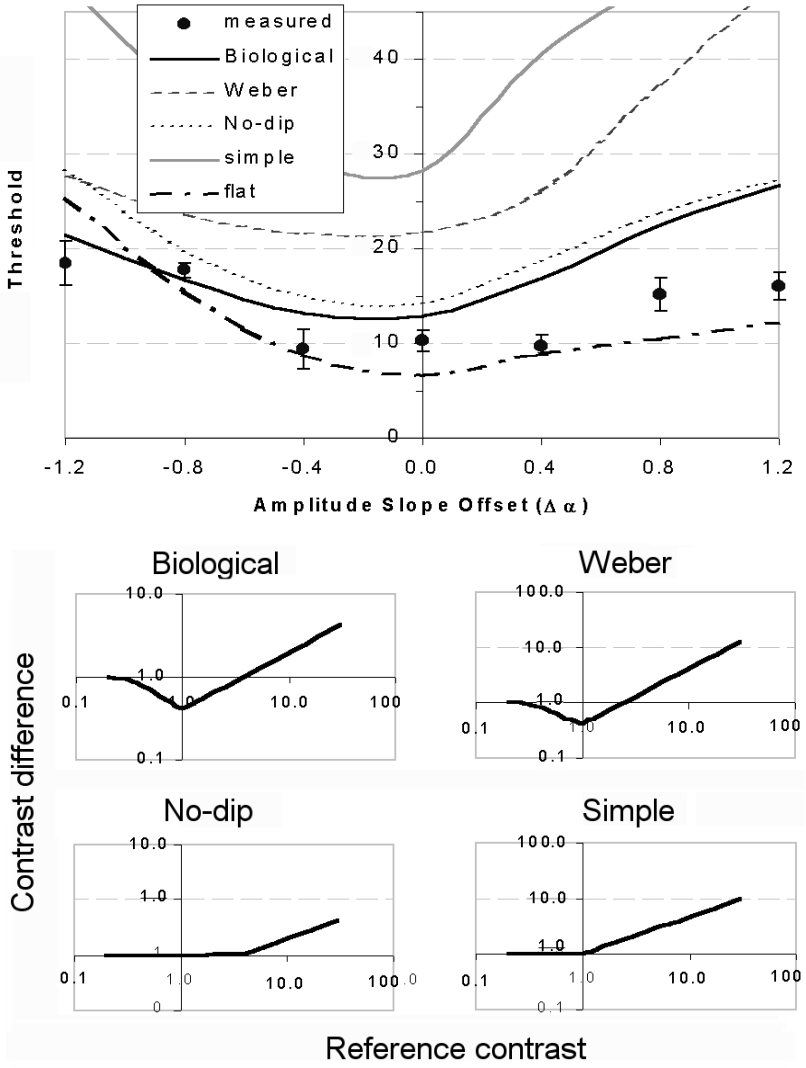

Figure 4.22: Different dipper template shapes. Top part shows the model's predictions when the dipper template is modified as shown in the bottom part of the figure. The morph sequence considered was woman-to-man for observer CAP. Corresponding experimental results are also shown (black dots). Rule 1a considered.

The first or "biologically plausible" one (see also Figure 4.1) is the familiar dipper function proposed by Legge and Foley (Legge 1981; Legge and Foley 1980; Nachmias and Sansbury 1974) and used in all our previous modelling. The second one has a dip but its

positive slope has been adjusted to follow Weber's law. The third one is a simplified version of the "biological" dipper where the "dip" has been eliminated and the fourth follows a simple function (a line in *log-log* coordinates) where the dip has also been eliminated. The fifth option ("flat" -not shown in the figure) is just a line where contrast difference is equal to 1 for all reference contrast values. As before, all other values (bandwidth, overall sensitivity, etc.) remain unchanged and the model rule used is Rule 1a.

Figure 4.22 also shows that for some of the dipper templates considered, the model's predictions do not get dramatically worse. The largest worsening effect is produced by the "simple" and the "Weber" version of the dipper. Surprisingly, the "flat" dipper does not make the model fail but quite the opposite. The reason for this may be that the flattening of the dipper function effectively lowers the contrast difference necessary for each receptor to signal and makes this value the same independently of the reference contrast. An organism with such contrast detection mechanism will always be able to detect the same changes in local contrast independently of the reference contrast, and is expected to have smaller morph discrimination thresholds than our average observer. The fact that their discrimination thresholds are similar to those measured psychophysically may show the need for some kind of calibration of our model algorithm.

Different CSF shape – influence of a shallower CSF:

The shape of the corresponding CSF (for observer CAP, see Chapter 2 on binocular data) was altered so that sensitivity for high spatial frequencies improves. Figure 4.23 shows what happens to the model's predictions when this occurs. The bottom part of the figure shows four different CSFs, which were produced by multiplying the curve 2, 3.3 and 10 times so that it looks more and more flat, increasing the observer's sensitivity to high spatial frequencies only.

Predictably, this produced lower thresholds for the images that have a higher content of detail (negative values of $\Delta\alpha$), keeping the threshold for the other sequences unmodified. All other values (bandwidth, dipper, overall sensitivity, etc.) remain unchanged and the model rule used is Rule 1a.

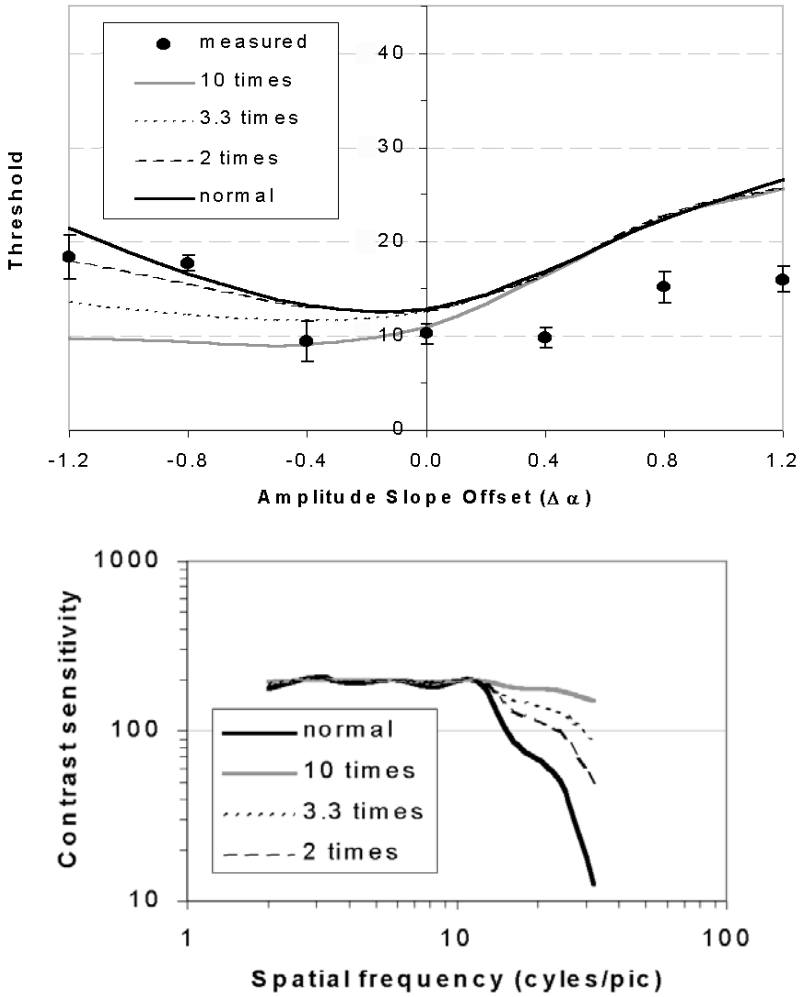

Figure 4.23: Influence of a shallower CSF. Top part shows the model's predictions when the shape of the CSF function is altered as shown in the bottom part of the figure. The morph sequence considered was woman-to-man for observer CAP. Corresponding experimental results are also shown (black dots). Rule 1a considered.

Different overall contrast sensitivity:

The overall value of the CSF was increased or decreased, to test the effects having different observers in our model. This was achieved by multiplying the observer's CSF data by $\sqrt{3}$, 3 and 1/3.

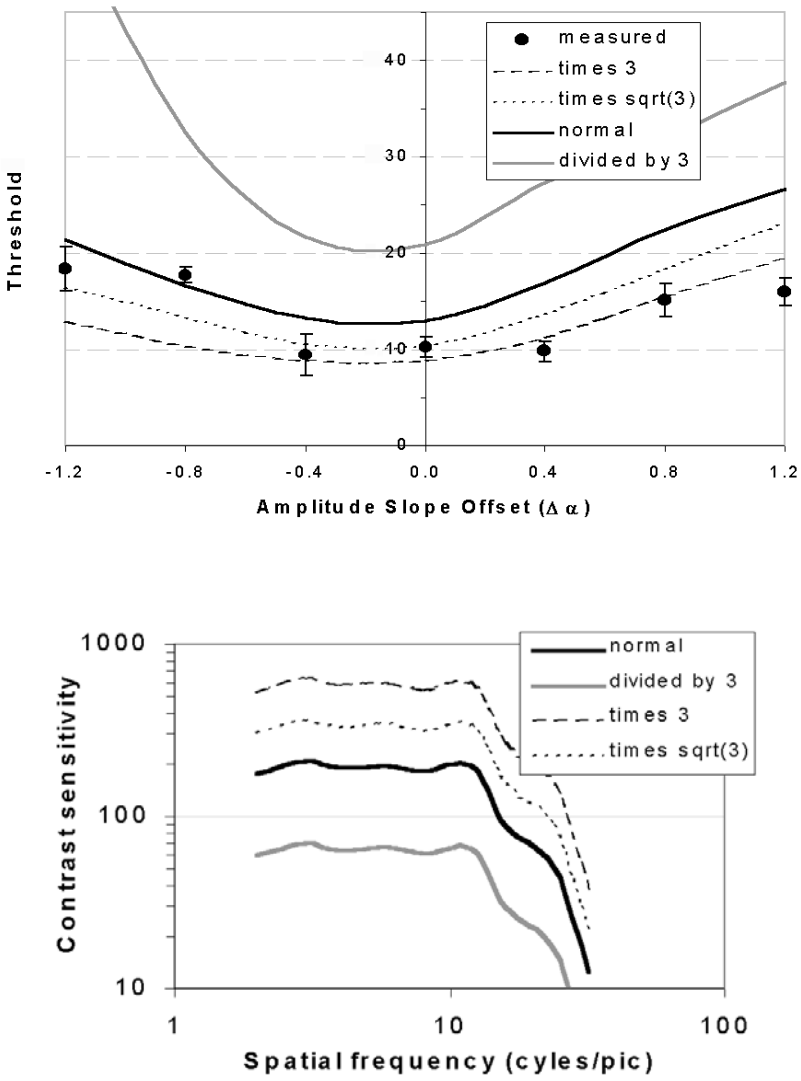

Figure 4.24: Different overall contrast sensitivity. Influence of a change in the overall CSF. Top part shows the model's predictions when the overall CSF function is multiplied by a parameter as shown in the bottom part of the figure. The morph sequence considered was woman-to-man for observer CAP. Corresponding experimental results are also shown (black dots). Rule 1a considered.

The top part of Figure 4.24 shows the effects of increasing or decreasing the overall CSF on the model's results, which is similar to altering the discrimination threshold value[19] in Section 4.2.5. In fact, the addition of a free parameter in Rule 1b was equivalent to finding the best multiplicative coefficient for the CSF so that our model fits optimally the experimental data. Here we can examine in more detail the effects of altering the overall CSF in our modelling.

From the figure, we can see that an increase in the observer's overall sensitivity produces lower thresholds in general, but for large values of the multiplicative coefficient, model predictions are too low in the left side of the graph. While the model improves its predictions on right side of the plot, it does poorly on the left side. This may indicate a balance where our simplest model is not far from the optimal point.

4.4.6. Future directions

Our attempt to model our previous psychophysical results has been an interesting and engaging exercise. We have been able to reasonably predict psychophysically measured results using extremely simplified versions of low-level visual mechanisms. These mechanisms include the output from "receptors" and "SF-channels", and the building of a multi-resolution image representation. This image representation is then compared to others before a decision is made to whether the images are similar or not. The only input from the human observer in this model is the characteristics of its CSF. Some important properties of the stimuli such as its Fourier orientation or phase spectrum were deliberately ignored. The rules that govern the model's decision have also varied from low

[19] Increasing the overall CSF produces a change in the position of the dipper function and therefore it modifies the difference between the predicted contrast difference and the dipper (see Equation 4.8). All this amounts to a vertical shift in the discrimination threshold curve, in the same way as adding a free parameter does.

levels of complexity to more elaborated ones. All our results indicate that a task such as local contrast discrimination can be performed robustly by the simplest set of decision rules. In fact, increasing the complexity of the model does not yield a clear improvement of these results. All this points to the lack of critical involvement of "higher" cortical functions.

The fact that the model seems to works even when any of its basic assumptions (channels bandwidths, CSF or dipper shape, etc.) is altered implies a degree of robustness in the interactions between the system and its parts: the whole system seems to be able to carry on doing its work despite these changes.

However, the predictions show that there is room for improvement of the model. This might be done by including some mechanisms that were not explored here: for example, we need to investigate the effects of calibrating each of the channels combined in Rule 4 as mentioned in Section 4.4.4. We may also contemplate the effects of splitting the SF-channels according to spatial orientation. Another improvement would be to make the model independent of the observer by considering a CSF that is the result of many contrast sensitivity measurements (e.g. Barten's CSF formula) (Barten 1993).

A more complex model should also account for lateral interactions among channels (Foley 1994). This approach assumes that the contrast gain of one channel can be reduced by activity from the others. Rohaly *et al* (Rohaly *et al.* 1997) added a single overall contrast gain mechanism and found this assumption to be very important compared to others. M. Chirimuuta and D.J. Tolhurst (personal communication) have tried to model the effects of non-specific suppression. They found that masking of one channel by others might result in a flattening of the dipper function. As shown before (in Figure 4.22), the flattening of the dipper function did improve our model's predictions slightly.

Our model has also been successfully tested in a different task (detection of objects on a background) and it is capable of predicting object visibility on synthetically produced *"natural backgrounds"*. This is to be explored in more detail in the future.

A future expansion of the model contemplates the prediction of discrimination thresholds in colour morph pictures and the modelling of avian vision.

4.5. Conclusions

a) Our simple model, in its two variations (or Rules) is capable of predicting the thresholds for discriminating between slightly different pictures from a morph sequence. The model also predicts optimal performance for natural slopes of the amplitude spectra. These predictions agree with psychophysical results.

b) The simplest version of the model (Rule 1a, 'winner takes all') performs significantly better than the other, more complex version (Rule 4a, with inter-receptor and inter-channel masking effects) suggesting that our morph discrimination task can be accomplished by the HVS using only simple low-level mechanisms. The addition of a degree of freedom to our simplest rule (Rule 1b) produces a significant improvement in the model's predictions, suggesting that there might be other mechanisms not taken into account here.

c) To gain a better understanding of the workings of the model we explored situations in which its parameters differed markedly from those assumed to be true for the normal (foveal, band-pass channel-based) models of vision.

Our results suggest that the robustness of our model might be the result of a combination of factors rather than the product of any single feature. No single characteristic of the model (filter bandwidths, shape of the contrast discrimination function, shape of the CSF function, etc.) could account by itself for the model's predictions. These results suggest a robust interaction among several aspects of spatial vision and the relevant properties of natural scenes.

Chapter 5.

Chapter 5

SPATIO-CHROMATIC PROPERTIES OF NATURAL IMAGES AND THEIR RELATIONSHIP TO HUMAN CONTRAST SENSITIVITY

Is there an optimisation?

5.1. Overview

So far, our work has explored the HVS's optimisation for encoding the *spatial* statistics of natural scenes in terms of *luminance*. However, we are also capable of encoding information in the wavelength domain by means of three different wavelength-sensitive receptors called cones; thus, we see the world in colour. In this chapter we extend our research into some of the *colour* properties of natural scenes, considering their relationship to the *spatial* filtering the HVS performs on this information, trying to find which of these properties may have had a greater role in shaping the way colour vision works.

It is well known that the HVS shows a relatively greater response to low spatial frequencies of chromatic spatial modulation than of luminance spatial modulation (Mullen 1985). Figure 1.17 (in Chapter 1) shows this imbalance. However, previous work (Brelstaff and Troscianko 1992; Párraga *et al.* 1998a; Webster and Mollon 1997) (see Figure 5.1) has shown that this differential sensitivity to low-SFs is not reflected in any differential luminance and chromatic content of general natural scenes. In one of these earlier studies (Párraga *et al.* 1998a), the spatio-chromatic properties of 29 natural scenes were analysed, but they mostly consisted of views of foliage and gardens.

These results are contrary to the prevailing assumption that the spatial properties of human vision ought to reflect general structure of natural scenes (Atick 1992; Barlow 1961, 2001; Field 1987, 1994; Laughlin 1981; Srinivasan *et al.* 1982; Van Hateren 1992b). Essentially, a match between the scene content and the visual system's response maximises the signal-to-noise ratio of the neural representation. That is, since HVS prefers low chromatic spatial frequencies, we would have expected natural scenes to be dominated by lower chromatic SFs than luminance SFs. They are not. However, these results may reflect the unsuitability of *general* natural scenes databases as the basis of such analysis, since no particular visual *task* for HVS was considered. As discussed in the introductory chapter, we would like to take a different approach here, considering that some of the strongest evidence for an optimisation between the visual world and the colour properties of the HVS does not come from analysis of general scenes but from colorimetric measures of a *specific* type of scenes. These suggest that red-green (and perhaps blue-yellow) colour discrimination in primates is particularly suited to the encoding of specific objects: reddish or yellowish objects (such as fruit or edible leaves) on a background of leaves (Dominy and Lucas 2001; Osorio and Vorobyev 1996; Regan *et al.* 1998; Sumner and Mollon 2000a). We therefore ask whether the *spatial*, as well as chromatic, properties of such fruit scenes are matched to the different spatial encoding properties of colour and luminance modulation in human vision. Given that general scenes are not matched, perhaps this special class of ecologically-relevant visual scenes *is* matched.

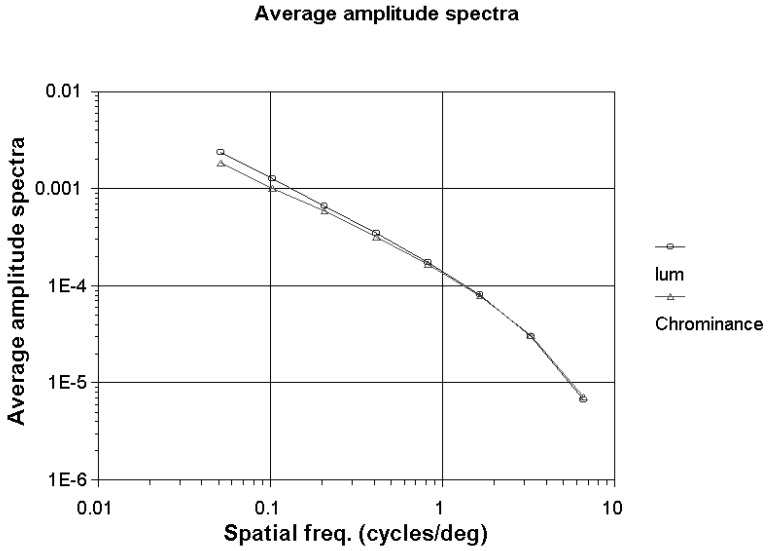

Figure 5.1: Fourier amplitude averaged across orientation for the luminance and chromatic component. The results show the average values for a dataset of 29 hyper-spectral natural images (Párraga *et al.* 1998a).

5.2. Methods

In order to extend our analysis of the spatial properties of natural scenes to the chromatic domain, we produced a colour database of natural scenes. To do this, a portable digital camera (Nikon Coolpix 950) was calibrated to have a linear response in its R, G and B primaries. The camera primaries were then transformed into three human cone primaries using an approximate transformation. Separate luminance and chrominance planes where it is possible to analyse the Fourier content were obtained. The calibration of the digital camera is not straightforward and will be discussed in the next section.

5.2.1. The Nikon Coolpix 950 colour digital camera specifications

Figure 5.2 portrays the Nikon Coolpix 950 digital camera. It has two recording modes: automatic (A-REC) and custom (M-REC which allows manual adjustments to focus, shutter speed, aperture, sensitivity, white balance, metering, brightness and contrast). We

used the M-REC mode, so that we had full control of the images. The camera's high-density image sensor (½ " CCD) has an effective pixel count of 1.92 million pixels (1,600 x 1,200 pixels). The camera supports several image sizes: 1,600 x 1,200 (full), 1,024 x 768 (XGA), and 640 x 480 pixels (VGA). The camera's lens is a X3 zoom Nikkor aspherical glass lens (38 mm – 115 mm), F2.6 – F4.0. Its focal length ranges from 30cm — ∞. It has 2 focus modes: auto focus and manual (ten steps from 10cm — ∞). The shutter is a combined mechanical and charge-coupled electronic shutter and its speed range is 8 sec – 1/750 sec. The aperture is electro-magnetically controlled. The camera's sensitivity is ISO 80 equivalent. It is possible to choose among three exposure-metering modes: 256-element matrix, centre-weighted and spot.

Figure 5.2: The Nikon 950 camera used in this research.

The camera also provides several automatic white balance programs for sunny and overcast conditions and for incandescent, fluorescent, and flash lighting. It has a self-timer (3 sec./10 sec.). The pictures are stored on a removable CompactFlash memory card (8 – 128 Mb) and the format could be chosen between JPEG and tagged-image format (TIFF-EXIF 2.0). We used uncompressed TIFF, since JPEG compression introduces spatial and

chromatic artefacts into the image. Video output was available to monitor the images in either NTSC or PAL external TV screens.

5.2.2. Default camera setup

We considered several parameters before taking a picture with the Nikon Coolpix camera. For example, in a fully automatic mode, the camera has a built-in algorithm that measures the exposure value (EV) in the central area of the picture frame. After this, the algorithm determines the diameter of the aperture opening (aperture) and shutter speed (expressed in terms of the integration time or IT) so that the amount of light captured fits within the dynamic range of the CCD sensor in the measuring area, ignoring the rest. In our case, we decided to avoid possible artefacts introduced by changes to the depth of field, focal length, etc. by keeping all optical parameters the same. In addition, the CCD sensitivity was fixed, so that the camera's built-in algorithm could only determine the best integration time. Since our choice of subject for the images did not include a strong presence of sky or very bright surfaces, the chances of having large saturated portions of the image were very low. The choice of "spot metering" settings allowed us to best fit the dynamic range of the CCD sensor to "interesting" parts of the image such as shady areas, vegetation, letting extremely bright portions (which were not very common) saturate.

Having in mind the most common situations that we might encounter in the natural world and in the laboratory, we choose three white balance settings to calibrate. These settings were: *Sunny*, *Cloudy* and *Incandescent*. All other conditions/settings were kept the same throughout the calibration and collection of the data (except stated otherwise). These were:

a) Metering type: *Spot*, which takes into consideration only the central portion of the picture, shown in the LCD monitor. This central portion has rectangular shape and occupies about 5% of the LCD screen.

b) Image Adjustment: *0*. This setting does not change the brightness and contrast to digitally "enhance" pictures.

c) Image quality: *Hi*. The camera does not use any kind of image compression and gathers uncompressed TIFF pictures of 5.5 Mb in size each.

d) Sensitivity: *Default*. The camera's sensitivity is roughly equivalent to an ISO 80 film.

e) Exposure: *Aperture priority*. The aperture is set to a fixed value (usually F11.4, which maximises the camera's depth of field, allowing most of the scene to be in focus) and the camera manipulates the shutter speed to fit the image within its dynamic range.

f) Zoom: *115 mm*.We usually took pictures using the telephoto lens' maximum value, which allows the camera (hardware, tripod etc.) to be far from the photographed objects, minimising interactions, such as reflections, shadows, etc.

g) Timer: Always *On* (10 seconds)

In addition, we checked that the image adjustment setting did not produce any kind of automatic contrast enhancement or normalisation. We took two pictures of a standard Kodak grey card. In the first picture, the card was on a white background and in the second, it was on a black background. We checked that the RGB values for the card were

the same in both cases. All images where captured while the camera was mounted on a fixed tripod.

5.2.3. Camera calibration (Overview)

To be able to gather images of the natural world and convert them to L, M and S cone primaries we need first to know the physical (electro-optical) properties of the Nikon 950 digital camera. To do this we proceeded to photograph objects, measure the physical properties of the light reflected from these objects, and compare the digital output of the camera with these properties (calibration). The measurements of spectral radiance were made using a spectroradiometer (Topcon Model SR1, calibrated by the National Physical Laboratory). This instrument was capable of measuring spectral radiance within the 380-760 nm wavelength range using a photomultipler tube as a photo-detector and a diffraction grating as dispersive element. Its measuring field could be selected manually to be 0.2 or 2º. Its viewing field was 5º and its sampling interval could be selected from 1, 5 or 10 nm in equal intervals. The spectroradiometer (SR) lens system's measuring distance was 400 mm— ∞ from the focal point (F4, f= 80 mm). Its calibration report (National Physical Laboratory report QMS112) states that '*the measured radiance values at all wavelengths and all luminance levels applied to the instrument are within the 4% limits as specified by the instruction manual*'. The instrument provides both physical and photometric measures: spectral radiance, radiance, luminance, tristimulus values (X, Y, Z), chromaticity coordinates (x, y) (u', v'), colour temperature, etc. It was connected to a PC using a GP-IB interface and the warm up time was approximately 60 min. A custom program was written in Delphi 3 to control this instrument from a Windows NT platform. Constant illumination was provided by a stabilised DC power supply (Type SP020 - Vinculum Products LTD, Royston, Herts, 0-50V, 0-20Amps) accurate to 30 parts per million in current according to

the manufacturer's specifications and a tungsten-halogen lamp (Osram HLX 64657FGX-24V, 250W) with fixed 10 Amps current. This was necessary to avoid variations in the light source produced by fluctuations of the mains during our measurements.

Figure 5.3 shows the schematics of our camera calibration. Its starting point is the estimation (later confirmed by accurate measurements) of the sensor's response linearity with large changes in light intensity. This response is defined by the IT values and RGB values registered by the camera when light intensity is increased a few orders of magnitude.

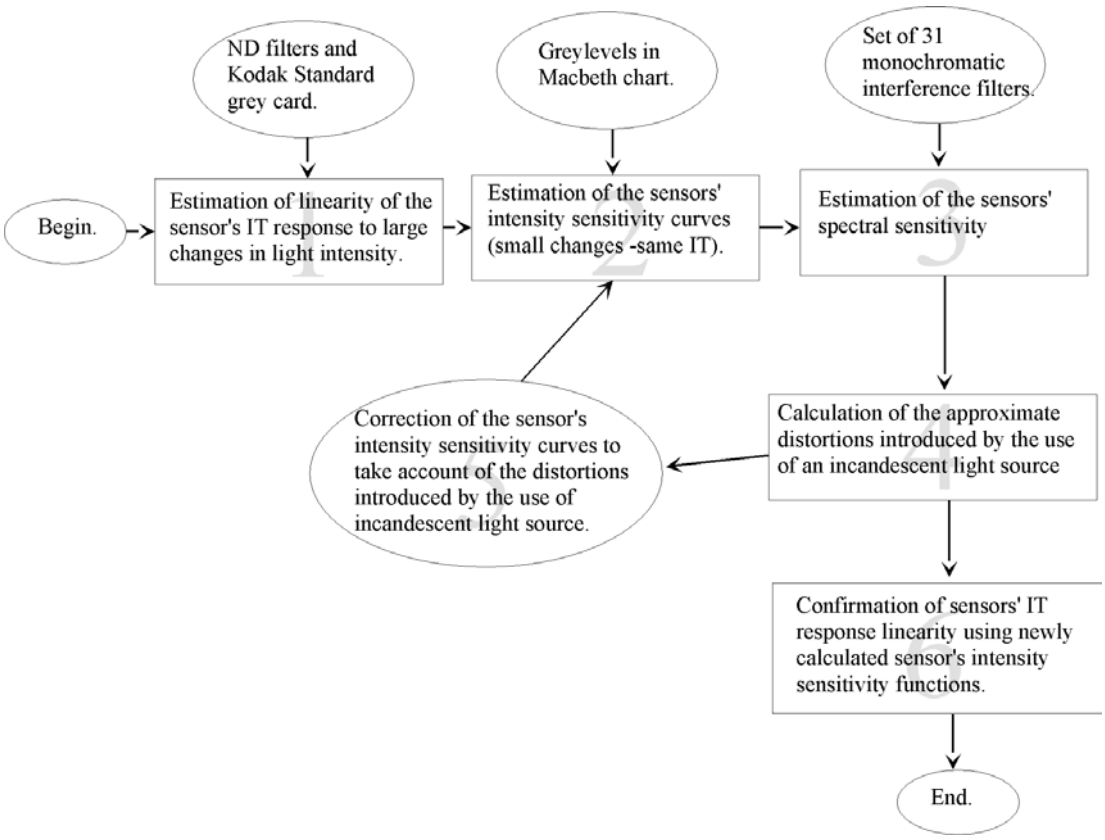

Figure 5.3: Schematic of the 6 basic steps taken to calibrate the Nikon camera used in this work.

