On the Perception of Transparency: Psychophysics and Applications to Medical Image Visualisation

REZA KASRAI

Department of Biomedical Engineering

McGill University, Montréal

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

Your file Votre référence
ISBN: 0-612-85715-8
Our file Notre référence
ISBN: 0-612-85715-8

Abstract

TRANSPARENCY is used routinely as part of a host of visualisation functionalities in software applications for image–guided procedures, though little research is devoted to the rigorous validation of the use of transparency in clinical visualisation. This thesis presents three psychophysical studies aiming to understand how the human visual system interacts with transparent stimuli. The first sets out to measure the performance of users in a 3–D manual segmentation task. Visualising the stimuli in stereo improved performance, though no effect of transparent surface rendering was revealed. In addition, subjects performed better using a standard 2–D mouse compared to a 3–D tracking device.

The next two studies explore the intensity and figural conditions for perceptual transparency using a novel six–luminance stimulus. While a number of models of intensity conditions have been previously proposed, it was found that the luminance–based formulation of Metelli's episcotister model, and a model based on ratios of Michelson contrasts best predicted the subjects' settings, which were found to be very precise. The results also showed that there exists a reasonably wide range of stimuli that give rise to at least some degree of perceived transparency. It was demonstrated that the relative arrangement of the colours around contour crossings (X–junctions) was a salient feature indicating to the visual system the plausibility of a transparent filter and the depth ordering of layers. In addition, the occlusion of X–junctions and perturbation of the orientation of a transparent filter's contours at the junction gave rise to reductions in performance, indicating the importance of junctions in transparency perception.

Résumé

BIEN que la transparence fasse partie intégrante d'une gamme d'outils de visualisation utilisés par les logiciels de thérapie guidée par l'image, peu d'efforts ont été investis afin de valider rigoureusement son utilisation en milieu clinique. Cette thèse présente trois études de psychophysique qui visent à améliorer la compréhension de l'interaction du système visuel humain avec des stimuli transparents. La première étude mesure la performance des observateurs lors d'une tâche de segmentation manuelle tri–dimensionnelle. Il a été prouvé que la visualisation des stimuli en stéréoscopie améliore la performance, même si aucun effet du rendu transparent des surfaces n'est révélé. Qui plus est, les sujets obtiennent une performance supérieure avec une souris standard plutôt qu'avec un appareil tri–dimensionnel.

Les deux études suivantes explorent les conditions d'intensité et de formes géométriques dans le contexte de la perception de la transparence en utilisant un nouveau stimulus à six luminances. Parmi les nombreux modèles de conditions d'intensité déjà proposés, il est établi que le modèle d'épiscotistère de Metelli en formulation basée sur la luminance, ainsi que le modèle selon les proportions de contraste Michelson constituent les meilleurs prédicteurs de réponses des sujets. Ces dernières sont par ailleurs très précises. Les résultats montrent qu'il existe toute une gamme de stimuli provoquant, dans une certaine mesure, une perception de transparence. L'arrangement relatif des couleurs autour d'un croisement de contours (une jonction en X) est une caractéristique saillante indiquant à la fois l'existence plausible d'un filtre transparent et l'ordre des couches en profondeur. De surcroît, une occlusion des jonctions en X ou une perturbation d'orientation des contours d'un filtre transparent à la jonction causent une réduction de performance, révélant ainsi l'importance des jonctions pour la perception de la transparence.

Acknowledgments

I am deeply grateful to my parents for their unfailing support throughout my studies.

I thank Robert Flanz and Roch Comeau for giving me a crash course in programming and managing large–scale graphical software platforms. Philippe St–Jean and Diego Clonda were my inspiring collaborators during the first incarnation of the surgical navigation software (VIPER) at the Image–Guided Neurosurgery Laboratory (Montréal). David Gobbi, Kirk Finnis, and Yves Starreveld of the Image–Guided Surgery Laboratory (London, ON) were instrumental in developing and supporting the second visualisation and navigation platform (ASP). I extend my gratitude also to all members of the Brain Imaging Center (Montréal Neurological Institute), the McGill Vision Research Unit, and the Imaging Research Laboratories (Robarts Research Institute, London, ON) whose expertise and friendship have spoiled me over the years.

None of this work would have been possible, however, were it not for the (unofficial) supervision of Bruce Pike and Louis Collins, and especially the guidance of Terry Peters. I am indebted to Fred Kingdom whose integrity and honesty are matched only by his wit and reason.

My studies at McGill University were supported financially by the Natural Sciences and Engineering Research Council and the Medical Research Council of Canada (now Canadian Institutes of Health Research), as well as the Governments of Ontario, Québec and Canada, who are to be encouraged for their continued investments in public post–secondary education and research.

Contents

List of figures

List of tables

Glossary

1–D, 2–D, 3–D ..... One-, two-, three–dimensional

2AFC ..... Two–alternative forced–choice

ANOVA ..... Analysis of variances

API ..... Application programming interface

AR ..... Augmented reality

ASP ..... Acronym still pending

AV ..... Augmented virtuality

BCE ..... Before the common era

CNS ..... Central nervous system

CRT ..... Cathode ray tube

CT ..... Computed tomography

DSA ..... Digital subtraction angiography

EEG ..... Electroencephalography

fMRI ..... Functional magnetic resonance imaging

IGS ..... Image–guided surgery

IGNS ..... Image–guided neurosurgery

LE ..... Luminance episcotister (model)

MEG ..... Magnetoencephalography

MIP ..... Maximum intensity projection

MNI ..... Montréal Neurological Institute

MRA ..... Magnetic resonance angiography

MRI ..... Magnetic resonance image or imaging

MRSI ..... Magnetic resonance spectroscopy imaging

PDF ..... Probability density function

PET ..... Positron emission tomography

Pixel ..... Picture element

rf ..... Radiofrequency

ROC ..... Receiver operating characteristic

ROI ..... Region of interest

RRI ..... Robarts Research Institute

SPECT ..... Single photon emission computed tomography

TPS ..... Thine–plate spline

VIPER ..... Visual integration platform for enhanced reality

Voxel ..... Volume element

Preface

Original contributions and organisation of the thesis

This thesis consists of six chapters, three of which describe empirical findings. Chapter 1 serves as a general introduction to the concept of transparency both in the context of computer graphics and human vision research. An overview of the history of transparency from ancient times through the Renaissance and including the present state of knowledge is provided.

Chapter 2 provides an overview of visualisation techniques and the most prevalent uses of transparency in the context of image–guided surgery. It reviews the sources of different types of medical images and the ways in which they are combined, highlighting the uses of transparency, and providing in the process a background for the experiments of Chapter 3. Chapter 3 explores some of the visualisation parameters (including stereoscopy and transparency) used during manual segmentation by presenting a series of psychophysical experiments aimed at identifying the specific combination most appropriate for the task.

In Chapter 4, a set of luminance–based transparency models are compared to find the one that best describes human observers' responses when processing stimuli containing illusory transparency. The accuracy and precision, as well as the flexibility in the model are also investigated. Chapter 5 subsequently uses the stimulus and model developed in Chapter 4 to look at the figural conditions, and specifically the role of contour crossings (junctions), in achromatic transparency.

Finally, Chapter 6 discusses and summarises the main findings of the thesis and suggests avenues for future research.

The significant contributions made by this thesis are as follows:

- Development of a comprehensive image–guided neurosurgery software application (VIPER) for clinical use;
- Adaptation of software modules for interaction with virtual surfaces using 2–D and 3–D input devices;
- Finding that users are more adept at performing a specific manual segmentation task using a standard 2–D mouse than a 3–D tracking device;
- Observation that stereoscopic visualisation improves segmentation accuracy in the above task;
- Use of a novel circular six–luminance stimulus for interrogation of the human visual system's internal transparency model;
- Empirically proving the validity of two particular luminance–based models for transparency perception;
- Derivation of the modified method of pair comparisons for two–alternative forced–choice studies where the 'correct' stimulus is either unknown or not present in every trial;
- Showing that there is a wide range surrounding an observer's ideal transparency point, wherein some degree of transparency is perceived;
- Revelation of a significant role of X–junctions in transparency perception; and,
- Showing empirical evidence of the visual system's sensitivity to the polarity of contours which form X–junctions.

Contributions of other authors

The contributions of the authors to the papers which are part of this work should be clarified as follows. Reza Kasrai, in his capacity as doctoral candidate, developed the methodology, conducted the experiments, carried out the statistical analysis, and wrote the bulk of the text of the reports, while Drs. Peters and Kingdom as the candidate's research supervisors provided mentorship and guidance in the projects. Reza Kasrai, Philippe St–Jean, and Robert Flanz contributed equally to the design and development of the image–guided neurosurgery software (VIPER), described briefly in §2.5. David Gobbi, Kirk Finnis, and Yves Starreveld were doctoral students in the Image–Guided Surgery Laboratory (London, ON) who had developed the visualisation software modules described in Chapter 3. The adaptation of these modules for stimulus generation and psychophysics, as well as the design of the experiments, were carried out by Reza Kasrai.

Chapter 3 is an expanded version of a conference proceeding presented at a workshop during the conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2001. Chapters 4 and 5 are based closely on the text of two manuscripts, one in press and the other published by peer–reviewed scholarly publications. The full references are as follows.

- R. Kasrai, K. Finnis, D. G. Gobbi, Y. P. Starreveld, and T. M. Peters, "Advantages of visualisation and manual segmentation in 3–D: an empirical study," in *Workshop on Interactive Medical Image Visualisation and Analysis, MICCAI 2001*, Utrecht, Netherlands, October 2001.

- R. Kasrai and F. A. A. Kingdom, "The precision, accuracy, and range of perceived achromatic transparency," *Journal of the Optical Society of America (A)*, vol. 18, no. 1, pp. 1–11, 2001.

- R. Kasrai and F. A. A. Kingdom, "Achromatic transparency and the role of local contours," *Perception* (in press).

Other publications

The following are additional contributions that arose from the work described in this dissertation.

Peer-reviewed articles

- P. St-Jean, A. F. Sadikot, D. L. Collins, D. Clonda, R. Kasrai, A. C. Evans, and T. M. Peters, "Automated atlas integration and interactive three–dimensional visualization tools for planning and guidance in functional neurosurgery," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 672–680, 1998.

- M. C. Preul, R. Leblanc, Z. Caramanos, R. Kasrai, S. Narayanan, and D. L. Arnold, "Magnetic resonance spectroscopy guided brain tumour resection: Differentiation between recurrent glioma and radiation change in two diagnostically difficult cases," *Canadian Journal of Neurological Sciences*, vol. 25, pp. 13–22, 1998.

Book chapters and proceedings

- F. Kingdom and R. Kasrai, "Discriminating material from illumination changes in complex scenes," in *54th Annual Spring Conference of the Society for Imaging Science and Technology: PICS 2001*, Montréal, Canada, pp. 5–8, 2001.

- R. Kasrai, F. A. A. Kingdom, and T. M. Peters, "The perception of transparency in medical images," in *Medical Image Computing and Computer Assisted Intervention—*

MICCAI'99, C. Taylor and A. Colchester, eds., no. 1679 in Lecture Notes in Computer Science, pp. 726–733, Springer-Verlag, Cambridge, UK, September 1999.

- K. Finnis, P. St-Jean, R. Kasrai, D. Clonda, and T. Peters, "Creation of a non–linearly warpable 3–D functional atlas for image-guided neurosurgery," *Medical Physics (Proceedings of the 44th Annual Meeting of the Canadian Organisation of Medical Physicists)*, vol. 25, no. 6, p. 1094, London, Canada, June 1998.

- P. St-Jean, A. F. Sadikot, D. Clonda, R. Kasrai, A. C. Evans, and T. M. Peters, "Interactive 3–dimensional visualization tools for stereotactic atlas–based functional neurosurgery," in *4th International Conference on Functional Mapping of the Human Brain*, Montréal, Canada, June 1998.

- P. St-Jean, R. Kasrai, D. Clonda, A. F. Sadikot, A. C. Evans, and T. M. Peters, "Interactive 3–dimensional visualisation tools for stereotactic atlas–based functional neurosurgery," in *Medical Imaging 1998: Image Display*, Y. Kim and S. K. Mun, eds., vol. 3335 of *Proceedings of SPIE*, pp. 198–207, San Diego, U.S.A., March 1998.

- T. M. Peters, R. M. Comeau, R. Kasrai, P. St-Jean, D. Clonda, M. Sinasac, M. Audette, and A. Fenster, "A comprehensive approach to image–guided surgery," in *Photonics West 1998: Surgical–Assist Systems*, A. Katzir *et al.*, eds., vol. 3262 of *Proceedings of SPIE*, pp. 244–252, San Jose, USA, January 1998.

- R. Kasrai, P. St-Jean, M. C. Preul, S. Narayanan, D. L. Arnold, and T. M. Peters, "Multimodality image visualization for image–guided neurosurgery," *Medical Physics (Proceedings of the 43rd Annual Meeting of the Canadian Organisation of Medical Physicists)*, vol. 24, no. 7, p. 1205, Charlottetown, Canada, June 1997.

- M. C. Preul, R. Kasrai, R. Comeau, S. Narayanan, A. Olivier, R. Leblanc, and D. L.

Arnold, "Using integrated multimodal imaging for neurosurgical guidance in the resection or biopsy of brain tumours," in *3rd International Conference on Functional Mapping of the Human Brain*, Copenhagen, Denmark, May 1997.

- M. C. Preul, Z. Caramanos, R. Kasrai, R. Leblanc, and D. L. Arnold, *Advanced Neurosurgical Navigation*, ch. Proton Magnetic Resonance Spectroscopic Imaging in Neurosurgery: Applications for Brain Tumors. New York, USA: Thieme Medical Publishers, Inc., 1999.

Abstracts and presentations

- R. Kasrai and F. A. A. Kingdom, "Articulated chromatic backgrounds facilitate detection of illumination boundaries," in *Perception*, vol. 30 (suppl.), 24th European Conference on Visual Perception, Kusadasi, Turkey, August 2001.

- F. A. A. Kingdom and R. Kasrai, "Discriminating illumination from material changes on mondrian backgrounds," *Investigative Ophthalmology and Visual Science (suppl.)*, vol. 42, no. 4, 2001.

- R. Kasrai and F. A. A. Kingdom, "Are X-junctions necessary for the accurate perception of transparency?," *Investigative Ophthalmology and Visual Science (suppl.)*, vol. 41, no. 4, p. S218, 2000.

- R. Kasrai, F. A. A. Kingdom, and T. M. Peters, "The dichoptic perception of achromatic transparency," in *Perception*, vol. 28 (suppl.), 22nd European Conference on Visual Perception, Trieste, Italy, August 1999.

- A. F. Sadikot, P. St–Jean, D. Clonda, R. M. Comeau, R. Kasrai, M. Panniset, G. Bertrand, and T. M. Peters, "An MRI–based computerized system for the surgi-

cal treatment of movement disorders," in *Proceedings of the Canadian Congress of Neurological Sciences*, Montréal, Canada, June 1998.

- P. St-Jean, D. Clonda, R. Kasrai, A. C. Evans, C. J. Holmes, A. F. Sadikot, and T. M. Peters, "Automatic integration of a thalamic atlas with 3–D MRI for image-guided neurosurgery," in *Proceedings of the Annual Meeting of the International Society for Magnetic Resonance in Medicine*, p. 559, Sydney, Australia, April 1998.

FOR the truths are obscure, the ends hidden, the doubts manifold, the minds turbid, the reasonings various; the premisses are gleaned from the senses, and the senses (which are our tools) are not immune from error. The path of investigation is therefore obliterated and the inquirer, however diligent, is not infallible. Consequently, when inquiry concerns subtle matters, perplexity grows, views diverge, opinions vary, conclusions differ and certainty becomes difficult to obtain.

—Ibn al-Haytham (preface to *The Optics* [1])

Chapter 1

Introduction

1.1 Purpose

THIS work aims, in part, to bridge the gap between the fields of computer graphics, medical image visualisation, and visual perception. A series of experiments is introduced which develops and uses psychophysical techniques to focus on different aspects of the perception of transparency. The purpose of these studies is twofold. First, to find a combination of visualisation parameters in a specific surgical planning task which facilitates the performance of observers interacting with medical images. The second aim is to understand under simple stimulus conditions the effects of form and luminance on the accuracy and precision of transparency perception by the human visual system. Whereas the former is in the context of a complex task involving chromatic stimuli which contain many cues—including surface shading, texture, and stereo disparity—the latter concerns itself only with the minimal intensity and figural conditions necessary to give rise to a transparent percept.

1.2 Motivation

With the continual increase in the processing power of relatively affordable computers, and the increase in the variety of their applications to medicine, the questions regarding how well computers and algorithms perform are slowly being replaced by those examining the nature of the interaction of humans with machines. Computers are now involved in numerous aspects of medicine, from simply recording bioelectric signals to controlling the acquisition of medical imaging data. Computers are also employed in manipulating medical images by using image–processing algorithms to render images for diagnosis, therapeutic planning, or surgical guidance. It is these final images which become the input stimuli in visualisation tasks performed by observers (typically clinicians). These tasks are as varied as the clinical procedures which exist, and can range from detection of pathology on an x–ray to measuring the size and position of a therapeutic lesion. It is clearly desirable then that these images convey not only the most accurate information, but that they represent it in a fashion an observer can best understand. As will be discussed in this chapter and the next, transparency plays a role in a number of visualisation techniques used in surgical and diagnostic practice, providing the motivation for a better understanding of the manner in which the human visual system interacts with stimuli containing transparent surfaces.

The questions addressed here encompass many existing domains of inquiry, and define new, multi–disciplinary ones. The appropriate background knowledge needs to include aspects of human vision research, computer graphics, and psychometric and psychophysical techniques. The work here was originally motivated by the desire to inform the recent dearth of medical visualisation techniques with rigorous analyses. Indeed, the application of some of the techniques and conclusions reached herein will be to tasks involving the visualisation and manipulation of neurological images. As a result, the review of current

knowledge contained in this and the next chapter will also include some neuroanatomy and the description of some neurosurgical techniques. In the interest of conciseness, however, in areas where no more than a cursory background is required, the survey will be limited to the most relevant materials.

1.3 Organisation of this chapter

The present chapter provides the necessary understanding of what is meant by transparency and how techniques for simulating transparent surfaces have developed over time. After describing the different types of transparency and presenting a few basic definitions, the chapter discusses the observations and theories concerning perceptual transparency throughout the history of optics. This will include an overview of the use of a device called an *episcotister*, originally used in experiments on colour mixing. The 18th and 19th century observations on colour led to the formulation of the Talbot–Plateau Law which rose out of the independent work of two vision scientists in the 1830s. This law formed a natural basis in the 1970s for the first algebraic formulation of perceptual transparency by the Italian Fabio Metelli. The chapter finishes by following the way the simulation of transparency using computer graphics developed out of the requirements of the motion picture industry in the 1970s, simultaneously yet independently from the vision research community.

1.4 Transparency

Transparency is a property of surfaces through which some amount of light is passed. The proportion of total incident light which passes through a surface is its *transmittance*, represented by t[1], and the proportion of light reflected from the surface is usually denoted as r.

[1]Following computer graphics nomenclature [2], $t = 1 - \alpha$, where α is the *opacity*.

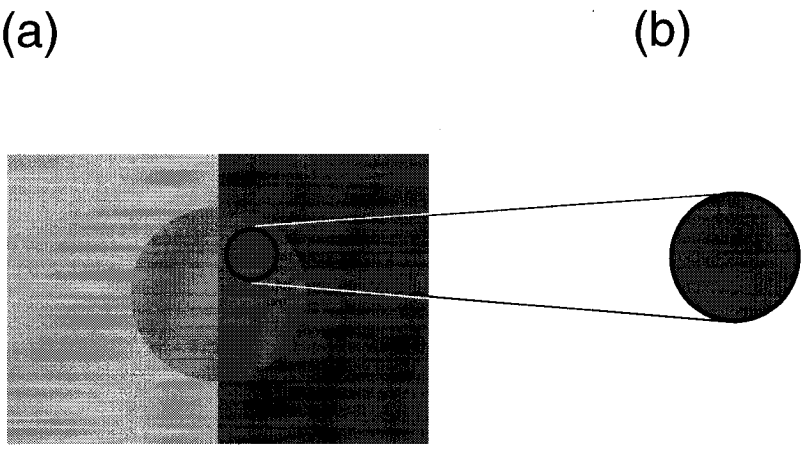


Figure 1.1: (a) Classical layer transparency stimulus with four luminances, and (b) region of superposition viewed in isolation, showing that physical transparency is not required for perceptual transparency.

Given that most surfaces also absorb some of the incident light,

$$r + t + a \equiv 1, \tag{1.1}$$

is an identity, where a is the proportion of absorbed light. In practice, neither is the absorption of light by different surfaces generally a known quantity, nor is it used in models of the way the visual system interprets transparent surfaces.

1.4.1 Seeing the medium

In the absense of a reflective component the transparent medium is not directly observed and distinguished from light travelling through it, since air, ether (or empty space) and water, not being particularly reflective media, do not significantly distort light except at their interfaces.

And yet humans do perceive transparent media, as when viewing a riverbed through the water's surface, or when looking through a dirty window at a field. From a formal

viewpoint, transparency constitutes a special case of a class of changes in a scene caused by variations in *illumination*. The latter is complemented by a class of variations in luminance and spectral reflectance due to changes in the *material* properties of surfaces in an image. How does the visual system detect the existence and compute the properties of transparent media, and more generally, changes in illumination? Subsequently, how does one create an image to give the best possible impression of an artificial world consisting of one or more transparent objects? These are the types of broader questions that motivate the particular psychophysical studies discussed in this thesis.