The second stage corresponds to an estimation of the sensor's sensitivity to light intensity variations within the same picture (same IT value) and is detailed in Section 5.2.5. The

third stage is the estimation of the sensor's spectral (colour) sensitivity using a set of monochromatic narrowband filters spanning over the visible spectrum (see Section 5.2.6). Once we have estimated the sensors' sensitivities both to light intensity changes and chromatic changes we can draft the shape of each of the sensor's spectral sensitivity functions. The basic problem with these measurements is that they are obtained using light that is not spectrally flat (incandescent light contains more energy in the "reddish" side of the spectrum than in the "bluish" side). Our next step is to calculate the amount of distortion introduced by the spectrally uneven light source and correct our measurements taking into account these variations. This is done by using the spectral sensitivity functions estimated in the previous step and the light source's spectral sensitivity plus some assumptions detailed in Section 5.2.6. Once the distortions introduced by the light source are calculated, both the luminance and spectral sensitivities of the camera's sensors are recalculated. The linear dependence of camera's IT with light intensity is also confirmed (Section 5.2.7) using these corrected values. Each of the steps mentioned above had its own difficulties derived from either the camera's own limitations or restrictions due to the method employed. The following sections explain how these were dealt in detail.

5.2.4. Estimation of the linearity of the sensor's IT response when exposed to large changes in luminance.

When a picture is taken, the camera's RGB sensors are active for a certain amount of time (called IT) enough to register a measure of the luminous energy falling on them before sending this information to the camera's processor unit. Since each sensor has a limited dynamic range, an algorithm estimates (in a short time before the picture is taken) the amount of IT necessary so that the luminous energy is large enough to be registered and does not saturate the sensor. These calculations are usually done based on the properties

of the G (green) camera sensor. According to this, if the camera is exposed to two different amounts of light energy, one very small and the other large, the sensor may register the same set of RGB values, but different IT.

To be able to calibrate our camera, we first need to have some certainty that this IT value changes linearly with light intensity (in our case, a convenient measure of this intensity is *luminance*). Since we have yet to estimate the response of the sensor's RGB values to changes of light intensity present in the same picture, we can only aim at having a rough estimation of the linearity of the sensor's response to *large* changes of luminance.

If we want to compare the values of RGB obtained from the camera across different lighting conditions, we need to divide the by the corresponding IT. A preliminary estimation of the relationship between these "corrected" RGB values and luminance was obtained as follows:

a) A measurement of the light reflected from the central part of the target was taken using the SR. Neutral density (ND) filters of values 0.5, 1.0, 1.5, 2.0, 2.5, 3.0 and 3.5 *log* units were added in front of the SR lens and new measurements were taken to calibrate the ND filters.

b) The SR was removed and the Nikon camera was set up on a fixed tripod sharply focused at the target. The distance between the camera and the target was 200mm. A picture was taken without a filter and through each of the ND filters. The first picture (without filter) was repeated at the end to check whether any of the conditions had changed during the session. This was repeated three times (once for each white balance).

c) All 27 pictures were downloaded to a PC. A custom made program (written in Matlab v.5.0) calculated the values of the central square (50x50 pixels) of all pictures. The values of R G and B within each of these squares were averaged. Typical values of SD were between 1.4 and 4% (R and G sensors) and between 3.8 and 6% (B sensor). The worst single case was the B sensor with the 3.5 ND filter (17% SD). The integration time corresponding to each of the pictures was obtained from the picture's header.

d) From the SR readings, we obtained the values of luminance in all the conditions (ND= 0.5, 1.0, 1.5, 2.0, 2.5, 3.0 and 3.5).

e) Camera values far from the middle of the camera's dynamic range (0-255 greylevels) *were not considered* in this preliminary measure. "Normalised RGB values" were obtained (by dividing the original RGB values on the corresponding IT) and the relationship between the camera's output and the SR measurements was plotted.

From these preliminary measures, we concluded that there is an approximate linear relationship between the camera's IT values and luminance, but this needs to be confirmed later using all the data points (see Section 5.2.7).

5.2.5. Calibration of the intensity levels of the R, G, and B sensors (Gamma correction)

In a picture taken with our camera, each pixel value (triplets of RGB values) is related to the local distribution of light intensity at the corresponding point in the real world. In an ideal situation, the function relating these values would be linear. In practice, this function

is found to be non-linear and the form of this relationship is unknown to us. To measure this relationship we did the following:

a) We used a ColorChecker colour rendition chart (Macbeth® chart – Kollmorgen Instruments Corporation) as target for both the SR and the camera. The lower part of the chart has six squares (labelled white, light-grey, medium grey, etc.), which have the same CIE *x,y* coordinates but different reflectance values (i.e. when illuminated by the same light source, they reflect light with the same wavelength composition but different intensity and luminance (Y) values). The chart was illuminated by incandescent light from a modified slide projector, supplied by constant current. The SR was pointed to the central part of each of the six grey squares of the chart to measure Y and *xy* for each square. Three pictures of the colour chart were taken (one for each of the white balances tested: incandescent, sunny and cloudy) and processed separately. Figure 2.4 shows the schematics of the measurement.

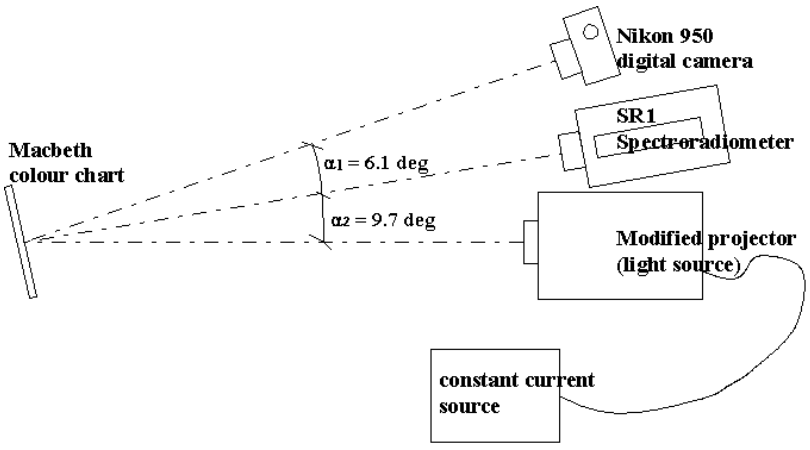

Figure 5.4: Disposition of the instruments for the Nikon 950 Gamma correction. The distances between the chart and the light source, SR and camera are 115 cm, 122 cm and 123.5 cm respectively. Indirect lighting from the projector to the camera or the SR was reduced using black cardboard screens.

b) From the camera pictures, (1600 x 1200 pixels) the six grey squares were cut out. The central part (10 x 10 pixels) of these grey squares was cropped out and the values of R, G and B averaged within these (their corresponding standard deviation was also calculated). The size of these (10 x 10 pixels) squares corresponds approximately to the size of the SR measurement spot. Typical values of STD were below 5% for RGB values larger than 50 greylevels and increased to about 8% for values below 50 greylevels.

c) The average values for each (10 x 10 pixels) square in each of the R, G and B image planes obtained were plotted against luminance (measured by the SR). The result was a set of three functions (one for each camera sensor) in each of the three white balance conditions (see Figure 5.5). We were able to find a single *family of curves* that fitted to the measured data in all cases:

$$L = a.b^{S}$$

Equation 5.1

where L is luminance; S is the grey level value obtained for each of the camera sensors b is a constant and a is the parameter that determines the family of curves. The functional relationship between dependent and independent variables (L as a function of S instead of the most usual S as a function of L) was decided based on convenience (it is easier to fit). The value of b was found to be equal to 1.015 for all conditions (Sunny, Incandescent and Cloudy) and all sensors (RGB) and the value of a was found to be the same (a= 11.26) for the G sensor regardless of the white balance employed. In summary, each of the 9 different curves of the family represented by Equation 5.1 is defined by only one parameter (a) and this parameter (or gain factor) is used by the camera's manufacturer to weight the output of the R and B sensors to compensate for the different lighting

conditions. Figure 5.5 shows the plots of the nine curves corresponding to these (preliminary) measurements.

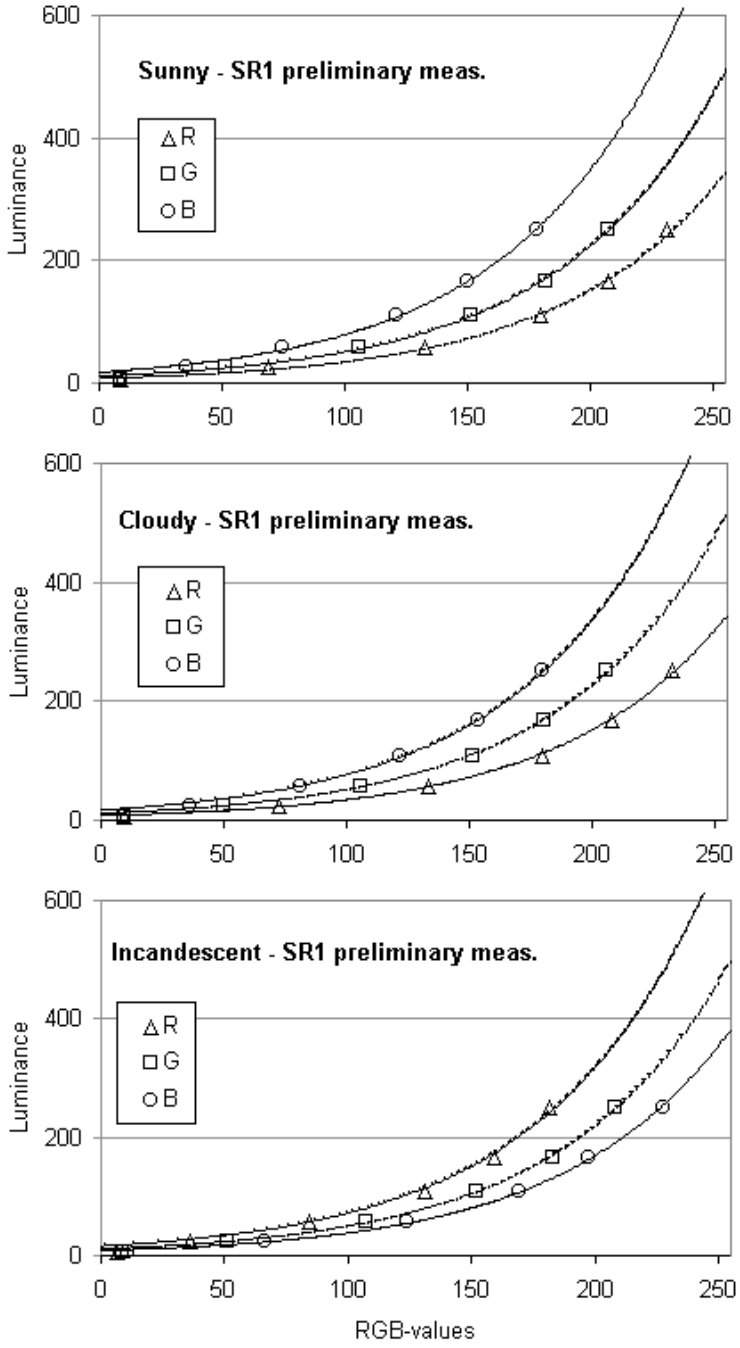

Figure 5.5: Plots of Luminance (as measured on the Macbeth chart by the SR) against the R,G, and B values produced by the camera.

Given that the curves represent each sensor's output against a measure of light energy (luminance) obtained around the central part of the visual spectrum (corresponding to the CIE luminous efficiency, $V(\lambda)$) using a light source that was not spectrally flat, they need to be corrected so that the energy stimulating all three (RGB) sensors is the same.

To be able to obtain the gain factors corresponding to the R and B sensors we need to estimate the ratio between the amounts of light captured by each of the sensors across the colour spectrum (sensor's spectral sensitivity) and the spectral properties of the incandescent light that we were using.

5.2.6. Estimation of the Nikon camera's sensor spectral sensitivity

After the preliminary measurements of section 5.2.5, we established the basic structure and functioning of the Nikon camera's RGB sensors as a function of light intensity. This can be summarised as follows:

The camera possesses three types of sensor (R, G and B), which have different spectral sensitivities but share the same sensitivity to increments of light energy (as described by Equation 5.1). The sensitivity of these sensors are modified according to parameter *a* in Equation 5.1. This parameter changes the sensor's sensitivity curve to produce curves R and B in all "manufacturer-defined" white balance conditions.

Knowing the above, we can now proceed to measure the R, G and B sensors sensitivity to wavelength changes (spectral sensitivity) *assuming* that all sensors behave basically in the same way (they share the same G sensitivity plots in Figure 5.4) before their output is modified by gain factor *a*.

To measure the Nikon camera's sensors spectral sensitivity we used a "target" which consisted of a black box with two holes (see Figure 5.6). The internal part of the black box was covered in black velvet. Opposite the front hole, there was a circular target, which consisted of Cyanoacrylate adhesive powder (Kodak-Eastman "standard white" with 99% reflectance throughout the visible spectrum). This substance produces a Lambertian (diffuse) pattern of reflected light and has approximately constant reflectance across the visible spectrum (i.e. "white" object).

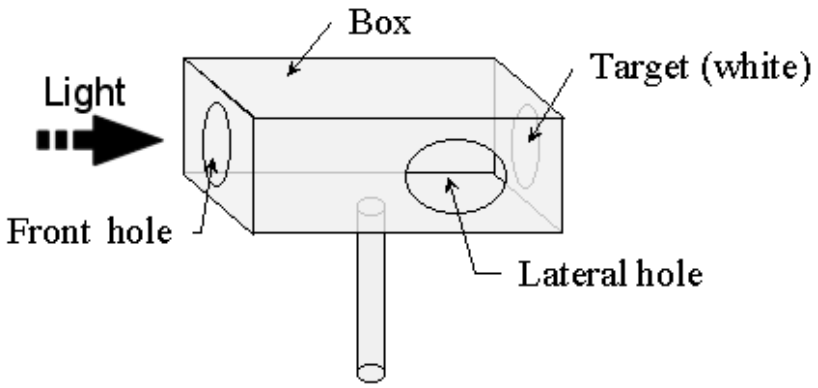

Figure 5.6: Scheme of the box containing the target that we used for our calibrations. The actual target consisted of Cyanoacrylate which has a fairly constant reflectance (0.989 to 0.992) in the range 400 to 700 nm.

The target was illuminated by the constant-current (10 Amps) incandescent light source, and the light reflected by the (white) target was measured using the SR. Indirect lighting from the projector to the camera or the SR was reduced using black cardboard screens. The arrangement of all elements is shown in Figure 5.7.

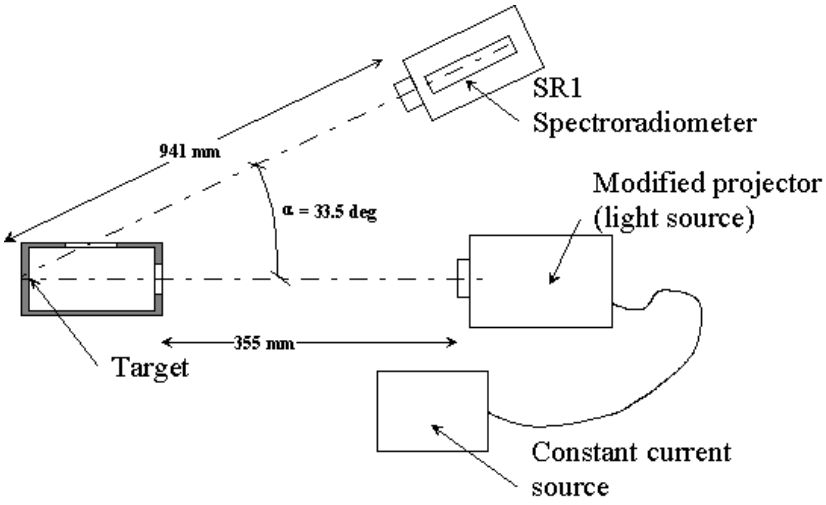

Figure 5.7: Positions of the elements including light source and spectroradiometer to measure the light reflected by the white target.

All dimensions were determined by the physical size of the laboratory equipment and by the minimum distance that the SR lenses were capable of focusing.

The Nikon Camera's spectral sensitivity within the visible spectrum (400-700 nm) was measured using a set of 31 colour filters (each of 10 nm bandwidth - Ealing Electro-Optics, Watford) spanning from 400 to 700 nm and with peaks at 10 nm intervals. The procedure was as follows:

a) The white target was set and illuminated with a constant-light incandescent source as shown in Figure 5.7. A measurement was also made of the central part of the target using the SR at 1 and 10 nm intervals. The stability of the light source and the SR were tested by measuring light from the white target on different times and different days.

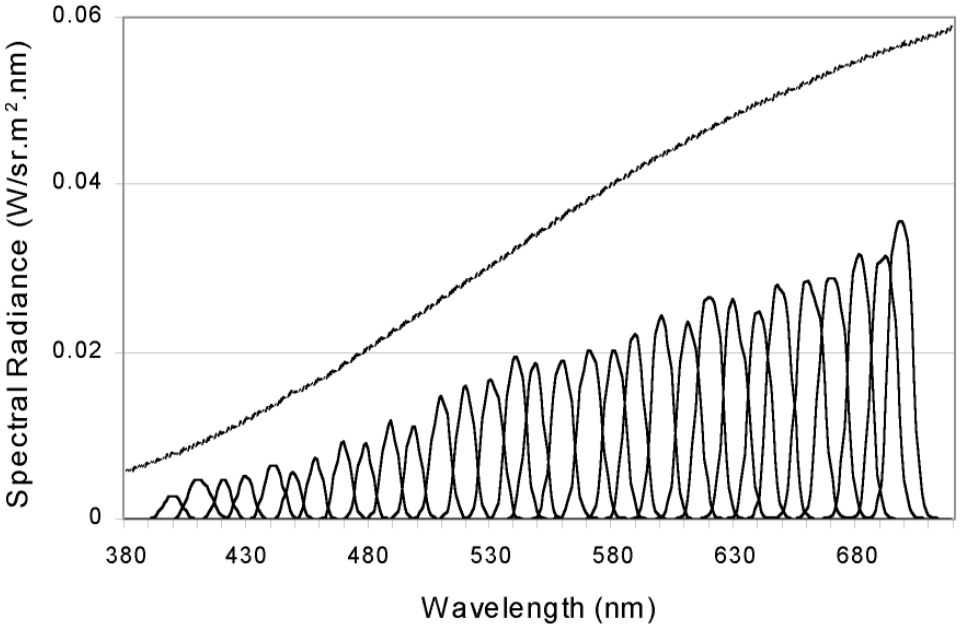

Figure 5.8: Spectral radiance of the 31 filters and the incandescent light source (broken curve at the top).

b) One by one, the filters were added in front of the SR lens and measurements were taken. From the 1nm-interval SR data, it was possible to reconstruct the transmitted radiance of the filters and their transmittance. Figure 5.8 shows the radiance measured through each of the filters and the incandescent light source (light reflected from the white target). Spectral transmittance was calculated by dividing the spectral radiance for each filter by the spectral radiance of the source (Figure 5.9) at each wavelength.

c) A Gaussian curve was fitted (using least-squares method) to the measured radiance for each of the filters and their bandwidth and areas calculated. Where the filter manufacturer's data was available, we compared the shapes of the transmittance plots (including peak, half-height bandwidth, etc.) to those of the fitted Gaussian curve. The agreement was good in all cases.

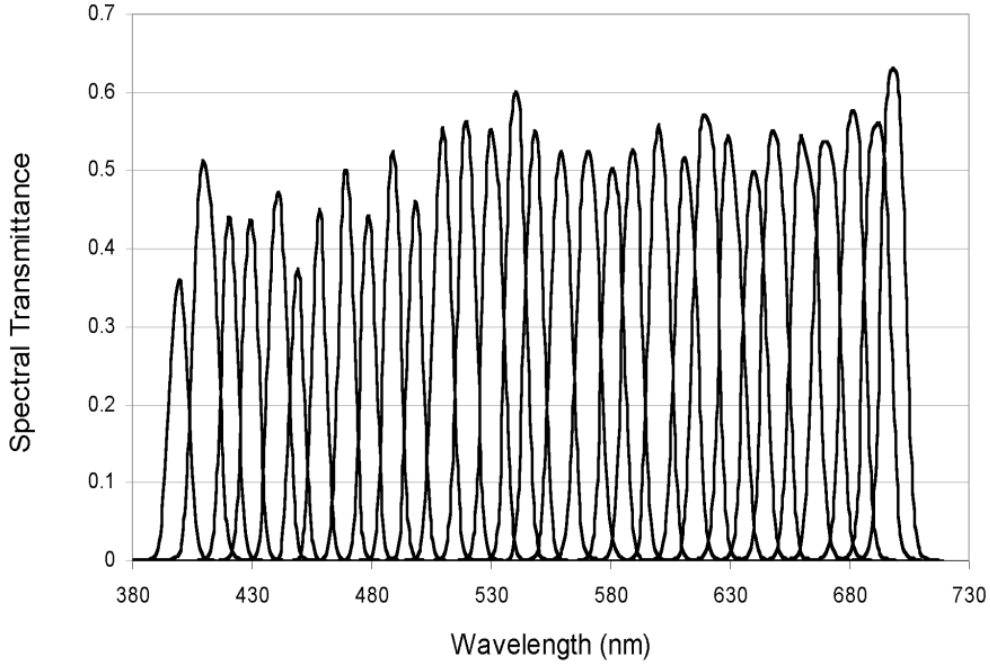

Figure 5.9: Spectral transmittance of the 31 colour filters.

d) A series of pictures of the white target was then taken (replacing the SR with the Nikon camera) through the whole set of colour filters. Two control pictures (at the beginning and the end of the series) were also taken without using any filter. This sequence was repeated three times, each time using a different white balance mode (Sunny, Cloudy and Incandescent).

e) The R, G and B values of the centre of each colour-filtered picture and their IT were again extracted. Values of B taken with filters within 400 and 500 nm were found to be saturated (the camera's IT was determined by the dynamic range of the G sensor, thus the B grey-level values were to equal to 255). A new set of pictures was taken to correct those (see below).

f) Considering that the relationship between luminance and the values of RGB, is stated by Equation 5.1, which is the same for all three colours and white balance

conditions (only the gain factor *a* changes), the RGB values (with and without filters) were linearised using Equation 5.1 and divided by the corresponding IT. At this point, the gain factor *a* was considered to be the same for all three RGB sensors and equal to the value previously measured from the G sensor plots in Figure 5.5. A constant was added to account for the fact that a sensor value of "0" should yield a linearised value of "0". Equation 5.2 show these relationships (*a*= 11.26 and *b*= 1.015).

$$LinearisedR = \frac{a \cdot (b^{R} - 1)}{IT};$$
$$LinearisedG = \frac{a \cdot (b^{G} - 1)}{IT};$$
$$LinearisedB = \frac{a \cdot (b^{B} - 1)}{IT};$$

Equation 5.2

g) These are now called "linearised" RGB values. These "linearised" values of RGB were then divided by the corresponding area under the radiance curve thus obtaining a measurement of the RGB sensors' sensitivity at the filter's peak. At this point we need to comment on the limitations of this method, since the ideal would be to have infinitely narrowband interference filters. Given that our interference filters do have transmittance functions of finite spectral bandwidths which somewhat overlap one another (see Figure 5.9), each calculated point on the sensors spectral sensitivity functions will have some "contamination" of energy belonging to neighbouring points.

h) Given that the values of B obtained by the camera were saturated (equal to 255), we had to adopt a different strategy to measure the B sensor's spectral sensitivity within the range 400-500 nm. To avoid saturating the B sensor, we took pictures

of the same target in the same conditions, only adding ND filters in front of the corresponding colour filter. At this point, the camera tried to increase the IT so that the G sensor was within range, but it reached its maximum value (IT= 8 sec). This allowed us to obtain values within the B sensor's dynamic range. SR measurements were repeated replicating these conditions.

i) The B sensor's "spiky" nature was confirmed to be a property of the sensors and not due to measurement noise, by repeating the measurements, altering the conditions (distance from the light source and target, etc.) in order to reduce the amount of noise.

Figure 5.10 shows the RGB sensor's spectral sensitivities calculated from values derived from Equation 5.2. Each RGB value in the figure was divided by the total spectral radiance transmitted by the corresponding interference filter (the area under each interference filter's radiance curve in Figure 5.8).

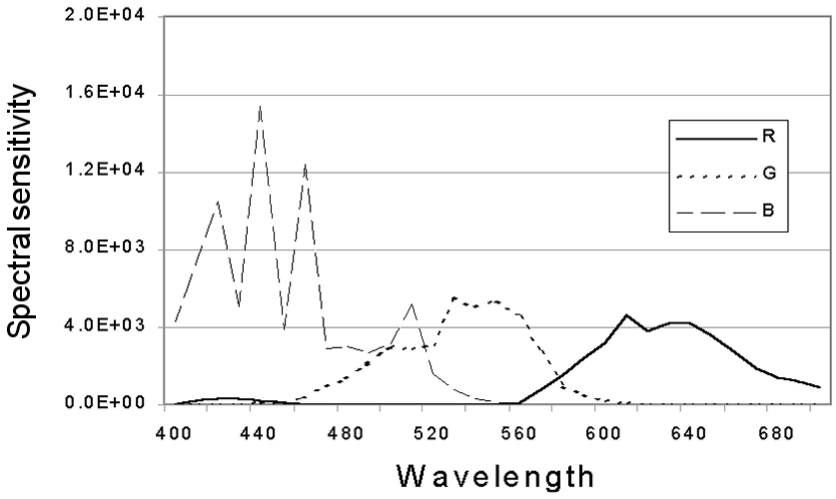

Figure 5.10: Spectral sensitivity of the Nikon camera's RGB sensors.

Once the spectral properties of the R, G and B sensors are known, it is possible to estimate how the use of an incandescent light source has introduced distortions to the data displayed in Figure 5.5. From the RGB curves in Figure 5.10 we can see that light with spectral distributions such as that shown in Figure 5.8 would have excited more the R sensor and less the B sensor, compared to the middle-wavelength sensor. Since our measure of light intensity in Figure 5.10 is based on *luminance* (which favours middle-wavelength intensity distributions such as that of the G sensor), we need to plot again the same data compensating for the non-even spectral energy distribution produced by our incandescent light source.

The correction factors were obtained by multiplying the RGB sensitivities (shown in Figure 5.10) by the spectrum of the light reflected from the unfiltered white target (as shown in Figure 5.8) and operating as follows:

$$C_{R/G} = \frac{\sum_{\lambda=400}^{700} R_\lambda \cdot E_\lambda}{\sum_{\lambda=400}^{700} G_\lambda \cdot E_\lambda} = 1.53$$

$$C_{B/G} = \frac{\sum_{\lambda=400}^{700} B_\lambda \cdot E_\lambda}{\sum_{\lambda=400}^{700} G_\lambda \cdot E_\lambda} = 0.51$$

Equation 5.3

In the previous equation E_λ represents the spectral radiance of the white target and R_λ, G_λ, B_λ are the spectral sensitivities described in Figure 5.10. These correction factors were used to recalculate the luminance measured previously in Figure 5.5.

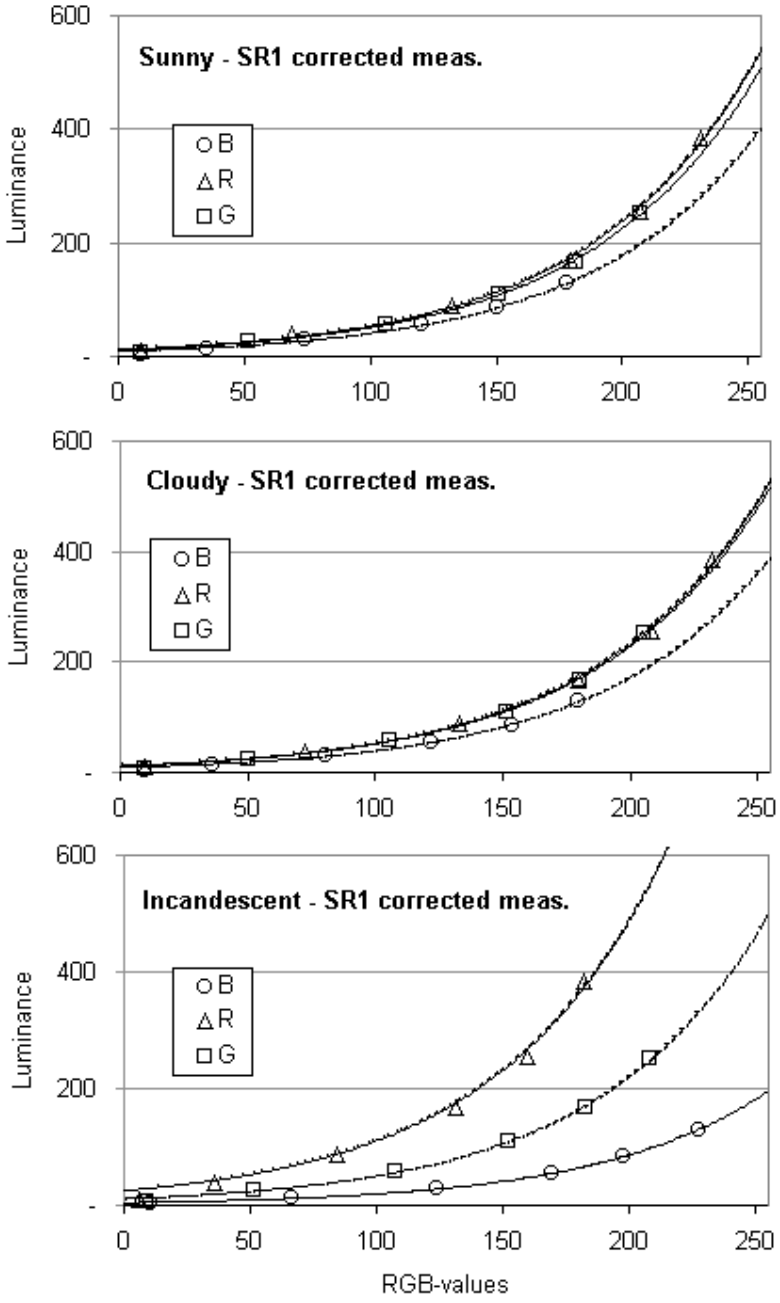

Figure 5.11: R, G, and B values (obtained from the Nikon camera) versus the measured luminance (obtained from the spectroradiometer). Exponential fitting curves are also shown for each of the sensors. Error bars are smaller than the corresponding symbols in this plot.

Figure 5.11 shows the "corrected" family of curves (produced by multiplying the luminance values of the R and B curves by the $C_{R/G}$ and $C_{B/G}$ correction factors respectively) corresponding to all the white balance conditions.

From the graphs in Figure 5.11 we were able to calculate the gain factors a applied by the camera manufacturers to all sensors in all white balance settings. Using this information, we can now write Equation 5.2 for all conditions as follows:

Sunny white balance

$$LinearisedR = \frac{12.10 \cdot (1.015^{R} - 1)}{IT};$$
$$LinearisedG = \frac{11.42 \cdot (1.015^{G} - 1)}{IT};$$
$$LinearisedB = \frac{9.08 \cdot (1.015^{B} - 1)}{IT};$$

Cloudy white balance:

$$LinearisedR = \frac{11.88 \cdot (1.015^{R} - 1)}{IT};$$
$$LinearisedG = \frac{11.62 \cdot (1.015^{G} - 1)}{IT};$$
$$LinearisedB = \frac{8.79 \cdot (1.015^{B} - 1)}{IT};$$

Incandescent white balance:

$$LinearisedR = \frac{23.80 \cdot (1.015^{R} - 1)}{IT};$$
$$LinearisedG = \frac{11.26 \cdot (1.015^{G} - 1)}{IT};$$
$$LinearisedB = \frac{4.39 \cdot (1.015^{B} - 1)}{IT};$$

The previous equations allow us to "gamma-correct" the sensor's output and convert RGB values into a linear measure of the light intensity stimulating each sensor.

In other words, after applying these equations to the camera's RGB output we are able to remove all the manufacturer's "white-balance" cosmetic enhancements to the image and obtain a measure of the light reflected from every small part of the scene into the camera.

5.2.7. Integration time (IT) linearity calibration

Our last concern regarding the Nikon camera's linearity has to do with the relationship between the camera's shutter speed (or IT) and the intensity of the light falling on its sensors (R, G and B). Since the aperture settings for the camera are fixed, the camera varies its IT to obtain a picture within its dynamic range and stores this IT value in the header of the corresponding (TIFF) picture file. The values of RGB obtained have to be corrected for this IT if we want to compare them across different lighting conditions. Preliminary measures (see Section 5.2.4) showed a linear relationship between IT values and luminance but now we can improve these measures by using all the data collected and not only points near the middle of the sensor's dynamic range. The data measured previously was re-analysed and Figure 5.12 shows a *log-log* plot of the relationship obtained between the camera's output and the SR measurements. The ordinate represents the values of RGB. These were linearised using Equation 5.1 and converted into "absolute RGB values" by dividing them by the corresponding IT. The abscissa represents the various values of luminance as measured with the SR through the filters. The *log-log* plot was necessary given the logarithmic nature of the ND filters. The lines on these plots represent the fitting in the *log-log* space. As Table 13 shows, all the slopes are close enough to one to assume a linear relationship in the linear space.

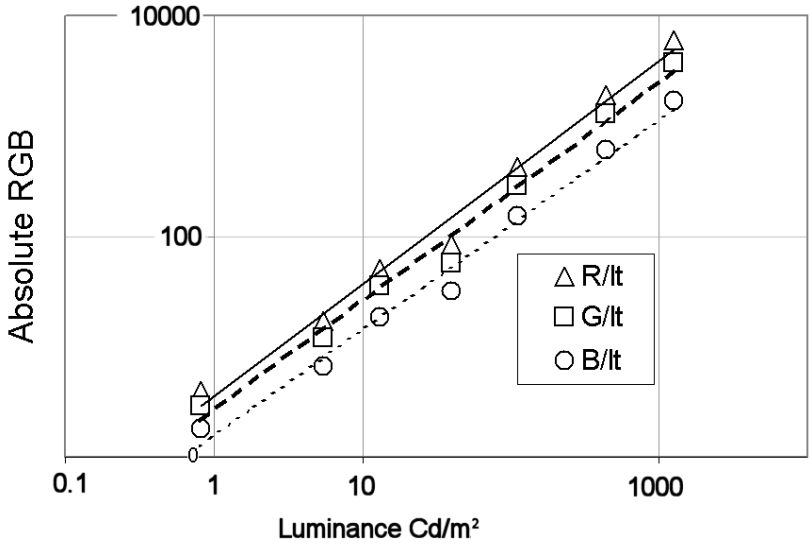

Figure 5.12: Relationship between luminance (as measured by the SR) and the values obtained for each of the RGB sensors (divided by the corresponding IT). The lines show the corresponding fittings in *log-log* space. Incandescent white balance settings.

We repeated these measurements on four off-centre squares. We took the central (256 x 256) part of these large (1600 x 1200) pictures and measured the RGB values on four squares (50x50) called north, south, east and west. The average values of RGB within these off-centre squares were close to the values of the central part (the difference was approx. the same as the STD of the average within each square).

Slope (α)	**Cloudy**	**Incandescent**	**Sunny**
R	1.042	1.016	1.045
G	1.036	0.996	1.044
B	1.021	0.960	1.038

Table 13: Slope values for the lines plotted in *log-log* co-ordinates showing the relationship between luminance and the RGB sensor value divided by IT, for each white balance setting. A linear relationship is indicated by the proximity of the slope to 1.

5.2.8. Conversion to the LMS cone sensitivities

Once the properties of the camera's sensors were known, we were able to linearise our digital pictures to obtain a reasonable measure of the red, green, and blue light captured by the camera's sensors for each of its constituent pixels. Our next step was to transform these RGB components into the equivalent amount of light that would be captured by the human cones. For this it was necessary to find the matrix that transforms the set of values $[R, G, B]$ into the corresponding set $[L, M, S]$:

$$\begin{pmatrix} L \\ M \\ S \end{pmatrix} = T \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

Equation 5.4

Where

$$T = \begin{pmatrix} t_1 & t_2 & t_3 \\ t_4 & t_5 & t_6 \\ t_7 & t_8 & t_9 \end{pmatrix}$$

Equation 5.5

Another way of modelling LMS and RGB spectral image acquisition is using matrix-vector notation. For example, the spectral intensity distribution of the light reaching the sensor can be expressed as:

$$\mathbf{E} = \begin{vmatrix} e_1 & e_2 & \dots & \dots & e_{31} \end{vmatrix}$$

where e_n is the radiance reaching the camera with a wavelength corresponding to the peak of each of the 31 filters. The camera's sensors RGB spectral sensitivity can be expressed as vectors, where each component represents the sensitivity of the sensor at the peak of a narrowband colour filter:

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ \dots \\ \dots \\ r_{31} \end{bmatrix}; \mathbf{g} = \begin{bmatrix} g_1 \\ g_2 \\ \dots \\ \dots \\ g_{31} \end{bmatrix}; \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ \dots \\ b_{31} \end{bmatrix}$$

and the same can be done with the LMS sensors:

$$\mathbf{l} = \begin{bmatrix} l_1 \\ l_2 \\ \dots \\ \dots \\ l_{31} \end{bmatrix}; \mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ \dots \\ \dots \\ m_{31} \end{bmatrix}; \mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \dots \\ \dots \\ s_{31} \end{bmatrix}$$

Using this notation, we can write:

$$R = \mathbf{E} \cdot \mathbf{r}$$
$$G = \mathbf{E} \cdot \mathbf{g}$$
$$B = \mathbf{E} \cdot \mathbf{b}$$

$$L = \mathbf{E} \cdot \mathbf{l}$$
$$M = \mathbf{E} \cdot \mathbf{m}$$
$$S = \mathbf{E} \cdot \mathbf{s}$$

and Equation 5.4 can be rewritten as:

$$L = t_1 R + t_2 G + t_3 B$$
$$M = t_4 R + t_5 G + t_6 B$$
$$S = t_7 R + t_8 G + t_9 B$$

which is equivalent to:

$$\mathbf{E} \cdot \mathbf{l} = t_1 \mathbf{E} \cdot \mathbf{r} + t_2 \mathbf{E} \cdot \mathbf{g} + t_3 \mathbf{E} \cdot \mathbf{b}$$
$$\mathbf{E} \cdot \mathbf{m} = t_4 \mathbf{E} \cdot \mathbf{r} + t_5 \mathbf{E} \cdot \mathbf{g} + t_6 \mathbf{E} \cdot \mathbf{b}$$
$$\mathbf{E} \cdot \mathbf{s} = t_7 \mathbf{E} \cdot \mathbf{r} + t_8 \mathbf{E} \cdot \mathbf{g} + t_9 \mathbf{E} \cdot \mathbf{b}$$

Where finally we can remove the intensity distribution of the light reaching the sensor, (**E**) as follows:

$$\mathbf{l} = t_1 \mathbf{r} + t_2 \mathbf{g} + t_3 \mathbf{b}$$
$$\mathbf{m} = t_4 \mathbf{r} + t_5 \mathbf{g} + t_6 \mathbf{b}$$
$$\mathbf{s} = t_7 \mathbf{r} + t_8 \mathbf{g} + t_9 \mathbf{b}$$

Equation 5.6

Equation 5.6 tells us that we need to find a set of 9 values t_1, t_2,…t_9 which convert each triplet RGB corresponding to a given wavelength in Figure 5.10 to the corresponding triplet in the LMS human spectral sensitivity curves (Figure 1.8). This means that our set of 9 t_i coefficients should satisfy a set of 31 equations (one for each sampled wavelength). It is easy to see that such a matrix only exists for a very limited set of RGB and LMS sensitivity functions, therefore, some kind of approximation to the problem's solution is needed. In practice, a working approximation to matrix T was obtained using both, the camera's sensors spectral sensitivities (Figure 5.10) and the Smith and Pokorny cone sensitivity functions as shown in Figure 1.8.

The LMS curves (DeMarco *et al.* 1992) shown in Figure 1.8 were obtained from a publicly available web site[20] and scaled so that they were compatible with psychophysical evidence and mathematical constraints for our transformations. These constraints were:

a) The wavelength of maximum sensitivity for the L-M and S-(L+M) mechanisms have to be close to 580 and 506 nm respectively. This means that the value of L(580 nm) must be approximately equal to M(580 nm) and that S(506 nm) should be approximately equal to the combined values of L(506 nm) and M(506 nm). See introductory chapter for a discussion on these properties of the HVS.

b) The rows of the transformation matrix T need to add up to the same number. This is necessary to keep the colour coordinates of a white object the same after the matrix transformation. It was implemented by scaling the L and S functions.

Since these two constraints could not be satisfied simultaneously, we reached a compromise by selecting scale factors that would satisfy the second condition while not departing from the first by more than 2%.

Each coefficient of the matrix T was obtained as follows:

$$t_1 = \sum_{400nm}^{700nm} L_i \cdot R_i; \qquad t_2 = \sum_{400nm}^{700nm} L_i \cdot G_i \qquad t_3 = \sum_{400nm}^{700nm} L_i \cdot B_i$$

$$t_4 = \sum_{400nm}^{700nm} M_i \cdot R_i \qquad t_5 = \sum_{400nm}^{700nm} M_i \cdot G_i \qquad t_6 = \sum_{400nm}^{700nm} M_i \cdot B_i$$

[20] Color Vision Lab data page, University of California-SD: http://www-cvrl.ucsd.edu/database/text/cones/dps.htm

$$t_7 = \sum_{400nm}^{700nm} S_i \cdot R_i \qquad t_8 = \sum_{400nm}^{700nm} S_i \cdot G_i \qquad t_9 = \sum_{400nm}^{700nm} S_i \cdot B_i$$

This technique is commonly used in neuroscience to map the spectral sensitivities of LMS cone receptors to the output of colour displays (Conway 2001) and in imaging science to calculate the transformation matrix between a device-independent colour space (such as CIE XYZ) and a device-dependent colour space (such as that determined by the RGB guns of a colour monitor) (Travis 1991a). A more detailed account of these calculations is presented in Appendix B.

5.2.9. Manipulation of the LMS planes to create luminance and chrominance planes

Once a picture was obtained, its central part (512 x 512 pix) was cropped out in order to reduce any spatial distortions that the camera's lenses may introduce into the periphery of each image (although no evidence was found of this in subsequent analysis). This was also done to reduce significantly the processing time for our PC. After that, the picture was loaded in a specialised matrix manipulation software (Matlab v.5.0) and converted to *double precision* (64-bit) format. Then it was linearised using Equation 5.1 with its corresponding parameters (depending of the *white balance* mode chosen) and divided by the corresponding IT. Then it was transformed into the $[LMS]$ components using Equation 5.4. The result was saved as a 64-bit image.

For some of the analyses we needed to obtain the chrominance and luminance components of the picture. Previous (Buchsbaum and Gottschalk 1983; Burton and Moorhead 1987; Ingling and Tsou 1988; Ruderman *et al.* 1998,; Travis 1991b; Wyszecki and Stiles 1967) studies have motivated various definitions of these components in terms

of the Smith-Pokorny (Smith and Pokorny 1975) cone responses. In this case, the definitions were as follows:

$$Lum = L + M$$
$$RG\_Chrom = (M - L)/Lum$$
$$BY\_Chrom = \left(S - \tfrac{Lum}{2}\right)/\left(S + \tfrac{Lum}{2}\right)$$

Equation 5.7

Our definitions of chrominance (*RG_Chrom,* and *BY_Chrom*) are based on the two chromatic channels that have been identified physiologically in the early visual system (see introductory chapter). *Luminance* corresponds to the subsystem that encodes achromatic (Derrington and Lennie 1984) visual information. *RG_Chrom* corresponds to the subsystem that compares the quantum catches of the L and M cones (Derrington *et al.* 1984; Krauskopf *et al.* 1982) and *BY_Chrom* corresponds to the subsystem that mediates the signals of the S cones (Derrington *et al.* 1984). Previous work (Párraga *et al.* 1998a) has shown that adopting a divisive (shadow-removing) definition of *RG_Chrom*, where sharp shadows are avoided may provide a clue to some of the distinctive features of visual processing needed for this task, such as the ability to perceive the same colour despite changes in light intensity (shadowing). For the *BY_Chrom* signal we adopted a definition that gives approximately equal numerical values to the *S* and the *Lum* signals in the numerator. Since we expect the signals for *L* and *M* planes to be quite large and not extremely different from each other (in our particular dataset of natural fruit and foliage), *Lum* is divided by a factor of 2, which accounts for its larger numerical value (roughly twice the value of *S*). Rearranging the definition:

$$BY\_Chrom = \frac{\frac{\left(S - \frac{Lum}{2}\right)}{2}}{\frac{\left(S + \frac{Lum}{2}\right)}{2}}$$

we can interpret *BY_Chrom* as a modulation between two signals, *S* and *Lum/2* divided by the average value between them.

It is possible that our definition of *BY_Chrom* is not the most appropriate and other manipulations of the LMS cone outputs (especially those less contaminated by luminance signal) yield larger differences of Fourier amplitude slopes between the Luminance and *BY_Chrom* channels.

5.3. Image gathering

The collection of images for our study was done under several conditions. All of the camera calibration was done inside our laboratory, where all parameters were controlled and the room could be darkened except for the target (illuminated by incandescent light). Indirect lighting from the projector to the camera was reduced using dark screens. Apart from these images, the pictures that we used for statistical analysis were taken outside the laboratory either in open spaces or inside the glasshouses of the Bristol Botanical Gardens. In all cases, the illumination was natural. On a typical day, we started around midday and waited until the sky was either completely overcast or clear. After that, the camera was fixed on a tripod, the set up checked and the picture taken using the camera's timer (to avoid shaking by the operator's hand). Particularly windy days were avoided. The usual places were private gardens (including the author's garden), the University of Bristol Royal Fort Gardens and Botanical Gardens. Some images were completely natural and some were arranged to locate fruits or other vegetables against a background of leaves.

Although the idea behind this natural scenes analysis is that ecologically-relevant visual tasks (particularly searching for food among leaves) have contributed to shape the spatio-chromatic characteristics of the HVS, we were not concerned too much in finding exactly the same types of fruits and leaves as primates would eat in the forest. The main reason for this was practical, since a collection of such pictures of fruit and leaves would imply a set of resources beyond the scope of this project. The second reason has to do with the fact that our analysis of the slope of the Fourier spectra of natural scenes is not extremely critical of the particular shape or coloration of the objects present in the image. For example, we believe that our findings would be equally valid for a whole range of images of reddish objects on a cluttered background of greenish patches illuminated by natural light, independently of the exact shape, colour and position of the objects. However, since this is difficult to prove with a limited picture database (and beyond the scope of this work), we restricted our image collection to pictures of objects that resembled those that are a valid target for a foraging primate. These included reddish and yellowish fruits (such as tomatoes, peppers, wild berries, fruit found in the tropical section of the Botanic Gardens, etc.) and red, yellow and blue flowers. To see the effects of different shapes and colorations in our image analysis we constructed a sequence of pictures of a human face on a background of leaves, seen from different distances, various sets of artificially-coloured berries, etc. Some pictures of general landscape were taken. In general, we found that the effects of the exact colour and shape in such images were negligible (e.g. the Fourier statistics of a human face on a background of leaves are similar to those of a red pepper on the same background) as far as our analysis is concerned. Perhaps it would be interesting, in the future, to compare the Fourier properties of this database to those of a similar database, collected in the natural environment where primate colour vision has evolved. Our pictures were taken during the year 2000 and part of 2001.

5.3.1. The distribution of colour and the number of "red" pixels present in our dataset

We obtained 124 images with a spatial resolution of 1600x1200 pixels by 24 bits (8 bits in each of RGB). These were taken under different conditions of illumination (sunny and cloudy), at different distances from the objects, and with many different kinds of non-green objects among foliage (large red fruit such as apples and tomatoes; small red berries; flowers). Almost all were taken using the minimum aperture (F11 – F11.4), i.e. the maximum depth of field available to avoid differences in focus within the same picture. Sixty-six images contained "red" pixels and fifty-eight did not. We defined a "red" pixel as one where the estimated activation of the L cone was 1.5 times or more the activation of the M cone. This was done after the bimodal distribution of colours obtained in a CIE style plot (see next Section), with a ratio of 1.5 in the through. Twenty-seven images were of landscapes where objects were at distances in the range 20 - 500 m. 32 images were close-ups of plants with no red objects. Most of the pictures were taken with the zoom lens set to its telephoto (19 – 20.4 mm focal length) setting. This avoided having to bring the camera very close to the objects, thus affecting illumination. Linear perspective is somewhat affected by this procedure compared to natural viewing but this was found not to affect the spectral slope values in any systematic way (see Section 5.4.1 for a more complete analysis). Square images with angular subtense of 13.6º were cropped from the rectangular images, and these were reduced to a size of 512 by 512 pixels.

Figure 5.13 shows the distribution of pixels according to their L and M cone catches for all our dataset, which is bimodal (it has two peaks). To produce Figure 5.13, all images were converted to $[LMS]$, as described above, and then the relative values l= L/(L+M+S) and m= M/(L+M+S) computed. This distribution is optimally bisected by

the criterion l>1.5 m, which separates "red" from "green" pixels as shown in the plots. Pixels that comply with the condition l>1.5 m are shown in dark grey and the others are shown in light grey.

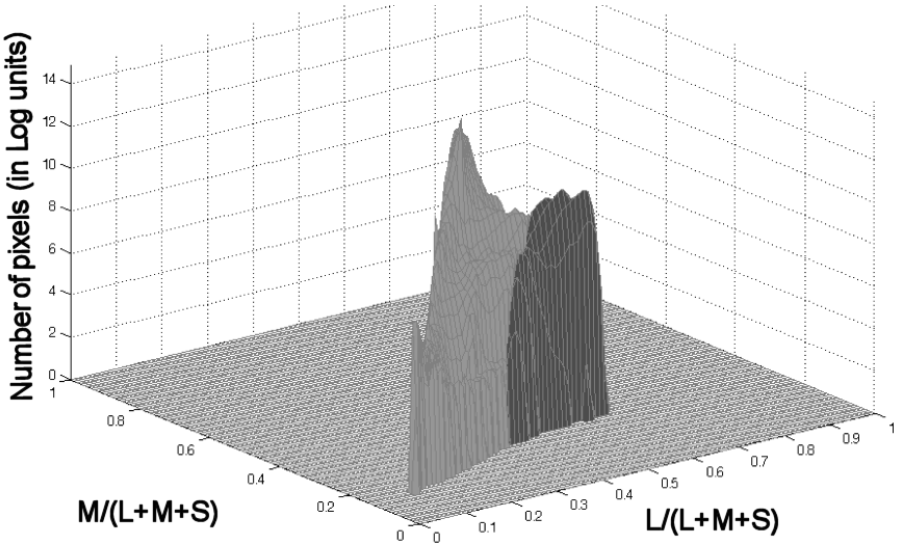

Figure 5.13: Distribution of the pixels for all the image dataset in a CIE-style diagram where the axis correspond to the L and M cone ratios. Here it is possible to see the bimodal distribution of "red" (darker grey surface) and "green" (lighter grey surface) pixels in our dataset.

5.3.2. Common practical problems

The most common practical problems were caused by the weather. Many of our pictures were taken under completely overcast conditions and, in some extreme cases, in rainy weather (glass houses pictures). Since the values of aperture were small (F11, F11.4) the integration times tended to be quite long for hand-held operation of the camera. This problem was sorted out by using a field tripod and 10 sec timer in all occasions and taking special care in avoiding effects of wind. Another practical problem was to find suitable fruit in locations where it could be photographed. In some cases, this was set-up by obtaining the fruit from a local shop and placing it on some leaves or on the grass. Another practical problem had to do with the size of our digital pictures (about 5.5 Mb

each) and the capacity of the storage device (64 Mb), which meant that we could only take 11 of them in one go, and would have to go back to the laboratory to download them onto a PC before continuing.

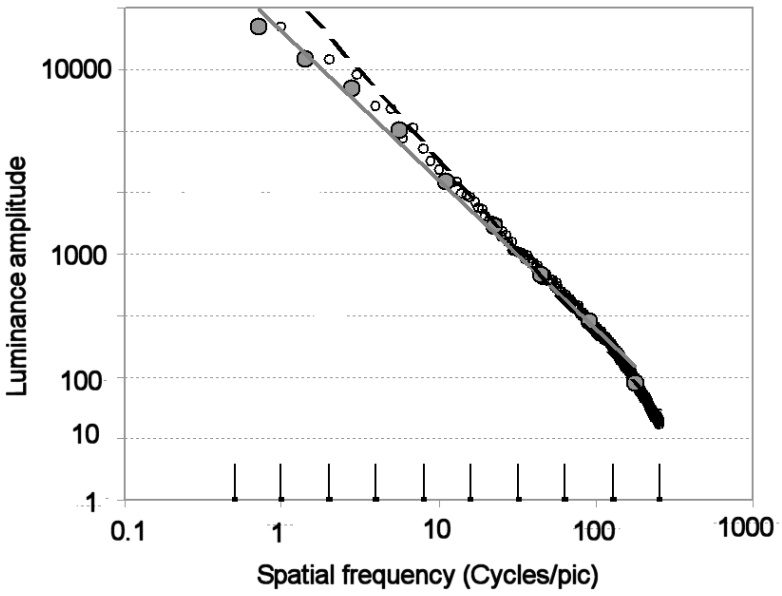

Figure 5.14: Example of a *log-log* plot representation of Fourier amplitude averaged across orientations. The empty circles represent the average amplitude for each of the 256 sample SF values. The broken line is the regression, dominated by the many points at high-SF (slope α=-1.42). The filled circles show the average amplitude when the sample values are only 9 (averaging regions are represented by the bands marked on the SF axis). The grey line represents the regression calculated only for these points (slope α=-1.20).

5.4. Computation of the slope of the Fourier amplitude spectra (α)

A standard two-dimensional fast-Fourier transform algorithm was used to derive the amplitude spectrum for a given *luminance* or *chromatic* (either red-green or blue-yellow) image representation. The DC component was removed. The usual way of calculating the amplitude slope requires that all SF-components of the picture are averaged regardless of

orientation and this average is plotted on *log-log* axes as shown by the empty circles in Figure 5.14.

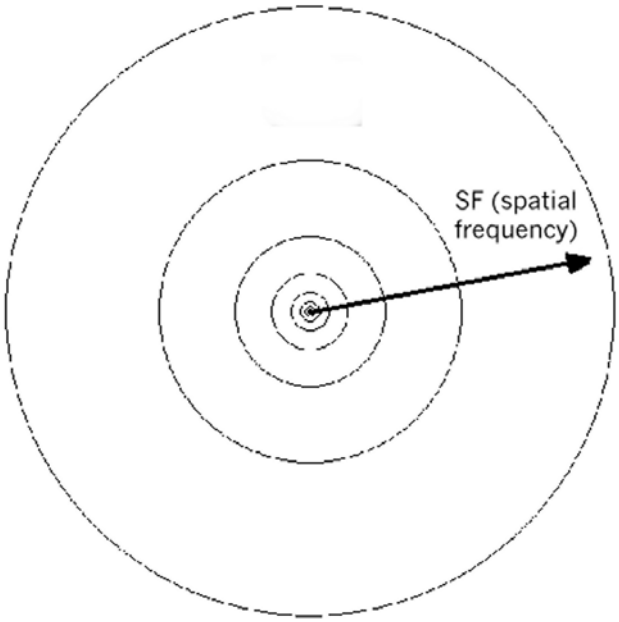

Figure 5.15: Scheme of the logarithmically-spaced one-octave SF bands used to calculate the slope of the amplitude spectrum (α).