1.4.2 Types of transparency

Physical transparency phenomena may be divided into different categories [3, 4] such as media, substances, and layers. The former category refers to air, fog, water, as well as rich, heterogeneously–lit environments with different sources of lighting and shadows. Substances include coloured liquids and glasses which may distort light by refraction or absorption. Whereas *translucent* objects allow light to pass through in a diffuse manner [5], *transparent* surfaces do not distort or blur the contours of the background. In this sense, substances are sometimes translucent, implying they have a non–zero transmittance. Conventionally—at least in the most common vision research usage—transparency refers to the last of the categories, namely layers. Layers have been studied more widely than the other categories, presumably because their homogeneous nature lends itself to generation of simpler stimuli for psychophysical research. This dissertation is concerned with this type of layer transparency (sometimes called 'filter' transparency), as shown in Figure 1.1(a).

It must be emphasised that all of these types of transparency may be simulated in the laboratory to give rise to illusory–transparency percepts. Just as media such as clouds may be displayed in a flight–simulation application, the impression of spatially–homogeneous

layers (or filters) may be created by arranging pieces of appropriately–coloured paper next to one another. On the other hand, physical transparency is not always perceived, as we occasionally walk into glass doors. Thus, as will be discussed in detail (in §4.6; see also [6]), neither is physical transparency necessary for perceptual transparency nor vice versa. The stimuli for all of the studies reported in this dissertation were displayed on CRT displays, and any perceived transparency is illusory in nature.

1.4.3 Lightness and brightness, reflectance and luminance

In order to describe the different types of percepts associated with grey–scale images, the perceptual concepts of *lightness* and *brightness* must be defined. In turn, this requires definitions of the physical properties *reflectance* and *luminance*. Reflectance (r) is simply the ratio of reflected light to total light incident on a surface. r, t, and a (from Equation 1.1) are typically scalar functions of wavelength. An image or surface is said to be achromatic if the reflectance spectrum is flat, and chromatic otherwise. Neutral–density (ND) filters, as a special case, have no reflective component ($r \equiv 0$), and are optically indistinguishable from shadows. Lightness, then, is the perceived reflectance of a surface. This is sometimes also described as the "colour of ink" on a page, or the perceived shade of grey if the surface is achromatic. Lightness and reflectance are often not linearly related—lightness depending additionally on other factors such as surround effects [7]—although their values are both typically expressed as fractions of unity or as a percentage.

Similarly, brightness is the perceptual (or proximal) counterpart of luminance, which is in turn a strictly physical (distal) quantity describing the amount of light coming from a surface. Luminance is measured in units of luminous intensity per surface area, such as *candelas per square meter* (cd/m^2). Luminous intensity is measured in candelas (one *lumen* per steradian). All the studies presented here were performed on a calibrated cathode–ray

tube (CRT). Consequently, it is simplest to describe the stimuli either by the luminance or relative reflectance of their different parts (i.e., contrast), regardless of the ambient lighting, which was generally kept constant throughout all the studies.

1.4.4 Contrast

Contrast is a measure of intensity differences within an image. Depending on the type of image and the number of different intensities involved, contrast may be defined in a number of ways. The standard definition of contrast between a patch on a relatively large, uniform background is the Weber contrast:

$$C = \frac{\Delta I}{I}, \tag{1.2}$$

where ΔI is the luminance difference between the patch and the background, and I is the luminance of the background.

The Michelson contrast is defined as

$$C_M = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}, \tag{1.3}$$

where I_{max} and I_{min} are the luminances of the two adjacent patches. Michelson contrast is typically used for images containing periodic patterns such as sinusoids, but can also be used in models describing transparency [8].

1.5 History of transparency

1.5.1 Pre-history

The phenomenon of transparency "which is that, beyond which nothing is hidden" [9] was implicated early among the multitudinous ancient theories of vision. Competing Greek philosophers proposed alternatively extramissive (after Euclid, where rays are projected

from the eyes) or intromissive (after Epicurus, whereby the eye receives light projected from objects) theories of vision. Another interesting conception, which was not widely adopted, began with Aristotle and his pupil Theophrastus (*ca.* 300 B.C.E.). They suggested that light was generated by the sun, but that it required a medium, called the transparent (or the diaphanous) through which it travelled before being received by the eye ([10], p. 12). These ideas were, in turn, handed down by the Greeks and preserved by Muslim scholars, and included such original insights as a crude version of the anatomy of the eye and the fact that light travels in straight lines. Due to strictures against dissection, however, the Muslims did not make significant progress on other theories inherited from the Greeks, particularly with respect to the anatomy of the eye.

The Muslim scholars did, however, make strides in optical theory. Notably, Ibn al–Haytham (latinised to Alhazen) being familiar with the Greek writings, agreed with Aristotle in that the medium was paramount. As al–Haytham (*ca.* 1040 A.D.) observed, "...the air and the transparent bodies are not tinged by the colours and the lights nor are they permanently altered by them; rather, the property of lights and colours is that their forms extend in straight lines, and it is a property of the transparent body that it does not prevent the penetration of the forms of lights and colours through its transparency." ([1], p. 89). His *Book of Optics* (Kitab al–Manzir [1]) along with the seminal works of a number of other Persian and Arab scholars (such as Ibn Sina or Avicenna, Kamal al–Farisi, and al–Kindi) were introduced to Western Europe at the beginning of the second millennium, and provided a physical and mathematical starting point for the scholars of the Renaissance.

Aside from some indirect observations with lenses and eyeglasses[2], transparency was

[2]Perhaps one of the longest–standing barriers to the discovery of the correct geometry and anatomy of the eye was the lack of understanding of the laws of refraction which describe the behaviour of light as it passes through different transparent media. At the same time, unavailability of a correct anatomical model delayed the deduction of the path of light rays as they travel through the lens to the retina. It was not until the late 13th century that reading glasses were, probably accidentally, first developed.

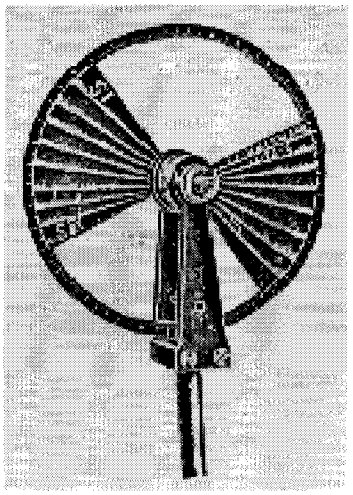


Figure 1.2: An episcotister (from the Greek *epi-*, upon or above + *skotos* darkness produced by a shadow).

not studied as a perceptual phenomenon until the late 19th century. This may have been due to the relative paucity of examples of transparency in nature. Nonetheless, the fact that an image depicted on a single plane can create a compelling multiplanar impression, suggests the question is worthy of study. As Gerbino [4] has suggested, "the obvious discrepancy between perceptual multidimensionality and stimulus unidimensionality accounts for the interest for transparency phenomena shown by many theorists."

1.5.2 Episcotisters and colour wheels

In the absense of present–day CRT displays, 19th and 20th century vision researchers used an episcotister, shown in Figure 1.2, to simulate transparency. An episcotister is simply a disk with cut out sectors, such that when the disk is spun faster than the visual system's critical fusion velocity, it appears as a transparent filter through which an underlying surface may be seen. Though it was Metelli's experiments in the 1970s using an episcotister which led him to formulate an algebraic model for transparency [11], similar studies in colour mixing had been performed beginning much earlier using colour wheels.

The simple idea of a colour wheel (or top) is to combine colours by painting them on a disk which rotates sufficiently rapidly that the colours fuse. Ptolemy (*ca.* 150) was perhaps the first to describe the phenomenon: "For when colour rotates about a distance perceptible to sight in the same perceived temporal moment, it is deemed to spread itself over all places through which it travels. For the phenomenon that occurs in the first rotation is always followed later by repetitions of the same sort" ([12], pp. 60–61). And later al-Haytham observed, "If the top is painted in different colours forming lines that extend from the middle of its visible surface, close to its neck, to the limit of its circumference, then forcefully made to revolve, it will turn round at great speed. Looking at it the observer will now see one colour that differs from all the colours in it, as if this colour were composed of all the colours of those lines." ([1], p. 145)

First used in the 19th century by Musschenbroek [13], and later by Plateau [14] and Talbot [15] for photometry, the episcotister was redesigned by Maxwell [16] "with the radii cut so that the disks [could] be fitted together and the resultant sectors varied at will" [17] thus coining the moniker 'Maxwell's disks'. Talbot (1834) observed that "...a regularly intermittent luminary whose observations are too frequent and too transitory for the eye to perceive, loses so much of its apparent brightness from this cause, as is indicated by the proportion between the whole time of observation, and the time during which it disappears." ([15], pp. 328–329) Almost simultaneously with Talbot, Plateau (1835) proposed "...to report the results of experiments that I have conducted, without knowledge of the work of M. Talbot, which establishes the same principle in a direct manner. This principle can be described in the following way: *When a luminous object excites the eye in a regular and intermittent manner, and its successiveness is not distinguishable because of their close temporal proximity, resulting in it being perceived as continuous, then the apparent brightness of this object is found to be diminished in the ratio of the sum of the durations of*

each period of light and dark, to the duration of a single period of light" ([18], pp. 52–53). As a result of the discovery of these relationships, the phenomenon was more generally formalised as what came to be called the Talbot–Plateau law.

The Talbot–Plateau law

The law may be summarised as stating that the effect of a brief presentation of a colour is proportional to its intensity and the duration of presentation. Since all points on a colour wheel travel at the same angular velocity, the effective intensity of a colour can be controlled by the size of the sector painted with that colour, as well as its physical intensity (reflectance). Thus,

$$tx + (1 - t)y = z, \tag{1.4}$$

where x and y are the reflectances of the two mixing colours, t and $(1 - t)$ the relative sector sizes (mixing proportions) respectively, and z the resulting fused colour.

The application of Talbot's law to transparency follows naturally if one replaces one of the mixing colours with the colour of the background seen through the episcotister's open sectors. The open sectors allow the surface behind the episcotister to be seen, and the proportion of open and closed sectors controls the mixing of figure and ground colours.

1.5.3 Figural and intensity conditions

In the latter part of the 19th century, Helmholtz, in keeping with his high–level or top–down philosophy, believed transparency to be akin to an illumination effect. He thought of it as a judgement requiring previous experience of similar surfaces under different illuminants. According to Helmholtz we can "[eliminate] the differences of illumination by which a body is revealed to us" ([19], p. 287). Similarly, he argues, when an observer views an object through a "coloured mantle", the visual system can separate the effects due to the

object's reflectance pattern from those due to the mantle. This idea of a putative transparent stimulus being split into figure and ground led eventually to a formal model based on *scission*. This model, described first by the Italian vision scientist Fabio Metelli almost a century later, proposes that "with the perception of transparency the stimulus color splits into two different colors, which are called the scission colors. One of the scission colors goes to the transparent layer and the other to the surface of the figure below" ([6], p. 93). This model set the groundwork for other algebraic models (e.g. [20, 4]), and is described more fully in Chapter 4.

Meanwhile Helmholtz' contemporary and student, Ewald Hering, disputed Helmholtz's description of transparency as "seeing through". To support his claims of a low–level description where prior experience with similar stimuli plays no role, he showed that when an observer fixated only the region of superposition, as demonstrated in Figure 1.1(b), it was the fusion colour which was perceived. Fuchs [21], and later Kanizsa [22], showed that beyond simply superposing materials of different colours, the geometry of the objects was also important. Simply put, "transparency depends on form as well as on colour" ([6], p. 92).

The experiments by Koffka and his students, Grace Heider [23] and B. Tudor–Hart [24], at Smith College (Massachussettes) in the 1930s continued Fuchs' investigations. Among other things, they found that transparency on completely homogeneous ground is not possible. While these researchers were aware of the interplay between form and colour, they did not fully explore stimuli with articulated backgrounds where the edge of the episcotister disk transected reflectance changes on the ground. As discussed in Chapter 4, the relationship between pairs of adjacent patches on the background as seen directly and indirectly (through the layer) are the key parameters in the model based on scission.

Since the early 20th century, the relationship between the two necessary conditions of

form (configuration or shape) and colour (both luminance and chromaticity) in perceptual transparency has sparked a debate on the primacy of one over the other. Studies can be found that demonstrate, according to the stimulus conditions, that the lack or disruption of one condition destroys the transparency percept. Namely, under some figural arrangements, transparency is never perceived [22] no matter what colours are chosen, and vice versa [25]. Just as with the Helmholtz–Hering debate, in all likelihood the truth lies somewhere in between. Since both conditions are constitutive, absence of either one will block the percept; moreover, under some circumstances, the constraints of one of the conditions may be looser. The last two studies contained here (Chapters 4 and 5) address each of these topics separately.

1.6 Transparency using computer graphics

1.6.1 Compositing images

Consider the following simple equation:

$$\alpha x + (1 - \alpha)y = z. \tag{1.5}$$

In 1977, two researchers at the New York Institute of Technology developed the above equation to describe the mixing of a partially opaque foreground with an opaque background [26]. Here, x is the colour of the foreground, y the colour of the background, α the opacity of the foreground, and z the resulting mixture of the two images[3]. This is the origin of the term 'alpha–blending' as it is known in computer graphics today. Opacity (α) in this context is defined as zero if the image is completely transparent, and unity if it is fully opaque. The quantities x and y are also typically expressed as fractions of unity, implying

[3] In practice, the operation is applied pixel–wise to the images.

that $z \in [0, 1]$. This operation, however, is not associative, meaning that the outcome is dependent on the *order* in which the images are blended together.

In 1981 Bruce Wallace [27], working at the Hanna–Barbera production studios, developed a more generalised formulation whereby both the contributing images and the final image could be semi–opaque:

$$\begin{aligned} \alpha_{\text{out}} &= 1 - (1 - \alpha_{\text{fgd}})(1 - \alpha_{\text{bgd}}) \\ z &= \alpha_{\text{fgd}} x + \frac{1 - \alpha_{\text{fgd}} y \alpha_{\text{bgd}}}{\alpha_{\text{out}}}, \end{aligned} \tag{1.6}$$

where α_{out} is the opacity of the final image, and α_{fgd} and α_{bgd} are respectively the opacities of the foreground and background images. While this description is more cumbersome, it is independent of depth order and thereby associative.

A simplification of the above equations was shown in 1984 by Porter and Duff [28], then working at Lucasfilm, by pre–multiplying x and y by their respective opacities, resulting in

$$\begin{aligned} \alpha_{\text{out}} &= \alpha_{\text{fgd}} + (1 - \alpha_{\text{fgd}})\alpha_{\text{bgd}} \\ z' &= x' + (1 - \alpha_{\text{fgd}})y', \end{aligned} \tag{1.7}$$

where $x' = \alpha_{\text{fgd}} x$, $y' = \alpha_{\text{bgd}} y$, and $z' = \alpha_{\text{out}} z$.

Equation 1.5 can in principle be thought of as a way of gradually blending one image into another as the relative weighting parameter t varies from 0 to 1, without any requirement that one image be perceived as transparent. But, as it was discussed above (see Equation 1.4), the first generation of algebraic relationships describing perceptual transparency developed in a different and independent manner to give rise to essentially the same formulation.

1.6.2 Spatial and temporal interlacing

In the absense of hardware acceleration of graphics, two early methods of compositing two images consisted of interlacing them pixel–wise either spatially or temporally. In the former, a mixed image was produced by alternating the pixels of the two images into a fine checkerboard pattern [29]. In the latter, the two images were displayed in the same spatial location, but refreshes of the frame buffer would alternate between the two contributing images. Using temporal interlacing, the relative weighting of the two images could be controlled by varying the relative display times for each frame.

Though in the strictest sense these computer graphics manipulations do not implement any type of perceptual transparency model, they suffice to the extent that they provide the impression of a blended final image composed of two others. With the introduction of more sophisticated video hardware and a more-or-less common application programming interface (API), implementing transparency using computer graphics has become routine and almost trivial.

1.6.3 Alpha–blending today

With the use of most high–level graphics libraries nowadays (e.g., Vtk[4], OpenInventor[5]) it is sufficient to set the opacity of an image or a surface (per–pixel or per–polygon if desired) in order to render objects transparently. Of the lower–level libraries available, the most widely used is OpenGL [2], originally developed by Silicon Graphics, Inc. The library's routines may be accessed either directly or indirectly through 'wrapper' routines of a mid– or high–level package.

In the context of digital image compositing, Equation 1.7 is in effect a formulation

[4]Available from `www.kitware.com`

[5]Silicon Graphics Inc., Mountain View, CA, USA

which may be generalised to any logical operator, and graphical libraries such as OpenGL are capable of implementing 12 operators producing all manner of related effects such as antialiasing, billboarding (see also [30]), simulated blur, and strokes of a paintbrush.

1.7 Summary

This introductory chapter places the use and generation of stimuli depicting illusory transparency in an historical context. Phenomenal transparency has been studied since at least the time of the Greek philosophers. Interest in the subject continued through the Arab and Persian scholars into 19th century Europe when experiments on the mixing of colours using episcotisters resulted in the Talbot–Plateau law. This in turn provided the foundation for the first algebraic formulation of perceptual transparency in the early 1970s. The motion picture and animation industry contemporaneously began to provide the impetus for the development of hardware and algorithms for compositing digital images.

It was not, however, until the latter part of the 20th century that psychophysicists began empirical studies on the human visual system's internal models of transparency. While the use of CRT–based transparency is prevalent in many domains (including medical image visualisation), comparatively few empirical studies have set out to ascertain whether the simple equations presented here correspond optimally to the way the visual system encodes and processes transparency. In the next chapter, some of the most prevalent uses of transparency in medical image visualisation are discussed. Although the chapter focuses almost entirely on virtual models of brain structures (for use during image–guided neurosurgery), the questions are in principle equally applicable to the visualisation of the layers of the earth's crust or the different sub–components of an automobile engine.

Chapter 2

Transparency in image–guided surgery: an overview

"Although information visualization is one aspect of the problem, the paradigm of interaction between the physician and the information is an equally significant one..."

—M. Solaiyappan [31]

2.1 Introduction

The investigation of transparency in this thesis was originally motivated by its routine use in various clinical visualisation applications, as discussed in §1.2. In order to familiarise the reader with the types of visualisation techniques and applications which make use of transparency in some way, this chapter presents an overview of the different imaging modalities and rendering techniques used in the visualisation of medical images. Issues related to the visualisation of images incorporating information from multiple modalities are also discussed, using examples of images generated by the software platforms developed in the laboratory. The chapter continues with a brief survey of current uses of image–guidance

in different therapeutic and interventional procedures, as well as some of the perceptual considerations regarding depth cues in mixed reality displays.

While the results of specific psychophysical experiments are not reported, it is important to note that the design of software prototypes developed at the Montreal Neurological Institute (MNI) and the Robarts Research Institute (RRI) was carried out in conjunction and consultation with clinicians in an iterative manner. Nevertheless, psychophysical experiments are valuable in demonstrating the viability and clinical utility of image–guidance systems, and so the chapter concludes with an exposition of the use of such techniques at some other centers, and the role played by transparency is highlighted throughout. This chapter thus sets the context for Chapter 3 in terms of the types of clinically relevant tasks which may be examined more closely using psychophysical techniques.

2.2 Image–guided surgery

Image–guided surgery (IGS) refers to a practice or an environment whereby surgical or therapeutic procedures are aided by images [32, 33]. Even though the term image–guided surgery implies the use of any sort of diagnostic image as an adjunct to surgery, it typically refers to the use of recently–developed, more sophisticated and comprehensive software platforms. In the stricter sense, however, IGS has a century–old history which started with the discovery of x–rays by Röntgen in 1895.

The first reported North American use of x–rays for guidance of a surgical procedure appeared in the Montréal Medical Journal [34]. John Cox, a professor of physics at McGill University produced a radiograph (or a new Röntgen photograph, as it was then known) of the leg of a patient with a bullet lodged against the outer edge of the tibia. Robert Kirkpatrick, a surgeon at the Montréal General Hospital removed the bullet with the aid of

the image. Since the nascency of these early projection radiographs, diagnostic imaging has evolved into a rich and mature field assisting procedures in nearly every part of the body. As such " 'image–guided surgery' is hardly specific as operations have been guided by images for decades," as Dorward [35] has pointed out.

The use of image–guidance in the different surgical specialties has not developed at the same rate, however. The elasticity of soft tissues as well as physiologic motion (breathing, heart–beat) create serious challenges for procedures relying on pre–operative images, such that interventions within or near bony structures (such as orthopaedic surgery [36], neurosurgery, and craniofacial reconstruction [37, 38, 39]) have benefited relatively more than others from IGS techniques. These bony structures can be used as landmarks for finding the correspondence between the the pre–operative image data and the patient intra–operatively. The overriding importance and delicacy of the central nervous system (CNS), however, requires such accuracy and minimal invasiveness that image–guided neurosurgery (IGNS) systems are by far the most prevalent and sophisticated in the research/academic [40] and commercial sectors ([41, 42, 43, 44] to name a few). As a result, most of the examples discussed here will draw from neurosurgical cases encountered in the Image–Guided Surgery Laboratory at the MNI and the RRI, as well as in other research groups.

The images used to guide surgical interventions may or may not reflect the exact geometry of the surgical patient. For example, before the widespread use of computers in neurosurgery, surgeons relied upon stained sections (compiled into anatomical atlases) reflecting the cytoarchitechtonic pattern of the regions of interest (ROIs), which were sometimes overlaid with contours outlining the anatomical structures. These atlases, however, are only faithful to the anatomical specimen used and do not correspond exactly to the anatomy of the surgical patient. Due to recent advances in intersubject registration algorithms [45, 46], these atlases may now be made patient–specific either manually [47, 48] or

automatically [49]. Interventions that use live intra–operative (such as video, ultrasound, and so–called open MRI) images have also enjoyed significant advances in IGS during the last decade.