The usual next step would be to calculate the slope of the line and report the number as α. In this approach, the high-SF range of the spectrum is over-represented when we measure the slope and these points will determine the value of α (the exemplary line fitted to the empty circles in Figure 5.2 has a slope of –1.42). Our approach is somewhat different (Párraga *et al.* 1998a). Spectral slopes were measured by dividing the Fourier space into 9 circularly symmetric, logarithmically-spaced, one-octave SF bands (see Figure 5.15), and averaging the Fourier content within each of the bands. The averages were then plotted against the mid-SF of the band in *log-log* co-ordinates, and the slope (α) of this line calculated by linear regression (as shown by the filled circles and grey line Figure 5.14). This prevented any bias due to oversampling of the high-SFs and reduced the overall effect of noise, which is primarily high-SF. The regression line fitted to the exemplary data

in Figure 5.15(b) has a slope of –1.20. The ordinate axis in Figure 5.14 shows the distribution of these bands in the SF domain.

5.4.1. Validation 1: estimation of the variability of the amplitude slope with optical characteristics of the camera

Optical characteristics of the camera (aperture, zoom settings, etc.) were varied as little as possible to avoid introducing artifacts in the estimation of α. The smallest possible aperture (around F11) was used in most cases in order to maximise depth of focus and the preferred focal length was "telephoto" to avoid interference between the camera operator and the scene. The use of the smallest aperture ensured that most of the objects in the scene are in focus (largest depth of field). Of the scenes in our dataset, 52 had a focal length of 19 mm (aperture F11) and 67 had a focal length of 20.4mm (aperture F11.4). Only five had a 7.2 mm focal length (aperture F7) and here we explore how this may have influenced our results.

To estimate the variability in the measurement of the luminance, red-green, and yellow-blue alpha slopes we recorded a typical image of a red fruit and green foliage at different viewing distances. We used different zoom settings (focal length = 7.2, 9.3, 10.8, 12.4, 14.1, 15.8, 17.6, 19.2, and 20.4mm) in such a way that the resultant subtense of the fruit was the same in all of the images. What varies is the degree of foreshortening of linear perspective in the images. We wished to know whether these effects on linear perspective would have any effect on our dependent variable, the spectral slope α. The values of α of the luminance, and chromatic representations were measured in all cases. Figure 5.16 shows these results. It is clear that there are small fluctuations of α around the average value (possibly due to small changes in the distribution of the objects in the different

pictures or changes in the illumination) but it is also clear that there is no correlation with focal length. Therefore, there is no significant effect of linear perspective on the spectral slope of this kind of image.

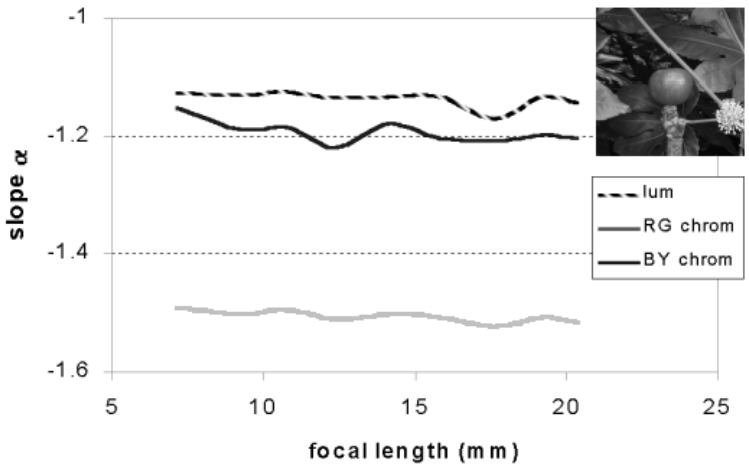

Figure 5.16: Variation of the slope α with the focal length of the camera.

5.4.2. Validation 2: errors introduced by the camera lens

To estimate the effects of chromatic aberration introduced by the camera lens, we took pictures (under controlled incandescent light conditions and under cloudy skylight) of a target that consisted of a white piece of paper with some black letters in bold characters. The lens was set to small aperture (F11.4). This produced pictures with a broad SF content. The three R, G and B planes that constitute the raw picture were separated and their contents Fourier-analysed. The slope (α) of the amplitude spectrum was measured for the three planes separately. The plots of Fourier amplitude versus spatial frequency in *log-log* co-ordinates were very similar for the three planes in each of the pictures. Table 14 (below) shows the values of α for two lighting conditions.

Results in Table 14 show that the differences in amplitude slope between the different planes are very small. In the case of the cloudy skylight illumination, this difference is about 3 % in the worst case. Moreover, a comparison of the slope sets for the two illuminants suggests that the small differences in α are a result of signal-to-noise ratio effects arising from the illuminant rather than from of the lens. A very similar set of numbers was also obtained for a much larger aperture (F4) suggesting that the lens is generally well designed to eliminate chromatic aberration (bottom row in Table 14).

Slope (α)	**R**	**G**	**B**
Incandescent light	-1.054	-1.0368	-0.995
Cloudy skylight	-1.077	-1.087	-1.111
Large aperture	-1.010	-1.015	-1.036

Table 14: Values of the three R,G and B amplitude slopes for the two pictures taken under different illumination and the third taken using incandescent light and a large lens aperture (F4).

5.4.3. Validation 3: errors occurring when the images were converted from the camera's RGB colour space to the human LMS cone representations

Since the spectral sensitivities of the RGB elements in the CCD camera do not exactly match the spectral sensitivities of the LMS cone fundamentals, the three LMS cone catches cannot be computed unambiguously (see Appendix B for more details on the subject). This becomes evident, as camera metamers will differ from human metamers. Given the way our camera was calibrated, we cannot compensate for this, but we can provide an estimation of how large the errors are likely to be. To estimate the camera's errors in determining the L, M and S cone catches for typical objects (fruits/leaves) photographed in these experiment, we photographed eight different objects under controlled illumination conditions in our laboratory. These objects were 5 different green

leaves and 3 red fruits (commonly found in our picture dataset). Figure 5.17 shows these images.

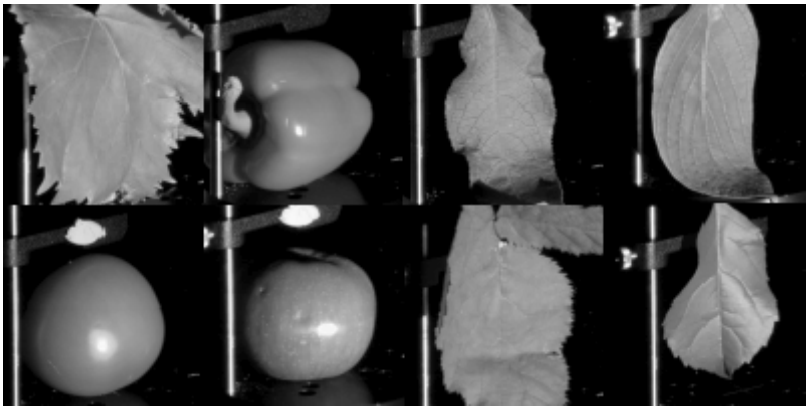

Figure 5.17: Images of the 6 different red and green objects (leaves and fruits) used in our estimation of errors due to metamerism.

The pictures were then converted to LMS cone representations using our standard procedure. The central 50x50 pixels (avoiding strong specularities, especially in the red fruit) were averaged and the standard deviation calculated. The value of this average was normalised by dividing by the total cone response (L+M+S). A similar value was obtained by measuring approximately the same spot with the SR and converting its spectral radiance output to the equivalent L, M and S cone catches. Both normalised values (obtained with the camera and the SR respectively) were compared and their mean differences found to be 4 % for the L signal and 8.7 % for the M signal (Table 15 show these results).

The largest differences in both measurements coincided with the largest standard deviation of the average of pixels within the 50x50 central box. This is due to the fact that the red surfaces were producing some specular reflections and their colour was not uniform. The same experiments performed on a non-specular, evenly coloured surface

(Macbeth colour chart) gives smaller differences (less than 5%) for the L, M and S signals obtained with the two methods (Table 16 shows these results).

Object	Calculated from **SRmeasurement**		Calculated from **Nikon 950 picture**	
	l	**m**	**l**	**m**
Vine leaf	0.461	0.444	0.441	0.417
Red pepper	0.659	0.322	0.637	0.292
Buddleia leaf	0.472	0.457	0.442	0.404
Oval leaf	0.450	0.432	0.437	0.414
Tomato	0.636	0.343	0.627	0.301
Apple	0.638	0.323	0.611	0.299
Bramble leaf	0.455	0.443	0.434	0.401
Small leaf	0.462	0.448	0.438	0.405

Table 15: Normalised values of l and m cone responses obtained with both the Nikon camera and the SR pointing to the set of fruits and leaves shown above.

Square number	**Colour**	Calculated from **SR measurement**		Calculated from **Nikon 950 picture**	
		l	**m**	**l**	**m**
7	orange	0.555	0.397	0.574	0.342
9	moderate red	0.539	0.345	0.574	0.315
11	yellow-green	0.481	0.460	0.457	0.423
13	blue	0.234	0.243	0.225	0.239
14	green	0.435	0.469	0.400	0.434
15	red	0.593	0.332	0.608	0.303
16	yellow	0.526	0.437	0.516	0.386
19	white	0.429	0.388	0.413	0.367
20	neutral8	0.428	0.387	0.416	0.367
21	neutral6	0.426	0.386	0.418	0.367
22	neutral5	0.421	0.385	0.417	0.367
23	neutral3	0.421	0.384	0.427	0.367
24	black	0.418	0.384	0.449	0.367

Table 16: Normalised values of l and m cone responses obtained with both the Nikon camera and the spectroradiometer pointing to different squares of a Macbeth Colour chart. The first column shows the number of each square according to the chart enumeration.

5.5. Results

Figure 5.18 shows two very different examples of photographs of natural scenes: (a) a landscape view of a British garden and (b) a close-up of a single ripe tomato seen against foliage. Picture (b) is representative of the type of images that form the majority (66 pictures) of our dataset, containing a red object on a background of leaves, illuminated by natural light with shadows cast across. Picture (a) is representative of the type of natural scenes that is used in the various natural scenes analysis in the literature (Chiao *et al.* 2000a; Párraga *et al.* 1998a; Ruderman *et al.* 1998) and constitute a substantial part of our dataset (58 pictures).

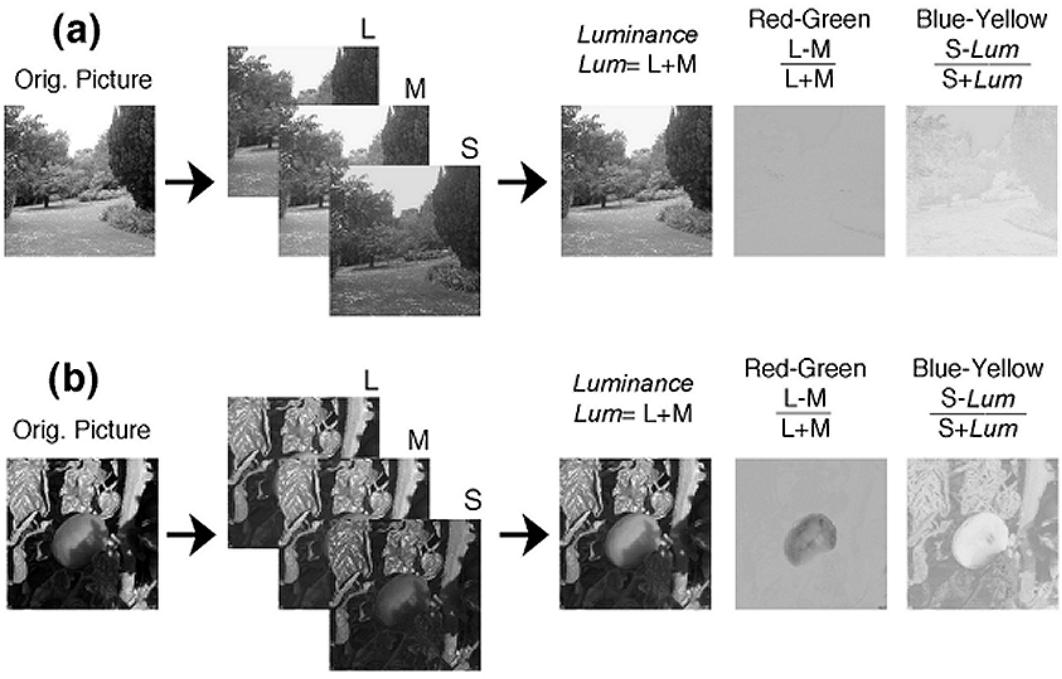

Figure 5.18: Two different examples of photographs of natural scenes. (a) Picture of an English garden and (b) picture of a ripe tomato against a background of leaves. The figure also shows the schematics of the process used to separate the luminance and the two chromatic information channel representations.

We calculated how the human L, M and S cones would have responded at each point in each of the scenes, and we then calculated the spatial form of the luminance signal and the red-green and blue-yellow chromatic signals within the HVS.

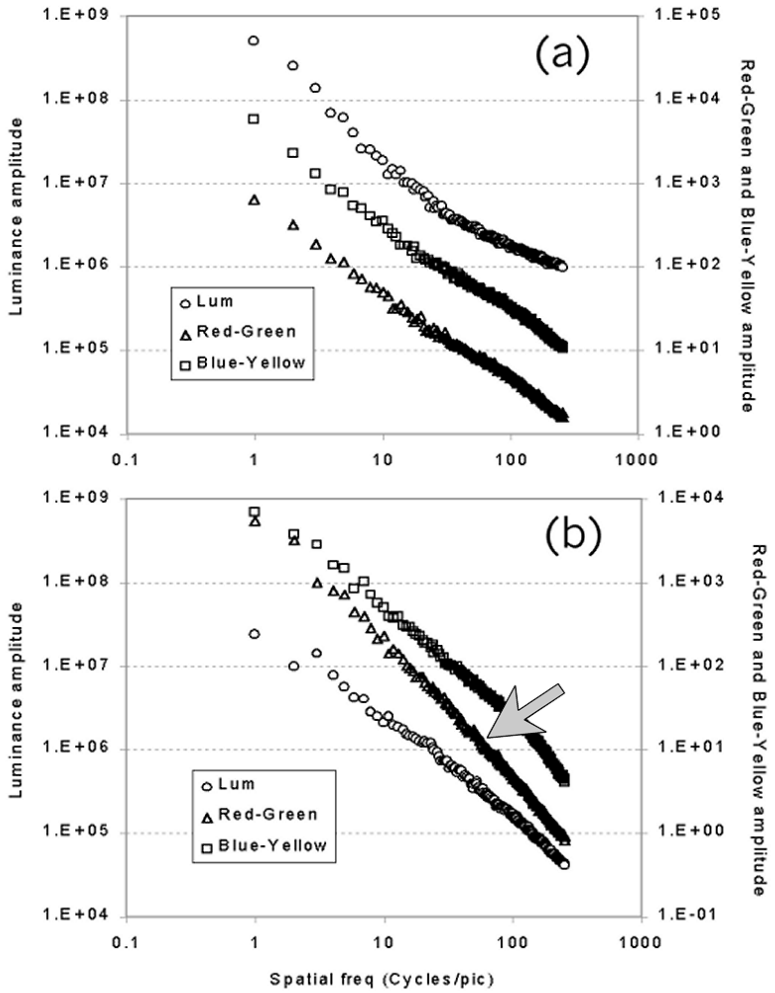

Figure 5.19: *Log-log* amplitude spectra of the pictures shown above: (a) garden scene and (b) ripe tomato. Fourier amplitude values for the luminance representations are shown on the left *y*-axis and amplitude values for chromatic representations are shown on the right *y*-axis. The grey arrow points at the Fourier amplitude plot of the red-green chromatic signal for the tomato picture, which is steeper.

Figure 5.19(a) shows the amplitude spectra plots for the luminance and for the two chromatic signals for the garden scene of Figure 5.18(a). It is clear that the slopes (α) are

very similar for the three graphs, though it should be noted that the magnitude of the amplitude in the red-green signals was considerably less than in the blue-yellow signal (the luminance has to be plotted on a different scale). This is not surprising since the values of M and L are similar so L-M is closer to zero.

Since human colour vision is biased more towards low-SFs than is luminance vision, we might have expected the spectral slope of the chromatic spatial signals to be steeper than that of the luminance signal, given that a steeper spectral slope gives a greater weight to low-SFs. The failure to find such an effect for the scene in Figure 5.18(a) confirms the earlier study of Párraga *et al* (Párraga *et al.* 1998a) using hyperspectral images.

Figure 5.19(b) shows plots of the Fourier amplitude spectra of the luminance and the two chromatic spatial signals of the close-up image of the ripe tomato "split" by a luminance shadow in Figure 5.18(b). The grey arrow points at the Fourier amplitude plot of the red-green chromatic signal for the tomato picture. It is, indeed, clear that the spectral slope (α) of the red-green spatial signal is now steeper than those of the luminance or blue-yellow spatial signals; the total amplitude for the red-green signal is, not surprisingly, much greater than for the general garden scene.

Figure 5.20 and Figure 5.21 show the generality of our finding that the amplitude spectrum of the red-green spatial signal is steep for close-up images of reddish objects viewed against foliage. In Figure 5.20, a plot of the amplitude slope for the luminance representation versus that of the red-green chrominance representation is shown. The graph shows two groups of points corresponding to pictures containing red objects (hollow circles) and pictures without them (grey triangles). In this graph, the majority of the scenery containing red objects lies below the diagonal and the majority of the "regular" scenery without red objects lie on or above the diagonal.

Figure 5.21(a) shows the ratio between the spectral slopes obtained for the luminance spatial signal and the red-green spatial signal for each scene as a function of how many "red" pixels (in %) there were in the scene (see section 5.3.1 above for a definition of a "red" pixel).

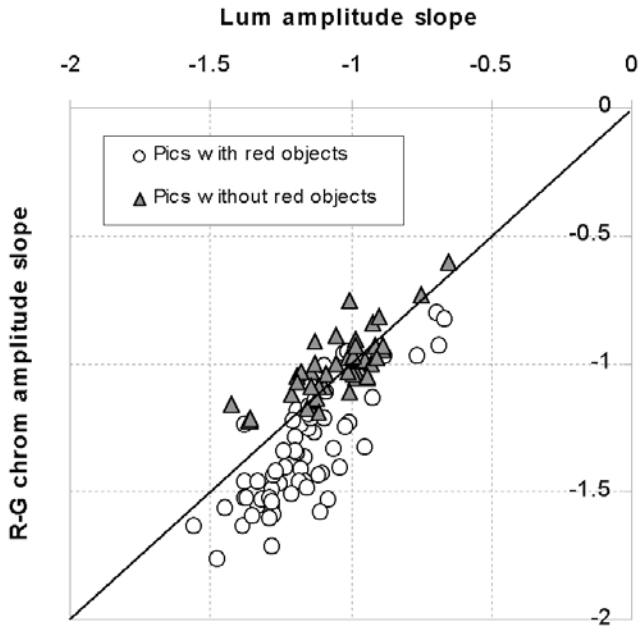

Figure 5.20: Plot of the amplitude slopes (α) for the luminance representations against the red-green chrominance representation slope. The line shows the diagonal of the plot. Round symbols correspond to pictures containing red objects and triangles correspond to pictures without any red object.

It can be seen that the results for all scenes follow a similar trend, and are differentiated on the basis of red pixel content. The data plotted on the left correspond to pictures of scenery without a strong presence of red objects (normal scenery—distant landscapes or close ups of foliage) and their $\alpha_{lum}/\alpha_{chrom}$ slope ratio near unity is consistent with previous findings (Párraga *et al.* 1998a). Scenes consisting of a background of foliage and a reddish object such as fruit appear on the right side of the graph. The important predictor of slope

ratio appears to be the proportion of red pixels on the leafy background. In general, the spectral slope of the red-green signal is greater than that of the luminance signal whenever the scene contains many red pixels.

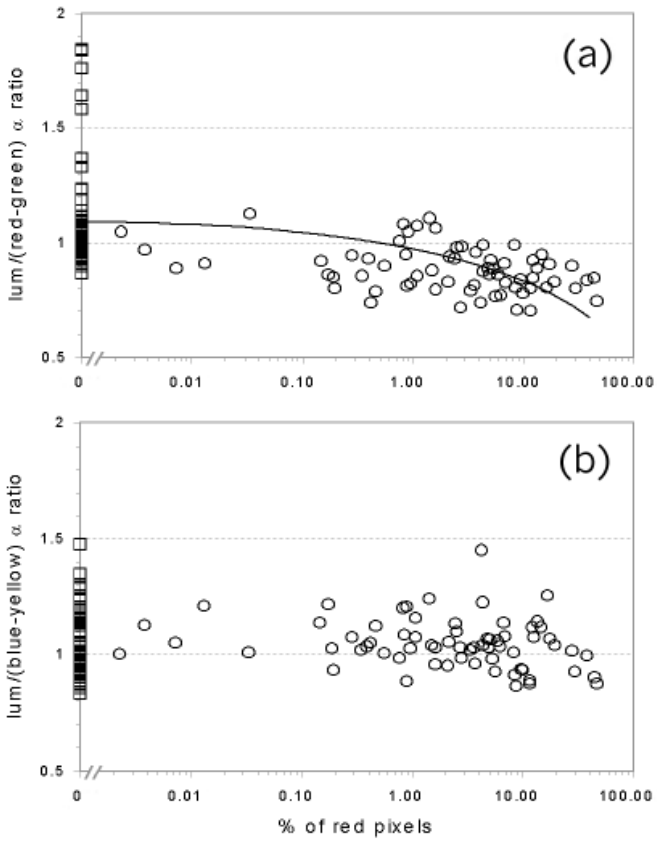

Figure 5.21: Ratio of the spectral slopes of the luminance and chromatic representations, for different image types, as a function of the percentage of pixels signalling the presence of a "red" object. The squares correspond to scenes with no presence of "red" pixels (i.e. they are plotted on the *y*-axis). The line in (a) represents the approximate data trend and reflects the fact that the average $\alpha_{lum}/\alpha_{chrom}$ slope ratio for normal scenery is about 1.1

Measures of the slope α of synthetic pictures with a large red object (like the tomato in Figure 5.18(b)) or many smaller red objects totalling the same area, show that the amplitude slope does not depend directly on the number of red objects. There seems to be a slight tendency to larger spectral slopes when red pixels are grouped together to

represent a single object, but many other unknown factors determine this result. When the pixels are clustered into several groups, representing several distinct fruits, the slopes of the amplitude spectra are not as steep, but the slope for the red-green signal is always steeper than for the other two signals.

In contrast with the previous results, Figure 5.21(b) shows that the spectral slope of the blue-yellow spatial signal does not differ systematically from that of the luminance signal. Table 17 shows the numerical means and STD of the α slope ratios for the different groups of images. There is no evidence in this set of images to suggest that there is any advantage in the bias of the human blue-yellow system to low-SFs. There may be other constraints on the development of the spatial characteristics of blue-yellow colour vision in primates. For example, it is possible that low sensitivity to high blue-yellow-SFs is a consequence of chromatic aberration in the eye and/or the sparse array of short-wave sensitive cones (see section 5.7 -Conclusions).

Numerical means and STD	Normal scenery (not including red objects); n=58.	All other images (including red objects); n= 66
Lum/(Red-Green) slope ratio	1.10 ± 0.22	0.88 ± 0.09
Lum/(Blue-Yellow) slope ratio	1.04 ± 0.14	1.05 ± 0.10

Table 17: The Table shows the numerical means and standard deviations of the slope ratios. 'n' refers to the number of pictures analysed in each case.

5.6. Discussion

5.6.1. Optimality issues. Is there an optimal viewing distance?

The prediction from human contrast sensitivity measures is that, if chromatic images contain more low-SF amplitude than luminance images, the ratio ($\alpha_{lum}/\alpha_{chrom}$) should be less than unity. We calculated the optimal value for this ratio, assuming optimal capture by the CSFs (Mullen 1985).

To calculate this, we fitted functions to the CSF data. Figure 5.22 shows these. The SF axis (*x*-axis in Figure 5.22) was divided into logarithmic SF-bands similar to those in Figure 5.15(a). We used the angular subtense of the camera optics to convert the bands from cycles/picture to cycles/deg. After this, we calculated the areas under each of the human CSFs for every band. The total areas were normalised to unity in both cases.

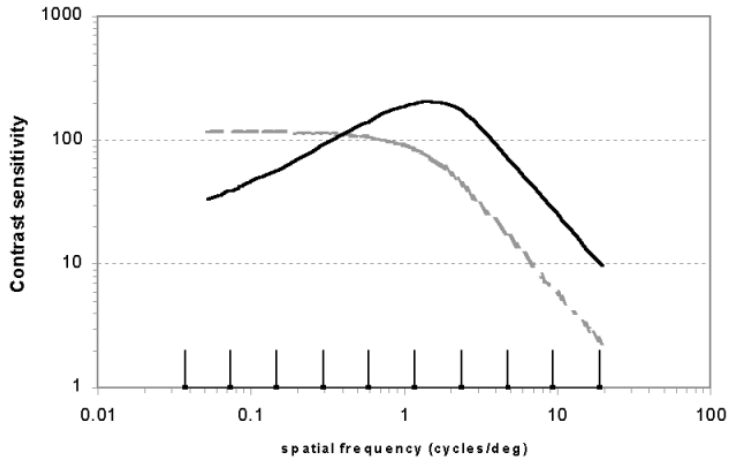

Figure 5.22: Functions fitted to the Mullen (1985) contrast sensitivity data for luminance (solid black lines) and red-green isoluminant (broken grey lines) gratings. The lines on the *x*-axis show the divisions of the Fourier space (similar to the bands shown in Figure 5.15)

Following this, we proceeded in similar fashion with a pair of exponential functions (that are similar to the linear Fourier amplitude functions when plotted in *log-log* space –see Figure 5.23). One of these (plotted in solid black in Figure 5.23) had a slope of –1.1 corresponding to our average value of α for "normal scenery". The slope of the second function (plotted as a broken grey line in Figure 5.23) was our free parameter. We found that a ratio of slopes of 0.76 for these two curves makes the ratio of the areas under them fit optimally with the ratios of the areas under Mullen's CSF data.

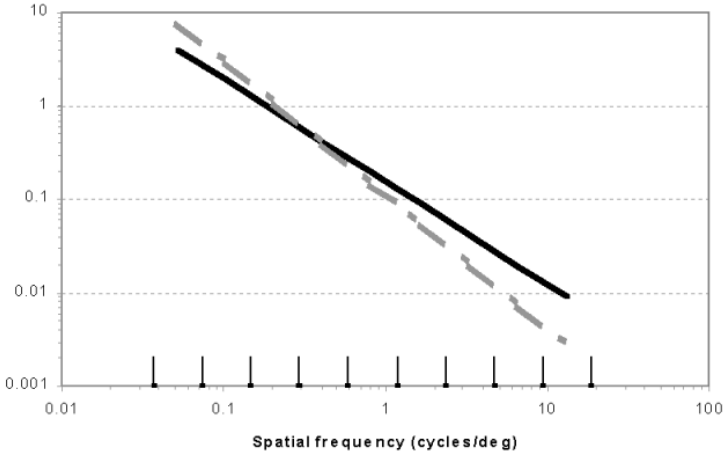

Figure 5.23: Plot of two exponential functions (lines when plotted in *log-log* space). The slope of the solid black line was fixed to –1.1. The slope of the other line was varied so that the ratio of areas under each of the lines fit optimally the ratio of the areas under the CSF data. The SF bands were chosen to be the same as in the previous figure. The areas were normalised to unity and calculated for each of the SF bands in all cases.

Once we had established this value for the optimal α slope ratio, we plotted the ratio of the spectral slopes, (for the luminance image and the red-green chromatic image), against the distance of the main objects in the scene from the camera in Figure 5.24. Distances were estimated from the real size of the object (in cm), the size of the object on the final picture (in pixels) and the subtended angle of our pictures. It is clear that the distant

scenes like Figure 5.18(a) have slope ratios around unity, but close views of red objects against foliage have lower slope ratios. The ratio matches the calculated optimal value for the human CSF at a viewing or grasping distance of about 0.4 m. This estimate does require some assumptions and would be influenced, for instance, if we had taken images with a different focal length of camera lens. However, our estimate of 0.4 m is of the same order of magnitude as the distance at which the human red-green spatio-chromatic system will be optimised for identifying red or yellow fruits against foliage. It could therefore be conservatively argued that this calculation suggests that the optimal viewing distance is of the order of magnitude of typical primate grasping distance.

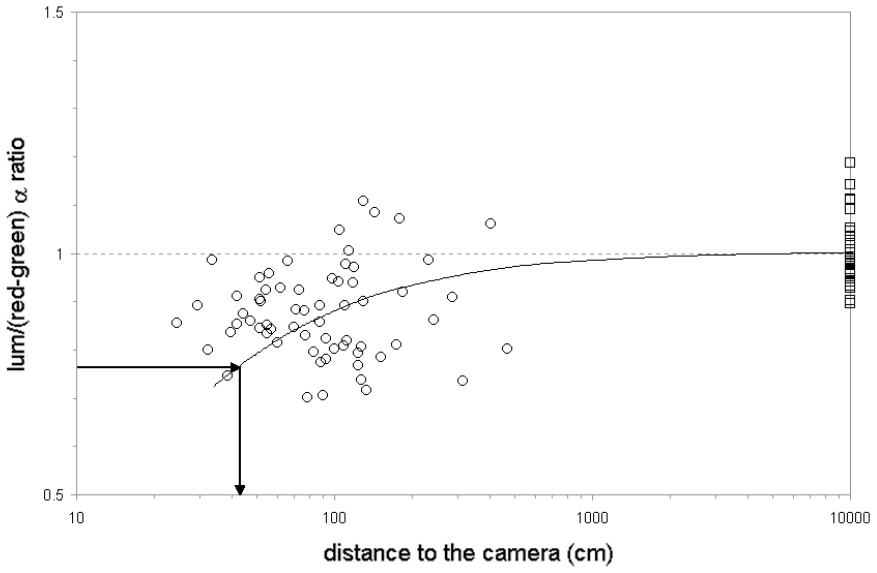

Figure 5.24: The relationship between the ratio of the spectral slopes (of the luminance and red-green chromatic representations) and distance from the camera to the main object in that scene, assuming the camera's (13.6°) angular subtense. The squares on the right represent "distant landscape" scenes, i.e. scenes where objects (trees, rocks, bushes, etc.) were numerous and their location was further away than 50 m. The line represents the data trend and reflects the fact that average values of $\alpha_{lum}/\alpha_{chrom}$ slope ratio are equal to 1.0 for these landscapes. The *log* *x*-axis was chosen to best show the geometrical relationship between the variables.

5.6.2. What makes the amplitude slope for the R-G representations of pictures containing fruit steeper?

Our divisive definition of RG_chrom (Equation 5.7) leads to the removal of light intensity variations (shadows), thus producing a particular effect on pictures with red objects. For example, in the case of a single red fruit against a background of leaves (shown in Figure 5.18(b)) with luminance variations (shadows) across the scene, our definition of RG_chrom produces an image where the fruit is represented by a central "red" region surrounded by a *uniform* "green" region (leaves). This causes the fruit to "pop out" since all background structure (luminance variations) is removed. The final RG_chrom picture is essentially similar to a "portrait" of a single darker object on a uniformly grey background. Our simulations show that monochrome portraits of single objects over uniform backgrounds tend to have steeper amplitude slopes. In Figure 5.25 a dark object is set against a cluttered background (panel a) which is then reduced in contrast by half (panel b) until it becomes halftone grey (panel c). Measurements of the Fourier amplitude slope show that it becomes more negative (steeper) as the background clutter is removed (for picture (a), α= -0.97, for picture (b), α = -1.075 and for (c), α =-1.26).

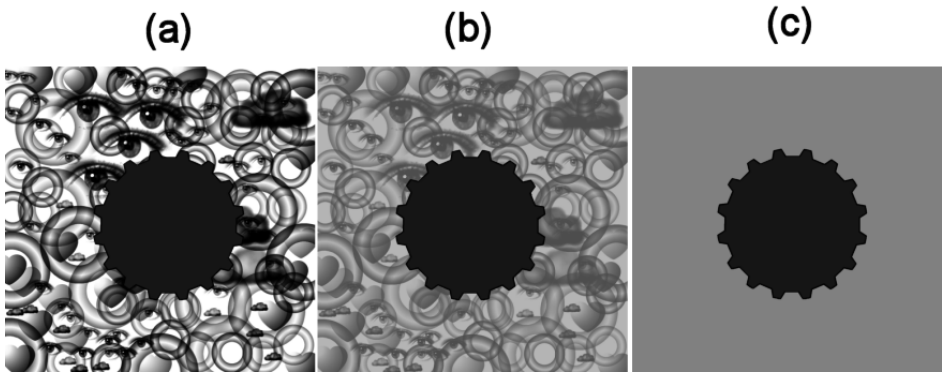

Figure 5.25: Synthetic pictures produced to show that Fourier amplitude slope is steeper for less cluttered backgrounds.

The "clutter removal" effect of the red-green chromatic system does not seem to be very strong for the yellow-blue system. Figure 5.18(b), shows how the red-green system eliminates the dappled background of leaves, giving a representation of a *uniform* fruit on a *uniform* background, allowing the fruit to "pop out". The same figure shows that the blue-yellow chromatic image does not provide the same uniform background. There are two main possible reasons for this. One is that shadow regions have indirect illumination from bluish skylight, and the other is that green leaves differ in their blueness (Dominy and Lucas 2001).

5.6.3. Generality of our results

We have collected stimuli in England and not in a tropical forest with the exact illumination, vegetation, etc. in which the HVS is thought to have evolved. However, we believe that these scenes are part of a wider variety of stimuli that can match the human contrast sensitivity data well, regardless of local environmental factors, and that our results would extend to any similar imaging situation (e.g. another garden or relevant forest). This generalisation implies that we cannot argue for the optimal encoding of any particular type of reddish or yellowish object on a leafy background. We have seen that any such object would do just as well, whether it is a red fruit (Osorio and Vorobyev 1996; Regan *et al.* 1998; Sumner and Mollon 2000a), a reddish human face, or a leaf which is more yellow than the rest (Dominy and Lucas 2001).

Since our results can be replicated with such different scenes as those of a human face, and a reddish blob, they do not specify exactly which kinds of images and tasks were associated with the evolution of colour vision in primates. What we have found is that human (and presumably other primate) colour vision is efficient at encoding images of reddish or yellowish fruit, viewed up close, against a leafy background. Our findings do

not rule out the idea that the low-SF bias of the red-green chromatic system is determined by optical factors such as chromatic aberration, but we show that such a bias would not disadvantage a foraging primate and it might even aid the task of picking ripe fruit or leaves.

This behaviour of the blue-yellow system implies that it is not optimised for detecting objects in foliage. Many questions remain to be answered. For example, what is the role of the low-pass transfer characteristic of the human blue-yellow system? There seems to be little connection between these and the efficient detection of fruit or any other target that we have discovered. However, it may be that a different class of images will reveal similar optimisation of the blue-yellow system, but it seems likely that other constraints such as the sparse sampling by the mosaic of short-wavelength sensitive cones in the retina, defocus of short wavelengths on the retina due to chromatic aberration, etc. are responsible for the blue-yellow system's properties. We have not investigated all the possible issues here.

5.7. Conclusions

a) We argue that the spatio-chromatic properties of a particular group of natural scenes (those including reddish or yellowish objects on a background of leaves) are relevant for understanding the HVS since they may represent stimuli that trichromatic primates and humans have evolved to optimally encode.

b) Our results suggest that the spatio-chromatic properties of the red-green chromatic system (in particular, the higher response of this system to low-

SF) are optimised for detecting reddish or yellowish objects against a background of foliage.

c) The blue-yellow chromatic system retains the same spectral slope as the luminance system and therefore, does not show any optimisation to the properties of the natural scenes investigated in this work.

d) The spatio-chromatic properties of the red-green chromatic system may be optimised for the encoding of any reddish or yellowish objects on a background of foliage, at viewing distances consistent with near (within reach) space.

Chapter 6.

Chapter 6

CONCLUSIONS

Is the human visual system optimised for encoding the statistical information of natural scenes?

6.1. Summary

Optimisation theories in biology are often based on three assumptions (Maynard Smith 1978). The first assumption considers all the possible observable features of an individual (regarded as a consequence of the interaction of its genotype with the environment), which in our case are the spatial and chromatic properties of the HVS. The second assumption regards what is maximised. Ideally, this should be the fitness of one individual, but in most cases is just some component or indicator of fitness. In our case, we consider maximisation of performance in a visual discrimination task and a foraging task. The third assumption is often about the mode of inheritance and population structure (natural selection cannot produce adaptation unless there is heredity), which we have not examined here. Throughout this work, we have tested how adequate the two first hypotheses are and to what extent they account for the spatio-chromatic properties of the HVS.

Our first approach to the optimisation problem has been to compare the performance of the HVS in an ecologically-relevant visual task in a range of visual conditions that differ slightly from those encountered in the visual environment. Our predictions are that best performance in the task will coincide with the properties met in those real visual

environments. Our second approach has been to consider the visual environment similar to that encountered during a primate foraging task and compare its statistical contents to the modulation transfer function of the HVS for chromatic and achromatic signals. As a control, we did analyse the same statistical properties of the general visual environment (it did not correspond to that encountered in a foraging task). Here we predicted that the modulation transfer function of both chromatic mechanisms would be matched to the environmental signal in the case of images corresponding to the foraging environment. The results obtained lead us to conclude that the HVS is optimised to efficiently capture some the properties of the environment in the presence of ecologically-relevant visual tasks. The overall conclusion must be treated with caution since we have shown optimisation for only a set of the HVS spatio-chromatic properties, and considered only a subset of all possible visual tasks.

Our results considering achromatic (foveal) spatial vision and chromatic vision seem to point to some degree of optimisation between the HVS and the properties of the environment. The situation is more complex in the case of peripheral vision, where some of the previous hypothesis might not be valid. It might be that peripheral vision does not satisfy the same criteria as foveal vision and we considered optimisation for the wrong visual task (image discrimination) when we should be testing a completely different variable, such as performance in movement detection, etc. This might explain the consistently superior foveal performance in this discrimination task despite compensatory resizing of the stimuli, etc.

In Chapter 5, we hypothesised that the main evolutionary reason of the chromatic properties of the HVS was to aid in foraging tasks (Mollon 1989; Mollon and Regan 1999; Regan *et al.* 1998, 2001; Regan *et al.* 1996; Sumner and Mollon 2000a, b). This hypothesis

is reinforced by our findings that optimal correspondence between the image statistics and the HVS modulation transfer function occurs at distances of the order of magnitude of grasping distance. We should also point out that in the case of blue-yellow chromatic vision, we did not find any relationship between its spatio-chromatic properties and those of the visual environment for our picture dataset. This may be because of two reasons (a) the subset of natural scenes considered (and therefore the task) is not adequate to the features of the blue-yellow mechanisms and (b) there are strong physical and physiological constraints (e.g. imperfect eye optics) which have played a key role in determining the limits of the system.

Much of our psychophysical results can be predicted from a simple multi-resolution model of local contrast discrimination involving only visual processes occurring at the level of the retina and area V1 of the visual cortex (i.e. "low level" visual processes). This points to the possibility that visual optimisation for the tasks proposed in Chapter 2 and Chapter 3 occurs at these levels. There are ways to further explore this hypothesis: for example by designing an experiment where the observer has to compare features of the reference and test image which do not correspond to *local* changes of contrast (e.g. by inverting or reversing the reference image). This will completely disrupt the model's performance and we predict it may possibly do the same with the observer's.

Our modelling results show that complex ways of combining the output of receptors and channels do not necessarily represent an improvement over the simplest approach, which is not completely in conflict with the literature (Rohaly *et al.* 1995; Watson 1993b; Watson and Solomon 1995). Another important feature of the model is its "robustness" to changes in the different parameters (filter bandwidths, shape of the contrast discrimination function, shape of the CSF function, etc.), which is consistent with the

variability found in nature. This suggests that it is the combination of many features and not only one of them, what is responsible for the optimal performance of the HVS in this visual discrimination task.

6.2. Answers to the specific questions of Section 1.10

a) Our results show that foveal achromatic vision in humans is optimised for performing a local contrast discrimination task, which we believe is ecologically-relevant in the presence of natural second-order statistics.

b) There are differences in terms of performance and optimisation between foveal and peripheral achromatic vision for the same ecologically-relevant discrimination task. Peripheral vision performs consistently worse than foveal vision and does not seem to be optimised to perform this task in the presence of natural statistics in the same way as foveal vision is, despite compensatory scaling factors.

c) A simple multi-resolution model of local contrast discrimination involving only "low level" visual processes can replicate the performance of the HVS in our contrast discrimination task, suggesting the lack of involvement of "higher" cortical processes.

d) We have found that the spatial properties (in particular, the physiological imbalance found between the shape of the modulation transfer functions) of the red-green chromatic and achromatic mechanisms of the parvocellular pathway are reflected in the statistical spatial and chromatic properties of

natural scenes. Our results show that this is not the case for the blue-yellow chromatic mechanism.

6.3. Future work

Throughout this research, we have encountered several ideas for future work, which may improve our understanding of the complex and fascinating relationship between the HVS and the properties of the visual environment.

The incorporation of the hypothesis of task-related optimisation has been a powerful tool, improving our capacity to make quantitative predictions using the optimisation paradigm. Following this, we have found that there are cases where the properties of the HVS do not match the properties of the environment (e.g. peripheral spatial vision and blue-yellow chromatic modulation transfer functions). However, we think that with the incorporation of new factors, such as other (more specific) ecologically-relevant visual tasks it is possible to account for such failures. Particularly, we believe that evaluating peripheral vision's performance in a different task, (e.g. object detection or movement detection) in the presence of natural scenes statistics may reveal some kind of optimisation, closer to what foveal vision has produced for discrimination tasks. Something similar may happen when we analyse a dataset of natural scenes that reflect some ecologically-relevant task which the blue-yellow visual system has evolved to perform. May be that in the case of the blue-yellow chromatic system, the relevant task is related to the presence of blue wavelengths in the light scattered by the atmosphere and in the shadows. We believe that it would be interesting to explore these possibilities in the future.

Another interesting possibility to explore is the role of "higher" visual processes in our visual discrimination task. For example, it might be interesting to try to disrupt the

performance of "lower" level visual mechanisms by either resizing, inverting or contrast-reversing the reference image and ask observers to perform the same discrimination task when local contrast comparisons are not possible. This might provide a disruption with a considerable increase of the discrimination thresholds and a change in the characteristics of the optimal stimuli.

Regarding the model discussed in Chapter 4, there is much room for improvement. The availability of faster and cheaper computers mean that we can now add many layers of complexity to our calculations (e.g. multiple orientation channels, different combinational rules, etc.) which may or may not improve the model's predictions. Another improvement, already discussed in Section 4.4.6, regards a possible "calibration" of the model.

We hypothesised that the image database collected in Chapter 5 has the same characteristics in terms of Fourier amplitude slope as similar databases collected anywhere in the world. This hypothesis could be tested by gathering images in the rainforest environment, where primate vision has evolved. At the moment, we have collected such images in Kibale Forest, Uganda and we are in the process of analysing them. Preliminary results show that natural scenes statistics may be the same regardless of changes in illumination, fruit and leaf coloration, etc.

So far we have applied different paradigms (spatial contrast discrimination and foraging) to study the achromatic and chromatic properties of the HVS. A combination of these paradigms by creating a morph colour discrimination task where the observer has to discriminate between two slightly different morph pictures based only on their *colour* differences may provide an insight into any optimisation of the chromatic channels to natural scenes. This can be done both ways, by creating a morph dataset with three

directions of change: (a) the morph spatial change, (b) the slope (α) of the Fourier spectrum in the chromatic plane and (c) the slope (α) of the Fourier spectrum in the luminance plane. A typical experiment may consist of an observer discriminating changes along the morph axis while one of the other axes varies, keeping the rest constant. Our predictions in this case would be that optimisation along the chromatic axis may not correspond to that along the luminance axis.

APPENDIX A.

Appendix A.

ESTIMATION OF ERRORS

A.1. Fitting a cumulative normal to the psychophysical data

The performance of an observer in our psychophysical experiments is reported as *morph thresholds* (expressed in % of morph). This is classically a measure of the morph change in the stimulus sequence necessary to produce a given level of performance. The results are also characterised by the rate at which performance improves with increasing morph change in the sequence. These two parameters are derived from a *psychometric function*, which describes the dependency of the observer's performance to the stimulus' change. Psychophysical data are obtained by sampling the observer's performance on our task (morph discrimination) at a number of different stimulus levels (see Figure A.1). In Figure A.1, each of the empty squares represents the average performance on a group of n trials and the line represents the fitted psychometric function.

To obtain the psychometric function, we applied the standard procedure (Finney 1971; Watson 1979) of fitting to the measured data the integral of a normal distribution (cumulative normal) which was constrained to fall between 50% (the guess rate in a 2AFC) and 98% (allowing for a 2% "finger error").

The fitting procedure cannot be a normal least squares fit since the errors on each morph percent value are different and, more importantly, they are asymmetrical about the measured probability. Figure A.1 shows that confidence bars at higher morph change values cannot go over 100% but they can stretch down up to smaller values. The fitting

method employed here consists of maximising the logarithm of the likelihood function (Collet 1991; Dobson 1990).

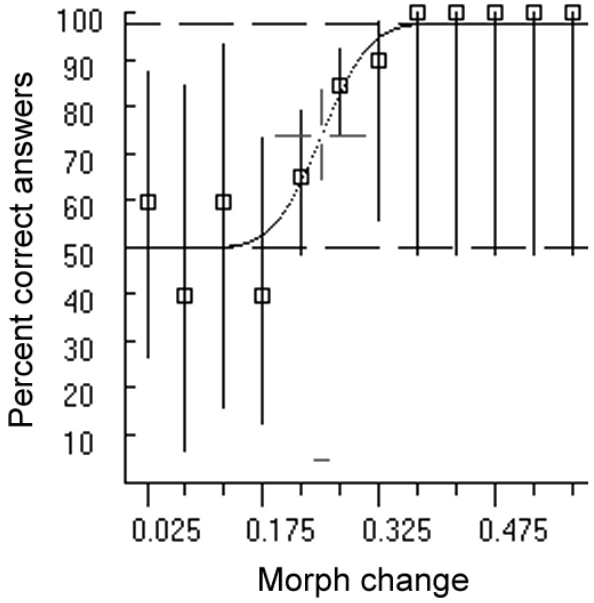

Figure A.1: Example of a typical set of psychophysical data obtained from a morph discrimination experiment. The ordinate shows the stimulus change (as a % of morph) and the abscissa the percent of correct answers. The curve shows the fitted psychometric function and the cross at the centre indicates the estimated threshold (74% correct answers)

If we suppose that in a typical example (like that shown in Figure A.1) each block consists of *n* of experimental trials and *K* is the total number of such trials, then we could model the psychometric function as (Wichmann and Hill 2001a):

$$\psi(x;\varepsilon,\beta,\gamma,\lambda) = \gamma + (1-\gamma-\lambda)F(x;\varepsilon,\beta)$$

Equation A.1

Where *x* is the morph change corresponding to each stimulus group (boxes in Figure A.1), γ is the base rate of performance (in our case fixed at 0.5 –or 50%) and λ is the miss rate or "finger error" in our 2AFC experiment. The parameters ε and β determine the

threshold of the psychometric function and its slope respectively. *F* is a function chosen to have a range (0,1). Together, all four parameters $\{\varepsilon, \beta, \gamma, \lambda\}$ determine the shape and position of ψ. We may call this set of parameters θ and rewrite Equation A.1 as $\psi(x;\theta)$.

A.2. Maximum-likelihood estimation

Given that in our case the values of *y* (observer's performance) have been generated by Bernoulli processes[21], it is possible to compute the likelihood value for a set of parameters θ, given the observed values of *y* as:

$$L(\theta; y) = p(y/\theta)$$

Equation A.2

Where the first term is the likelihood function and the second term is the probability function. The reversal in the order of notation points out the fact that once the data has been collected, *y* is fixed and θ is the variable. For simplicity, we assumed that *y* represents the fraction of correct answers and varies between 0 and 1. After expanding Equation A.2 we have:

$$L(\theta; y) = \sum_{i=1}^{K} C_{y_i}^{n_i} \left(p(x_i;\theta)^{y_i n_i} + (1 - p(x_i;\theta))^{(1-y_i)n_i} \right)$$

Equation A.3

[21] A Bernoulli process is a stochastic process consisting of finite or infinite sequence of independent random variables $X_1, X_2, X_3, \ldots$, such that (a) for each *i*, the value of X_i is either 0 or 1 and (b) for all values of *i*, the probability that $X_i=1$ is the same number *p*.

To find the maximum likelihood of L we have to find the set of parameters $\hat{\theta}$ for which the likelihood value is maximum:

$$L(\hat{\theta}; y) \geq L(\theta; y)$$

Equation A.4

As it is frequently the case, *log* functions are easier to maximise numerically, and since the *log* function is monotonic, it will be maximised by the same set of parameters as *L*. We can write Equation A.3:

$$l(\theta; y) = \sum_{i=1}^{K} \log \binom{n_i}{y_i n_i} + y_i n_i \log \psi(x_i; \theta) + (1 - y_i) n_i \log(1 - \psi(x_i; \theta))$$

Equation A.5

The easiest way to determine the set of parameters $\hat{\theta}$ that maximises Equation A.5 is to do it iteratively maximising the terms of the *l* that depend on θ. The implementation of this (called the Nelder-Mead SIMPLEX search algorithm) can be found in the literature (Press *et al.* 1992).

A.3. Estimation of the goodness of fit (χ^2)

To assess the goodness-of-fit of our psychometric function, we use a measure called *deviance* (*D*), defined as twice the difference between the *log*-likelihood of a perfectly fit function and $l(\hat{\theta}; y)$. A perfectly fit function would have many parameters and go though all the data points, thus its residual error (between the data and the predictions) is zero. $\theta_{\max}$ denotes the set of parameters for that perfectly fit function and *K* (the number of block presentations) represents now the number of free parameters. Following this:

$$D = 2 \cdot \left[l(\theta_{\max}; y) - l(\hat{\theta}; y) \right]$$

$$D = 2 \cdot \log\left[\frac{L(\theta_{\max}; y)}{L(\hat{\theta}; y)} \right]$$

Equation A.6

Since our fitting has two free parameters (ε and β) and the perfectly fit function has *K* free parameters, our measure of deviance has *K*-2 degrees of freedom. Deviance is preferred for assessing goodness of fit rather than, lets say, likelihood because for binomial data its distribution is asymptotical (Wichmann and Hill 2001a).

A.4. Effect of stimulus-independent errors (*lapses*)

Maximum-likelihood parameter estimation is extremely sensitive to stimulus-independent errors (Harvey Jr. 1986; Swanson and Birch 1992). This errors, (also called "observer lapses" or "finger errors") can bias dramatically our thresholds and slopes estimates. This misestimation is a direct consequence of the binomial *log* likelihood error metric represented in Equation A.5. When $\psi(x;\theta) \to 1$; the third term $(1 - y_i) n_i \log(1 - \psi(x_i;\theta)) \to -\infty$; unless the coefficient $(1 - y_i) n_i$ is 0. This coefficient will be equal to 0 only when performance is perfect (remember that y_i is the proportion of correct answers). In case of an observer lapse, the coefficient becomes non-zero and the large negative term influences the sum. This problem is solved by allowing the coefficient λ in Equation A.1 to take a non-zero value (Wichmann and Hill 2001a). We adopted a fixed value of $\lambda = 0.02$.

A.5. Estimation of the standard error (SE)

Standard errors were estimated from Edwards (Edwards 1972). In this approach, we know the value of α (our threshold estimate) and want to know what range of psychometric functions (each with its underlying threshold) could have given us that. If we fix the value of β to be the one giving the biggest likelihood and make γ and λ known constants, we can plot $l(\alpha, \hat{\beta})$ against α and obtain a function with a peak at the best fit $\hat{\alpha}$. The goodness-of-fit it is related to the shape of this peak (e.g. very sharp peak means we know the threshold precisely). Consequently, the standard error can be estimated as the second differential of the likelihood function $l(\alpha, \hat{\beta})$.

$$SE = \sqrt{-1 \Big/ \frac{\partial^2 l(\alpha, \hat{\beta})}{\partial \alpha^2}}$$

Equation A.7

This technique has been criticised for two different reasons: (a) lapses (stimulus independent errors) produced by human observers can bias the estimates of the psychometric function and (b) psychophysical datasets are usually rather small by the standards required for this statistical test. The estimation of the standard errors from staircase simulations (Monte Carlo resampling techniques) seeded from the original data fit has been advocated in the literature (Wichmann and Hill 2001a, b).

An estimation of the difference between our method for estimating standard errors and the staircase simulations was done by Tolhurst, D.J. and Chirimuuta, M. (unpublished observations). They concluded that standard errors estimated by staircase simulations are slightly bigger than those estimated by Equation A.7. They also pointed out that error bars

in our standard errors estimations should be asymmetrical (i.e. the thresholds measured are more likely to be too small than too high).

APPENDIX B.

Appendix B.

CALCULATION OF THE RGB TO LMS TRANSFORMATION MATRIX

B.1. Device-dependent colour space

In the course of the calibration of the Nikon 950 digital camera, we encountered a colour space defined by the three R, G, and B camera sensors. Each of these sensors produced values for each of the component pixels in the range 0-255 (8 bit values). For example, light (which is a combination of wavelengths) could stimulate the three camera sensors and produce a colour defined by the triplet $[RGB]$=(129, 152, 240). This triplet determines the position of a point (or colour) in a cube where each of the axes represents how much the input light has stimulated each of the three RGB receptors (which in turn depends on the receptor's spectral sensitivity). Since the receptor's stimulation depends on the Nikon Camera setup (camera's sensors sensitivity, white balance, etc.), the colour space produced by each combination of camera parameters is called a "device-dependent" colour space. On the other side, a "device-independent" colour space is one whose colour coordinates are fixed and produce the same triplet every time they are applied to the same colour. An example of a device-independent colour space is the CIE 1931 colour space discussed in Section 1.3.6, which is based on the $\bar{x}(\lambda), \bar{y}(\lambda)$ and $\bar{z}(\lambda)$ standard colour matching functions. Another device independent colour space is that determined by the Smith and Pokorny human cone spectral sensitivity functions (L(λ), M(λ) and S(λ)), also discussed in Section 1.3.6.

B.2. Calculation of the CIE 1931 tristimulus values

An example of how the spectral sensitivity functions that determine a colour space interact with input light to produce the colour space coordinates is the calculation of the CIE 1931 tristimulus values (also called XYZ values). The broken curve (plotted on the right side *y*-axis) in Figure B.1 shows a typical distribution of spectral energy, (radiance) reflected from an object illuminated by a tungsten light source (dark thick curve). The grey curve (plotted on the left *y*-axis) in Figure B.1 illustrates the spectral reflectance corresponding to the same object.

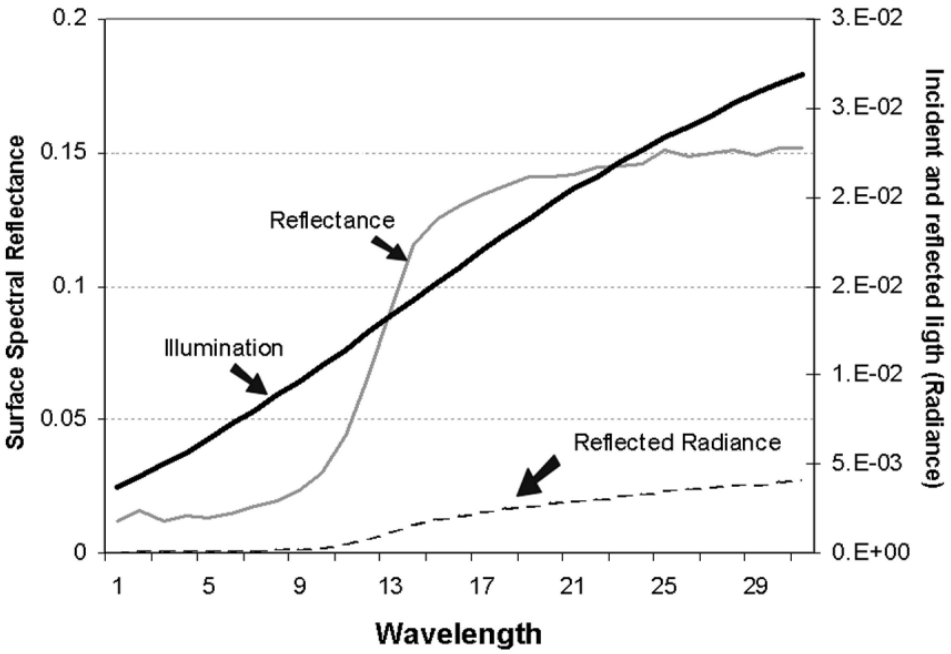

Figure B.1: Incident and reflected spectral radiance (right side *y*-axis) from an object with a given spectral reflectance (left side *y*-axis).