2.3 Sources of images in neuronavigation

The images employed in IGNS may in general be morphological or functional in nature. IGNS, or neuronavigation, in the context of modern neurological surgery connotes the use of pre–operative morphological {e.g. computed tomography (CT), magnetic resonance imaging (MRI) or angiography (MRA), ultrasound [50, 51, 52]}, functional {e.g. electroencephalography (EEG), magnetoencephalography (MEG) [53], positron emission tomography (PET), single–photon emission computed tomography (SPECT), functional MRI (fMRI) [54]}, and/or biochemical {e.g. magnetic resonance spectroscopy imaging (MRSI) [55, 56, 57]} data.

It is worth noting that while for the most part imaging studies used in neurosurgery are still acquired pre–operatively, the surgical planning is more frequently continued intra–operatively and more interactively [58]. As mentioned above, when the pre–operative imaging modality does not reveal the necessary underlying functional or morphological boundaries, atlases [59, 60, 61, 62] are sometimes matched to the patient's pre–operative brain geometry and integrated into the visualisation and navigation software [49].

Surgeons may also map out the different functional areas of a patient's brain intra–operatively by recording electrical signals from sensory or motor activity of the patient, using a micro–recording electrode [63]. Responses may also be induced through electrical stimulation of the cortex or the basal nuclei using a macro–electrode [64, 65, 66]. Moreover, these recording and stimulation data from an ensemble of patients may be ac-

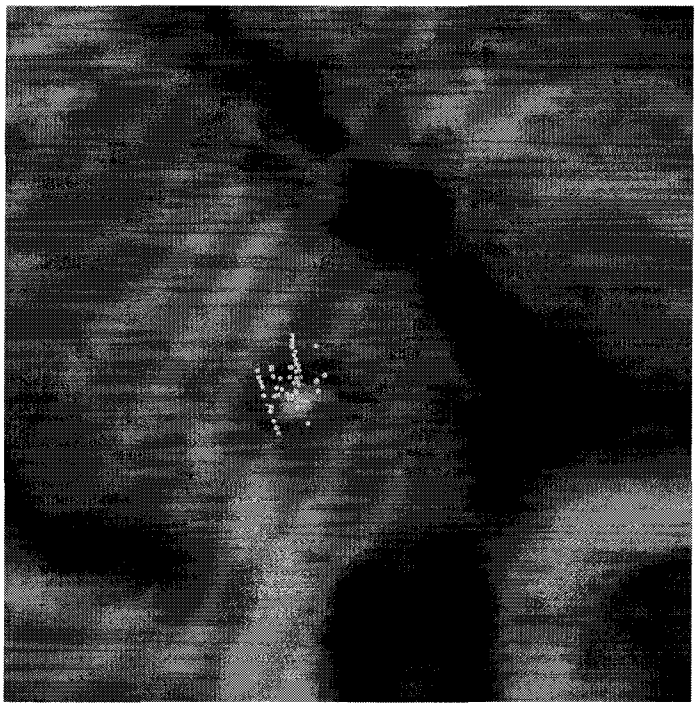


Figure 2.1: Macrostimulation point–cloud (yellow). Note also the hyper–intense therapeutic radiofrequency lesion on the post–operative MRI corresponding to the position of the point–cloud.

Figure 2.2: Cut-aways of deep brain structures revealing point–clouds of physiologic stimulation. Figures 2.1 and 2.2 courtesy of K. Finnis, Robarts Research Institute.

cumulated into a probabilistic functional atlas [67, 68]. One way this latter type of data may be visualised is by colour–coding a point cloud according to the responses elicited and displaying it in the 3–D scene along with patient–specific anatomical data, as in Figures 2.1 and 2.2.

2.4 Transparency and multimodal integration

Though most of the advanced visualisation applications employ image–processing algorithms to extract clinically relevant information, the so–called raw image data still represent a crucial part of the ensemble of intra–operative information. Recent visualisation systems allow the registration and integration of data from many different imaging modalities in order to provide the maximum amount of information to the clinician [69, 70, 71, 49, 72]. The integrated visualisation techniques associated with these images can be broadly distinguished into 2–D and 3–D presentations, even though the final images are always 2–D projections onto a computer display (or light box). In 2–D displays, the images are viewed as planar cuts (slices) through the volume of interest. 3–D displays generally refer to the visualisation of pre–segmented surface and volume renderings (see §2.5.1).

Multimodal visualisation may further be divided into integrated and adjacent displays [73]. Adjacent displays show corresponding slices of the different modalities separately but side-by-side [74]. In these types of presentations a linked cursor is quite useful in comparing corresponding structures or contours [51, 75]. Since adjacent displays do not actually superimpose images, they are not considered further here. Integrated displays, on the other hand, use either spatial or temporal interlacing (see §1.6.2) to form the combined image. In integrated 2–D displays, sections of volumetric anatomical data are typically overlaid with the corresponding functional or biochemical images, such that the underly-

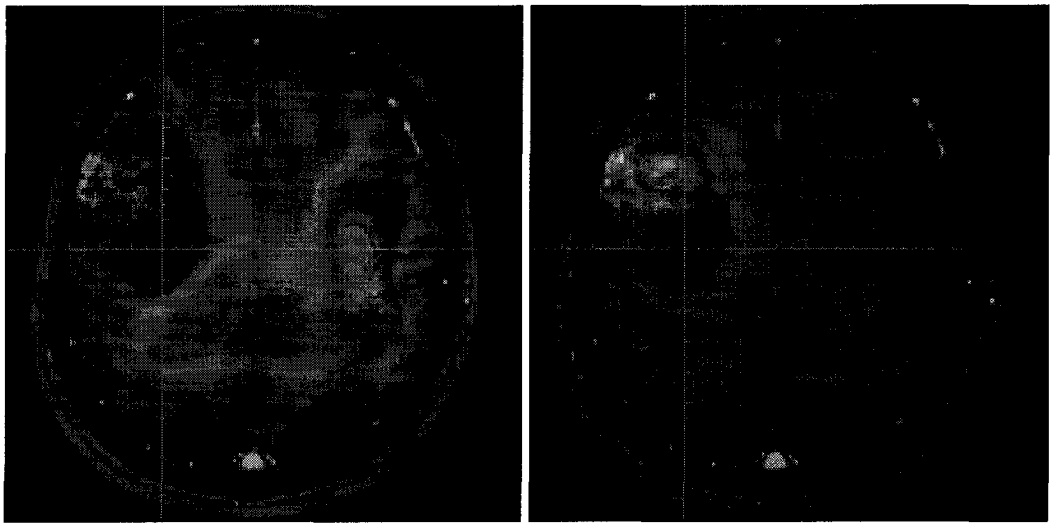


Figure 2.3: MRSI biochemical concentration images superimposed on the corresponding anatomical MR image. High choline concentration (left) indicates area of hyper–cellularity, whereas low N–acetylaspartate (NAA) concentration (right) in the tumour area indicates lack of viable neuronal tissue.

ing structure can be "seen through" the functional image. The simultanous visualisation of structural and functional data is important not only for localising and navigating within often pathological anatomy, but also for reducing the need for intra–operative stimulation with the patient awake. Two examples of diagnostic biochemical and anatomical images are shown in Figure 2.3.

2.5 A taxonomy for visualisation in biomedicine

Given the wealth of different modalities providing information pertinent to surgical and therapeutic procedures, it is little wonder that visualisation of composite images has become an avenue of investigation in its own right. To better track the course of their development, it is helpful to categorise the types of displays and applications based on both their visual characteristics and inherent complexity.

To this end, Solaiyappan [31] has grouped visualisation systems into several gener-

ations. For example, while 1–D waveforms comprised the first generation systems and were used primarily in patient monitoring systems (i.e. electrocardiography and blood pressure), second generation systems performed 2–D image processing such as contour stacking. Similarly, third generation systems refer to those using surface and volumetric rendering, and fourth generation devices include a fourth component (e.g. time) in addition to the three spatial dimensions. Fifth and later generation systems are the state of the art and incorporate additional features such as images from multiple modalities, 3–D spatial tracking systems such as the optically–based POLARIS[1], the magnetically–driven devices such as the Flock of Birds[2], the ultrasound–based 3–D Logitech mouse[3], and sensory (haptic and tactile) feedback devices for more immersive virtual reality. In order of technological sophistication and diagnostic power, all visualisation applications may be described as *illustrative* (first, second, and third) *investigative* (fourth), or *imitative* (fifth generation) [31].

Transparency plays a role in a number of third and later generation displays. For example, Figure 2.4 shows a screen–shot of a fifth generation IGNS system developed for clinical use at the MNI by the author, in collaboration with other members of the IGNS Laboratory. The software (VIPER) incorporated a number of functionalities, including: support of 3–D tracking systems for frameless stereotaxy; support for stereotactic frame–based procedures, along with the ability to record tag points corresponding to regions of physiological stimulation; stereoscopic visualisation; visualisation of multiple pre–registered images; and, surface and volume rendering. ASP, the successor to VIPER, developed by members of the laboratory at the RRI, is able to integrate many of the same and other functionalities into a highly modular, object–oriented and extensible application (see Figure 2.6). Mul-

[1]Northern Digital, Mississauga, Canada

[2]Ascension Technology Corporation, Burlington, VT, USA

[3]Fremont, CA, USA

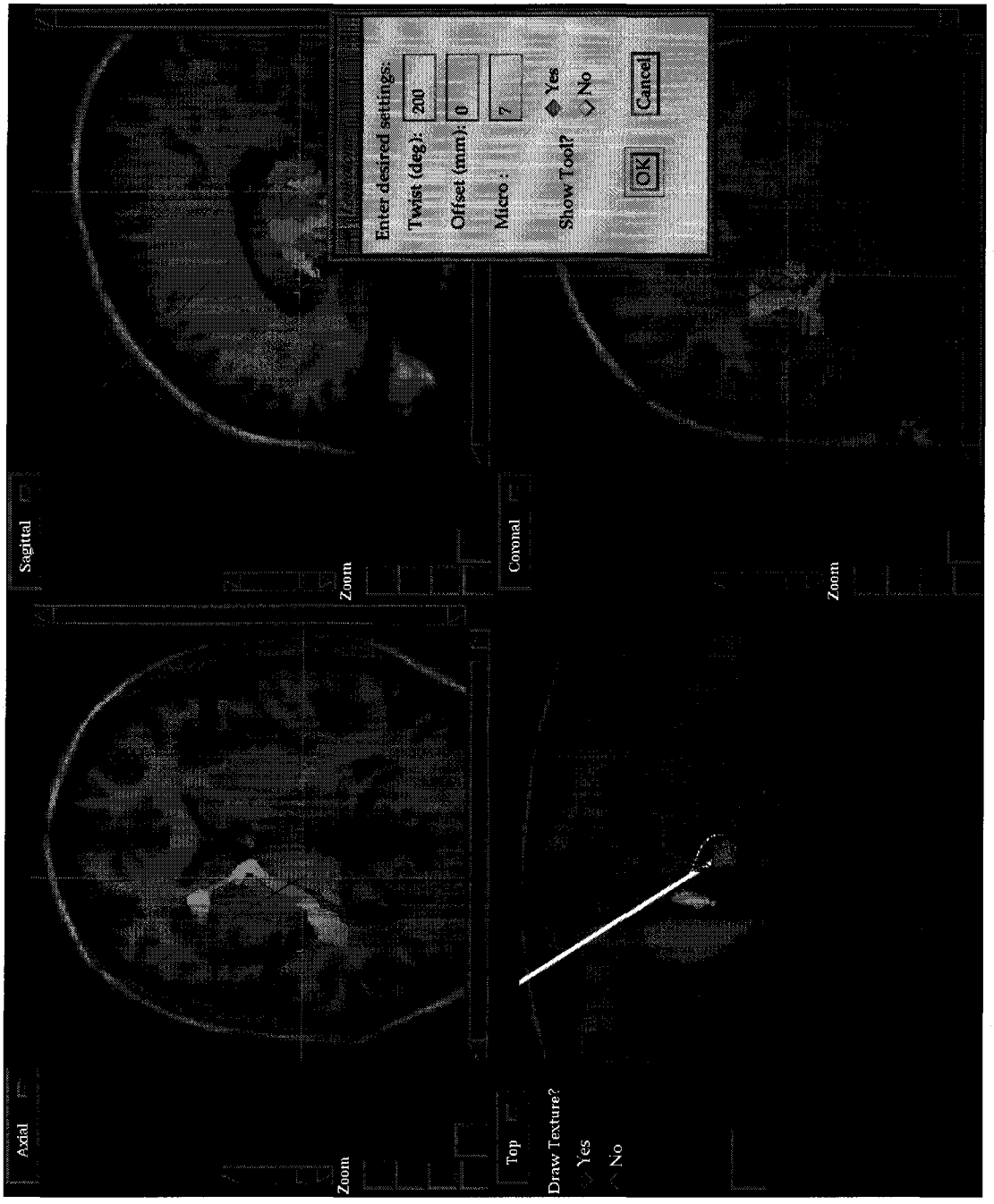


Figure 2.4: A screenshot of the IGNS Laboratory surgical guidance software (VIPER) displaying the Schaltenbrand and Wahren atlas overlaid on an MRI. The trajectory and position of the leukotome is also modelled. The bottom–left panel shows surface–renderings of some deep brain nuclei, including the thalamus (red mesh) and the *vim* (green).

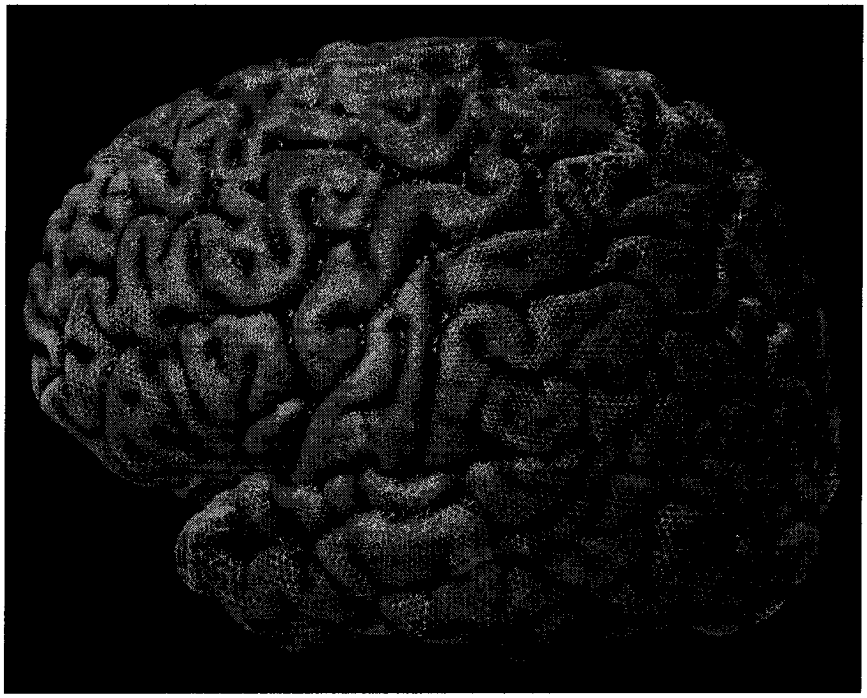


Figure 2.5: A polygonal–mesh representation of a cortical surface.

timodal visualisation, as described in §2.4, involves the superposition of more than one co–registered image, where all but the background image (typically the anatomical scan) are assigned an opacity of less than one. As the next section discusses, surface and volume rendering are two volumetric visualisation techniques which also involve the use of transparency.

2.5.1 Surface rendering

Two of the most prevalent ways of representing volumetric data include surface rendering and ray–casting (including volume rendering), each with its own advantages and drawbacks. Using volumetric data derived from CT or MRI, surfaces corresponding to structures of interest can be extracted (or *segmented out*, see [76, 77], for example). These surfaces are mesh–like representations approximating the geometry of underlying surfaces of interest [78] (see Figure 2.5). In many cases, depending on the neuroanatomical nomenclature

used, a surface is enclosed within, or is transected by, another structure. For example the *vim* (*ventralis intermedius nucleus*), a thalamic nucleus typically targeted in Parkinson's surgery, is part of the thalamus as shown in Figure 2.4 (bottom left panel).

Other surfaces of interest one may wish to visualise include prospective lesions (see Figure 2.7). Therapeutic lesions may be created in brain tissue using one of two methods. A simple mechanical technique makes use of a leukotome [79], which is a hollow shaft approximately 20 cm in length from the end of which a narrow wire protrudes in a "D" shape to cut the tissue as it is rotated about the shaft's axis (see Figure 2.8).

Lesions are more typically caused by heating brain matter using a radiofrequency (rf) probe [80]. The extent of the resulting rf lesions is a function of probe temperature and the length of time the probe is active. Regardless of the technique used, however, the planned lesion may either transect or envelope a target structure. Should one wish to visualise both the lesion (either prospectively during pre–surgical planning, or segmented out of a post–operative scan) and the targetted anatomical structures, one or both must be rendered transparently.

One type of 'transparent' surface rendering (shown in Figures 2.4 and 2.5) makes use of the underlying polygonal representation by drawing only the mesh, with no interpolated surface shading, except at the mesh lines themselves. The possible advantages of different modes of surface rendering is discussed in Chapter 3.

2.5.2 Ray–casting

The main advantage of surface rendering is the compression of large data sets into compact lists of polygons, reducing computational load during visualisation. Segmentation of the surfaces of interest, often requiring user intervention (also discussed in the next chapter), represents a large operational overhead. Moreover, the surface of interest is determined

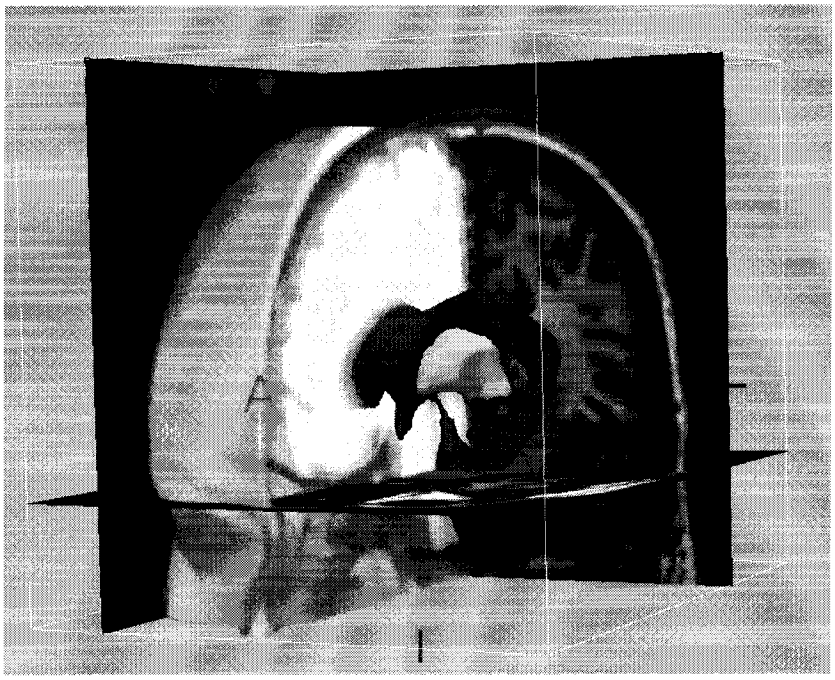


Figure 2.6: Example of a fifth generation visualisation system (ASP). Volume rendering of the brain's right side is mixed with surface rendered ventricles (red) and texture mapped MR data (greyscale). Figure courtesy of IGS Lab, RRI.

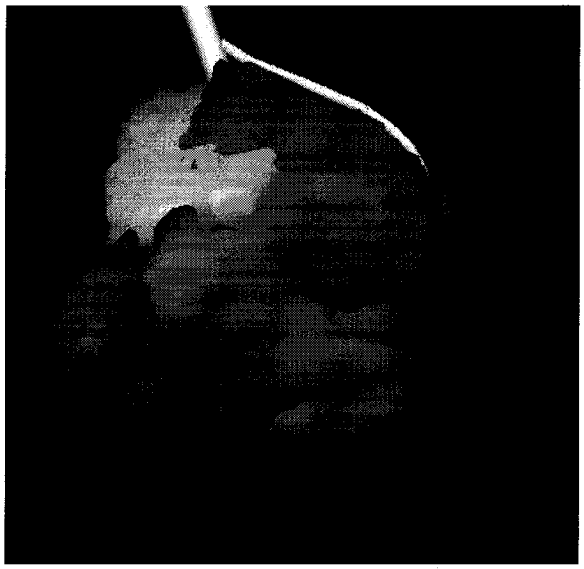


Figure 2.7: A prospective leukotome lesion (red) along with surface rendered surrounding thalamic nuclei, visualised using the VIPER application.

via a one–time binary decision, and cannot typically be manipulated interactively during visualisation. Ray–casting techniques, on the other hand, make use of all the volumetric information and require no calculation of structural boundaries. The drawback is, of course, the large computations required for the reconstruction of each view during visualisation.

As the name implies, ray–casting entails shooting virtual rays through a volume to calculate the intensity of incident pixels on the image (projection) plane, which can be arbitrarily placed to simulate different viewpoints. The resulting images can be thought of as resembling standard projection x–rays. Whereas intensities on a standard x–radiograph are a function of the log–sum of the attenuation coefficients of the different materials within the irradiated volume, in ray–casting the original voxel intensities can be combined in a variety of ways to highlight different structures. In maximum intensity projection (MIP) images, for example, the maximum voxel intensity encountered along each ray is recorded at the corresponding pixel location on the projection plane. MIP is typically used to visualise CT or MR angiograms (see [81], for example).

Another ray–casting technique, called volume rendering [82, 83, 84, 85], allows the observer to view parts or all of a volumetric data set by assigning opacity values to voxels (volume elements), both as a function of voxel intensity (luminance) and as a function of depth (distance from viewer). In all these instances, the surface or reconstructed slices simulated to be closer to the observer are rendered as semi–opaque, or transparent. Some more sophisticated volume rendering methods also weight opacities with the gradient volume (a sort of 3–D contrast volume) [86] to enhance sharp surfaces.

Although some investigators recommend using transparency in volumetric renderings only when the surfaces of interest are too ill–defined to segment [87], use of a particular rendering algorithm (volume or surface) in the end depends on the particular application. For example, for lower resolution functional (SPECT or PET) data, volume rendering might

be more appropriate [88] than surface rendering. But volume renderings of rich anatomical MR images of brain have yet to find a clinical application, except perhaps to facilitate the identification of surface landmarks for patient–to–image registration for frameless image–guided neurosurgery.

2.5.3 Virtual endoscopy

The use of transparency is nowadays routine in many imitative visualisation applications. Surface and volume renderings of the walls of organs such as the cerebral ventricles [89, 90, 91, 92], the bronchi [93, 94] and the upper airways [95], the nasal sinus [96], the ear [97, 98, 99], the gastro–intestinal tract [100], and the lower urinary tract [101] are generated from different views to simulate real endoscopy. One advantage of virtual endoscopy [102] is that these walls can then be 'peeled away', or made transparent, to reveal the structures behind. Distances between rendered objects can be measured [103], and virtual biopsies can be made based on voxel intensity. Figure 2.9 shows a transparently rendered torso revealing the inner organs for minimally–invasive image–guided cardiac surgery [104].

Mixing polygonal surfaces with volumetric renderings (e.g. Figure 2.6), however, is a relatively complex operation. As a simpler example, how does one volume render a data set embedded with surface–rendered transparent objects? Since the transparency operation of compositing (or blending; see §1.6.3) may not be commutative depending on the implementation, the graphics primitives to be rendered (whether they be volume or polygon elements) must be sorted in depth and composited in the proper (back-to-front) order. Depending on the size of the volume and the number of polygons, this can be a very expensive operation in software, and so far the hardware solutions proposed deal only with images generated by pure ray–casting. Although sophisticated algorithmic solutions are beginning to be developed [105], the achievement of interactive or near-real-time performance still

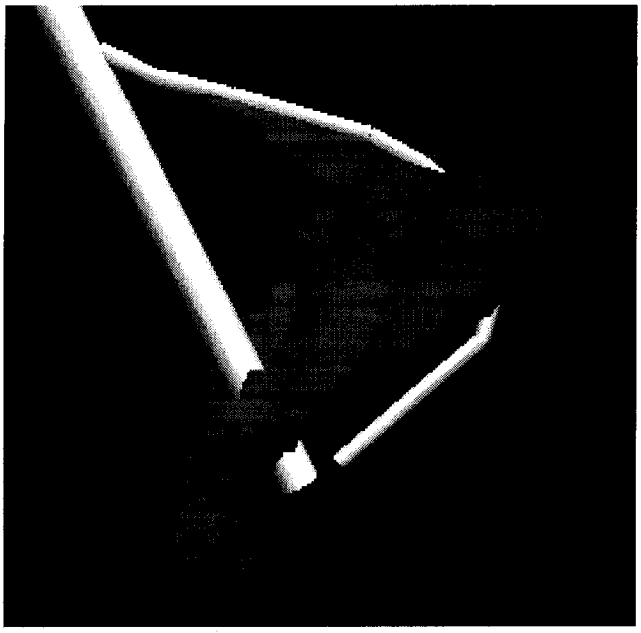


Figure 2.8: VIPER displaying a leukotome about to sever the connection between the *vim* (red) and surrounding tissue (not shown).

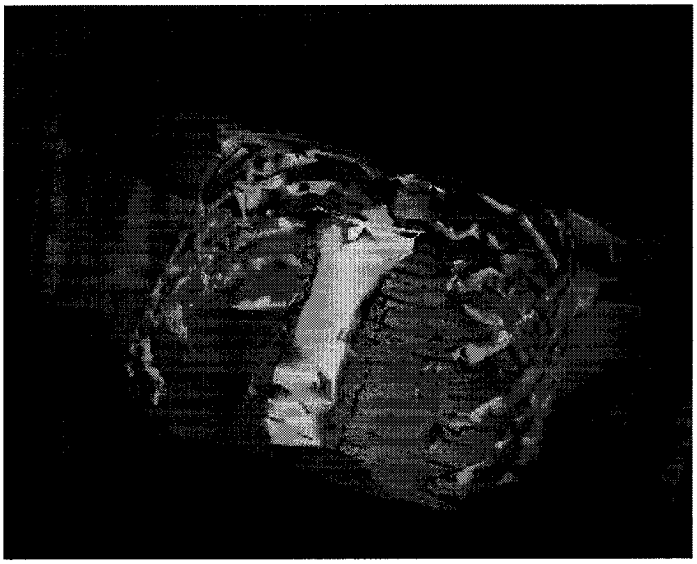


Figure 2.9: Peeling away surface layers for minimally–invasive cardiac surgery.

requires high–end (e.g. Infinite Reality, Silicon Graphics Inc., Mountain View, CA, USA) graphics systems.

Not all applications of data fusion, however, require the use of transparency. Video images from real endoscopes [106, 107] are sometimes used in some mixed reality (see below) visualisation environments as texture maps [108] to enhance reconstructed pre–operative images. Information from functional images may also be mapped as a texture onto pre–segmented cortical surfaces.

2.6 Mixed reality: From reality to virtuality

In discussions involving the integration of data from different sources it is helpful to develop a uniform terminology to describe the visual environment. Milgram [109] among others [110, 111] has defined a so–called taxonomy for environments which mix the real and virtual worlds. The continuum from strict reality to complete virtuality is subdivided according to three broad criteria: Extent of World Knowledge, Reproduction of Fidelity, and Extent of Presence Metaphor. Milgram gives examples of six cases which encompass the *mixed reality* portion of the continuum, from augmented reality (or AR, where some computer–generated objects are integrated into a mostly direct view of the world, such as Figure 2.10) to augmented virtuality (or AV, where some directly–viewed objects, such as video images of the surgical field are blended into a mostly artificial rendering of anatomy [112]). Most of the examples discussed in this work and in the field of medical image visualisation lean towards the strictly virtual extremum of the spectrum. Even though the volumes and surfaces depicted represent real objects (*viz.* a patient's neuroanatomy), they are viewed neither directly (i.e. with the eye) nor indirectly (i.e. using analog or digital video). These types of displays are often incorrectly referred to as AR displays, as most

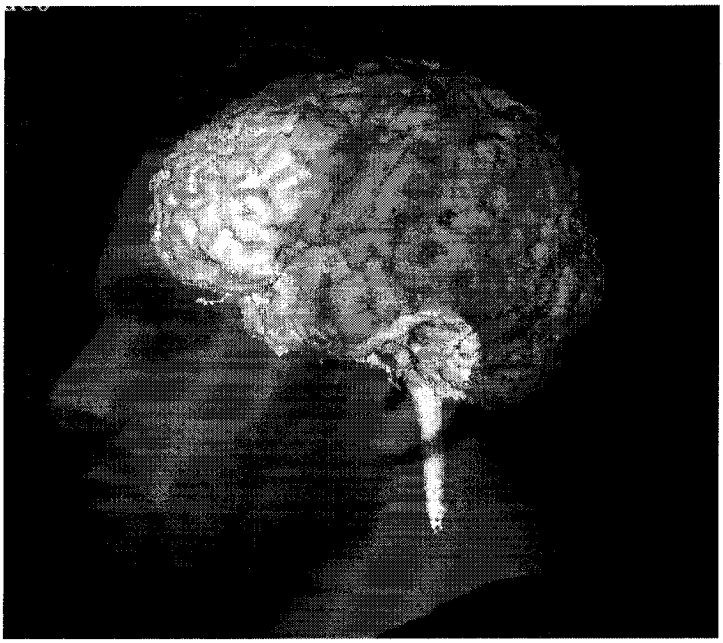


Figure 2.10: Mixing live video with surface rendering.

mixed reality displays are, while they are in fact AV displays.

2.6.1 Obstacles in mixed reality

Some clinical sites do, however, incorporate pre–segmented surfaces of interest into the ocular of an operating microscope [113, 70] or into an intra–operative video view of the surgical field [112, 114, 115, 116] for neuronavigation, resulting in an AV environment. Combining live video images with computer models may, however, result in an inconsistency in visual cues, leading to suboptimal perception of the scene. Drascic [117] lists a number of such issues.

Interposition failures, for example, are caused when near objects are occluded by far objects, even though the remaining depth cues (such are relative disparity, familiar size, etc.) are otherwise consistent. This creates difficulties, for example, in superimposing surface rendered subcortical objects with live video of the surgical field (shown in Figure 2.10). Another problem is a discrepancy in spatial frequency content between computer models

and objects viewed directly or indirectly. Drascic suggests that this resolution mismatch may be interpreted as a difference in accomodation, making the graphic object appear closer than the video object. A general problem relating to stereoscopic visualisation is the uncoupling of two physiological mechanisms relaying depth information: (lens) accomodation and (eye) vergence. When stereoscopic images are presented, the depth of the display (monitor) is constant whereas the vergence of the eyes changes to foveate different regions of the scene.

These considerations outline the difficulties involved in the fusion of images not just from different modalities, but from different parts of the mixed reality spectrum. No matter whether all registration and calibration problems are solved, unless the latter depth cue issues in visualisation are also addressed, perceptual errors will likely continue to occur.

2.7 Measuring truth and veridicality

In the context of the discussion of visualisation in this dissertation, the fidelity of the images (i.e., Truth) is taken as an article of faith. This being said, there are two basic ways to ensure the accuracy of images. One is to use cadavers [118, 119] or anthropomorphic phantoms [120, 121, 122] whose geometry is known, and the other is to simulate the objects on the computer (*viz.* a digital phantom) [123]. This work rather concerns itself with whether the images are perceived veridically by the observer. That is, an evaluation of visualisation is independent of the accuracy of either the image acquisition or processing techniques. This is not to say that all sources of error (from acquisition to intra–operative tracking [124, 125, 126] and visualisation) should not be characterised before bringing any surgical guidance system into the clinic. While many image–guidance systems are continually incorporating additional information and functionality into the sur-

gical navigation environment, it is a huge undertaking to quantify the overall clinical gains in terms of clinical outcome and time or money savings, and only a small number of researchers [127, 128, 129] have carried out such comparative studies.

Depending on the system developer's needs, measurements of an observer's acceptance of an application may be taken along different dimensions [130], such as *user satisfaction* or *performance*, among others (see [131] for a complete review). For example, while the observer may perform tasks using one tool more rapidly and accurately, the tool may demand a heavier cognitive load and have a less satisfying interface. While performance is generally easier to measure directly, user satisfaction may be equally important as a factor in clinically acceptability. More subjective factors may still be quantified using other statistical methods such as surveys with scaled responses (e.g. from 1 to 5 corresponding to "strongly agree" to "strongly disagree") and receiver operating characteristic (ROC) curves [132, 133, 134].

For example, Choyke *et al.* [81] used areas under the ROC curve, as well as κ (kappa) statistics to compare five methods for depicting hepatic MR angiograms. κ statistics were also used by Hans *et al.* [135] to measure interobserver agreement in the visualisation of the inner ear using various ray–casting and surface rendering algorithms.

A recent study by Stokking *et al.* [136] serves as another example where differing visualisation techniques were compared under controlled conditions. Three 2–D displays (a SPECT only display; a SPECT–MRI adjacent display; and a SPECT display superimposed by the 2–D cortical contour segmented from MRI) and one 3–D rendering method ("Normal Fusion", whereby cortical SPECT activation is mapped onto the segmented cortical surface along the direction of the surface normal) were presented to five clinical observers. The observers assigned pre–selected hot and cold spots on the functional image to their anatomical locations, specifying lobe and gyrus, along with a confidence rating for each

observation. Mean confidence rating and inter–observer agreement were used to order the four displays in terms of clinical utility. Although the superiority of the Normal Fusion technique may not be surprising considering the inherent additional pre–segmented three–dimensional anatomical information displayed, the adjacent SPECT–MRI and SPECT with contours displays showed no significant differences. This indicates that simply overlaying anatomical images with functional ROIs is not necessarily optimal for identifying the precise anatomical correlates of functional activation in 2–D displays.

Charland and Peters [137] conducted a series of experiments with a view to optimising a task that required positioning a 3–D cursor in depth. The stimuli consisted of stereoscopically displayed digital subtraction angiograms (DSA) of cerebral vasculature. The task in this multi–observer study was to place a cursor in the same depth plane as targets identified on the images. Among the salient parameters emerging from the results were (eye-to-monitor) viewing distance, contrast of the cursor with the vessels and with the background, and inclusion of a size versus distance cue in the cursor.

In a study intended to compare the difference in performance between mono- and stereoscopic viewing as well as the effect of contrast and viewing distance, Tendick *et al.* [138] used a laparoscopic imaging system in a knot–tying task. They found no difference in performance between stereo and mono views, but did show, in agreement with Charland and Peters, that "good contrast was found to be most important at longer working distances." From a vision research standpoint this is not a surprising discovery, considering the interrelationship of contrast and spatial frequency in visual acuity. At higher spatial frequencies contrast sensitivity, or the ability to resolve details of an image, decreases with increasing spatial frequency.

Even though the last section presented a few psychophysical studies in the context of medical image visualisation, validation is more typically described in terms of clinical

utility as reported anecdotally by a clinician, or in the context of a limited number of cases where a diagnosis is confirmed by biopsy or at surgery [139, 140, 141].

2.8 Discussion

A large number of unsolved problems remain relating to visualisation of volumetric data in medicine. Although further increases in computational power can certainly be expected, to the extent that static (or low frame–rate dynamic) images are still in widespread use, there is a need for continued investigation into the interaction of clinical observers with single–view (or non–dynamic) multimodal images. Höhne and Pommert [142] cite the "large number of rather technical parameters for controlling segmentation and shading" as an obstacle to the widespread acceptance of novel rendering techniques by the medical community. Similarly, Kaufman *et al.* [143] in outlining the complex parameter–space associated with volume rendering observe that "by far the most commonly used method for finding good transfer functions [which assign values to properties of the object being visualised, such as colour and transparency] is 'trial and error'."

An idea which is slowly emerging in light of the myriad of types of images available to clinicians, is that of context–dependent multimodal visualisation. Clearly not all the different images are required simultaneously and throughout a surgical procedure. Some investigators [144] are therefore beginning the important task of scripting the steps involved in different surgical cases. At each step, the surgeon defines the task along with the relevant images to build up the full script for the particular type of procedure. This mode of operational planning should enhance the clinical utility and effectiveness in incorporating images from different modalities. After all, the goal of multimodal image–guided surgery is not merely to display all available information in a single view, but rather to present it in

a clinically relevant and understandable manner. As Kundel [145] has pointed out, "A good display is one that emphasizes the diagnostic features and minimizes the camouflaging effects of the nondiagnostic anatomic features."

Given the number of "information–guided" therapy systems with a large number of common functionalities that are available, the single most salient point of distinction may indeed turn out to be user interface. Clinicians, and specifically surgeons, "must be able to focus on the patient and the procedure rather than on devices...The machine interface...needs to be designed in a way that conforms intuitively to the physicians' needs, not vice versa." [146]. Unfortunately there are few investigations [147] into novel input devices that are appropriate for the operating room and can be used by surgical staff. And with the majority of today's IGS systems requiring a dedicated operator (i.e. technician or engineer) in the surgical theater, it is clear that the majority of surgeons are far from interfacing directly and independently with the computer.

Chapter 3 Preview

While the scope and complexity of different rendering algorithms is ever increasing, relatively little work is dedicated to the rigorous psychophysical study of the nature of the interaction of the human visual system with the images generated. Clearly, given the complexity and richness of medical images, especially those integrating multiple imaging modalities, this is not a simple task. Nevertheless, an extensive psychophysical literature exists, exploring different aspects of human visual perception, which provides the beginnings of a course of investigation. Moreover, a rich array of psychometric and evaluative techniques has emerged not only to validate the fidelity of structures and volumes derived from imaging, but also to measure the performance of observers with respect to the various rendering methodologies.

The next chapter presents a series of psychophysical experiments using a manual segmentation editing task. An initial pilot study sets out to examine the performance of subjects using a 3–D pointing device, similar in functionality to the types of tracking devices used in clinical settings. Considering a particular use of transparency in the visualisation of medical images, the rest of the chapter's experiments aim to find an optimal set of visualisation parameters for this specific pre–operative clinical task. An earlier version of this report appeared in R. Kasrai, K. Finnis, D. G. Gobbi, Y. P. Starreveld, and T. M. Peters, "Advantages of visualisation and manual segmentation in 3–D: an empirical study," in *Workshop on Interactive Medical Image Visualisation and Analysis, MICCAI 2001*, Utrecht, Netherlands, October 2001.

Chapter 3

3–D Manual Segmentation of Brain Structures

"The gold standard for the value of any method of displaying a medical image is human performance on a medically relevant task."

—H. L. Kundel [145]

Abstract

This chapter presents a series of experiments examining the relative effects of two visualisation parameters (type of surface rendering and presence of stereo disparity) and one interaction parameter (2–D or 3–D mouse). These effects are measured as a function of subject performance on a manual segmentation task. The task involves interacting with and manipulating control points in order to completely capture a hyper–intense lesion on a grayscale volume within the segmentation surface. The volumetric stimuli were either ellipsoids or therapeutic lesions on a real MR volume. While the results indicate no significant effect of surface rendering type (wireframe vs. transparent vs. opaque), the subjects performed better when using a traditional 2–D mouse than when using a 3–D tracking device. A significant improvement of performance was also revealed in trials containing stereo disparity.

3.1 Introduction

One of the more common tasks in the visualisation of medical images is image segmentation. While medical images can be generated and acquired from a variety of sources (see §2.3), rarely are the regions of interest already identified or well circumscribed. In recent years most efforts in image segmentation have concentrated on either the automation of algorithms or their generalisation to a variety of modalities and applications. Perhaps justifiably, little or no attention has been given to manual techniques, since the ulimate goal of any digital image processing algorithm is to be free of supervision. Yet, the automation of such algorithms is a difficult problem unlikely to be fully and robustly solved in the near–term. Issues such as noise, sampling artifacts, spatial aliasing, variability within and across subjects, and the complexity in the shape of some objects of interest contribute to the difficulties involved [148]. Moreover, the noise and signal statistics of images vary significantly as a function of the imaging modality, hindering the generalisation of automatic algorithms originally developed for specific types of images. To the extent that such obstacles exist, there will continue to be a need for efficient and intuitive tools for user intervention or supervision.

In light of this, the goal of this study is to explore some of the factors implicated in visualising and interacting with medical images, with a view to finding a combination which gives rise to the best performance by subjects in a specific segmentation task. These factors include: two different modes of interacting with 3–D images (using either a 2–D or a 3–D input device); monoscopic versus stereoscopic visualisation; and, three types of surface rendering technique (see §2.5.1).

3.1.1 Manual and semi–automatic 2–D segmentation techniques

Segmentation algorithms may be roughly divided into two categories: image–based and model–based. Image–based (or low–level) techniques rely on operators that analyse local image statistics such as intensity (and/or gradient) and shape. Model–based methods, on the other hand, begin with templates incorporating global knowledge of shape or semantic context in an attempt to subsequently characterise the patient–specific geometry.

A number of both image- and model–based segmentation methods require some degree of supervison or interaction by a user (e.g. [149]). Some model–based methods may need a rough delineation of a few contours before beginning the optimisation process. High–contrast images such as CT scans of bony structures are easily segmented using a low–level thresholding operation where voxels are labelled either as background or object (e.g. bone) depending on user–defined ranges of intensity. Seeding or region growing is another low–level technique that requires the user to select a starting point within an object. Neighbouring pixels are then combined together, as the region grows out from the seed point (implemented, for example, in [150]), based on local image features such as texture or intensity.

3.1.2 2–D versus 3–D manual segmentation

Most classical tools designed for manual segmentation or user intervention rely on visualising 2–D images, which may be appropriate depending on the modality (e.g., ultrasound, x–ray) and acquisition technique. If, however, the structures of interest are three–dimensional in nature, segmenting them on 2–D cut–planes using contours (as shown in Figure 3.1) or by growing regions, has inherent drawbacks. Apart from the general problems associated with manual segmentation (reproducibility of results, operator bias, operator fatigue, etc.),

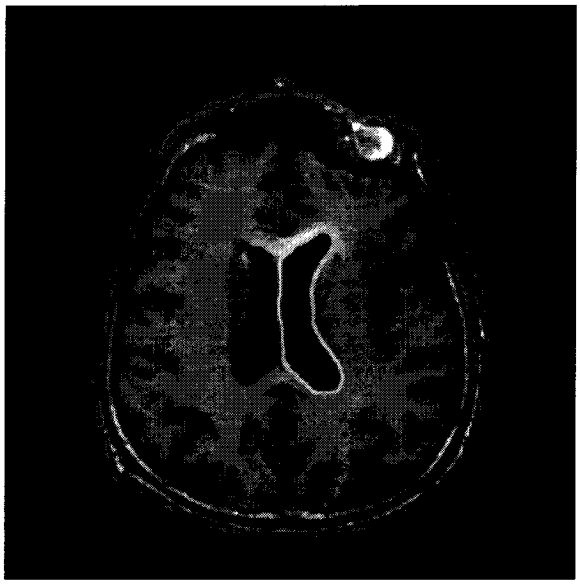


Figure 3.1: Segmenting a ventricle using a 2–D contour.

stacking the 2–D contours to reconstruct a 3–D surface or volume requires an additionnal processing step which may introduce further artifacts (see, for example, [49]).

On the other hand viewing objects in 3–D while ensuring the ensemble of depth cues including stereo disparity is self–consistent (see §2.6.1), is clearly more intuitive from a visualisation standpoint. In addition, a 3–D environment lends itself directly to interacting with and deforming 3–D structures, and does not force "the operator to view each 2–D slice separately to deduce and measure the shape and volume of 3–D structures" [148].

Although visualising the segmentation "field" stereoscopically as a 3–D rendered object is possibly advantageous, another limiting factor is that drawing or editing contours is typically performed with a conventional (2–D) mouse. A novel form of interacting with meshes is introduced here, using a true 3–D navigation device. Thus it is important to differentiate between 3–D visualisation and 3–D interaction, and to examine the relative and combined improvements in performance of the procedure using this approach.