Given a surface with spectral reflectance $R(\lambda)$ illuminated by light with spectral distribution $E(\lambda)$, the X, Y and Z tristimulus values are calculated as follows:

$$X = \frac{1}{K}\int_{\lambda} R(\lambda)E(\lambda)\bar{x}(\lambda)d\lambda$$
$$Y = \frac{1}{K}\int_{\lambda} R(\lambda)E(\lambda)\bar{y}(\lambda)d\lambda$$
$$Z = \frac{1}{K}\int_{\lambda} R(\lambda)E(\lambda)\bar{z}(\lambda)d\lambda$$
$$K = \int_{\lambda} E(\lambda)\bar{y}(\lambda)d\lambda$$

Equation B.1

Where $\bar{x}(\lambda), \bar{y}(\lambda)$ and $\bar{z}(\lambda)$ are the CIE 1931 standard observer functions and K is a normalising factor which makes Y= 1 for a surface that reflects 100% at all wavelengths. In the exemplary case shown in Figure B.1, XYZ= (173, 156, 12). Figure B.2 illustrates these interactions.

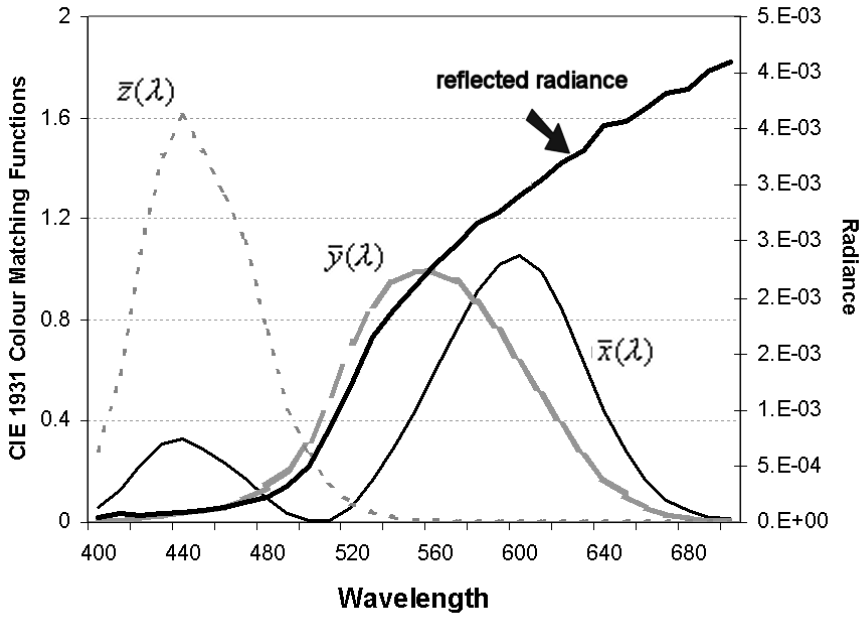

Figure B.2: The CIE 1931 standard observer colour matching functions (left *y*-axis) and the spectral radiance of light reflected from the same object shown in the previous figure.

B.3. Calculation of the chromaticity coordinates

The values of the *x, y* and *z* chromaticity coordinates can be calculated from the X, Y and Z tristimulus coordinates as follows:

$$x = \frac{X}{X+Y+Z}$$
$$y = \frac{Y}{X+Y+Z}$$
$$z = \frac{Z}{X+Y+Z}$$

Equation B.2

Where the third coordinate (*z*) can be calculated from the other two (*z=1-x-y*). This means that we can obtain the chromaticity coordinates of any complex light stimuli by integrating as described above and all possible colours can be represented by only two numbers. Plotting the values of *x* and *y* for monochromatic light produces a curve called *spectral locus*.

B.4. Calculation of the LMS cone excitation values

Similarly, we can calculate the amount of excitation of the three pigment types (L, M and S) for a complex light stimulus I(λ) by using the Smith and Pokorny (Smith and Pokorny 1972) cone spectral sensitivity functions ($l(\lambda)$, $m(\lambda)$ and $s(\lambda)$). Figure B.3 shows these functions.

$$L = \frac{1}{K}\int_{\lambda} I(\lambda)l(\lambda)d\lambda$$
$$M = \frac{1}{K}\int_{\lambda} I(\lambda)m(\lambda)d\lambda$$
$$S = \frac{1}{K}\int_{\lambda} I(\lambda)s(\lambda)d\lambda$$

Equation B.3

We could also calculate the amount of excitation that a primary such as $\bar{x}(\lambda)$ would produce:

$$L = \frac{1}{K}\int_{\lambda} \bar{x}(\lambda)l(\lambda)d\lambda$$
$$M = \frac{1}{K}\int_{\lambda} \bar{x}(\lambda)m(\lambda)d\lambda$$
$$S = \frac{1}{K}\int_{\lambda} \bar{x}(\lambda)s(\lambda)d\lambda$$

Equation B.4

Each of the CIE 1931 standard observer primaries $\bar{x}(\lambda), \bar{y}(\lambda)$ and $\bar{z}(\lambda)$ has a defined overlap with the absorption spectra of the three pigments and therefore leads to a defined sensation. A linear increase in one primary would traduce to scalar multiplication with the pigment sensitivities. Because of the principle of univariance, we can add the influences of the three CIE primaries to the resulting excitations of the three L, M and S pigment types. This implies that there is a linear transformation between the $[XYZ]$ tristimulus values and the $[LMS]$ colour space determined by the pigment isomerizations.

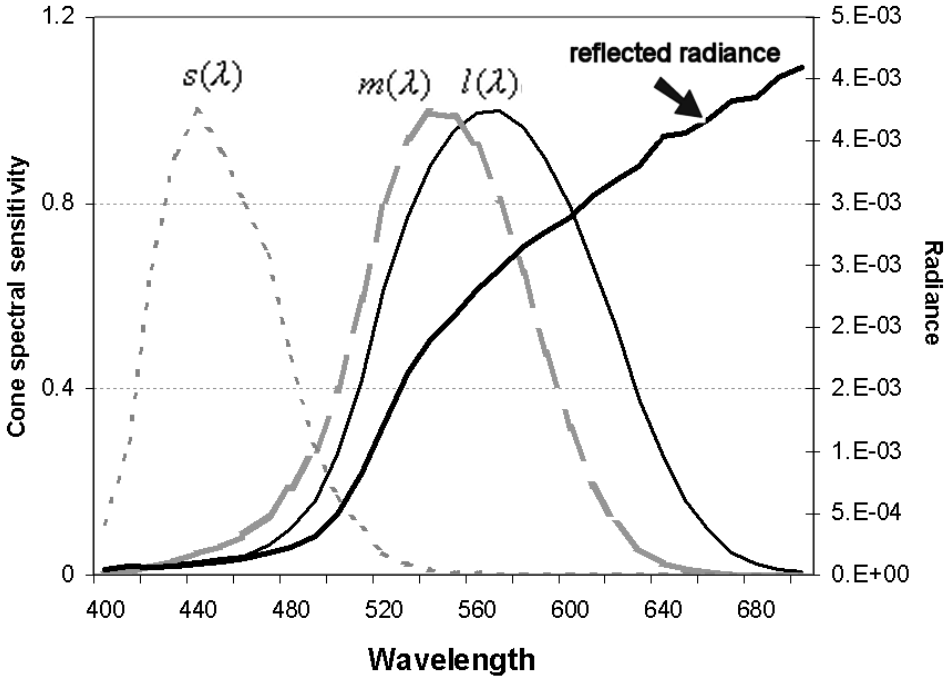

Figure B.3: The Smith and Pokorny (Smith and Pokorny 1972) cone spectral sensitivity functions (left *y*-axis) and the spectral radiance of light reflected from the same object shown in the previous figure.

B.5. Calculation of the RGB to LMS transformation matrix

The transformation between the camera's $[RGB]$ space coordinates and the $[LMS]$ cone space coordinates can be obtained by finding the best solution to the transformation matrix T in Equation B.5:

$$\begin{pmatrix} L \\ M \\ S \end{pmatrix} \approx T \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

Equation B.5

Since both systems do not fill the orthogonality requirements to make this transformation exact, Equation B.5 does not have an exact solution. An approximate solution can be found by projecting the axis of one coordinate system in the other as if they were

orthogonal. To do this we need to know the camera's R, G, B values produced by light with spectral profile similar to that of the $l(\lambda)$, $m(\lambda)$ and $s(\lambda)$ cone spectral sensitivity functions.

$$L_R = \int_\lambda l(\lambda)R(\lambda)d\lambda; \quad L_G = \int_\lambda l(\lambda)G(\lambda)d\lambda \quad L_B = \int_\lambda l(\lambda)B(\lambda)d\lambda;$$

$$M_R = \int_\lambda m(\lambda)R(\lambda)d\lambda; \quad M_G = \int_\lambda m(\lambda)G(\lambda)d\lambda; \quad M_B = \int_\lambda m(\lambda)B(\lambda)d\lambda;$$

$$S_R = \int_\lambda s(\lambda)R(\lambda)d\lambda; \quad S_G = \int_\lambda s(\lambda)G(\lambda)d\lambda; \quad S_B = \int_\lambda s(\lambda)B(\lambda)d\lambda;$$

Each coefficient of the matrix T corresponds to one of these "projections" of the R, G and B primaries into the L, M, S space.

$$T = \begin{vmatrix} L_R & L_G & L_B \\ M_R & M_G & M_B \\ S_R & S_G & S_B \end{vmatrix}$$

Equation B.6

Now, suppose that the camera's RGB sensors are illuminated by light with the same energy across the visible spectrum (white light). Because of the normalisation applied to the camera's spectral sensitivity functions, such light will produce a similar output in each of the RGB sensors.

$$\begin{aligned} L_w &= L_R + L_G + L_B \\ M_w &= M_R + M_G + M_B \\ S_w &= S_R + S_G + S_B \end{aligned}$$

Equation B.7

If we assume that similar light will also produce equal activity in the three cone types $[LMS]$, it is possible to map the white point from one space to the other. Mathematically this can be represented as:

$$L_w = M_w = S_w$$

Equation B.8

We can satisfy Equation B.6 and Equation B.8 simultaneously in two different ways. One option is to solve the following equation:

$$\begin{vmatrix} L_R & L_G & L_B \\ M_R & M_G & M_B \\ S_R & S_G & S_B \end{vmatrix} \cdot \begin{vmatrix} x_1 \\ x_2 \\ x_3 \end{vmatrix} = \begin{vmatrix} 1.0 \\ 1.0 \\ 1.0 \end{vmatrix}$$

Equation B.9

by finding the vector $\boldsymbol{X}$ that "normalises" the sum of the row so that they all add up to the same number. The other option is to compute $a_1 \cdot l(\lambda)$, $m(\lambda)$ and $a_2 \cdot s(\lambda)$ cone spectral sensitivity functions where a_1 and a_2 are normalising parameters, so that the rows of matrix T add up to the same number. We chose to explore the second possibility. Figure B.4 shows a plot of the Smith and Pokorny (Smith and Pokorny 1975) cone sensitivity functions after they were normalised so that Equation B.8 is true for the Nikon 950 camera (a_1= 0.856, a_2= 1.418).

Figure B.5 shows a plot of the opposing red-green (M-L) and blue-yellow (S-(M+L)/2) signals. From the same figure it could be seen that the wavelength of maximum sensitivity for the L-M and S-(L+M) mechanisms are close to 570 and 500 nm respectively, which is near the psychophysically measured values (580 and 506 nm) for colour matching experiments.

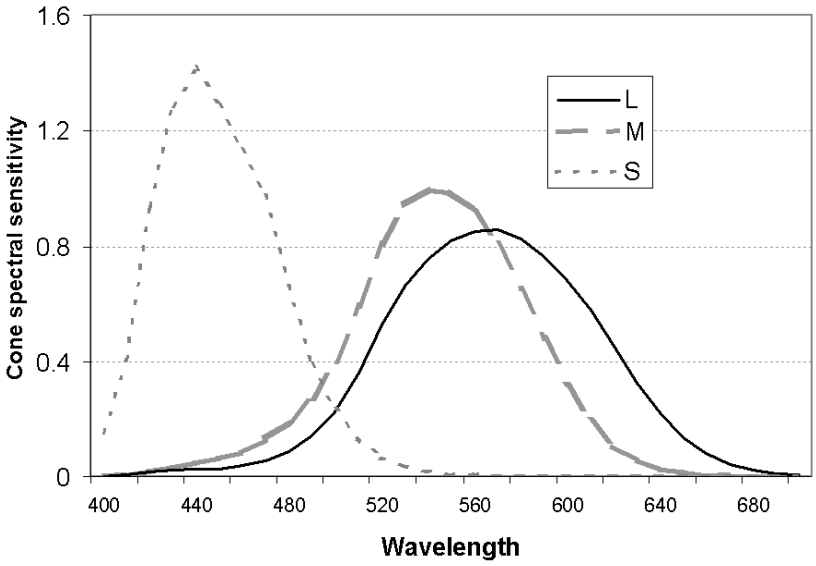

Figure B.4: Smith and Pokorny (Smith and Pokorny 1972) cone spectral sensitivity functions normalised.

The final transformation matrix T was obtaining after calculating the parameters in Equation B.6 using the functions plotted in Figure B.4 and the CIE 1931 standard observer functions ($\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$).

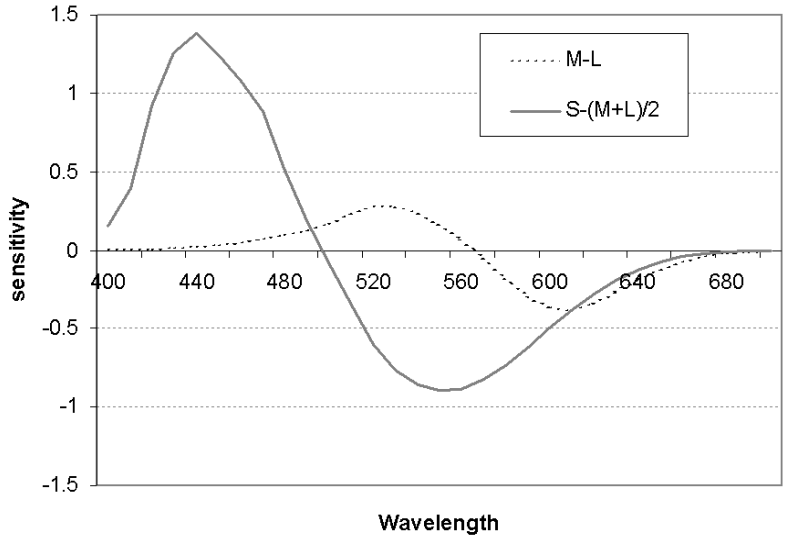

Figure B.5: Plot of the red-green (M-L) and blue-yellow (S-Lum/2) opposing signals as a function of wavelength.

B.6. Reliability of the matrix transformations

As mentioned before, our matrix T can only be estimated up to certain degree using Equation B.6 (its discrete for was presented in Equation 5.6). The reason for this uncertainty has its roots in the fact that R, G and B sensors of the camera do not respond to light on a wavelength-to-wavelength basis, instead they integrate colour responses along the wavelength spectra. Because of this, it is possible for two wavelength distributions to produce the same R, G and B responses. These wavelength distributions (or colours) are called metamers. Each set of primaries ($R(\lambda), G(\lambda), B(\lambda)$ and $l(\lambda)$, $m(\lambda)$, $s(\lambda)$) has its own set of metamers which are not necessary the same. For example, a pair of lights that produce the same camera $[RGB]$ output are not seen the same by a human observer. This failure of agreement in colour match is called camera-eye metamerism in the colour imaging community (Hong *et al.* 2000). The transformation matrix T between the device-dependent camera's colour space $[RGB]$ and the device-independent cone sensitivity colour space $[LMS]$ is only 100% accurate in the case when both sets of spectral sensitivity functions $R(\lambda), G(\lambda), B(\lambda)$ and $l(\lambda)$, $m(\lambda)$, $s(\lambda)$ have the same metamers. In all other cases, there are different degrees of inaccuracies. Figure B.6 illustrates this point with an example of a conversion between an arbitrary set of primaries $[ABC]$ and another $[PQR]$, which results in a highly inaccurate matrix conversion (all coefficients of matrix T will be close to 0).

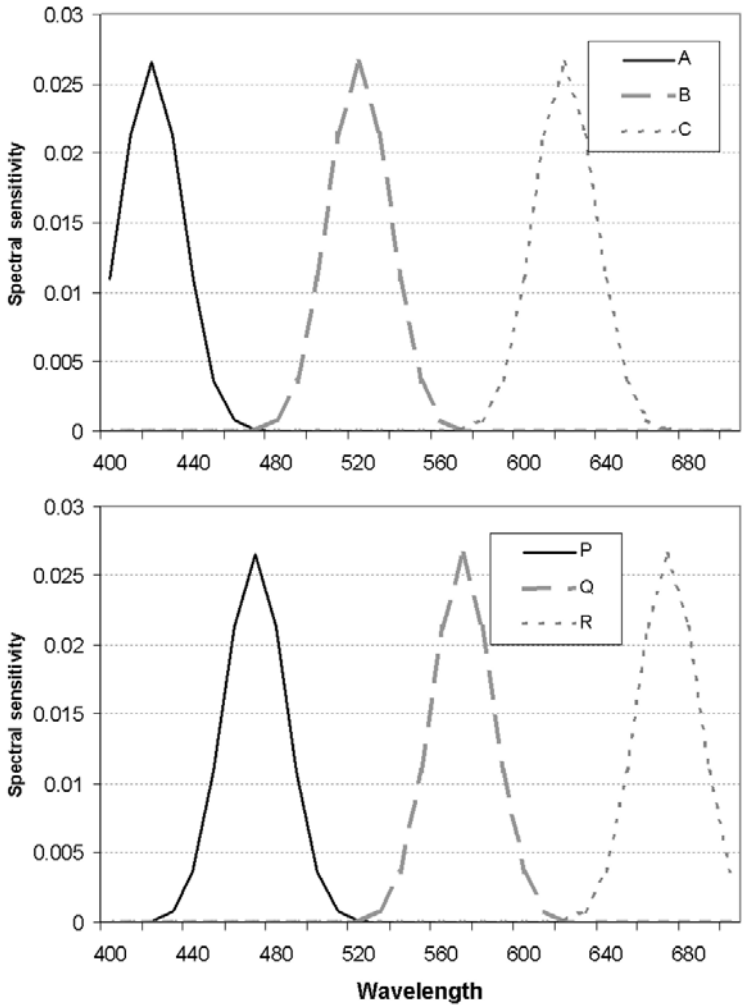

Figure B.6: Example of two hypothetical set of primaries (ABC and PQR) that would give a highly inaccurate matrix transformation T if calculated using the method above.

In this (rather extreme) example the set of curves do not overlap and their peaks are located in different positions along the visible spectrum. This type of inaccuracy is likely to cause the mismatch between the L, M and S values predicted from SR radiometric data and those calculated from camera R, G and B data in Section 5.4.3.

The matrix below is the (customised) transformation matrix T obtained for our Nikon Coolpix 950 digital camera using the procedure described in the previous sections:

$$T = \begin{vmatrix} 0.347 & 0.574 & 0.077 \\ 0.144 & 0.720 & 0.134 \\ 0.029 & 0.130 & 0.840 \end{vmatrix}$$

As a comparison, we estimated the coefficients of a matrix T' obtained from the International Telecommunications Union (ITU) standard XYZitu601-1 (D65). This standard matrix (used in the literature to convert between the device dependent $[RGB]$ space and the device-independent XYZ space) was modified so that its rows add up to 1. Matrix *X* was obtained by combining the normalised ITU standard matrix with the XYZ-to-LMS standard transformation matrix (Wyszecki and Stiles 1967). The result is shown below (Reinhard *et al.* 2001).

$$X = \begin{vmatrix} 0.3811 & 0.578 & 0.040 \\ 0.196 & 0.724 & 0.078 \\ 0.024 & 0.128 & 0.844 \end{vmatrix}$$

Using the matrix T obtained in this way and Equation 5.6 we calculated a set of LMS sensitivity functions, where each point is produced by a matrix transformation of the original sensor's RGB triplets at each wavelength. Figure B.7 shows a plot of these calculated LMS values (secondary *y-axis*) superimposed to the original LMS cone sensitivity functions (primary *y-axis*). The figure shows the limitations of this matrix transformation, given that the calculated LMS values still resemble the shape of the original camera's RGB functions from where they are derived. Figure B.7 also shows that the calculated L function has now shifted closer to the M function and is sensitive to middle-wavelength frequencies, more in accordance with the shape of the real LMS functions. However, given that the spectral distribution of light reflected from most natural objects is in general continuous across the visible spectrum and does not contain sharp peaks, we believe that these approximations are appropriate for the kind of stimulus

that we intend to study (see Section 5.4.3 for a quantitative comparison between the camera's transformation and a calibrated instrument).

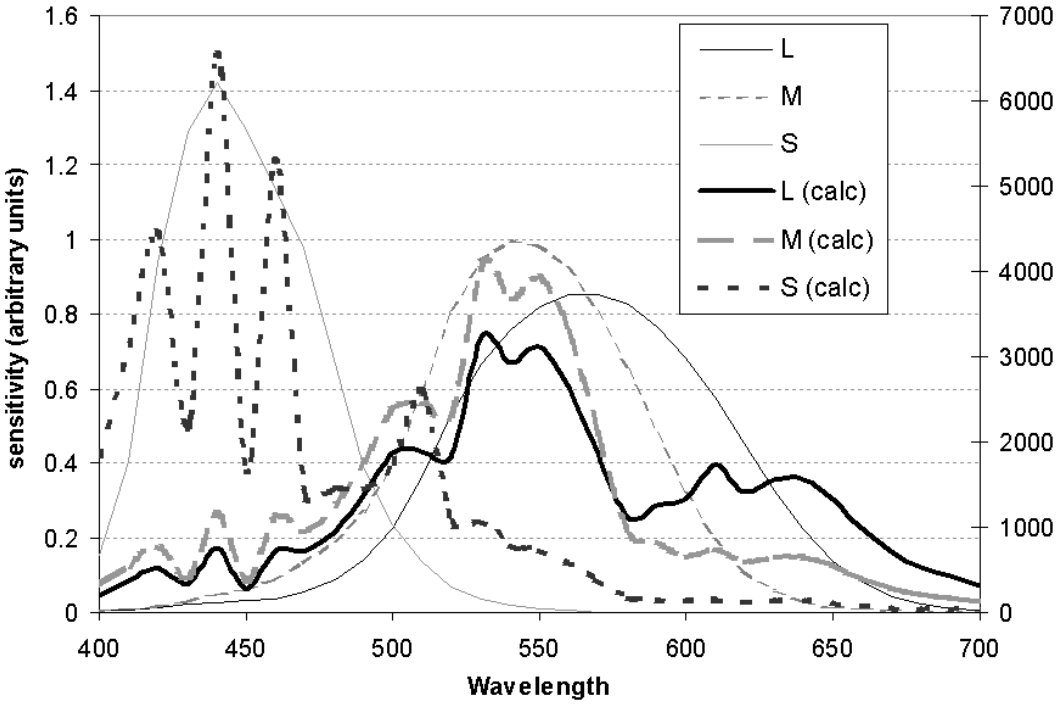

Figure B.7: Comparison between the LMS spectral sensitivity functions calculated from the camera's RGB spectral sensitivities and the human LMS sensitivity functions.

REFERENCES

References

Ahnelt P.K., Kolb H. and Pflug R. 1987. *Identification of a sub-type of cone photoreceptor, likely to be blue sensitive, in the human retina.* Journal of Comparative Neurology. 255, 18-34.

Aho A.C., Donner K., Hyden C., Larsen L.O. and Reuter T. 1988. *Low Retinal Noise in Animals With Low Body-Temperature Allows High Visual Sensitivity.* Nature. 334, 348-350.

Ahumada A.J. 1987. *Putting the Visual-System Noise Back in the Picture.* Journal of the Optical Society of America A. 4, 2372-2378.

Albrecht D.G. and De Valois R.L. 1981. *Striate cortex responses to periodic patterns with and without the fundamental harmonics.* Journal of Physiology-London. 319, 497-514.

Alpern M. and Pugh E.N. 1974. *The density and photosensitivity of human rhodopsin in the living retina.* Journal of Physiology. 237, 341-370.

Amundson R. 1994. *Two concepts of constraint: Adaptationism and the challenge from developmental biology.* Philosophy of Science. 61, 556-578.

Antes J.R. 1974. *The time course of picture viewing.* Journal of Experimental Psychology. 103, 62-70.

Atick J.J. 1992. *Could information theory provide an ecological theory of sensory processing?* Network: Computation in Neural Systems. 3, 213-251.

Atick J.J. and Redlich A.N. 1992. *What does the retina know about natural scenes?* Neural Computation. 4, 196-210.

Attneave F. 1954. *Informational aspects of visual perception.* Psychological Review. 61, 183-193.

Attwell D. and Laughlin S.B. 2001. *An energy budget for signaling in the grey matter of the brain.* Journal of Cerebral Blow Flood and Metabolism. 21, 1133-1145.

Aubert H. and Förster R. 1857. *Beiträge zur Kenntniss des indirekte Sehens (I) Untersuchungen über den Raumsinn der Retina.* Arch. f. Ophth. 3, 1-37.

Azzopardi P. and Cowey A. 1993. *Preferential Representation of the Fovea in the Primary Visual-Cortex*. Nature. 361, 719-721.

Badcock D.R. 1984a. *How do we discriminate relative spatial phase?* Vision Research. 24, 1847-1857.

Badcock D.R. 1984b. *Spatial phase or luminance profile.* Vision Research. 24, 613-623.

Baddeley R., Abbott L.F., Booth M.C., Sengpiel F., Freeman T., Wakeman E.A. and Rolls E.T. 1997. *Responses of neurons in primary and inferior temporal visual cortices to natural scenes.* Proceedings of the Royal Society of London Series B. 264, 1775-1783.

Baker C.L. and Braddick O.J. 1985. *Eccentricity-dependent scaling of the limits of short-range motion perception.* Vision Research. 25, 803-812.

Balboa R.M. and Grzywacz N.M. 2000a. *Occlusions and their relationship with the distribution of contrasts in natural images.* Vision Research. 40, 2661-2669.

Balboa R.M. and Grzywacz N.M. 2000b. *The role of early retinal lateral inhibition: More than maximizing luminance information.* Visual Neuroscience. 17, 77-89.

Balboa R.M., Tyles C.W. and Grzywacz N.M. 2001. *Occlusions contribute to scaling in natural images.* Vision Research. 41, 955-964.

Barlow H.B. 1957. *Purkinje shift and retinal noise.* Nature-London. 179, 225-256.

Barlow H.B. 1961. *Possible principles underlying the transformation of sensory messages.* in Sensory Communications. MIT Press, Cambridge, Mass. 217-234.

Barlow H.B. 1988. *Vision - the thermal limit to seeing.* Nature. 334, 296-297.

Barlow H.B. 2001. *Redundancy reduction revisted.* Network: Computation in Neural Systems. 12, 241-253.

Baylor D.A. 1987. *Photoreceptor signals and vision.* Investigative Ophtalmology and Visual Science, Supplement. 28, 34-49.

Baylor D.A., Nunn B.J. and Schnapf J.L. 1987. *Spectral sensitiviy of cones of the monkey Macaca Fascicularis.* Journal of Physiology-London. 390, 145-160.

Becken R. and Cervone D. 1983. *Selective looking and the noticing of unexpected events.* Memory and Cognition. 11, 601-608.

Bell A.J. and Sejnowski T.J. 1995. *An infromation-maximisation approach to blind separation and blind deconvolution.* Neural Computation. 7, 1129-1159.

Bell A.J. and Sejnowski T.J. 1997. *The 'independent components' of natural scenes are edge filters.* Vision Research. 37, 3327-3338.

Bennett A.T.D. and Cuthill I.C. 1994. *Ultraviolet vision in birds: what is its function?* Vision Research. 34, 1471.

Bennett P.J. and Banks M.S. 1987. *The orientation specificity of two visual after-effects.* Nature. 326, 873-876.

Benson P.J. 1994. *Morph transformations of the facial image.* Image and Vision Computing. 12, 691-696.

Biederman I. and Ju G. 1988. *Surface vs edge edge-based determinants of visual recognition.* Cognitive Psychology. 20, 38-64.

Biederman I., Mezzanotte R.J. and Rabinowitz J.C. 1982. *Scene perception: detecting and judging objects undergoing relational violations.* Cognitive Psychology. 14, 143-177.

Blakemore C. and Campbell F.W. 1969. *On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images.* Journal of Physiology-London. 203, 237-260.

Blakemore C. and Tobin E.A. 1972. *Lateral inhibition between orientation detectors in the cat's visual cortex.* Experimental Brain Research. 15, 439-440.