3.1.3 Specifc goals of the study

The motivation for the present study is twofold. Firstly, despite the putative advantages of 3–D renderings in inherently 3–D tasks, there are few studies which set out to show rigorously the superior performance of users under these conditions. Although the results of this or indeed any task involving so many parameters can in general rarely be applied to other tasks, it is nevertheless worth quantifying the effects of a few obvious variables. Moreover, with the increase in computational power (and relative decrease in cost) of personal computers, the limiting factor remaining in improving human–computer interfaces is the inadequacy of our understanding of the human visual system. Though much headway has been made in basic human vision research, few of the psychometric techniques used in that arena have been employed in the applied domains. As such, the second goal of this study is to serve as an example of how to devise a simple protocol to investigate quantitatively not only the performance of users under different conditions, but also by extension the usability or qualitative advantages of one technique over another.

Given the current interest in minimally–invasive image–guided procedures and therapies, there is no shortage of software environments that incorporate a number of functionalities. In addition, some low- and high–level non–automatic segmentation algorithms require at least some degree of supervision. The interaction of users with these software applications, however, is not necessarily optimal, and there is always a trade–off between usability and the number of tools and functions one would wish to include.

3.2 System description

The different aspects of the visualisation platform (ASP) have been previously described [40, 151] and only the relevant components are presented here. This system uses

the OpenGL–based Visualization Toolkit[1]. The different components of the software are written in C++ and the Python programming language[2]. Rigorous object–oriented software design methods have been applied to ensure hardware independence and a standard application programming interface (API).

3.2.1 Real–time nonlinear deformation of surfaces

The visualisation system used in this study incorporates a 3–D thin–plate spline (TPS) transformation module [152]. This module, implemented as part of a generalised nonlinear transform class, makes use of landmark (control) points to interactively modify a 3–D TPS transformation (see also [153] for a description). Considering a plane in space as an example, the area local to the control points (which need not necessarily be on the surface) may be warped or stretched, much as a rubber sheet. As shown in Figure 3.2, users may place control points with the mouse and drag them around to visualise in an interactive manner the result of applying the transformation to a 3–D image. The image may consist of either data along a regularly–sampled structured grid (e.g. an MRI volume), or a polygonal mesh describing a surface, or both.

3.2.2 Visualisation

Apart from the standard monoscopic 2–D display of the viewing volume, the software system supports stereoscopic visualisation using either liquid crystal glasses[3] or a liquid crystal polarizer on the monitor. In all the experiments described here, subjects wore liquid crystal goggles in both stereo and mono conditions with the difference that during the mono conditions the stimuli contained no stereo disparity. For rendering scenes containing

[1] Available at `www.kitware.com`.

[2] Available at `www.python.org`.

[3] CrystalEyes, Stereographics, San Rafael, CA, USA

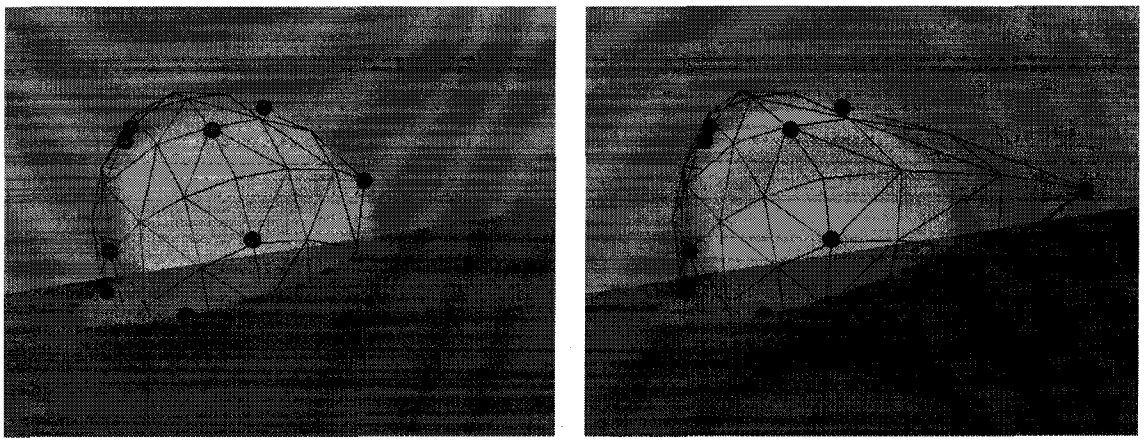


Figure 3.2: Deforming a wireframe surface using 3D spline control points.

mixtures of multiple transparent and opaque surfaces, care was taken to depth–sort the polygons in the correct order for proper alpha–blending ([105]; see also §1.6.1). Surfaces can also be displayed either by rendering only the underlying mesh or by Gouraud–shading polygons and setting a uniform opacity as desired.

3.2.3 Interaction

Input devices

Users have the option of interacting with the display using either a traditional (2–D) computer mouse or a true 3–D tracking device. Among the 3–D tracking systems tested and supported are the POLARIS, the Flock of Birds, and the 3–D Logitech mouse. Optical tracking systems, including the POLARIS, have been found to have a operating accuracy and precision of about 1 mm [125, 126]. The Flock of Birds system has a reported [154] accuracy of $\approx$1.8 mm for translation and 0.5° for rotations. The stated accuracy of the Logitech mouse is 2% of the transmitter–receiver distance (on the order of 1 mm in this case). For the experiments described here the 3–D Logitech system was used, as shown in Figure 3.3. The hardware uses ultrasound emitted from transducers to track the 3–button mouse equipped with small microphones.

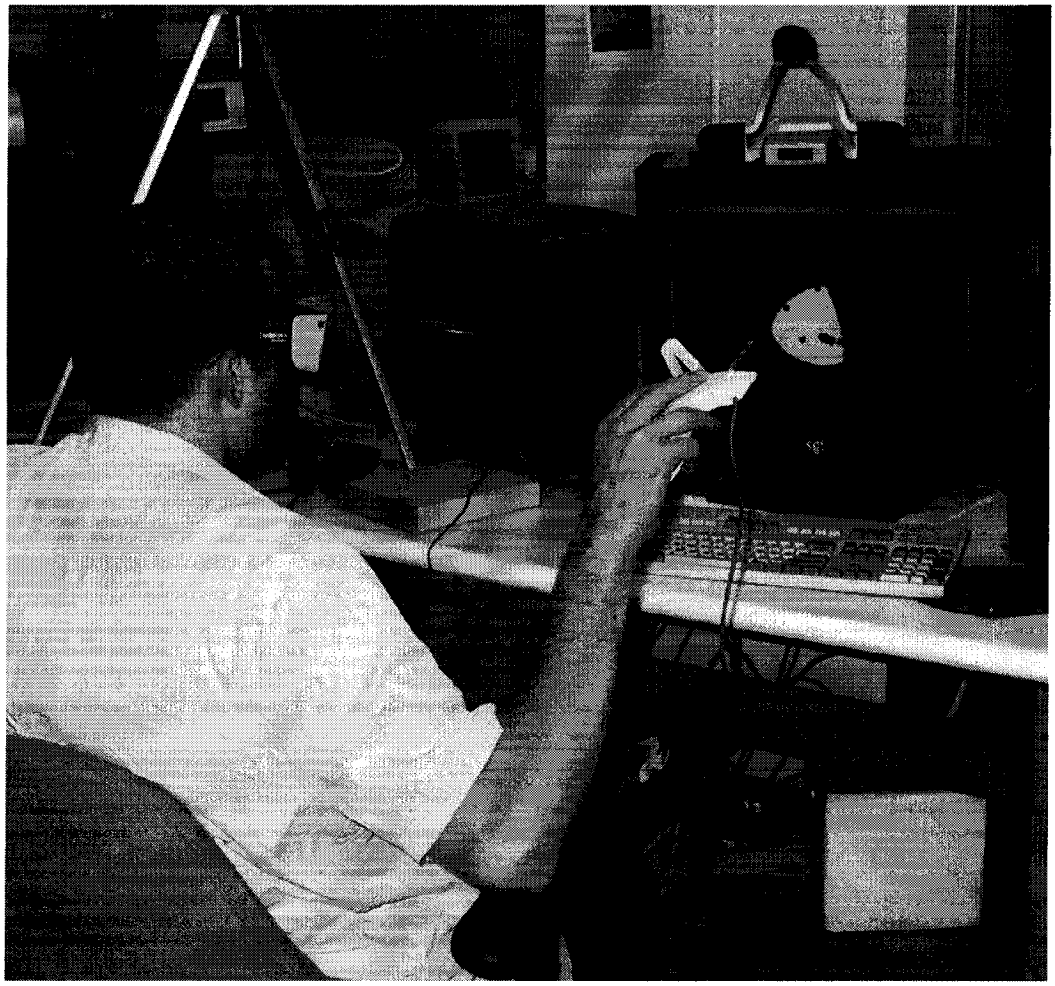


Figure 3.3: Subject using the 3–D mouse and wearing stereo goggles.

Camera

A general problem associated with navigating a 3–D scene using a 2–D translational input device is that of mapping the 2–D mouse position to points in 3–D space, simulating different camera viewpoints. Because of the many–to–one nature of the 2–D to 3–D mapping, the trajectory of these camera viewpoints is path–dependent, such that the reconstruction of the virtual camera's movements does not depend solely on the start and end positions of the 2–D cursor on the image. Even though this may seem non–intuitive, most users are able to adapt rapidly.

When the camera rotation functionality was selected, left–right translations of the 2–D

mouse resulted in rotations of the camera about the vertical axis, and forward–backward translations rotated the camera about the left–right horizontal axis. Camera rotations using the 3–D mouse enjoy the benefit of an additional degree of freedom as compared to the 2–D mouse. All rotations with the 3–D mouse were computed so as to produce mirrored rotations of the objects in the scene. For example, (clockwise) rotations of the 3–D mouse about the vertical and left–right horizontal axes naturally resulted in reverse (counter clockwise) rotations of the scene about the same axes.

Navigational functionalities such as zooming and rotating the camera were bound respectively to the right and left buttons on both 2–D and 3–D mice. Panning of the camera, although available in the laboratory version, was not permitted in the experimental version, and the focal point of the camera was fixed to the centre of the segmentation target.

The position of the current point was indicated in 2–D mouse mode by a cross–hair cursor which travelled along the closest surface falling on the line between the camera and the focal point. In 3–D mouse mode, a simple rendering of a probe indicated the position of the pointer tip. It is important to note that the perceived position of the 3–D probe (even during stereoscopic visualisation) did not correspond to the physical position of the 3–D mouse.

Control points

The 3–D TPS control points were selected and dragged around using the middle button of the 3–D mouse. In 2–D mouse mode the spline control points were selected and moved along the surfaces of the objects in the scene also using the middle mouse button. In 2–D mode, the control points may also be moved towards or away from the centre of rotation (the camera's focal point) using a keyboard modifier (`Shift` or `Ctrl`) along with the middle mouse button. In both 2–D and 3–D mouse modes the control points changed

colour when the cursor or pointer was close enough to be selected and moved.

Cut–planes

Although the 3–D mouse may also be overloaded with additional functionality using key-board modifiers in combination with button presses, this was explicitly avoided in order to leave the user's other hand free (e.g. during intraoperative use) for other instruments or devices. Thus, moving the three mutually orthogonal (axial, sagittal, coronal) cut–planes was also accessed by the middle mouse button, such that if the pointer tip was within a certain threshold of one of the planes, that plane could be selected and dragged when the middle button was pressed. When using the 2–D mouse, the cut–planes (as well as the control points) could be dragged while depressing the middle mouse button.

3.3 Experimental setup

An example of the type of application which would benefit from the functionalities of the system described above is the segmentation of lesions on post–operative clinical images. Specifically, the Image–Guided Surgery Laboratory has a database of images containing therapeutic lesions created using radiofrequency thermocoagulation probes (see [68] for a detailed overview). These lesions, as shown in Figure 3.4, are created in deep–brain structures such as thalamic nuclei to relieve Parkinsonian symptoms. Subsequently, the segmentation of these lesions on post–operative MR images allows clinicians to correlate clinical outcome with the position and size of the lesions. Segmentation of other objects such as tumours or demyelination lesions due to multiple sclerosis in longitudinal studies is useful in observing and quantifying the progress of disease [155].

The following experiments consist of two separate types of stimuli. The first is a simple

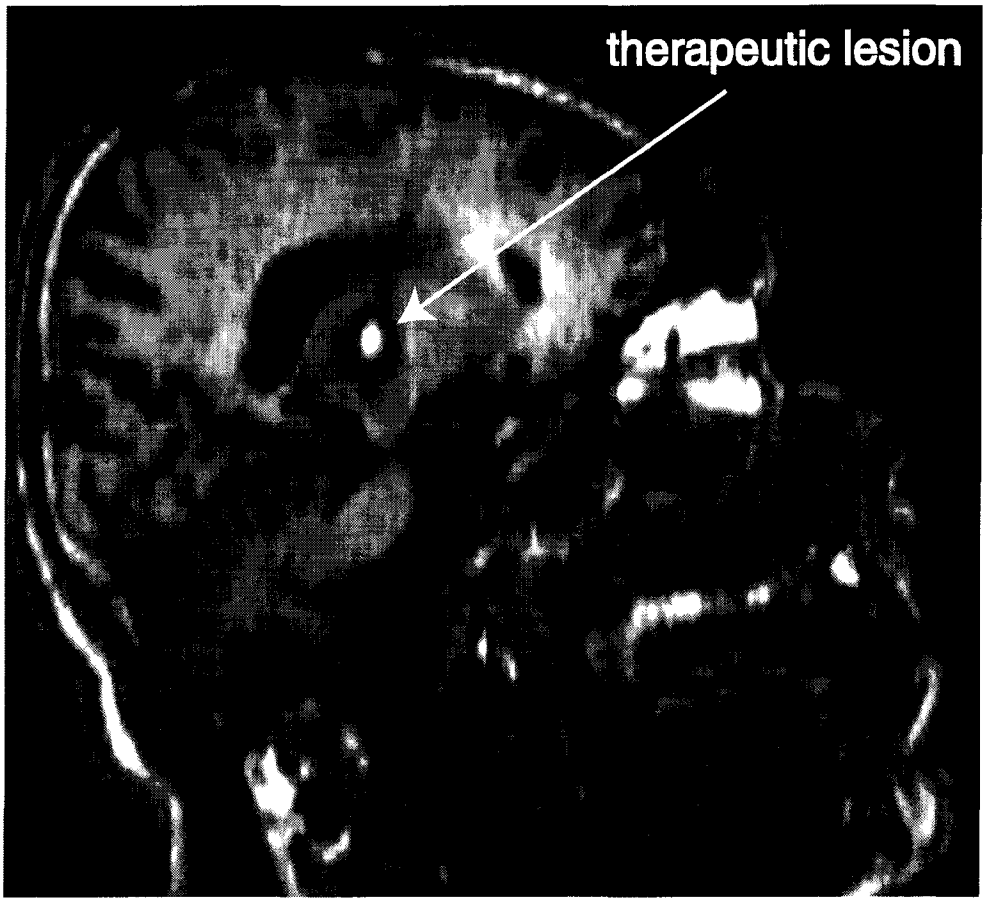


Figure 3.4: Sagittal MRI section showing hyper–intense post–operative lesion.

'white' ellipsoid in a 'black' surround, which resembles the clinical lesions in its high contrast with the surround, but only roughly in its shape. In the second type, a post–operative MR volume of a patient who had undergone functional surgery was used. While the first type of stimulus has the advantage that the exact geometry of the target region is known, the latter stimulus type is more representative of the kind of images used in clinical practise.

3.3.1 Subjects

Subjects with normal or corrected–to–normal vision were allowed some time (1–2 hours) to become familiar with the 2–D and 3–D interfaces. Subjects were selected who had

prior experience with application software for visualising, interacting with, and segmenting digital medical images, but were kept naïve with respect to the experimental hypotheses.

3.3.2 Generation of stimuli

Ellipsoid phantom

A digital phantom (a binary–valued *volume*) containing an ellipsoid with non–equal principle axes was constructed, along with another ellipsoid *surface* of slightly different geometry (see Figure 3.5). The two ellipsoids' axes were rotated in space with respect to each other such that the two shapes did not correspond perfectly in any of the three cut–planes. The semi–axes of the target volume and initial surface were $7.5 \times 8.5 \times 10$ and $9 \times 10 \times 8$, respectively. The subjects' task was to drag around 20 pre–positioned control points to match the surface to the underlying volume within a 5–minute time limit. For the purposes of the following experiments, subjects did not have the option of placing additional control points. At the end of the allotted time for each trial, the subject's manually–edited mesh was written to disk for analysis.

Post-operative lesions

These images contained a hyper–intense deep–brain therapeutic lesion bounded by a hypo–intense region of edema, as shown in Figure 3.6. The subjects' task was to capture as much as possible of the lesion by deforming the initial surface (in this case a sphere centred on the lesion) as above, using the 20 control points. The surfaces segmented by the subjects were compared to a segmentation performed by an expert in deep–brain neuroanatomy, using the same interface and a 2–D mouse, but with an unlimited number of control points. The radius of the roughly–spherical lesion was about 5 mm.

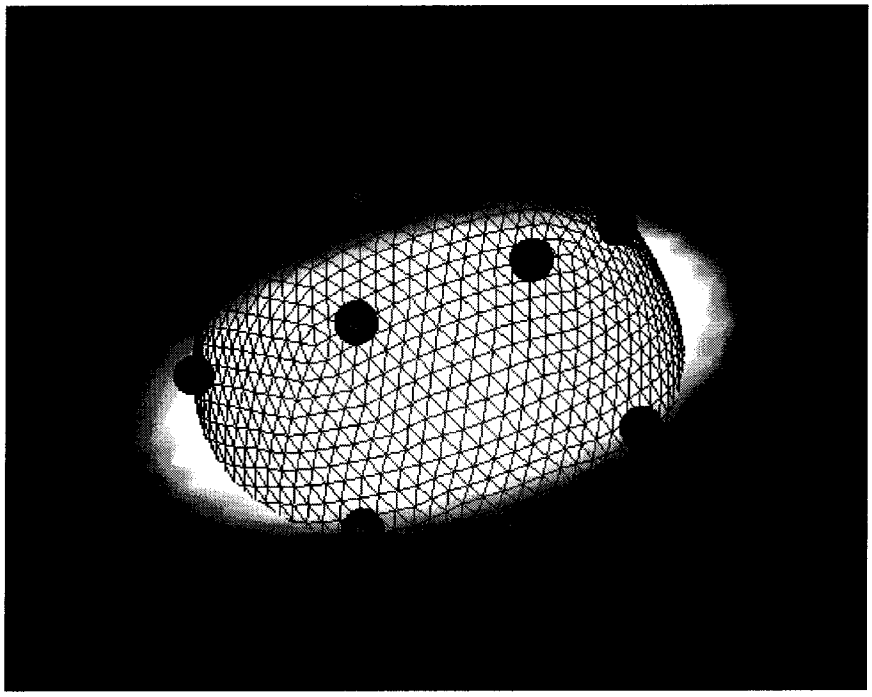


Figure 3.5: Ellipsoid stimulus.

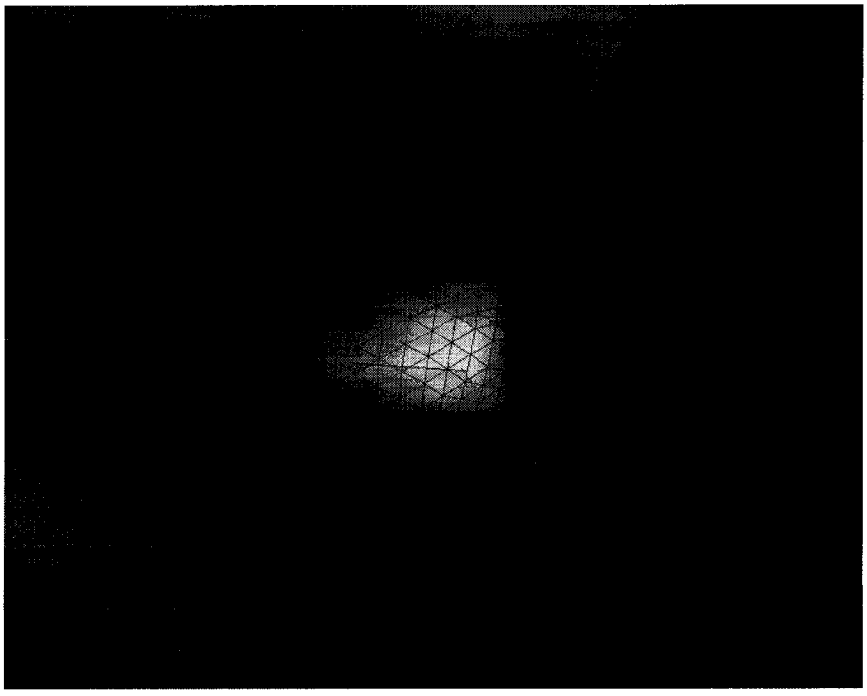


Figure 3.6: Post–operative lesion stimulus.

3.4 Analysis

For experiments involving the ellipsoid stimuli, the resultant meshes were compared with a reference surface (not displayed during the experiments) corresponding exactly to the boundary of the test volume. The nodes of the polygonal surfaces generated from each trial were matched to the reference surface using a nearest–neighbour search algorithm. Node–wise difference maps between the reference surface and each of the subjects' meshes were generated [156, 157]. The mean and maximum difference vectors for each mesh were extracted and their magnitudes (the ℓ_2–norm) used as a figure of merit indicating the subject's accuracy in each trial. If the two surfaces were significantly different in shape the mean difference vector would have a larger magnitude. A long maximum difference vector, on the other hand, reflects significant deviations from the reference surface at a particular region of the surface.

The mean and maximum magnitudes were then averaged across all of a particular subject's repeated measurements. In the case of single measurements, the data were averaged across subjects.

3.5 Pilot experiment:

Interaction using 2–D vs. 3–D mouse

The first experiment carried out aimed to test (among other factors) the effect of (2–D vs. 3–D) interaction mode. The combination of (2–D vs. 3–D mouse) × (mono vs. stereo viewing) × (Gouraud–shading with opacity of 0.7 vs. 1.0 vs. wireframe rendering) resulted in 12 conditions. Using only the ellipsoid stimulus, three subjects (JM, PS, and RK) performed 12 trials each.

The raw results, showing the deviation vector magnitudes (both the average over each segmentation surface, as well as the maximum), are displayed in Table 3.1. Note that in this table (and the following similar ones) all values are dimensionless distances. For a sense of the scale of these error distances, refer to §3.3.2 describing the extent of the corresponding target surfaces.

The mean and maximum deviation vectors were averaged across subjects. A three–factor analysis of variance (ANOVA) test, summarised in Table 3.2, revealed an enhancement of overall performance in stereoscopic compared to monoscopic viewing, but a degradation of performance in 3–D compared with 2–D mouse interaction.

Given that interaction using the 3–D mouse did not improve performance, the following two experiment focussed on the visualisation aspects of the study using only the 2–D mouse.

3.6 Experiment 1: Segmentation of ellipsoids

This experiment involved three subjects (AD, HW, and RK) who performed three repetitions of six conditions: (mono vs. stereo viewing) × (Gouraud–shading with opacity of 0.7 vs. 1.0 vs. wireframe rendering). The raw results are shown in Table 3.3.

A three–factor ANOVA, summarised in Table 3.4, revealed a significant effect due to stereoscopic visualisation, but no effect of rendering type. The results also show a statistically significant difference between the subjects.

3.7 Experiment 2:

Segmentation of post–operative lesions

The same subjects as in Experiment 1 performed segmentations of a post–operative hyper–intense lesion under the same six experimental conditions as above. The raw results are shown in Table 3.5. The ANOVA results, displayed in Table 3.6, show a significant improvement due to stereoscopic visualisation in the maximum deviation vector magnitudes, as well as a significant difference between subjects.

3.8 Discussion

The experimental results illustrate a number of points worth highlighting. The assumption that interaction using a 3–D tracking device afforded inherent advantages was not borne out by the data from the pilot experiment. There are a number of factors which may have contributed to this result. Firstly, users have had years, if not decades, of experience using traditional 2–D mice. Issues such as a limited workspace (*viz.* the mouse pad) and mapping movements in the horizontal plane to cursor movements in a vertical plane (the monitor) are either not fundamental problems or have been resolved through years of practice using the device. 3–D tracking systems, on the other hand, are relatively new additions to the host of available input devices. A fairer experimental design would either have allowed subjects months or years to become as familiar with the 3–D mouse as they are with 2–D mice, or have attempted to find subjects who were as unfamiliar with traditional computer mice as they were with 3–D devices. The latter may well be an exercise in futility more than 20 years after the introduction of the first personal computers. Even if such subjects could be found in North America, they would likely have little experience with computers, let alone

medical imaging applications.

Another concern from the human factors viewpoint is the lack of limb support when using 3–D devices. It was demonstrated in the pilot study that it is quite difficult to achieve positional accuracy and precision while holding one's arm out in space. Attempts were made to limit the required physical workspace to less than 1 m^3 by selecting an appropriate gain factor for zooming, rotating, and translating the camera, while subjects rested their elbows either in their laps or on the backrest of another chair. It is unlikely that using this set–up subjects were as comfortable as they were using a standard 2–D mouse. As a result, the error in positioning due to arm fatigue or normal tremor is much greater relative to the positional accuracy of the 3–D mouse. This is despite the fact that the desired target position of a control point is determined through visual feedback and not by the absolute position of the pointer in space. It is thus conceivable that given enough time the subjects could have performed equally as well as when using the 2–D mouse. If so, the factor limiting their performance was arm instability rather than inferior target planning.

If inferior performance using the 3–D mouse was in fact due to differences in the apparatus, an appropriate control condition might have been to program the 3–D mouse to be functionally identical to the 2–D mouse, with the exception that it would be operated in 3–D space. In other words, the subjects would hold and operate the 3–D mouse as before, but without one of the degrees of freedom (e.g. twist rotations, or translations in the Z–direction). The results from this condition would, in principle, distinguish between performance differences due to the particular apparatus and those due to perceptual or motor factors.

The data revealed a consistent improvement in performance in trials where the scene was rendered stereoscopically. This is a result which would be expected in a task that requires an appreciation of the three–dimensional position and shape of points and surfaces.

As is well known, stereo disparity is a powerful depth cue which, if used in a consistent manner with other monocular depth cues (such as occlusion, parallax, and kinetic depth effect occuring when the scene is rotated [117]) can convey a compelling sensation of depth. Again, it is possible that subjects would have performed equally well without the stereo disparity cue provided unlimited time. Given the 5–minute time limit, however, the trade–off between speed and accuracy is revealed by a statistically significant drop in performance in monocular trials.

The data failed, however, to show a significant effect of surface rendering method. This is a surprising result, considering one might have expected performance of the task to have been facilitated by transparent surfaces. The task can essentially be boiled down to the simultaneous matching of contours. As the subject moves the cut–plane through the volume, at each position a comparison is made between the 'white–black' contour of the underlying volume (whether ellipsoid or post–operative lesion) and the portion of the blue surface which intersects the three cut–planes. The subject then manipulates the position of the nearest control point to overlap the two sets of contours as closely as possible. It would follow then, if the blue surface is opaque and occludes the volume's contour, that the surface must be pushed past the volume's contour to make it visible, and then pulled back again to the overlapping position. If the blue surface is transparent (or rendered as a polygonal mesh) it can be moved directly to the appropriate position, thus saving time and increasing accuracy.

Nevertheless, the data showed consistently that in this task there were no gains in performance associated with transparency. One possible explanation might be the following, illustrated in Figure 3.7. Assume that at a particular position of the cut–plane the two contours are matched and the blue surface is Gouraud–shaded but transparent. Since the underlying volume is more or less binary–valued (either 'bright' or 'dark'), the increase in

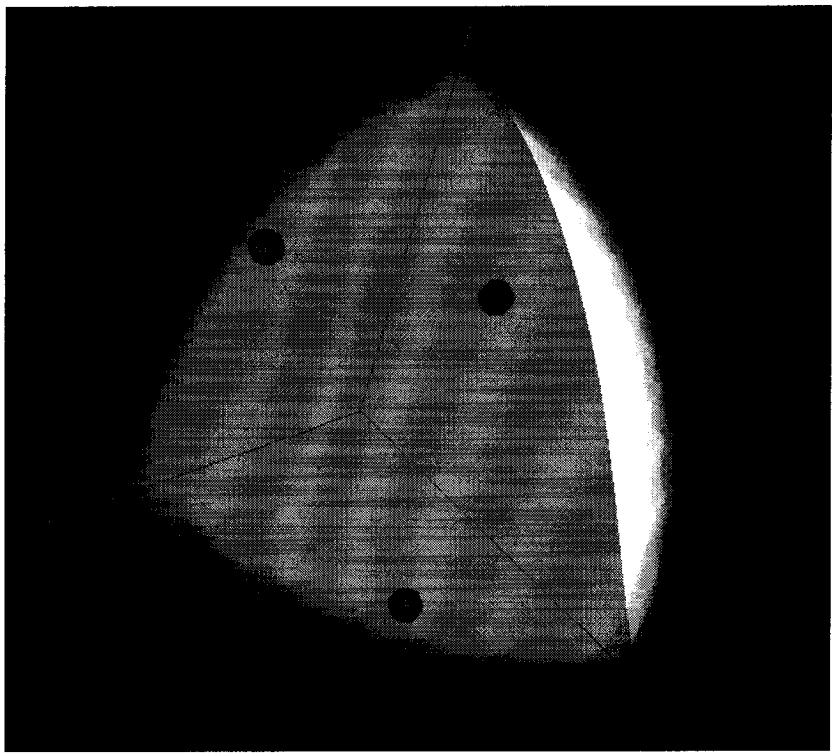


Figure 3.7: Region where the underlying lesion may be masked by the transparent surface.

brightness of the overlapping region (bright lesion seen under the transparent blue surface) might be all attributed to a brighter blue surface. In that case the subject might perceive no lesion at all, and only a brighter transparent blue surface, prompting the subject to move the surface to (re)detect the underlying contour, as in the case of an opaque surface.

3.9 Summary and future work

In this chapter, a novel software system has been introduced which allows for visualisation of and interaction with images, both in 2–D and 3–D. An experiment has been described which evaluates the effectiveness of different visualisation and interaction parameters using a manual segmentation task. The results show that while comparable segmentations can be achieved using a 3–D tracking device, neither did the subjects feel as comfortable nor did they show superior performance when compared with performing the same tasks using a

standard 2–D mouse. The subjects did show an improvement in performance with stimuli containing stereo disparity, but no effect of surface rendering type was revealed.

Extensive experience using this system with clinical images in the laboratory leads to the following general observations and suggestions. Although using control points to interact with surface meshes is intuitive and convenient, depending on the starting surface many such control points may be required to match the surface locally to the underlying volume. These control points may in turn occlude the volume of interest which may be quite small relative to the total volume. In these cases a feature to toggle the visibility or opacity of the control points as well as the visibility of the segmented surface is useful. Moreover, though thin–plate splines were used to warp the space (and thus the nodes in the meshes), other nonlinear 3–D schema (including direct interaction with a subset of mesh nodes) could equally well have been implemented. Thus for complex shapes the system as described is better suited to *editing* the results of automatic or semi–automatic algorithms rather than actual segmentation from a generic starting surface (e.g. a sphere or ellipsoid).

3.10 Acknowledgements

The invaluable expertise of members of the Image–Guided Surgery Laboratory at the John P. Robarts Research Institute (David Gobbi, Kirk Finnis, and Yves P. Starreveld) in developing the surgical planning and visualisation platform is gratefully acknowledged. In addition, Jeremy Gill provided assistance in developing a methodology for comparing polygonal surfaces.

	mono			stereo			
subjects	JM	RK	PS	JM	RK	PS	
	average						
opaque	1.85	1.80	1.48	1.37	1.63	1.34	2D mouse
	1.61	1.94	1.51	1.68	1.69	1.49	3D mouse
transparent	1.70	1.49	1.48	1.47	1.55	1.55	2D mouse
	1.58	1.95	1.57	1.52	1.92	1.55	3D mouse
wireframe	1.54	1.58	1.64	1.46	1.68	1.43	2D mouse
	1.45	1.71	1.65	1.53	1.69	1.59	3D mouse
	maximum						
opaque	5.39	4.76	5.41	5.68	5.62	4.99	2D mouse
	7.41	6.09	6.38	5.16	5.82	5.40	3D mouse
transparent	7.85	5.38	5.66	5.43	5.53	5.71	2D mouse
	6.12	6.44	6.57	6.11	5.41	5.64	3D mouse
wireframe	6.27	5.81	5.65	5.95	5.50	4.99	2D mouse
	5.62	6.01	7.09	6.49	5.79	5.97	3D mouse

Table 3.1: Raw data from pilot experiment: average and maximum error (deviation) vector magnitudes.

	average p	max. p
2–D vs. 3–D mouse	0.10	0.03
rendering type	0.83	0.39
stereo vs. mono	0.15	0.02

Table 3.2: Results of pilot experiment: p-values for average and maximum deviation vector magnitudes.

subjects	RK			AD			HW			
trial #	1	2	3	1	2	3	1	2	3	
	average									
opaque	1.16	0.97	0.96	0.65	0.59	0.88	0.75	0.79	1.05	mono
	0.91	0.83	0.82	0.44	0.66	0.92	0.57	0.69	0.72	stereo
transparent	1.15	0.95	0.90	0.58	0.74	0.78	0.74	0.83	0.86	mono
	1.10	0.92	0.93	0.52	0.78	0.77	0.62	0.79	0.86	stereo
wireframe	1.57	0.99	0.86	0.69	0.64	0.65	0.67	0.79	0.68	mono
	0.92	0.83	0.77	0.54	0.50	0.65	0.75	0.79	0.74	stereo
	maximum									
opaque	2.65	1.86	1.63	1.64	1.57	1.63	1.66	2.37	1.80	mono
	1.69	1.61	1.66	1.38	1.67	1.57	2.14	1.86	1.98	stereo
transparent	2.57	1.94	1.49	1.56	1.69	1.52	1.94	1.81	1.61	mono
	1.84	1.66	2.10	1.26	1.70	1.56	1.67	1.64	2.45	stereo
wireframe	1.57	1.60	1.80	1.76	1.75	1.50	2.00	2.03	1.70	mono
	1.88	1.53	1.71	1.36	1.28	1.55	1.67	1.67	1.74	stereo

Table 3.3: Raw data from ellipsoid stimuli: average and maximum error (deviation) vector magnitudes.

	average p	max. p
stereo vs. mono	0.02	0.15
rendering type	0.64	0.31
subjects	< 0.01	0.001

Table 3.4: Results of Experiment 1: p-values for average and maximum deviation vector magnitudes.

subjects	RK			AD			HW			
trial #	1	2	3	1	2	3	1	2	3	
	average									
opaque	0.50	0.53	0.43	0.42	0.45	0.33	0.40	0.40	0.68	mono
	0.51	0.57	0.36	0.34	0.43	0.38	0.46	0.60	0.62	stereo
transparent	0.63	0.56	0.49	0.32	0.34	0.41	0.54	0.59	0.65	mono
	0.56	0.44	0.41	0.37	0.34	0.37	0.55	0.49	0.65	stereo
wireframe	0.69	0.62	0.67	0.28	0.28	0.48	0.59	0.42	0.61	mono
	0.59	0.54	0.50	0.30	0.33	0.34	0.53	0.43	0.49	stereo
	maximum									
opaque	1.36	0.93	0.89	1.27	1.65	1.13	0.98	1.28	1.52	mono
	1.56	1.33	1.02	0.88	0.91	1.12	1.31	1.62	1.64	stereo
transparent	1.58	1.48	1.58	0.98	0.89	1.27	1.23	1.61	1.59	mono
	1.43	1.03	1.00	0.87	0.93	0.79	1.18	1.62	1.53	stereo
wireframe	1.43	1.58	1.95	0.88	0.77	1.00	1.62	1.13	1.66	mono
	1.41	1.51	1.11	0.90	0.81	1.00	1.45	1.00	1.40	stereo

Table 3.5: Raw data from lesion stimuli: average and maximum error (deviation) vector magnitudes.

	average p	max. p
stereo vs. mono	0.16	0.07
rendering type	0.73	0.99
subjects	$\ll 0.001$	$\ll 0.001$

Table 3.6: Results of Experiment 2: p-values for average and maximum deviation vector magnitudes.

Chapter 4 Preview

The previous chapter examined the role of transparency in an applied context. In the next chapter a significantly simpler stimulus is presented in order to find the model which best represents subjects' internal model of achromatic transparency. Chapter 4 is an adaptation of the following article:

R. Kasrai and F. A. A. Kingdom, "The precision, accuracy, and range of perceived achromatic transparency," *Journal of the Optical Society of America (A)*, vol. 18, no. 1, pp. 1–11, 2001.

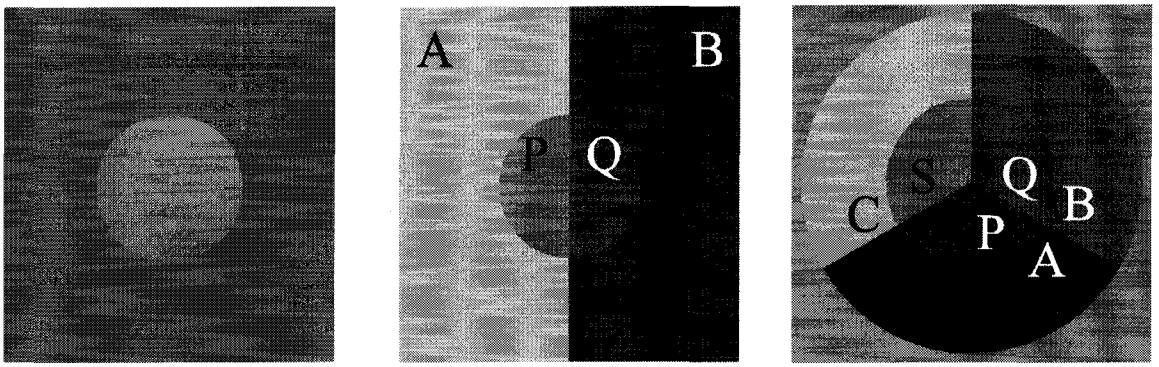
Chapter 4

The precision, accuracy, and range of perceived achromatic transparency

"At the thought of statistics, the Collector, walking through the chaotic Residency garden, felt his heart quicken with joy... For what were statistics but the ordering of a chaotic universe? Statistics were the leg–irons to be clapped on the 'thugs' of ignorance and superstition which strangled Truth in lonely byways." —J. G. Farrell, *The Siege of Krishnapur*

Abstract

How accurately do human observers perceive the properties of an achromatic transparent filter with both reflective and transmissive components? To address this question a novel six–luminance stimulus was employed, consisting of three transparent layer luminances set against three background luminances, which satisfied the conventional constraints of perceptual transparency. In one experiment, subjects adjusted one of the three layer luminances to complete the impression of a uniform transparent disk. It was found that the luminance–based formulation of Metelli's episcotister model, and a model based on ratios of Michelson contrasts best predicted the subjects' settings, which were both accurate and precise. In another experiment, pairs of stimuli selected from a range with various values of the adjustable layer luminance were presented in a series of forced–choice trials. A novel implementation of the pair comparisons method was employed to recover the distribution which describes each subject's preference pattern. Results showed that there exists a reasonably wide range of stimuli which give rise to at least some degree of perceived transparency.



(a) **(b)** **(c)**

Figure 4.1: Transparency stimuli. (a) Two–luminance stimulus does not provide a strong transparent percept; (b) four luminances simulating an episcotister in the classical bi–partite display; and, (c) six–luminance stimulus with mid–grey background.

4.1 Introduction

4.1.1 The reflectance formulation

The perception of transparency has intrigued vision researchers for at least a hundred years. Investigators from Helmholtz [19] to Metelli [6] have proposed different theories to explain the phenomenal experience associated with perceiving a background through a transparent surface, and employed different techniques for testing and proving those theories. Metelli used an episcotister—a rotating opaque disk with open sectors of varying size giving the effect of different illusory opacities. When the alternation of the episcotister's open and opaque sectors exceeds the critical flicker frequency, it appears as a transparent disk through which a bipartite background can be seen, as in Figure 4.1(b). Metelli's observations led to the formulation of his so–called episcotister reflectance model [6, 11]

$$\begin{aligned} p &= ta + (1-t)r \\ q &= tb + (1-t)r, \end{aligned} \tag{4.1}$$

where r is the reflectance of the opaque episcotister surface, t is the fraction of the disk

that is open, a and b are the reflectances of the background surfaces, and p and q are the resulting layer reflectances[1], such that $0 \leq \{a, b, p, q, r, t\} \leq 1$. The episcotister reflectance model is an expansion of Talbot's law of colour mixture (see also §1.5.2), $ax+(1-a)y = z$, where x and y are two colours expressed as reflectances, a and $1 - a$ are the relative linear mixing proportions, and z is the resulting fusion colour. In each case, reflectances from two sources contribute linearly to produce the desired mixture.

4.1.2 Illumination considerations and the luminance formulation

Metelli's original reflectance equations have since been reformulated by Gerbino [4] to allow for non–homogeneous illumination of the transparent layer and the background, since illumination considerations cannot be incorporated into a purely reflectance–based formulation. It has been shown the visual system is able to correctly perceive characteristics of surfaces in two differently-illuminated depth planes [158]. Similarly, an opaque surface and a transparent overlay may in general be differently illuminated. This consideration necessarily requires reformulating Eqs. 4.1 in terms of luminances rather than reflectances. The luminance formulation of any transparency model is also more amenable to psychophysical studies using CRT displays, where luminance (emitted pixel intensity) is the variable being manipulated. When luminance, instead of reflectance or lightness (perceived reflectance), is used as the independent variable to describe the necessary colour relationships for perceptual transparency, Eqs. 4.1 become

$$\begin{aligned} P &= taI + (1-t)rI' = tA + (1-t)rhI = tA + F \\ Q &= tB + F, \end{aligned} \tag{4.2}$$

[1]The terms "layer reflectance" and "layer luminance" are used throughout this paper to denote the absolute reflectance or luminance of the stimulus corresponding to the putative transparent layer, even though the variations in luminance would, in the physical case, be due to the background.

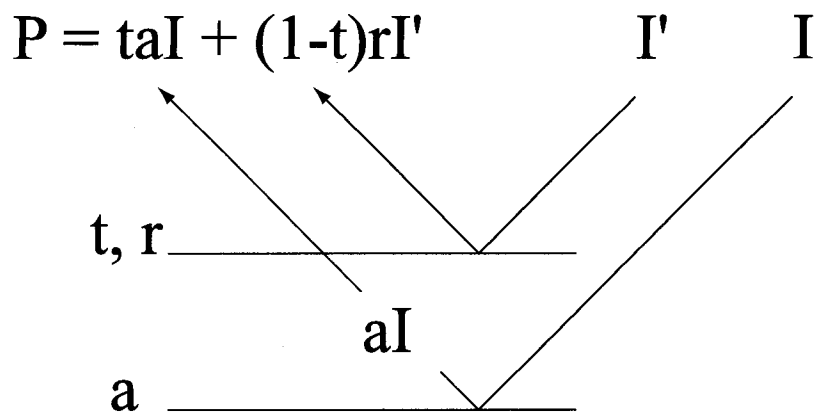


Figure 4.2: Optic array describing the episcotister luminance model. Note that the background is illuminated only directly and not through the layer.

where P and Q are the luminances of the portion of the image representing the layer (**P** and **Q**; see Figure 4.1), A and B are the luminances of the background (**A** and **B**), I and I' are the (generally non–equal) illumination components for the two surfaces with ratio h, such that $I' = hI$, and t and r are the transmittance and reflectance of the layer. The second term is collapsed into an overall additive component F, since with four known luminances (A, B, P, Q) and two equations, only two unknowns (t and F) can be extracted, and the product rhI cannot be disambiguated without additional assumptions or prior knowledge.