Bowmaker J.K. 1991. *Visual pigments, oil droplets and photoreceptors.* in The perception of colour. Vision and visual disfunction. J. R. Cronly-Dillon. 1991. The Macmillian press Ltd., London. 108-127.

Bowmaker J.K. and Dartnall H.J. 1980. *Visual pigments of rods and cones in a human retina.* Journal of Physiology. 298, 501-511.

Bowmaker J.K. and Mollon J.D. 1983. *Microspectrophotometric results of Old and New World primates.* in Colour vision: physiology and psychophysics. 1983. Academic Press INC., London. 57-68.

Boycott B.B. and Kolb H. 1973. *The horizontal cells in the rhesus monkey retina.* Journal of Comparative Neurology. 148, 115-140.

Braddick O. 1973. *A shrot-range process in apparent motion.* Vision Research. 14, 519-527.

Bradley A., Switkes E. and De Valois K.K. 1985. *Orientation and spatial frequency selectivity of adaptation to isoluminant color patterns.* Investigative Ophtalmology and Visual Science. 26 (suppl), 182.

Brady N. 1997. *Spatial scale interactions and image statistics.* Perception. 26, 1089-1100.

Brady N. and Field D.J. 1995. *What's constant in contrast constancy - the effects of scaling on the perceived contrast of bandpass patterns.* Vision Research. 35, 739-756.

Brady N. and Field D.J. 2000. *Local contrast in natural images: normalisation and coding efficiency.* Perception. 29, 1041-1055.

Brindley G.S. 1954. *The order of coincidence required for visual threshold.* Proceedings of the Physical Society, London B. 673-676.

Brindley G.S. and Lewin W.S. 1968. *The sensations produced by electrical stimulation of the visual cortex.* Journal of Physiology. 196, 479-493.

Buchsbaum G. and Gottschalk A. 1983. *Trichromacy, Opponent Colors Coding and Optimum Color Information-Transmission in the Retina.* Proceedings of the Royal Society of London Series B-Biological Sciences. 220, 89-113.

Burr D.C., Morrone M.C. and Ross J. 1994. *Selective supression of the magnocellular visual pathway during saccadic eye movements.* Nature. 371, 511-513.

Burton G.J. and Moorhead I.R. 1987. *Color and Spatial Structure in Natural Scenes.* Applied Optics. 26, 157-170.

Calkins D.J. 2001. *Seeing with S cones.* Progress in Retinal and Eye Research. 20, 255-287.

Campbell F.W. and Robson J.G. 1968. *Application of Fourier analysis to the visibility of gratings.* Journal of Physiology. 197, 551-566.

Campbell N.W. and Kulikowski J.J. 1966. *Orientational selectivity of the human visual system.* Journal of Physiology-London. 187, 437-445.

Campbell N.W., MacKeown W.P.J., Thomas B.T. and Troscianko T. 1997. *Interpreting image databases by region classification.* Pattern Recognition. 30, 555-563.

Carlson C.R. 1978. *Thresholds for perceived image sharpness.* Photographic Science and Engineering. 22, 69-71.

Carpenter R.H.S. 1988. *Eye movements and vision.* in Movements of the eyes. 2nd. Pion Ltd., London. 324-325.

Carpenter R.H.S. and Robson J.G. 1999. *Vision research : a practical guide to laboratory methods.* Oxford University Press 1999, Oxford.

Chacko L.W. 1948. *The laminar pattern of the lateral geniculate nucleous in the primate.* Journal of Neurology, Neurosurgery and Psychiatry. 11, 211-224.

Chiao C.C., Cronin T.W. and Osorio D. 2000a. *Color signals in natural scenes: characteristics of reflectance spectra and effects of natural illuminants.* Journal of the Optical Society of America A. 17, 218-224.

Chiao C.C., Vorobyev M., Cronin T.W. and Osorio D. 2000b. *Spectral tuning of dichromats to natural scenes.* Vision Research. 40, 3257-3271.

Christianson S.A., Loftus E.F., Hoffman H. and Loftus G.R. 1991. *Eye fixations and memory for emotional events.* Journal of Experimental Psychology: Learning, Memory and Cognition. 17, 693-701.

Chung S.T.L., Legge G.E. and Tjan B.S. 2002. *Spatial-frequency characteristics of letter identification in central and peripheral vision.* Vision Research. 42, 2137-2152.

Chung S.T.L., Mansfield J.S. and Legge G.E. 1998. *Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision.* Vision Research. 38, 2949-2962.

Cicerone C.M., Gowdy P.D. and Otake S. 1994. *Composition and arrangement of the cone mosaic in the living human eye.* Investigative Ophtalmology and Visual Science. 35 (suppl.), 1571.

Cicerone C.M. and Nerger J.L. 1989. *The relative numbers of long-wavelength-sensitive to middle-wavelength-sensitive cones in the human fovea centralis.* Vision Research. 29, 115-128.

Clark A.A., Troscianko T., Campbell N.W. and Thomas B.T. 2000. *A comparison between human and machine labelling of image regions.* Perception. 1127-1138.

Cleland B.G., Dubin M.W. and Levick W.R. 1971. *Sustained and transient neurones in the cat's retina and lateral geniculate nucleus.* Journal of Physiology. 217, 473-496.

Clement R. and Moorhead I. 2000. *An analogy between colour and spatial coding.* Perception. 29, 1447-1453.

Collet D. 1991. *Modelling binary data.* Chapman & Hall/CRC, New York.

Conway B.R. 2001. *Spatial structure of cone inputs to color cells in alert macaque primary visual cortex (V-1).* Journal of Neuroscience. 21, 2768-2783.

Coppola D.M., Purves H.R., McCoy A.N. and Purves D. 1998. *The distribution of oriented contours in the real world.* Proceedings of the National Academy of Sciences USA. 95, 4002-4006.

Cowey A. and Rolls E.T. 1974. *Human cortical magnification factor and its relation to visual acuity.* Experimental Brain Research. 21, 447-454.

Cowey A., Small M. and Ellis S. 1994. *Left visuo-spatial neglect can be worse in far than un near space.* Neurophysiologia. 32, 1059-1066.

Cronly-Dillon J.R. 1991. *Origins of invertebrate and vertebrate eyes.* in Evolution of the eye and the visual system. Vision and visual dysfunction. J. R. Cronly-Dillon. 1991. Macmillan, London. 15-51.

CRS 1995. *Palettes and LUT's.* in VSG user's guide - Version 4.02. Cambridge Research Systems Ltd., Kent, UK. 73-75.

Dan Y., Atick J.J. and Reid R.C. 1996. *Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory.* Journal of Neuroscience. 16, 3351-3362.

Daniel P.M. and Whitteridge D. 1961. *The representation of the visual field on the cerebral cortex in monkey.* Journal of Physiology. 159, 203-221.

Dartnall H.J., Bowmaker J.K. and Mollon J.D. 1983. *Human visual pigments: microspectrophotometric results from the eyes of seven persons.* Proceedings of the Royal Society of London Series B. 220, 115-130.

Daugman J.G. 1985. *Uncertainty relation for resolution in space, spatial-frequency and orientation optimized by two-dimensional visual cortical filters.* Journal of the Optical Society of America A. 2, 1160-1169.

Dawkins R. 1986. *The blind watchmaker.* Longman Scientific & Technical, Harlow.

De Monasterio F.M. and Gouras P. 1975. *Functional properties of ganglion cells of the rhesus monkey retina.* Journal of Physiology. 251, 167-195.

De Monasterio F.M., McCrane E.P., Newlander J.K. and Schein S. 1985. *Density profile of blue sensitivity cones along the horizontal meridian of macaque retina.* Investigative Ophtalmology and Visual Science. 26, 289-302.

De Valois K.K. 1977a. *Independence of black and white: phase-specific adaptation.* Vision Research. 17, 209-215.

De Valois K.K. 1977b. *Spatial frequency adaptation can enhance contrast sensitivity.* Vision Research. 17, 1057-1065.

De Valois K.K. and Switkes E. 1983. *Simultaneous masking interactions between chromatic and luminance gratings.* Journal of the Optical Society of America. 73, 11-18.

De Valois K.K. and Tootell R.B.H. 1983. *Spatial-frequency-specific inhibition in cat striate cortex cells.* Journal of Physiology-London. 336, 359-376.

De Valois R.L., Abramov I. and Thorell L.G. 1966. *Analysis of response patterns of LGN cells.* Journal of the Optical Society of America. 56, 966-977.

De Valois R.L., Albrecht D.G. and Thorell L.G. 1982a. *Spatial-frequency selectivity of cells in macaque visual cortex.* Vision Research. 22, 545-559.

De Valois R.L. and De Valois K.K. 1990. *Spatial Vision.* Oxford psychology series. D. E. Broadbent, J. L. McGaugh, N. J. Mackintosh, M. I. Posner, E. Tulving and L. Weiskrantz. Oxford University Press, Oxford, UK.

De Valois R.L., Yund E.W. and Hepler N. 1982b. *The orientation and direction selectivity of cells in macaque visual cortex.* Vision Research. 22, 531-544.

De Vries H.L. 1947. *The heredity of the relative numbers of red and green receptors in the human eye.* Genetica. 24, 199-212.

Deangelis G.C., Ohzawa I. and Freeman R.D. 1993. *Spatiotemporal Organization of Simple-Cell Receptive-Fields in the Cats Striate Cortex .1. General-Characteristics and Postnatal-Development.* Journal of Neurophysiology. 69, 1091-1117.

DeMarco P., Pokorny J. and Smith V.C. 1992. *Full spectrum cone sensitivity functions for X-chromosome linked anomalous trichromats.* Journal of the Optical Society of America A. 9, 1465-1476.

Derrington A.M., Krauskopf J. and Lennie P. 1984. *Chromatic Mechanisms in Lateral Geniculate-Nucleus of Macaque.* Journal of Physiology-London. 357, 241-265.

Derrington A.M. and Lennie P. 1984. *Spatial and Temporal Contrast Sensitivities of Neurons in Lateral Geniculate-Nucleus of Macaque.* Journal of Physiology-London. 357, 219-240.

Dobelle W.H., Turkel J., Henderson D.C. and Evans J.R. 1979. *Mapping the representation of the visual field by electrical stimulation of human visual cortex.* American Journal of Ophtalmology. 88, 727-735.

Dobson A.J. 1990. *Introduction to generalised linear methods.* Chapman & Hall, London.

Dominy N.J. and Lucas P.W. 2001. *Ecological importance of trichromatic vision to primates.* Nature. 410, 363-365.

Dong D.W. and Attick J.J. 1995. *Statistics of time-varying images.* Network: Computation in Neural Systems. 6, 345-358.

Dow B.M., Snyder A.Z., Vautin R.G. and Bauer R. 1981. *Magnification factor and receptive field size in foveal striate cortex of the monkey.* Experimental Brain Research. 44, 213-228.

Drasdo N. 1977. *The neural representation of visual space.* Nature. 266, 554-556.

Drasdo N. 1991. *Neural substrates and threshold gradients of peripheral vision.* in Limits of vision. Vision and visual dysfunction. J. R. Cronly-Dillon. 1991. Macmillan Press Ltd., London. 251-276.

Drasdo N. and Thompson C.A. 1989. *Do visibility and colour recognition isopters relate to the distribution of Pa and Pb ganglion cells in the human retina?* Ophthalmic and Physiological Optics. 9, 447-450.

D'Zimura M. and Lennie P. 1986. *Mechanisms of colour constancy.* Journal of the Optical Society of America. 3, 1662-1672.

Edwards A.W.F. 1972. *Likelihood : an account of the statistical concept of likelihood and its application to scientific inference.* Cambridge University Press, London.

Enroth-Cugell C., Hertz G. and Lennie P. 1977. *Cone signals in the cat's retina.* Journal of Physiology. 269, 273-296.

Enroth-Cugell C. and Pinto L.H. 1970. *Algebraic summation of centre and surround inputs to retinal ganglion cells.* Nature. 226, 458-459.

Enroth-Cugell C. and Robson J.G. 1966. *The contrast sensitivity of retinal ganglion cells of the cat.* Journal of Physiology. 187, 517-522.

Eskew R.T. and Boynton R.M. 1987. *Effects of field area and and configuration on chromatic and border discriminations.* Vision Research. 27, 1835-1844.

Fasick J.I. and Robinson P.R. 2000. *Spectral-tuning mechanisms of marine mammal rhodopsins and correlations with foraging depth.* Visual Neuroscience. 17, 781-788.

Fendick M. and Westheimer G. 1983. *Effects of practice and the separation of test targets on foveal and peripheral stereoacuity.* Vision Research. 23, 145-150.

Field D.J. 1987. *Relations between the statistics of natural scenes and the response properties of cortical-cells.* Journal of the Optical Society of America A. 4, 2379-2394.

Field D.J. 1994. *What is the goal of sensory coding?* Neural Computation. 6, 559 - 601.

Field D.J. and Brady N. 1997. *Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes.* Vision Research. 37, 3367-3383.

Findlay J.M. and Gilchrist I.D. 2003. *Active vision: the psychology of looking and seeing.* Oxford University Press, London.

Finney D.J. 1971. *Probit analysis.* 3rd. Cambridge University Press, Cambridge.

Foley J.M. 1994. *Human luminance pattern-vision mechanisms: masking experiments require a new model.* Journal of the Optical Society of America A. 11, 1710.

Foley J.M. and Legge G.E. 1981. *Contrast Detection and near-Threshold Discrimination in Human-Vision.* Vision Research. 21, 1041-1053.

Frome F.S., Buck S.L. and Boynton R.M. 1981. *Visibility of borders: separate and combined effects of colour differences, luminance contrast and luminance levels.* Journal of the Optical Society of America. 71, 145-150.

Fyfe C. and Baddeley R. 1995. *Finding compact and sparse-distributed representations of visual images.* Network: Computation in Neural Systems. 6, 333-334.

Gabor D. 1946. *Theory of communication.* Journal of the Institution of Electrical Engineers. 93, 429-459.

Geisler W.S., Perry J.S., Super B.J. and Gallogly D.P. 2001. *Edge co-occurrence in natural images predicts contour grouping performance.* Vision Research. 41, 711-724.

Georgeson M.A. and Sullivan G.D. 1975. *Contrast constancy: deblurring in human vision by spatial frequency channels.* Journal of Physiology-London. 252, 627-656.

Goldsmith T.H. 1991. *Evolution of visual pigments and colour vision.* in Evolution of the eye and visual system. Vision and visual dysfunction. J. R. Cronly-Dillon. Macmillan, London.

Gonzalez R.C. and Woods R.E. 1992. *Digital image processing.* Addison-Wesley, Reading, Mass. ; Wokingham.

Gorea A. and Sagi D. 2001. *Disentangling signal from noise in visual contrast discrimination.* Nature Neuroscience. 4, 1146-1150.

Gould S.J. and Lewontin R.C. 1979. *The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme.* Proceedings of the Royal Society of London Series B. 205, 581-598.

Gouras P. 1968. *Identification of cone mechanisms in monkey ganglion cells.* Journal of Physiology. 199, 533-547.

Gouras P. 1971. *The function of the midget cell system in primate colour vision.* Vision Research, Suppl. 3, 397-410.

Gouras P. 1991. *Precortical physiology of colour vision.* in The perception of colour. Vision and visual disfunction. J. R. Cronly-Dillon. 1991. The Macmillian press Ltd., London. 163-178.

Graham N., Robson J.G. and Nachmias J. 1978. *Grating summation in fovea and periphery.* Vision Research. 18, 815-825.

Grassmann H. 1854. *On the theory of compound colours.* Philosophy Magazine (Ser. 4.). 7, 254-264.

Gregory R.L. 1991. *Origins of eyes - with speculations on scanning eyes.* in Evolution of the eye and the visual system. Vision and visual dysfunction. J. R. Cronly-Dillon. 1991. Macmillan, London. 52-59.

Gregory R.L. 1998. *Eye and brain : the psychology of seeing.* 5th. Oxford University Press, Oxford.

Gunther K.L. and Dobkins K.R. 2002. *Individual differences in chromatic (red/green) contrast sensitivity are constrained by the relative number of L- versus M-cones in the eye.* Vision Research. 42, 1367-1378.

Guth S.L., Massof R.W. and Benzschawel T. 1980. *Vector model for normal and dichromatic vision.* Journal of the Optical Society of America. 70, 197-212.

Hampton D.R. and Kertesz A.E. 1983a. *The extent of Panum's area and the human cortical magnification factor.* Perception. 12, 161-165.

Hampton D.R. and Kertesz A.E. 1983b. *Fusional response to local peripheral stimulation.* Journal of the Optical Society of America. 73, 7-10.

Hancock P.J.B., Baddeley R.J. and Smith L.S. 1992. *The principal components of natural images.* Network-Computation in Neural Systems. 3, 61-70.

Harvey Jr. L.O. 1986. *Efficient estimationof sensory thresholds.* Behaviour Research Methods, Instruments and Computers. 18, 623-632.

Harwerth R.S. and Levi D.M. 1978. *Reaction time as a measure of suprathreshold grating detection.* Vision Research. 18, 1579-1586.

Hassler R. 1966. *Comparative anatomy of the central visual system in day -and night- active primates.* in Evolution of the forebrain. Georg Thieme Verlag, Stuttgart.

Heeger D.J. 1992. *Normalization of cells in cat striate cortex.* Visual Neuroscience. 9, 181-197.

Hendry S.H. and Calkins D.J. 1998. *Neuronal chemistry and functional organisation in primate visual system.* Trends in Neuroscience. 21, 344-349.

Hendry S.H. and Reid R.C. 2000. *The koniocellular pathway in primate vision.* Annual Review of Neuroscience. 23, 127-153.

Hess R.F., Dakin S.C., Kapoor N. and Tewfik M. 2000. *Contour interaction in fovea and periphery.* Journal of the Optical Society of America A. 17, 1516-1524.

Hess R.F. and Field J.F. 1993. *Is the increased spatial uncertainty in the normal periphery due to spatial undersampling or uncalibrated disarray.* Vision Research. 33, 2663-2670.

Holland J.H. 1992. *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence.* Complex adaptive systems. 1st MIT Press. MIT Press, Cambridge, Mass. ; London.

Hong G., Luo M.R. and Rhodes P.A. 2000. *A study of digital camera colorimetric characterization based on polynomial modeling.* Color Research and Application. 26, 76-84.

Horridge G.A. 1991. *Evolution of visual processing.* in Evolution of the eye and visual system. Vision and visual dysfunction. J. R. Cronly-Dillon. Macmillan, London. 229-270.

Howell D.C. 1997. *Statistical methods for psychology.* 4th. Duxbury Press, Belmont, CA.

Hubel D.H. and Livingstone M.S. 1987. *Segregation of form, color, and stereopsis in primate area 18.* Journal of Neuroscience. 7, 3378-3415.

Hubel D.H. and Wiesel T.N. 1959. *Receptive fields of single neurons in the cat's striate cortex.* Journal of Physiology. 148, 574-591.

Hubel D.H. and Wiesel T.N. 1960. *Receptive fields of optic nerves in the spider monkey.* Journal of Physiology. 154, 572-580.

Hubel D.H. and Wiesel T.N. 1962. *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.* Journal of Physiology. 160, 106-154.

Hubel D.H. and Wiesel T.N. 1968. *Receptive fields and functional architecture of monkey striate cortex.* Journal of Physiology. 195, 215-243.

Hubel D.H. and Wiesel T.N. 1974. *Uniformity of monkey striate cortex: a paralell relationship between field size, scatter and magnification factor.* Journal of Comparative Neuroscience. 158, 295-306.

Hubel D.H. and Wiesel T.N. 1977. *Functional architecture of macaque visual cortex.* Proceedings of the Royal Society of London Series B. 198, 1-59.

Hurlbert A. 1999. *Is colour constancy real?* Current Biology. 9, R558-R561.

Ingling C.R. and Martinez Uriegas E. 1983. *The Relationship Between Spectral Sensitivity and Spatial Sensitivity For the Primate R-G X-Channel.* Vision Research. 23, 1495-1500.

Ingling C.R. and Tsou B.H.P. 1988. *Spectral Sensitivity For Flicker and Acuity Criteria.* Journal of the Optical Society of America A. 5, 1374-1378.

Ingling C.R.J. and Martinez E. 1983. *The spatiochromatic signal of the r-g channel.* in Colour vision - physiology and psychophysics. 1983. Academic Press, London. 433-453.

Irvin G.E., Norton T.T., Sesma M.A. and Casagrande V.A. 1986. *W-Like Response Properties of Interlaminar Zone Cells in the Lateral Geniculate-Nucleus of a Primate (Galago-Crassicaudatus).* Brain Research. 362, 254-270.

Jennings J.A.M. and Charman W.N. 1981. *Off-axis image quality in the human eye.* Vision Research. 21, 445-455.

Jones J.P. and Palmer L.A. 1978a. *An evaluation of the two-dimensional Gabor filter model of receptive fields in cat striate cortex.* Journal of Neurophysiology. 58, 1233-1258.

Jones J.P. and Palmer L.A. 1978b. *The two-dimensional spatial structure of simple receptor fields in cat striate cortex.* Journal of Neurophysiology. 58, 1187-1211.

Kaplan E. and Shapley R.M. 1986. *The primate retina contains two types of ganglion cells, with high and low contrast sensitivity.* Proceedings of the National Academy of Sciences USA. 83, 2755-2757.

Kelly D.H. 1984. *Retinal inhomogenity. I.Spatiotemporal contrast sensitivity.* Journal of the Optical Society of America A. 7, 1113-1123.

Kersten D. 1987. *Predictability and Redundancy of Natural Images.* Journal of the Optical Society of America A. 4, 2395-2400.

Klein S.A. and Levi D.M. 1987. *Position sense of the peripheral retina.* Journal of the Optical Society of America A. 4, 1543-1553.

Knau H., Jagle H. and Sharpe L.T. 2001. *L/M cone ratios as a function of retinal eccentricity.* Color Research and Application. 26, S128-S132.

Knill D.C., Field D. and Kersten D. 1990. *Human Discrimination of Fractal Images.* Journal of the Optical Society of America A. 7, 1113-1123.

Krauskopf J., Williams D.R. and Heeley D.W. 1982. *Cardinal directions of colour space.* Vision Research. 22, 1123-1131.

Krebs W. and Krebs I.P. 1989. *Quantitative morphology of the central fovea in the primate retina.* American Journal of Anatomy. 184, 225-236.

Kremers J., Scholl H.P.N., Knau H., Berendschot T.T.J.M., Usui T. and Sharpe L.T. 2000. *L/M cone ratios in human trichromats assessed by psychophysics, electroretinography, and retinal densitometry.* Journal of the Optical Society of America A. 17, 517-526.

Kretzmer E.R. 1952. *Statistics of television signals.* Bell Systems Technical Journal. 31, 751-763.

Kuffler S.W. 1953. *Discharge patterns and functional organisation in the mammalian retina.* Journal of Neurophysiology. 16, 37-68.

Kulikowski J.J., Marcelja S. and Bishop P.O. 1982. *Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex.* Biological Cybernetics. 43, 187-198.

Kulikowski J.J. and Tolhurst D.J. 1972. *Psychophysical evidence for for sustained and transient mechanisms in human vision.* Journal of Physiology-London. 232, 149-163.

Kulikowski J.J. and Walsh V. 1993. *Colour vision: Isolating mechanisms in the overlapping streams.* Progress in Brian Research. 95, 417-426.

Land F.L. 1991. *Optics of the eyes of the animal kingdom.* in Evolution of the eye and the visual system. Vision and visual dysfunction. J. R. Cronly-Dillon. 1991. Macmillan, London. 118-135.

Langer M.S. 2000. *Large-scale failures of f^{α} scaling in natural image spectra.* Journal of the Optical Society of America A. 17, 28-33.

Latham K. and Whitaker D. 1996. *A comparison of word recognition and reading performance in foveal and peripheral vision.* Vision Research. 36, 2665-2674.

Laughlin S.B. 1981. *A simple coding procedure enhances a neuron's information capacity.* Zeitschrift für Naturforschung C. 36, 910-912.

Laughlin S.B. 1983. *Matching coding to scenes to enhance efficiency*. in Physical and biological processing of images : proceedings of an international symposium organized by the Rank Prize Funds, London, England, 27-29 September, 1982. Springer-Verlag, Berlin ; New York. 42-52.

Lauritzen J.S., Tolhurst D.J. and Pelah A. 1999. *A psychophysical test of a physiological model of the contrast in natural scenes.* Journal of Physiology. 515P, 102P.

Legge G.E. 1979. *Spatial frequency masking in human vision: binocular interactions.* Journal of the Optical Society of America. 69, 838-847.

Legge G.E. 1981. *A Power Law For Contrast Discrimination.* Vision Research. 21, 457-467.

Legge G.E. and Foley J.M. 1980. *Contrast masking in human vision.* Journal of the Optical Society of America. 70, 1456-1471.

Lennie P., Krauskopf J. and Sclar G. 1990. *Chromatic mechanisms in the striate cortex of the macaque.* Journal of Neuroscience. 10, 649-669.

Levi D.M. and Klein S.A. 1986. *Sampling in spatial vision.* Nature. 320, 360-362.

Levi D.M., Klein S.A. and Aistebaomo A.P. 1985. *Vernier acuity, crowding and cortical magnification.* Vision Research. 25, 963-977.

Levi D.M., Klein S.A. and Aitseboamo P. 1984. *Detection and discrimination of the direction of motion in central and peripheral vision of normal and amblyopic observers.* Vision Research. 24, 789-800.

Littman D. and Becklen R. 1976. *Selective looking with minimal eye movements.* Perception and Psychophysics. 20, 77-79.

Loftus G.R. 1972. *Eye fixations and recognition memory for pictures.* Cognitive Psychology. 3, 525-551.

Loftus G.R. and Mackworth N.H. 1978. *Cognitive determinants of fixation location during picture viewing.* Journal of Experimental Psychology: Human Perception and Performance. 4, 565-572.

Ludvig N., Tang H.M., Eichenbaum H. and Gohil B.C. 2003. *Spatial memory performance of freely-moving squirrel monkeys.* Behavioural brain research. 140, 175-183.

Lund J.S., Lund R.D., Hendrickson A.E., Bunt A.H. and Fuchs A.F. 1975. *The origin of efferent pathways from primary visual cortex, area 17, of the macaque monkey as shown by retrograde tranport of horseradish peroxidase.* Journal of Comparative Neurology. 164, 287-304.

Lythgoe J.N. 1984. *Visual pigments and environmental light.* Vision Research. 24, 1539-1550.

Lythgoe J.N. 1991. *Evolution of visual behaviour.* in Evolution of the eye and visual system. Vision and visual dysfunction. J. Cronly-Dillon. The Macmillan Press Ltd., London. 3-14.

Lythgoe J.N. and Partridge J.C. 1989. *Visual pigments and the acquisition of visual information.* Journal of Experimental Biology. 146, 1-20.

Lythgoe J.N. and Partridge J.C. 1991. *The modelling of optimal visual pigments of dichromatic teleosts in green coastal waters.* Vision Research. 31, 361-371.

Mack A., Tang B., Tuma R. and Kahn S. 1992. *Perceptual organization and attention.* Cognitive Psychology. 24, 475-501.

Mackworth N.H. and Morandi A.J. 1967. *The gaze selects informative details within pictures.* Perception and Psychophysics. 2, 547-552.

MacNichol E.F., Levine J.S., Mansfield R.J.W., Lipetz L.E. and Collins B.A. 1983. *Microspectrophotometry of visual pigments in primate photoreceptors.* in Colour vision: physiology and psychophysics. 1983. Academic Press INC., London. 14-38.

Makela P., Whitaker D. and Rovamo J. 1993. *Modeling of Orientation Discrimination across the Visual-Field.* Vision Research. 33, 723-730.

Maloney L.T. 1986. *Evaluation of linear models of surface spectral reflectance with small numbers of parameters.* Journal of the Optical Society of America A. 3, 1673-1683.

Marr D. 1982. *Vision : a computational investigation into the human representation and processing of visual information.* W.H. Freeman, San Francisco.

Martin P.R., White A.J.R., Goodchild A.K., Wilder H.D. and Sefton A.E. 1997. *Short communication: evidence that blue-on cells are part of a third geniculocortical pathway in primates.* European Journal of Neuroscience. 9, 1336-1541.

Maynard Smith J. 1978. *Optimization theory in evolution.* Annual Review of Ecology and Systematics. 9, 31-57.

Melmoth D.R., Kukkonen H.T., Makela P.K. and Rovamo J. 2000. *The effect of contrast and size scaling on face perception in foveal and extrafoveal vision.* Investigative Ophtalmology and Visual Science. 41, 2811-2819.