Figure 4.2 shows the optic array describing the luminance episcotister (LE) model from which Eqs. 4.2 are derived, which assumes that the background surface is illuminated directly and not through the layer. Others [20] have assumed that the background is illuminated exclusively through the transparent layer. From an ecological standpoint, isotropic diffuse illumination is the norm, and neither assumption is strictly valid. Rather, when the layer is close to or flush with the background, most of the light reflecting off the background (through the layer) will have already passed through the layer once, giving rise to a t^2 factor. This quadratic term, and multiple reflections between the filter and the background, are the defining characteristics of the filter model of Beck *et al.* [20] (the formulation for which

is given below; see Eqs. 4.4). When the layer is at a distance from the background, most of the light reflected from the background will not have passed through the layer first, and Eqs. 4.2 are more valid. It should be made clear that even though these models have a basis in optical physics, they should not be taken as physical models *per se*, rather as models of the assumptions the visual system makes when interpreting transparency stimuli.

Isolating t and F from the luminance–based Eqs. 4.2 gives

$$\begin{aligned} t &= \frac{P-Q}{A-B} \\ F &= \frac{AQ-BP}{A-B}, \end{aligned} \tag{4.3}$$

implying that given the four known luminances the visual system can extract the values of t and F. Even though it has been shown that some form of illusory transparency is possible with just three luminances under certain conditions [21, 159], this is generally not the case; four luminances are necessary to unambiguously render a given combination of t and F, as evident from Eqs. 4.3. Figure 4.1 shows the progression from two luminances (no phenomenal transparency) to four (the minimum necessary), to the six–luminance stimulus used in this study, the rationale for which is described below.

4.1.3 Measuring perceived transparency

It is unlikely that the visual system calculates the values of t and F explicitly, as one does not appear to have phenomenal access to the absolute values of these parameters when viewing a transparency stimulus. This is not to say that one cannot perceive a filter to be, say, more opaque or less reflective than another, but it would presumably be difficult to select a filter with, say, $t = 0.5$ from an array of filters with different t values. How might one measure the accuracy and precision with which human observers perceive and encode the properties of a simulated transparent surface? Accuracy is a measure of closeness to

an expected setting. This may be a setting according to the observer's internal model of transparency. Precision, on the other hand, is the reliability or reproducibility of the settings, which will be reflected in their variability. Consider the four–luminance display shown in Figure 4.1(b). Fixing both of the background luminances and one of the layer luminances leaves a range of the other layer's luminances that will produce valid transparencies, i.e. valid combinations of t and F. In other words, there is no unique solution with the four–luminance display, and therefore the four–luminance display is insufficient on its own as a tool for measuring perceived transparency. In their study of perceived transparency, Gerbino *et al.* [160] used *two* four–luminance displays (i.e., eight luminances in total): one a test stimulus, the other a comparison stimulus. Their subjects adjusted F (thus simultaneously altering both P and Q) on the test to perceptually match that of the comparison. The investigators showed that subjects were capable of making reasonably accurate settings according to the luminance version of the episcotister model.

A simpler and better suited stimulus for measuring transparency, however, is a six–luminance display,[2] one form of which is shown in Figure 4.1(c). Six luminances are the minimum necessary for a unique solution using just one adjustable luminance.[3] The stimulus in Figure 4.1(c) consists of three background and three layer luminances. With two of the background luminances and their two corresponding layer luminances appropriately set, a valid transparent surface can be created (as the minimum four–luminance requirement for transparency is satisfied). With the third background luminance then set arbitrarily as a test, the subject can adjust the remaining layer luminance to complete the transparent surface—to create a perceptually uniform t and F consistent with all three layer

[2]The principle of the six–luminance method has also been described by W. Gerbino in a conference presentation, though it was not mentioned in the associated conference abstract, [161].

[3]Additionally, in this display all the relevant luminances are in close proximity in central vision, such that no scanning of the image is required. In Gerbino *et al.*'s study, even though subjects were instructed to fixate halfway between the test and comparison patterns which were presented side–by–side, it is equivocal where the observers actually fixated.

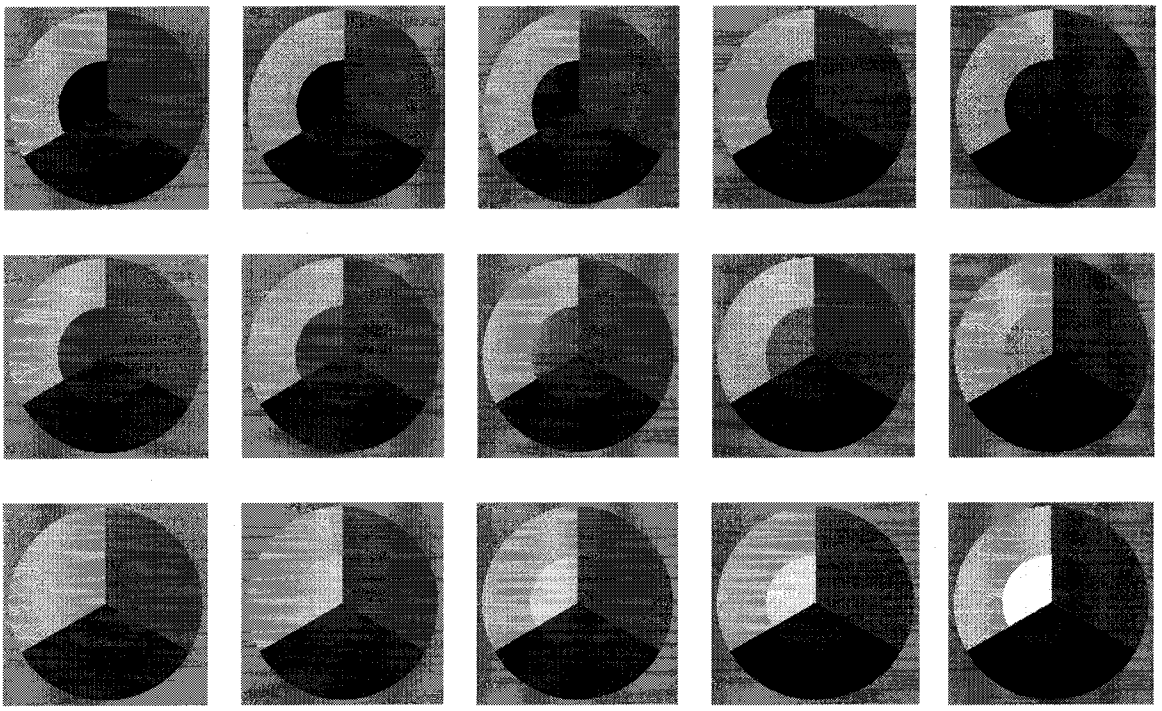


Figure 4.3: Range of possible test patch settings. The correct image, according to the luminance episcotister (LE) model, is in the centre.

luminances. In principle the subjects do not need to be phenomenologically aware of the distinction between t and F with this method, since they are only required to make the whole layer appear uniform in its transparency properties. It is worth emphasising here, however, that the subjects do not adjust t and F independently. Consequently, the results do not directly show how well these two properties of transparent filters are independently set. Rather, the results will speak to the overall salience of the transparency percept, and how well subjects perform while directly manipulating a single stimulus luminance.

To understand how the six–luminance stimulus is employed, consider Figure 4.3. All the stimuli are identical except for one of the layer luminances, which increases systematically as one progresses through the series. Inspection of the stimulus set reveals that at the extreme ends of the series, the layer does not appear to form a uniform transparent surface. However, in the middle of the series one does have the percept of a milky

($F > 0.0$), semi–transmissive ($0 < t < 1$), uniform transparent surface. Only one of the stimuli, however, is the correct one, at least according to the episcotister luminance model (midway in the middle row; provided the luminances have been correctly reproduced in the figure). It is important to understand that the correct third layer luminance depends on the particular model of transparency. Although there are certain restrictions on the possible settings of the two fixed layer luminances, within those restrictions a given combination of background and layer luminances will produce a different t and F combination depending on the transparency model assumed by the visual system. In Experiment 1, subjects were required to adjust this third layer luminance to provide the best perceived transparent surface, for a number of stimuli with various combinations of background luminances and layer t's and F's.

The first goal of this study is to employ the six–luminance stimulus to test between various models of perceived transparency, and to measure both the accuracy and precision with which observers make transparency settings with respect to their preferred model.

4.2 General methods

4.2.1 Stimulus

The stimulus consisted of two simulated concentric disks, each divided into three equal overlapping sectors, producing an illusory transparent layer on a tri–partite background [see Figure 4.1(c)], such that the sectors **P**, **Q**, and **S** were on top of the background portions **A**, **B**, and **C**. The layer luminances P and Q were calculated to produce a range of t's and F's according to the LE model. It must be emphasised, however, that this in no way precludes the subject from making settings according to another internal model. The difference in predictions between three of the models is quite small as will be shown be-

low; setting two of the layer luminances according to one of the models merely ensured that there was an adequate range of t and F values represented in the stimuli for any of the models. The luminance of the area surrounding the stimulus was fixed at an intensity of half the maximum luminance produced by the display (i.e., 20 cd/m^2). Stereoscopic image pairs with a maximum outer diameter of 7.5 $\times$ 7.5 cm^2 were viewed through a custom–built 8–mirror stereoscope with a principal ray path–length of 45 cm, for a maximum visual angle of 7.1°. Even though all parts of the image were in the plane of fixation, a stereoscope was used because these experiments were part of a larger project which included some conditions where the layer was set to a different depth from the background. The stereoscope also served to maintain a fixed viewing distance. However, only the data from flat–plane stimuli are reported here.

4.2.2 Display

All experiments were performed using an SGI (Silicon Graphics, Inc., Mountain View, CA, USA) O2 workstation (150–MHz R10000 processor) on a 17–in. monitor displaying 1280 $\times$ 1024 pixels at a vertical refresh rate of 72 Hz. The luminance output of the monitor was measured using a single channel optometer with photometric detector (Model S370, United Detector Technology), and calibrated to produce luminances between 0 and 40 cd/m^2.

4.2.3 Subjects

Three subjects participated in the experiments, one of whom was naïve to the purpose of the experiment (KW) and two of whom were the investigators (RK, FK). All three subjects were experienced psychophysical observers who had normal or corrected–to–normal vision.

4.3 Experiment 1: accuracy and precision

4.3.1 Procedure

A combination of luminances of A, B, C, and F (ranging from 0 to 30 cd/m^2), including four different values of t (0.2, 0.4, 0.6, 0.8) were pre–selected with the constraints that $\{A, B, C\} \neq 20$ cd/m^2 (surround colour), and $\{A, B, C, P, Q, S\} < 40$ cd/m^2 (maximum luminance output). The contrasts between adjacent stimulus regions were minimised in order to avoid potentially extreme non–linear behaviour of the visual system. This was done by limiting the (Michelson) contrast between the background values (A, B, C) to 30%-40%. The contrast between the layer luminances is, from the model definition (see Eqs. 4.2), always equal to (for $F = 0$) or smaller than the contrast between background luminances. The subjects' task was to adjust the luminance S, using the computer mouse to drag a slider on the display, such that they perceived the transparent layer as a contiguous disk with uniform transmissive and reflective characteristics on the tri–partite background. The luminance S was set to a random value at the beginning of each trial. An example of the range of possible settings for one stimulus is shown in Figure 4.3. Each subject made three sets of adjustments on the set of forty pre–computed stimuli.

4.3.2 Results and discussion

Subjects' final settings of the luminance S were compared first to the theoretically expected luminance for each stimulus, as calculated according to the LE model, $S = tC + F$. The results for the three subjects are presented in Figures 4.4(a)–(c). Each datum point is the average of three adjustment values, with the bars indicating standard error. Points falling on or near the $y = x$ (45°) line indicate good agreement with expected values, or high accuracy. For all three subjects, the coefficient of determination (r^2) was at least 0.99.

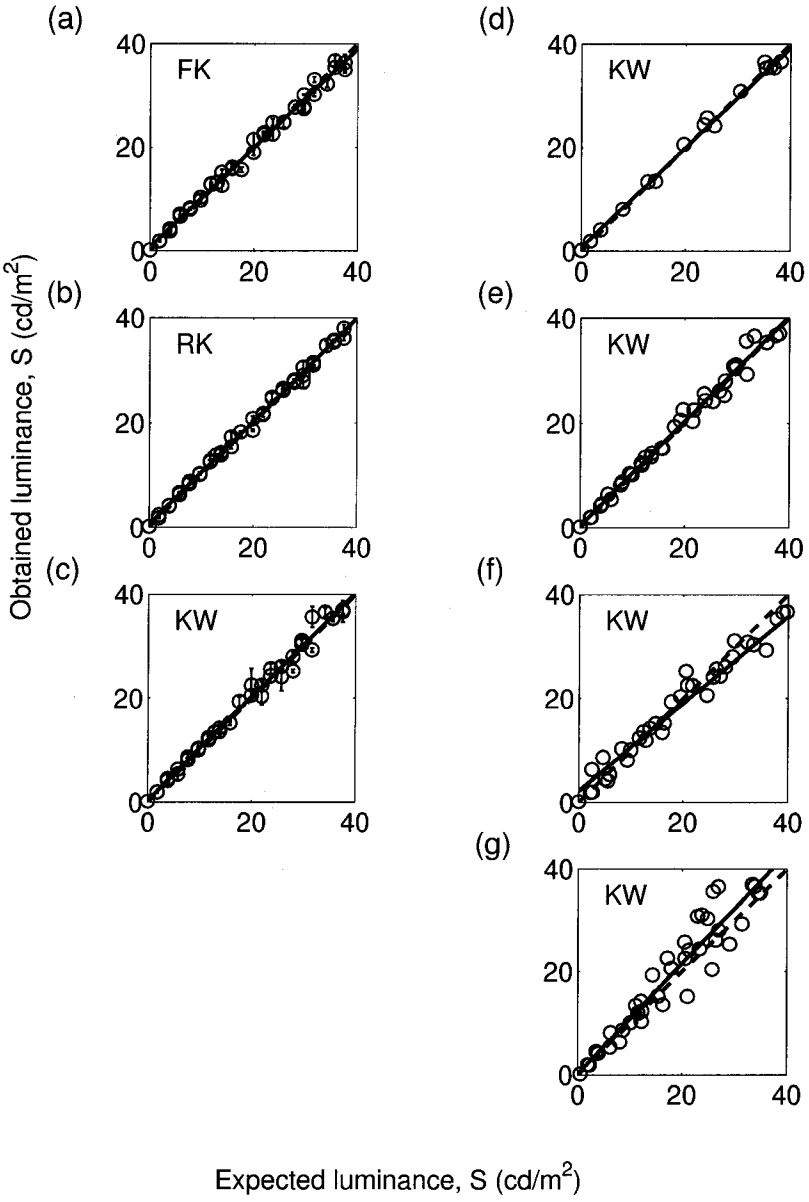


Figure 4.4: Left column (a-c) shows adjustment settings for three subjects (along with solid line depicting a linear regression) compared to the expected episcotister luminance settings (dashed 45° line). The right column shows the data from one subject (KW) compared to predictions according to alternate models: (d) filter model, (e) Michelson contrast model, (f) arithmetic mean model, (g) average brightness model. The performance of subject (KW) matches the first three models the most closely.

Small standard errors are the result of low variability in repeated measurements, or high precision. The results shown in Figures 4.4(a)–(c) clearly reflect very good performance to the extent that both accuracy and precision are high for all subjects.

The rest of the plots in Figure 4.4 compare the results from one subject (KW) with some alternative models. The first alternative model was the filter model of Beck *et al.* [20]. Whereas the LE model does not require an illumination component (it is implicit in the additive term F), the luminance version of the filter model includes an explicit illumination variable, I. For the purpose of comparison with the LE model, the highest luminance in the image was taken as the illumination component, which makes the implicit assumption that the highest luminance is 'white', with reflectance 1.0. From the luminance formulation of the filter model [4],

$$\begin{aligned} t_f &= \frac{\sqrt{(A-B)(P-Q)(I^2-AQ)(I^2-BP)}}{A(I^2-BP)-B(I^2-AQ)}, \\ F_f &= \frac{I^2(AQ-BP)}{A(I^2-BP)-B(I^2-AQ)}, \\ S_f &= t_f^2\frac{I^2}{I^2-FC}C+F_f, \end{aligned} \tag{4.4}$$

where t_f and F_f are the transmittance and the reflectance terms, and S_f is the expected luminance setting. After eliminating stimuli which gave negative t_f and F_f values, the expected settings for S_f given C were calculated. Even though in using Eqs. 4.4 one obtains different t_f and F_f values from the LE model, the resultant luminances, S_f, are still quite close to the LE predictions. This is not surprising, as it has been shown by Gerbino [4] that the reflectance and transmittance values of the filter model (Eqs. 4.4) approach those of the LE model (Eqs. 4.2) as higher and higher values of the illumination component are assumed. The fact that a number of stimuli had to be rejected due to illegal values of t_f (outside the range [0, 1]) and F_f ($F_f < 0$) suggests that the solution space of acceptable transparent percepts is smaller for the luminance filter model than the luminance

episcotister model. It should be noted that the results for the filter model apply only insofar as the assumption of the lightest patch being white is valid.

A model recently proposed by Singh and Anderson [8] employs ratios of Michelson contrasts between adjacent areas as a predictor of subjects' perceived transmittance. Instead of the luminance differences in Eqs. 4.3, the transmittance, t_M, is given by

$$t_M = \frac{\frac{P-S}{P+S}}{\frac{A-C}{A+C}} = \frac{\frac{P-Q}{P+Q}}{\frac{A-B}{A+B}}. \tag{4.5}$$

To compare this with other models, Eq. 4.5 was rearranged such that expected values of the luminance S could be directly plotted, giving

$$S_M = P\frac{A(AQ-BP)+C(AP-BP)}{A(AP-BQ)+C(AQ-BP)}. \tag{4.6}$$

Another hypothesis is that subjects may simply have used the average contrast between the layer luminances and their background when making a setting. This would be equivalent to making the assumption that the transparent layer is a neutral density filter i.e., that $F \equiv 0$, even though this assumption is not supported by the physical characteristics of the stimuli. There are a number of different indices that could measure average contrast. One such index would be the arithmetic mean, i.e., $S = \frac{1}{2}\,C(\frac{P}{A}+\frac{Q}{B})$. Another is the geometric mean—the equivalent of taking the average in log–space—which is given by

$$\begin{aligned} \log\frac{S}{C} &= \frac{1}{2}(\log\frac{P}{A}+\log\frac{Q}{B}) \\ S &= C(\frac{PQ}{AB})^{1/2}. \end{aligned} \tag{4.7}$$

Since the geometric mean of two values is in general close to the arithmetic mean when the two samples are similar (recall that an effort was made to avoid large stimulus contrasts), in this case the two means are almost identical, and only the comparison with the arithmetic mean is shown in Figure 4.4(f). Finally, the average luminance

$$S = \frac{P+Q}{2} \tag{4.8}$$

was also calculated and plotted against observed data, since it is conceivable that subjects were making some sort of average brightness setting of the three layer luminances, independent of the background intensities.

The results show that the LE model, the ratio of Michelson contrasts model, and to nearly the same extent, the luminance filter model ($r^2 = 0.99$ for all three for subject KW) account very well for subjects' performance. Neither of the other plotted predictions shown in Figure 4.4 match the results as well as the LE model ($r^2 = 0.96$ and 0.90 respectively, for the arithmetic mean and average brightness hypotheses), and both the slope of the regression and the goodness of fit are inferior. Taken together, the results from Experiment 1 demonstrate that with the six–luminance display, human subjects make accurate (i.e., close) and precise (i.e., reproducible) transparency judgements according to the luminance formulation of the episcotister model, and the Michelson contrast model.

The performance of the subjects in this experiment says little, however, about whether there is a range of settings around the ideal luminance S, given a set of $\{A, B, C, P, Q\}$ that nevertheless produces some degree of perceived transparency. The next experiment was designed to address this question.

4.4 Experiment 2: Range of perceived transparency

To understand the purpose of this experiment consider the following loose analogy. Imagine that one wishes to determine the wavelength perceived as unique yellow. In principle, one might adjust the wavelength of light until unique yellow was perceived. The estimate of unique yellow and the precision of the estimate would then be given by the mean and, say, standard error of a series of such adjustments. One might then compare the measured estimate of unique yellow with a prediction based, for instance, on some physiological

model of colour vision. This is analogous to Experiment 1, where the luminance needed to produce the best transparent percept was determined and compared to a putative physical/psychological model, and the accuracy and precision of a range of such settings ascertained. One could, however, ask a quite different question: over what range of wavelengths around unique yellow do subjects perceive at least some degree of yellowness? To estimate this range one might ask subjects to indicate the wavelength boundaries on either side of unique yellow, beyond which no yellowness was perceived. This, however, would be a somewhat crude measure, since it would not take into account the *pattern* of decline of perceived yellowness as one moved away from unique yellow.

In this experiment the range of luminances over which a degree of transparency is perceived was measured, using a novel technique that allows the determination of the pattern of decline of perceived transparency as one departs further and further from the point of ideal or best transparency.