Merbs S.L. and Nathans J. 1992. *Absorption spectra of human cone photopigments.* Nature. 356, 433-435.

Merigan W.H., Bryne C.E. and Maunsell J.H.R. 1991a. *Does primate motion perception depend on the magnocellular pathway?* Journal of Neuroscience. 11, 3422-3429.

Merigan W.H., Katz L.M. and Maunsell J.H.R. 1991b. *The effects of parvocellular lateral geniculate nucleus lesion on the acuity and contrast sensitivity of macaque.* Journal of Neuroscience. 11, 994-1001.

Miyahara E., Pokorny J., Smith V.C., Baron R. and Baron E. 1998. *Color vision in two observers with highly biased LWS/MWS cone ratios.* Vision Research. 38, 601-612.

Mollon J.D. 1989. *"Tho she kneeld in that place where they grew..." the uses and origins of primate colour vision.* Journal of Experimental Biology. 146, 21-38.

Mollon J.D. 1991. *Uses and evolutionary origins of primate colour vision.* in Evolution of the eye and the visual system. Vision and visual dysfunction. J. Cronly-Dillon. 1991. Macmillan Press Ltd., London. 306-319.

Mollon J.D. and Regan B.C. 1999. *The spectral distribution of primate cones and of the macular pigment: Matched to properties of the world?* Journal of Optical Technology. 66, 847-852.

Moore C.M. and Egeth H. 1997. *Perception without attention:evidence of grouping under conditions of inattention.* Journal of Experimental Psychology: Human Perception and Performance. 23, 563-581.

Morgan C.T. 1965. *Physiological Psychology*. McGraw-Hill series in Psychology. H. F. Harlow. third. Kogakusha Co, Ltd, Tokio, Japan.

Movshon J.A., Thompson I.D. and Tolhurst D.J. 1978a. *Spatial and temporal contrast sensitivity of neurons in areas 17 and 18 of the cat's visual cortex*. Journal of Physiology. 283, 101-120.

Movshon J.A., Thompson I.D. and Tolhurst D.J. 1978b. *Spatial summation in the receptive fields of simple cells in the cat's striate cortex*. Journal of Physiology. 283, 53-77.

Mullen K.T. 1985. *The contrast sensitivity of human color vision to red-green and blue-yellow chromatic gratings*. Journal of Physiology-London. 359, 381-400.

Mullen K.T. and Kingdom F.A.A. 1991. *Colour contrast in form perception*. in The perception of colour. Vision and visual dysfunction. J. Cronly-Dillon. 1991. Macmillan Press Ltd., London. 198-217.

Nachmias J. and Sansbury R.V. 1974. *Grating contrast discrimination may be better than detection.* Vision Research. 14, 1039-1042.

Nagle M.G. and Osorio D. 1993. *The tuning of human photopigments may minimize red-green chromatic signals in natural conditions.* Proceedings of the Royal Society of London Series B-Biological Sciences. 252, 209-213.

Nascimento S.M.C., Ferreira F.P. and Foster D.H. 2002. *Statistics of spatial cone-excitation ratios in natural scenes*. Journal of the Optical Society of America A. 19, 1484-1490.

Neumeyer C. 1991. *Evolution of colour vision*. in Evolution of the eye and visual system. Vision and visual dysfunction. J. R. Cronly-Dillon. 1991. Macmillan Press Ltd., London. 284-305.

Newby E.A. and Rock I. 1998. *Inattentional blindness as a function of proximity to the focus of attention.* Perception. 27, 1025-1040.

Norman R.A. and Werblin F.S. 1974. *Control of retinal sensitivity. I Light and dark adaptation of vertebrate rods and cones.* Journal of General Physiology. 63,

Norton T.T. and Casagrande V.A. 1982. *Laminar Organization of Receptive-Field Properties in Lateral Geniculate-Nucleus of Bush Baby (Galago-Crassicaudatus).* Journal of Neurophysiology. 47, 715-741.

Norton T.T., Casagrande V.A., Irvin G.E., Sesma M.A. and Petry H.M. 1988. *Contrast-Sensitivity Functions of W-Like, X-Like, and Y-Like Relay Cells in the Lateral Geniculate-Nucleus of Bush Baby, Galago-Crassicaudatus.* Journal of Neurophysiology. 59, 1639-1656.

Ocarroll D.C., Bidwell N.J., Laughlin S.B. and Warrant E.J. 1996. *Insect motion detectors matched to visual ecology.* Nature. 382, 63-66.

Olshausen B.A. and Field D.J. 1996. *Emergence of simple-cell receptive field properties by learning a sparse code for natural images.* Nature -London. 381, 607-609.

Olshausen B.A. and Field D.J. 1997. *Sparse coding with an overcomplete basis set: A strategy employed by V1?* Vision Research. 37, 3311-3325.

Osorio D. and Vorobyev M. 1996. *Colour vision as an adaptation to frugivory in primates.* Proceedings of the Royal Society of London Series B. 263, 593-599.

Paradiso M.A. and Carney T. 1988. *Orientation discrimination as a function of stimulus eccentricity and size: nasal/temporal retinal assymetry.* Vision Research. 28, 867-874.

Paradiso M.A., Carney T. and Freeman R.D. 1989. *Cortical processing of hyperacuity tasks.* Vision Research. 29, 247-254.

Parker G.A. and Maynard Smith J. 1990. *Optimality theory in evolutionary biology.* Nature. 348, 27-33.

Párraga C.A., Brelstaff G., Troscianko T. and Moorhead I.R. 1998a. *Color and luminance information in natural scenes.* Journal of the Optical Society of America A. 15, 563-569.

Párraga C.A., Kennedy G.M. and Tolhurst D.J. 1998b. *How amblyopic vision uses phase and amplitude information in natural images.* Perception. 27, 8.

Párraga C.A. and Tolhurst D.J. 2000. *The effect of contrast randomisation on the discrimination of changes in the slopes of the amplitude spectra of natural scenes.* Perception. 29, 1101-1116.

Párraga C.A., Tolhurst D.J. and Troscianko T. 1998c. *The human visual system is optimised for discriminating the spatial structure of scenes with natural second-order statistics.* Journal of Physiology. 515P, 102P.

Párraga C.A., Tolhurst D.J. and Troscianko T. 1998d. *Natural image statistics and visual processing. Are they matched?* Perception. 25 Suppl., 152-153.

Párraga C.A., Tolhurst D.J. and Troscianko T. 1999. *A computational model predicts discrimination thresholds for morphed objects in natural scenes.* Perception. 28 Suppl., 127b.

Párraga C.A., Troscianko T. and Tolhurst D.J. 2000a. *The human visual system is optimised for processing the spatial information in natural visual images.* Current Biology. 10, 35-38.

Párraga C.A., Troscianko T. and Tolhurst D.J. 2002. *Spatiochromatic Properties of Natural Images and Human Vision.* Current Biology. 12, 483-487.

Párraga C.A., Troscianko T., Tolhurst D.J. and Gilchrist I.D. 2000b. *Discrimination thresholds for morphed objects in peripheral vision.* Perception. 29, 58-58.

Pasternak T. and Merigan W.H. 1981. *The luminance dependence of spatial vision in the cat.* Vision Research. 21, 1333-1340.

Peli E. 1990. *Contrast in Complex Images.* Journal of the Optical Society of America A. 7, 2032-2040.

Peli E. 1996. *Test of a model of foveal vision by using simulations.* Journal of the Optical Society of America A. 13, 1131-1138.

Peli E. and Geri G.A. 2001. *Discrimination of wide-field images as a test of a peripheral-vision model.* Journal of the Optical Society of America A. 18, 294-301.

Pelli D.G. 1979. *The effects of noise masking and contrast adaptation on contrast detection, contrast discrimination and apparent contrast.* Investigative Ophtalmology and Visual Science. 18, 59.

Pelli D.G. and Zhang L. 1991. *Accurate control of contast on microcomputer displays.* Vision Research. 31, 1337-1350.

Petrov Y. and Zhaoping L. 2003. *Local correlations, information redundancy and sufficient pixel depth in natural images.* Journal of the Optical Society of America A. 20, 56-66.

Piotrowski L.N. and Campbell F.W. 1982. *A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase.* Perception. 11, 337-346.

Pointer J.S. 1986. *The cortical magnification factor and photopic vision.* Biological Reviews. 61, 97-119.

Pollehn H. and Roehring H. 1970. *Effect of noise on the modulation trasnfer function of the visual channel.* Journal of the Optical Society of America. 60, 842-848.

Press W.H., Flannery B.P., Tekulosky S.A. and Vetterling W.T. 1986. *Numerical recipes.* Cambridge University Press, New York.

Press W.H., Teukolsky S.A., Vetterling W.T. and Flannery B.P. 1992. *Numerical recipes in C : the art of scientific computing.* 2nd. Cambridge University Press, New York.

Quick R.F. 1974. *A vector magnitude model of contrast detection.* Kybernetik. 16, 65-67.

Rayner K. 1998. *Eye movements in rading and information processing: 20 years of research.* Psychological Bulletin. 124, 372-422.

Regan B.C., Julliot C., Simmen B., Vienot F., Charles-Dominique P. and Mollon J.D. 1998. *Frugivory and colour vision in Alouatta Seniculus, a trichromatic platyrrhine monkey.* Vision Research. 38, 3321-3328.

Regan B.C., Julliot C., Simmen B., Vienot F., Charles-Dominique P. and Mollon J.D. 2001. *Fruits, foliage and the evolution of primate colour vision.* Philosophical Transactions-Royal Society of London Series B Biological Sciences. 229-284.

Regan B.C., Vienot F., CharlesDominique P.C., Pefferkorn S., Simmen B., Julliot C. and Mollon J.D. 1996. *The colour signals that fruits present to primates.* Investigative Ophthalmology and Visual Science. 37, 2997.

Reinhard E., Ashikhmin M., Gooch B. and Shirley P. 2001. *Color transfer between images.* IEEE - Computer Graphics and Applications. 0272-1716/01, 34-41.

Rentschler I. and Treutwein B. 1985. *Loss of spatial phase relationships in extrafoveal vision.* Nature. 313, 308-310.

Ringach D.L. 2002. *Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex.* Journal of Neurophysiology. 88, 455-463.

Rizzolatti G., Mattelli M. and Pavesi G. 1983. *Deficits in attention and movement following the removal of postarcuate (area 6) and prearcuate (area 8) cortex in macaque monkeys.* brain. 106, 655-673.

Robson J.G. 1980. *Neural images: the physiological basis of spatial vision*. in Visual coding and adaptability. L. Erlbaum Associates, Hillsdale, N.J. 177-214.

Robson J.G. and Graham N. 1981. *Probability summation and regional vairation in contrast sensitivity across the visual field.* Vision Research. 21, 409-418.

Rodieck R.W. 1965. *Quantitative analysis of cat retinal ganglion cell responses to visual stimuli.* Vision Research. 5, 583-601.

Rohaly A.M., Ahumada A.J. and Watson A.B. 1995. *A Comparison of Image-Quality Models and Metrics Predicting Object Detection.* SID International Symposium Digest of Technical Papers. 26, 45-48.

Rohaly A.M., Ahumada A.J. and Watson A.B. 1997. *Object detection in natural backgrounds predicted by discrimination performance and models.* Vision Research. 37, 3225-3235.

Roorda A. and Williams D.R. 1999. *The arrangement of the three cone classes in the living human eye.* Nature. 397, 520-522.

Rovamo J., Makela P., Nasanen R. and Whitaker D. 1997. *Detection of geometric image distortions at various eccentricities.* Investigative Ophtalmology and Visual Science. 38, 1029-1039.

Rovamo J. and Virsu V. 1979. *An estimation and application of the human cortical magnification factor.* Experimental Brain Research. 37, 495-510.

Rovamo J., Virsu V. and Näsänen R. 1978. *Cortical magnification factor predicts the photopical contrast sensitivity of peripheral vision.* Nature. 271, 54-56.

Rubin N. and Hua X.L. 1998. *Perceiving occluded objects under conditions of inattention.* Investigative Ophtalmology and Visual Science. 39, S1113.

Ruderman D.L. 1995. *The statistics of natural images.* Network: Computation in Neural Systems. 6, 105.

Ruderman D.L. 1997. *Origins of scaling in natural images.* Vision Research. 37, 3385-3398.

Ruderman D.L., Cronin T.W. and Chiao C.C. 1998. *Statistics of cone responses to natural images: implications for visual coding.* Journal of the Optical Society of America A. 15, 2036-2045.

Rushton W.A.H. 1962. *Visual pigments in man.* Scientific American. 207, 120-132.

Rushton W.A.H. 1964. *Prentice lecture: Colour blindness and cone pigments.* American Journal of Optometry and Archives of American Academy of Optometry. 41, 265-282.

Sachs M.B., Nachmias J. and Robson J.G. 1971. *Spatial-frequency channels in human vision.* Journal of the Optical Society of America A. 61, 1176-1186.

Schade O.H. 1956. *Optical and photoelectrical analog of the eye.* Journal of the Optical Society of America. 46, 721-739.

Schien S.J. 1988. *Anatomy of macaque fovea and spatial density of neurons in foveal representation.* Journal of Comparative Neurology. 269, 479-505.

Schiller P.H., Logothetis N.K. and Charles E.R. 1989. *The function of the color-opponent (C-O) and broad-band (B-B) channels in perception at isoluminance.* Investigative Ophtalmology and Visual Science, Supplement. 30, 232.

Seger J. and Stubblefield J.W. 1996. *Optimization and adaptation.* in Adaptation. Academic Press, San Diego, CA. 93-123.

Seidemann E., Poirson A.B., Wandell B.A. and Newsome W.T. 1999. *Color signal in area MT of the macaque monkey.* Neuron. 24, 911-917.

Shannon C.E. and Weaver W. 1949. *The mathematical theory of communication.* University of Illinois Press, Urbana [Ill.] ; London.

Shapley R., Kaplan E. and Soodak R. 1981. *Spatial summation and contrast sensitivity of X and Y cells in the lateral geniculate nucleous of the macaque.* Nature. 292, 544-545.

Shapley R.M. and Brodie S. 1993. *Responses of human ERG to rapid color exchange: implications for M/L cone ration.* Investigative Ophtalmology and Visual Science. 34 (Suppl.), 911.

Shepherd J.W., Davies G.M. and Ellis H.D. 1981. *Studies of cue saliency.* in Perceiving and remembering faces. Academic Press, London. 105-131.

Shouval H., Intrator N. and Cooper L.N. 1997. *BCM network develops orientation selectivity and ocular dominance in natural scene.* Vision Research. 37, 3339-3342.

Simons D.J. and Chabris C.F. 1999. *Gorillas in our midst: sustained inattentional blindness for dynamic events.* Perception. 28, 1059-1074.

Smith V.C. and Pokorny J. 1972. *Spectral sensitivity of color-blind observers and the cone photopigments.* Vision Research. 12, 2059-2071.

Smith V.C. and Pokorny J. 1975. *Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm.* Vision Research. 15, 161-171.

Solomon J.A. and Pelli D.G. 1994. *The visual filter mediating letter identification.* Nature. 369, 395-397.

Srinivasan M.V., Laughlin S.B. and Dubs A. 1982. *Predictive Coding - a fresh view of inhibition in the retina.* Proceedings of the Royal Society of London Series B-Biological Sciences. 216, 427-459.

Stephenson C. and Braddick O. 1983. *Discrimination of relative phase in fovea and periphery.* Investigative Ophthalmolology and Visual Science (Suppl.). 24, 146.

Stiles W.S., Wyszecki G. and Ohta N. 1977. *Counting metameric object-colour stimuli using frequency-limited spectral reflectance functions.* Journal of the Optical Society of America. 67, 779-784.

Stroemeyer C.F., Cole G.R. and Kronauer R.E. 1985. *Second-site adaptation in the red-green chromatic pathways.* Vision Research. 25, 219-237.

Stroemeyer C.F., Kronauer R.E. and Cole G.R. 1983. *Adaptive mechanisms controlling sensitivity to red-green chromatic flashes.* in Colour Vision: Physiology and Psychophysics. Academic Press, London.

Stromeyer C.F. and Julesz B. 1972. *Spatial frequency masking in vision: critical bands and spread of masking.* Journal of the Optical Society of America. 62, 1221-1232.

Sumner P. and Mollon J.D. 2000a. *Catarrhine photopigments are optimized for detecting targets against a foliage background.* Journal of Experimental Biology. 203, 1963-1986.

Sumner P. and Mollon J.D. 2000b. *Chromaticity as a signal of ripeness in fruits taken by primates.* Journal of Experimental Biology. 203, 1987-2000.

Swanson W.H. and Birch E.E. 1992. *Extracting thresholds from noisy psychophysical data.* Perception and Psychophysics. 51, 409-422.

Tadmor Y. and Tolhurst D.J. 1989. *The effect of threshold on the relationship between the receptive-field profile and the spatial-frequency tuning curve in simple cells of the cat striate cortex.* Visual Neuroscience. 3, 445-454.

Tadmor Y. and Tolhurst D.J. 1993. *Both the phase and the amplitude spectrum may determine the appearance of natural images.* Vision Research. 33, 141-145.

Tadmor Y. and Tolhurst D.J. 1994. *Discrimination of changes in the second-order statistics of natural and synthetic-images.* Vision Research. 34, 541-554.

Tadmor Y. and Tolhurst D.J. 2000. *Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes.* Vision Research. 40, 3145 - 3157.

Tansley B.W. and Boynton R.M. 1976. *A line, not a space, represents visual distinctnessof borders formed by different colours.* Science. 191, 954-957.

Teo P.C. and Heeger D.J. 1995. *A general mechanistic model of spatial pattern detection.* Investigative Ophtalmology and Visual Science, Supplement. 36, 438.

Thompson P. 1980. *Margareth Thatcher: a new illusion.* Perception. 9, 483-484.

Thomson M.G.A. 1999a. *Higher-order structure in natural scenes.* Journal of the Optical Society of America A. 16, 1549-1553.

Thomson M.G.A. 1999b. *Visual coding and the phase structure of natural scenes.* Network-Computation in Neural Systems. 10, 123-132.

Thomson M.G.A. 2001. *Sensory coding and the second spectra of natural signals.* Physical Review Letters. 83, 2901-2904.

Thomson M.G.A. and Foster D.H. 1997. *Role of second- and third-order statistics in the discriminability of natural images.* Journal of the Optical Society of America A. 14, 2081-2090.

Thorell L.G., De Valois R.L. and Albretch D.G. 1984. *Spatial mapping of monkey V1 cells with pure color and luminance stimuli.* Vision Research. 24, 751-769.

Thorpe S.J., Gegenfurther K.R., Fabre-Thorpe M. and Bulthoff H.H. 2001. *Detection of animals in natural images using far peripheral vision.* European Journal of Neuroscience. 14, 869-876.

Toet A. and Levi D.M. 1992. *The two-dimensional shape of spatial interaction zones in the parafovea.* Vision Research. 32, 1349-1357.

Tolhurst D.J. 1972. *Adaptation to square-wave gratings: Inhibition between spatial frequency channels in the visual system.* Journal of Physiology-London. 226, 231-248.

Tolhurst D.J. 1973. *Separate channels for the analysis of the shape and the movement of moving visual stimulus.* Journal of Physiology. 231, 385-402.

Tolhurst D.J. and Barfield L.P. 1978. *Interactions between spatial frequency channels.* Vision Research. 18, 951-958.

Tolhurst D.J. and Ling L. 1988. *Magnification factors and the organisation of the human striate cortex.* Human Neurobiology. 6, 247-254.

Tolhurst D.J. and Tadmor Y. 1997. *Band-limited contrast in natural images explains the detectability of changes in the amplitude spectra.* Vision Research. 37, 3203-3215.

Tolhurst D.J. and Tadmor Y. 2000. *Discrimination of spectrally blended natural images: Optimisation of the human visual system for encoding natural images.* Perception. 29, 1087-1100.

Tolhurst D.J., Tadmor Y. and Chao T. 1992. *Amplitude spectra of natural images.* Ophthalmic and Physiological Optics. 12, 229-232.

Tolhurst D.J. and Thompson I.D. 1981. *On the variety of spatial-frequency selectivities shown by neurons in area-17 of the cat.* Proceedings of the Royal Society of London Series B-Biological Sciences. 213, 183-199.

Tolhurst D.J., Troscianko T., Benson P.J. and Párraga C.A. 1998. *The use of morphed images in the quantitative study of human spatial vision.* Journal of Physiology-London. 506P, 11-P12.

Tootell R.B.H., Silverman M.S., Switches E. and De Valois R.L. 1982. *Deoxyglucose analysis of retinotopic organisation in primate striate cortex.* Science (NY). 218, 902-904.

Travis D. 1991a. *Effective color displays.* Academic Press, London.

Travis D. 1991b. *The visual system.* in Effective color displays. Accademic Press, London. 33-71.

Troscianko T., Chirimuuta M., Párraga C.A. and Tolhurst D.J. 2000a. *Perceived contrast of filtered natural images.* Perception. 29, 19.

Troscianko T., Chirimuuta M., Párraga C.A. and Tolhurst D.J. 2000b. *Perceived contrast of natural images is mediated by higher- order image statistics.* Perception. 29, 82-82.

Troxler D. 1804. *Uber Das Verschwinden Gegenbener Gegenstande Innerhalb Unseres Gesichtskreises.* in Ophthalmol Bibliotek. 431-573.

Turiel A., Parga N., Ruderman D.L. and Cronin T.W. 2000. *Multiscaling and information content of natural color images.* Physical Review -Series E-. 62, 1138-1148.

Van Der Schaaf A. and Van Hateren J.H. 1996. *Modelling the power spectra of natural images: Statistics and information.* Vision Research. 36, 2759-2770.

Van Essen D.C., Newsome W.T. and Maunsell J.H.R. 1984. *The visual field representation in striate cortex of the macaque monkey: asymmetries, anysontropies and individual variability.* Vision Research. 24, 429-448.

Van Hateren J.H. 1992a. *Theoretical predictions of spatiotemporal receptive-fields of fly LMCs, and experimental validation.* Journal of Comparative Physiology A -Sensory Neural and Behavioral Physiology. 171, 157-170.

Van Hateren J.H. 1992b. *A theory of maximising sensory information.* Biological Cybernetics. 68, 23-29.

Van Hateren J.H. and Ruderman D.L. 1998. *Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex.* Proceedings of the Royal Society of London Series B-Biological Sciences. 265, 2315-2320.

Van Hateren J.H. and Van Der Schaaf A. 1998. *Independent component filters of natural images compared with simple cells in primary visual cortex.* Proceedings of the Royal Society of London Series B-Biological Sciences. 265, 359-366.

Vimal R.L., Pokorny J., Smith V.C. and Shevell S.K. 1989. *Foveal cone thresholds.* Vision Research. 29, 61-78.

Vinje W.E. and Gallant J.L. 2000. *Sparse Coding and Decorrelation in Primary Visual Cortex During Natural Vision.* Science -New York Then Washington-. 287, 1273-1275.

Virsu V., Näsänen R. and Osmoviita K. 1987. *Cortical magnification and peripheral vision.* Journal of the Optical Society of America. 4, 1568-1578.

Virsu V. and Rovamo J. 1979. *Visual resolution, contrast sensitivity, and the cortical magnification factor.* Experimental Brain Research. 37, 475-494.

Volkman F.C., Riggs L.A., White K.D. and Moore R.K. 1978. *Contrast sensitivity during saccadic movements.* Vision Research. 18, 1193-1200.

Von Helmholtz H. 1911. *Treatise on Physiological optics.* J. P. C. Southall. 1924. The Optical Society of America, Washington.

Vos J.J. and Walraven P.L. 1971. *On the derivation of the foveal receptor primaries.* Vision Research. 11, 799-818.

Wachtler T., Lee T.W. and Sejnowski T.J. 2001. *Chromatic structure of natural scenes.* Journal of the Optical Society of America A. 18, 65-77.

Walraven P.L. 1974. *A closer look at the tritanopic convergence point.* Vision Research. 14, 1339-1343.

Wandell B.A. 1995. *Foundations of vision.* Sinauer Associates, Sunderland, Mass.

Watson A.B. 1979. *Probability summation over time.* Vision Research. 19, 515-522.

Watson A.B. 1983. *Detection and recognition of simple spatial forms.* in Physical and biological processing of images: proceedings of an international symposium organized by the Rank Prize Funds, London, England, 27-29 September, 1982. Springer series in information sciences. Springer-Verlag, Berlin ; New York. 100-114.

Watson A.B. 1987. *Efficiency of a Model Human Image Code.* Journal of the Optical Society of America A. 4, 2401-2417.

Watson A.B. 1993a. *DCTune: A Technique for Visual Optimization of DCT Quantization Matrices for Individual Images.* Sid International Symposium Digest of Technical Papers. 24, 946.

Watson A.B. 1993b. *Digital images and human vision.* MIT Press, Cambridge, Mass.

Watson A.B. and Robson J.G. 1981. *Discrimination at threshold: labelled detectors in human vision.* Vision Research. 21, 1115-1122.

Watson A.B. and Solomon J.A. 1995. *Contrast gain model fits masking data.* Investigative Ophtalmology and Visual Science. 36, S438.

Watson A.B. and Solomon J.A. 1997. *Model of visual contrast gain control and pattern masking.* Journal of the Optical Society of America A. 14, 2379-2391.

Webster M.A. and Mollon J.D. 1997. *Adaptation and the color statistics of natural images.* Vision Research. 37, 3283-3298.

Weckstrom M. and Laughlin S.B. 1995. *Visual ecology and voltage-gated ion channels in insect photoreceptors.* Trends in Neurosciences. 18, 17-21.

Wertheim T. 1894. *Peripheral visual acuity (translated by I. Dunsky).* American Journal of Optometry and Physiological Optics. 57, 919-929.

Westheimer G. 1982. *The spatial grain of the perifoveal visual field.* Vision Research. 22, 157-162.

Whitaker D., Latham K., Makela P. and Rovamo J. 1993. *Detection and Discrimination of Curvature in Foveal and Peripheral-Vision.* Vision Research. 33, 2215-2224.

Whitaker D., Makela P., Rovamo J. and Latham K. 1992. *The Influence of Eccentricity on Position and Movement Acuities as Revealed by Spatial Scaling.* Vision Research. 32, 1913-1930.

White A.J.R., Solomon S.G. and Martin P.R. 2001. *Spatial properties of koniocellular cells in the lateral geniculate nucleus of the marmoset Callithrix jacchus.* Journal of Physiology-London. 533, 519-535.

Wichmann F.A. and Hill N.J. 2001a. *The Psychometric Function: I. Fitting, Sampling, and Goodness of Fit.* Perception and Psychophysics. 63, 1293-1313.

Wichmann F.A. and Hill N.J. 2001b. *The Psychometric Function: II. Bootstrap-Based Confidence Intervals and Sampling.* Perception and Psychophysics. 63, 1314-1329.

Wiesel T.N. and Hubel D.H. 1966. *Spatial and chromatic interactions in the lateral geniculate nucleous of the rhesus monkey.* Journal of Neurophysiology. 29, 1115-1156.

Williams D.R., MacLeod D.I.A. and Hayhoe M. 1989. *Foveal tritanopia.* Vision Research. 21, 1341-1356.

Williams D.R., McLeod D.I.A. and Hayhoe M. 1981. *Punctate sensitivity of the blue-sensitive mechanisms.* Vision Research. 21, 1357-1375.

Wood J.M., Wild J.M., Drasdo N. and Crews J.S. 1986. *Perimetric profiles and cortical representation.* Ophthalmic Research. 18, 301-308.

Wright M.W. and Bowmaker J.K. 2001. *Retinal photoreceptors of paleognthous birds: the ostrich (struthio camelus) and rhea (rhea africana).* Vision Research. 41, 1-12.

Wyszecki G. and Stiles W.S. 1967. *Color science : concepts and methods, quantitative data and formulas.* Wiley, New York.

Xu J., Pokorny J. and Smith V.C. 1997. *Optical density of the human lens.* Journal of the Optical Society of America A. 14, 953-960.

Xu X.M., Bonds A.B. and Casagrande V.A. 2002. *Modeling receptive-field structure of koniocellular, magnocellular, and parvocellular LGN cells in the owl monkey (Aotus trivigatus).* Visual Neuroscience. 19, 693-701.

Xu X.M., Ichida J.M., Allison J.D., Boyd J.D., Bonds A.B. and Casagrande V.A. 2001. *A comparison of koniocellular, magnocellular and parvocellular receptive field properties in the lateral geniculate nucleus of the owl monkey (Aotus trivirgatus).* Journal of Physiology-London. 531, 203-218.