4.4.1 Procedure

The subjects were presented with forced–choice[4] pairs of six–luminance stimuli in two temporal intervals and were asked to indicate (by a button press) which of each pair contained the most compelling transparency, insofar as it presented a layer with most uniform transmissive and reflective properties around the entire disk. Each stimulus was presented for 1250 ms, with a 250–ms interstimulus interval. When no stimulus was present the screen was a blank mid–grey colour. Whereas in the adjustment data of Experiment 1 the stimuli encompassed a wide range of different combinations of parameter values, in this experiment all the trials were based on a stimulus with the following fixed parameters:

[4]There is disagreement in the literature over the usage of the term 'forced–choice'. Some maintain it should only be used if there is a correct answer on every trial, which is not the case in this experiment. The term forced–choice is meant here to imply that there are two alternatives on each trial from which one must be chosen. See [162] for a discussion of this issue.

$A = 6.0$ cd/m^2, $B = 14.7$ cd/m^2, $C = 24.0$ cd/m^2, $t = 0.75$, and $F = 9.0$ cd/m^2, resulting in expected values of P, Q, and S of 13.5, 20.0, and 27.0 cd/m^2, respectively, according to the LE model.

In each trial two stimuli with, say, S_1 = 24.5 cd/m^2 and S_2 = 34.5 cd/m^2 were chosen from the list of nine potential luminances of S. The values of S were sampled every 2.5 cd/m^2, and the list was centered on the expected LE model value (i.e., 27.0 cd/m^2), such that $S_i \in \{17.0, 19.5, 22.0, 24.5, 27.0, 29.5, 32.0, 34.5, 37.0\}$ (see Figure 4.3, for example). Each stimulus pair was presented twice, in different presentation order. The angular orientation of the stimulus was randomised such that the test patch did not appear in the same location from one pair to the next, while the orientation within each pair was held constant. The trials were interleaved such that in one trial the luminances P and Q were held fixed while the luminance S was selected from an array of nine values (as above), while in the next trial Q and S were held fixed while the luminance P was selected from a list of nine values centred on its expected value, and so on. The values of A, B, C, t, and F, as well as the colour of the surround, were held fixed throughout.

Three sessions of 216 trials were performed by subjects FK and KW. Each session consisted of three interleaved stimulus sets—one set for each third of the circle: **P**, **Q**, and **S**. There are $\frac{8\times9}{2}$ ways of selecting pairs from a list of nine. With two presentations of each pair within each of the three session, this yields a total of $8 \times 9 \times 3 \times 3 = 648$ trials, which were grouped together and tallied.

The subjects were, in addition, asked to make three free–adjustment settings for the three parts of the circle (as in Experiment 1) to determine each subject's best transparency setting.

4.4.2 Analysis: Modified method of pair comparisons

The method of pair comparisons [163] is traditionally used to rank order objects (e.g., carrots, zucchini, celery) in a category (e.g., vegetables) in order of preference, and to subsequently find the objects' co–ordinate values along a unidimensional preference scale with an arbitrary origin, e.g. {-1.2, 0.2, 1.5} for {celery, carrots, zucchini}. The traditional method of pair comparisons cannot be used, however, to recover an underlying preference *distribution* where a unidimensional parameter space with equally spaced samples is already defined. In the case of this experiment, the parameter space is the luminance of the variable test patch (be it **P**, **Q**, or **S**) which is sampled strictly every 2.5 cd/m^2. This distribution describes not only how much more compelling a stimulus is than another (by comparing the two corresponding points on the function), but also delimits the range of stimuli regarded as transparent at all. Presumably, if two stimuli are equally undesirable, each will be selected an equal number of times, and their ordinate values will be equal.

The following modification to the classical method of pair comparisons was made in order to recover what will hereafter be referred to as the *a priori* probability density function (PDF). It is assumed that given a range of transparency stimuli ($i \in 1, 2, \ldots, N$) which vary along a single dimension (be the relevant parameter transmittance, reflectance, or luminance of the test patch, as in this case; see Figure 4.3), there exists a probability for each stimulus ($0 \leq q_i \leq 1$) to be preferentially selected over the others by an observer over a series of repeated trials. An assumption is made also that the sampled range of the parameter space is wide enough that the tails of the PDF tend to zero (that is, that the whole distribution is captured in the PDF), and that the area under the PDF is normalised, i.e. $\Sigma q_i = 1$. For example, the PDF of a series of seven stimuli might be $\{0.0, 0.1, 0.4, 0.3, 0.2, 0.1, 0.0\}$, as shown in Figure 4.5. Data could also have been collected by presenting all seven stimuli arranged randomly side by side over a number of trials, and by asking a subject to select

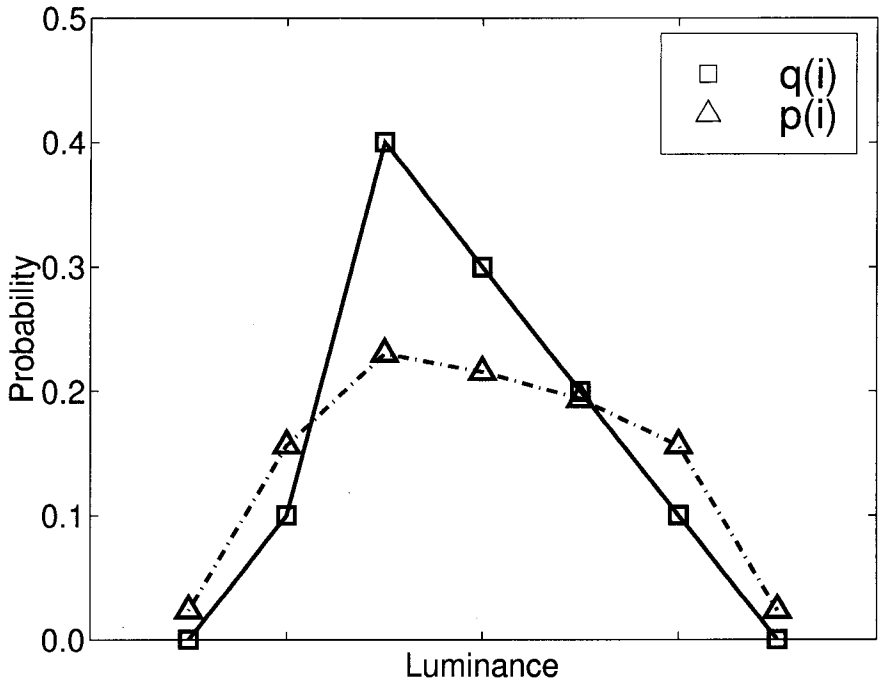


Figure 4.5: Simulated example of underlying *a priori* probabilities (squares) and tally scores (triangles).

the best one, or the one with most consistent transparency according to some criterion, or even by presenting single stimuli and asking for binary (yes/no) or scalar (a score from 1 to 10) judgements, as Beck *et al.* [20, 164] and others [165] have done.

To sample the parameter space using a forced–choice paradigm with the pair comparison method, random pairs were selected from the series and the observer was asked to choose the preferred one. A tally was kept of the number of times each stimulus was selected in favour of all the others. This list of tallies (the observed data) will be called p. The *a priori* distribution sought will be labelled q, where q_i is the probability of selecting the ith stimulus over all the other stimuli in the series. Given a pair of stimuli with underlying probabilities q_j and q_k presented in a given trial, the probability of choosing stimulus j over stimulus k is simply $p_{j \succ k} = q_j/(q_j + q_k)$. That is, the probability of choosing q_3 over q_5 in the above example on any trial where q_3 is compared with q_5 will be 0.4/(0.4 + 0.2) or 2/3. Another interpretation would be that on average, in two out of three trials where

stimuli 3 and 5 are compared, stimulus 3 would be selected. If both of q_j and q_k are zero, $p_{j \succ k}$ is assumed to be 0.5, since the observer is forced to make a choice, and will likely choose each stimulus an equal number of times. Over a series of N trials, the probability of selecting any particular stimulus becomes

$$p_i = \frac{1}{C_2^N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{q_i}{q_i + q_j}, \tag{4.9}$$

where

$$C_k^N = \frac{N!}{k!(N-k)!}$$

is the number of possible ways of picking k items out of N without replacement. Since one is selecting pairs, $k = 2$, and if one ensures that each pair is presented an equal number of times, $1/C_2^N$ becomes a common factor to all the terms in the sum. The terms where $j = i$ are excluded from the sum, because the same two stimuli were never compared in the same trial.

The array p_i then contains the observed tallies, expressed as proportions of the number of times stimulus i was selected across all trials, and q_i are the *a priori* probabilities one seeks to recover, as shown in Figure 4.5. Since each p_i is dependent upon all the q_i (i.e., the whole PDF), inverting the transformation is not a simple task. The problem can be easily solved numerically, however, if formulated as a constrained minimisation with Lagrange multipliers [166]. Specifically,

$$\sum_{i=1}^{N} \left(p_i - \frac{1}{C_2^N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{q_i}{q_i + q_j} \right)^2 \tag{4.10}$$

is minimised subject to the constraints

$$\sum_{i=1}^{N} q_i = 1, 0 \leq q_i \leq 1, |q_i - q_{i+1}| < 0.3. \tag{4.11}$$

P (cd/m^2)	FK	KW	Q (cd/m^2)	FK	KW	S(cd/m^2)	FK	KW
3.5	4	4	10.0	12	5	17.0	13	10
6.0	5	8	12.5	14	10	19.5	30	24
8.5	36	36	15.0	17	18	22.0	46	43
11.0	31	33	17.5	40	37	24.5	36	37
13.5	33	34	20.0	37	37	27.0	28	34
16.0	34	32	22.5	38	40	29.5	28	31
18.5	35	31	25.0	31	34	32.0	15	18
21.0	17	19	27.5	16	21	34.5	12	13
23.5	21	19	30.0	11	14	37.0	8	6

Table 4.1: Raw forced–choice results. The data for each subject is presented for each sector of the stimulus. The tallies show the number of times each stimulus was selected. Note that when the stimuli with, say, P = 3.5 cd/m^2 and P = 21.0 cd/m^2 are compared, the luminances, Q and S, of the other two sectors are held fixed at the LE model values. These raw tallies were normalised with respect to the column total (N = 216) to give p_i for each curve in Figures 4.6 and 4.7.

The constraints ensure that the PDF sums to unity, and that each individual probability is between zero and unity. A gradient constraint was also added to avoid overshoots and undershoots in the minimisation.[5] Visual inspection of the data confirms that large differences in adjacent bins are artefactual or are due to outliers.

4.5 Results and discussion

Table 4.1 lists the raw (nonnormalised) tallies for both subjects. Figures 4.6 and 4.7 show the recovered *a priori* probabilities (q_i) as well as the raw tally data (p_i) for subjects KW and FK. The solid vertical lines show the LE model predictions. The dashed vertical lines show

[5]The minimisation was performed using the `constr()` function in the MATLAB optimisation toolkit (The MathWorks, Inc., Natick, MA, USA).

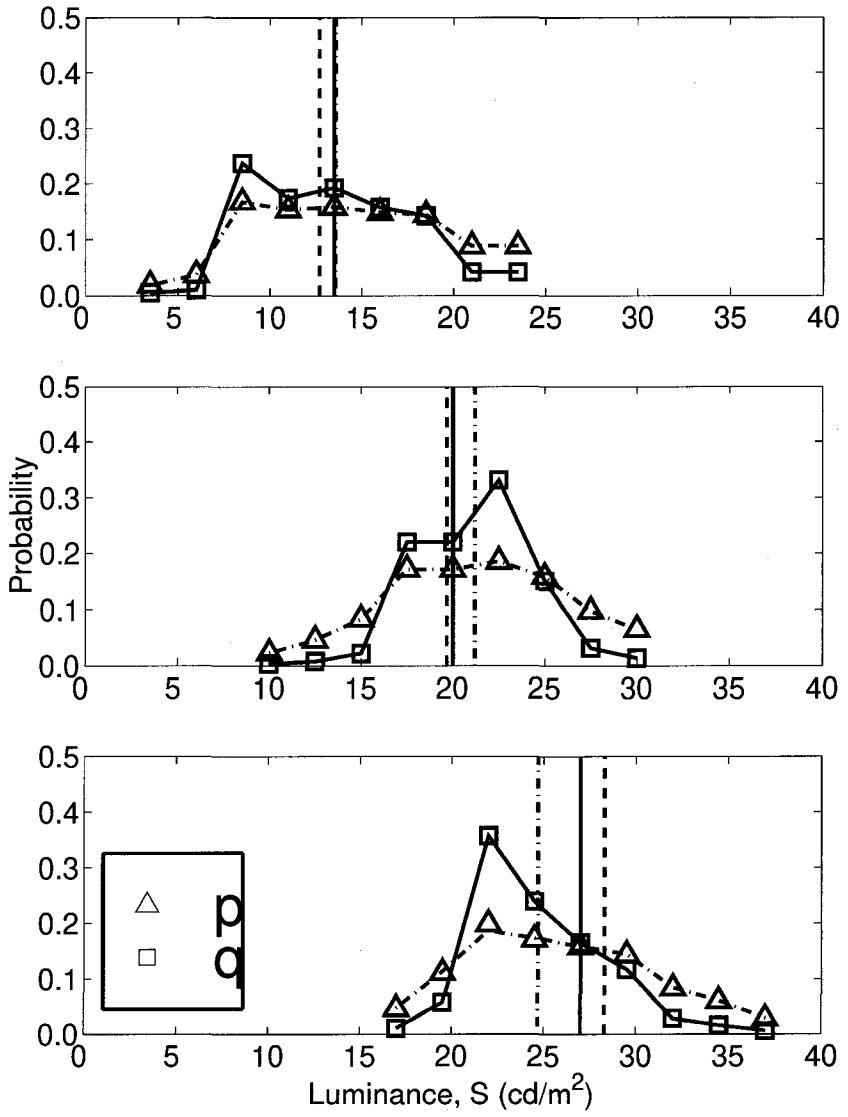


Figure 4.6: Forced–choice results for subject KW. Each plot shows the results from the luminance perturbation of one part (P, Q, and S, respectively, from top to bottom) of the stimulus around the value according to the luminance episcotister model (solid vertical line). The original tally data is shown by triangles. The squares show the recovered *a priori* probabilities, q. The dash–dot line joining the triangles shows a check of the solution by substituting q_i back into Equation 4.9 to compare with the observed tallies, p. The vertical dash–dotted line is the first moment (average) of the *a priori* data, and the dashed line is the average of three adjustment settings made by the subject.

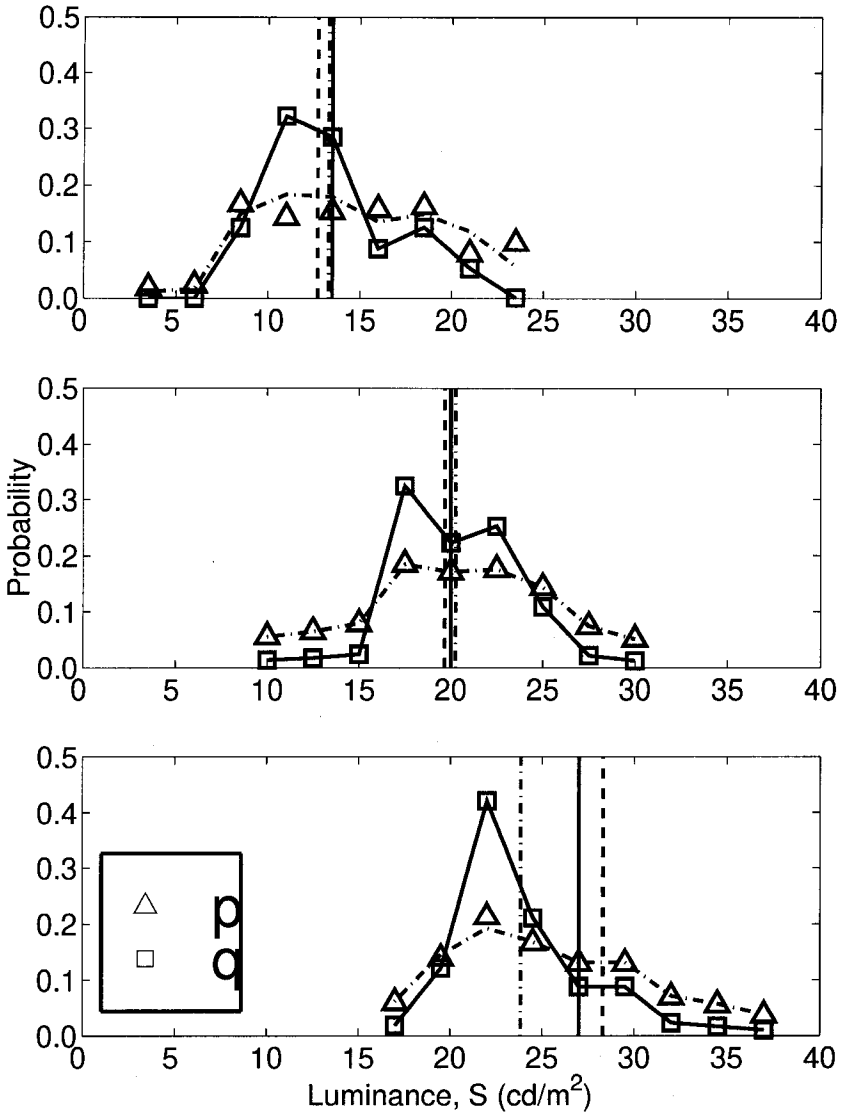


Figure 4.7: Forced–choice results for subject FK.

the subjects' average adjustment settings corresponding to each part of the stimulus, which were measured separately. Rather than imposing a functional form (i.e., a Gaussian) on the recovered results in order to obtain some measure of the central tendency and spread, the average and standard deviation of the distribution were computed. The dotted lines show the means of the three distributions. These results are summarised in Table 4.2.

The results show that both the subjects' free adjustments and the mean of the PDF are reasonably close to (generally within 3 cd/m^2 of) the predicted episcotister model lumi-

	FK	KW
P, LE model value	13.5	13.5
P, free adjustment (std err)	16.3 (1.0)	12.7 (0.6)
P, PDF mean (std dev)	14.2 (6.5)	13.6 (6.5)
Q, LE model value	20.0	20.0
Q, free adjustment (std err)	17.9 (0.2)	19.7 (0.3)
Q, PDF mean (std dev)	20.3 (6.5)	21.2 (6.6)
S, LE model value	27.0	27.0
S, free adjustment (std err)	21.2 (0.1)	28.3 (1.5)
S, PDF mean (std dev)	23.8 (7.2)	24.7 (6.9)

Table 4.2: Summary of forced–choice results. All values are luminances in cd/m^2. No distinction was made between P, Q, and S during the study since the angular orientation of the stimulus was randomised.

nances. The widths of the distributions also confirm the hypothesis that there is a series of luminances which give rise to at least some degree of phenomenal transparency. The distributions appear to be unimodal, confirming that each subject has an individual point of best transparency.

These results could also be understood in terms of balanced and unbalanced transparency. Unbalanced transparency refers to displays in which Metelli's rules regarding the limits on t and r (see Eqs. 4.1) are violated. Unbalanced transparency implies the presence of two *different* transparencies with different combinations of t and r for the two different background–layer pairs, rendering Eqs. 4.1 indeterminate [167]. It may be that the results of this experiment reflect, in part, the preference of observers for balanced rather than unbalanced transparency.

4.6 General discussion

In this chapter the accuracy, precision, and range of perceived achromatic transparency was investigated using a new stimulus and a new psychometric technique—the modified method of pair comparisons. In the first experiment, observers adjusted the luminance of a test patch on a putative transparency stimulus such that it created the most compelling impression of a uniform transparent disk overlaying a tripartite background. The results were compared to predictions from a number of models of perceived transparency and other hypotheses concerning subject performance. It was shown that Singh and Anderson's ratio of Michelson contrast model [8] and the luminance formulation of Metelli's episcotister model [4] (in agreement with a previous study by Gerbino [160] using an eight–luminance stimulus arrangement) best predicted the subjects' adjustments. Our analysis revealed that subjects did not appear to assume that the layer was a simple neutral density filter, reinforcing the idea that they were sensitive to the additive reflective component, F. In turn, this confirms the idea that a neutral density filter (which gives rise to multiplicative, subtractive [168], or film [3] transparency) is perceived as a limiting case of transparency where the layer has zero reflectance.

The filter model proposed by Beck *et al.* [20] takes into account internal reflections between the filter and the background, resulting in a slightly different set of equations (Eqs. 4.4). The protocol of Beck *et al.*, however, required subjects to make a simple yes/no judgement as to whether each surface, displayed separately, was perceived as transparent. This binary response method is a cruder version of the modified pair comparison technique presented in Experiment 2. Although Gerbino *et al.* [160] did not compare their results to predictions made by the filter model of Beck *et al.*, Gerbino [4] (and, presumably, Metelli [6]) has favoured models compatible with the theory of colour scission—where

image intensities can be attributed to two superimposable sources, namely a background and a layer. According to Metelli and Gerbino, the luminance episcotister model would be preferable to the luminance filter model. Irrespective of compatibility with scission theory, the results of Experiment 1 show that the LE model encompasses a wider set of perceptually valid transparency stimuli than the filter model, even though the episcotister and filter model results for the stimuli which did produce legal transmittance and reflectance values were similar. In addition, the results show that the three subjects made settings with both high accuracy and precision. This finding indicates that there is a definitive and reliable ideal transparency point for human observers.

The second experiment explored the parameter space surrounding the predicted LE model solutions on a specific exemplar stimulus to find whether there is a range of stimuli around this predicted ideal point that also gives rise to at least some degree of perceived transparency. It was found that there is a wide range of the variable layer luminance (average standard deviation ≈ 7 cd/m^2) within which a subject would report some phenomenal transparency, even though the subject's *most preferred* point might be quite rigorously and reliably reproduced to be near the mean of this distribution.

Few studies have endeavoured to determine the exact luminance or reflectance combination required to give rise to a perceptually transparent surface, whether by using a classical bipartite arrangement [such as in Figure 4.1(b)] or otherwise. For example, Gerbino *et al.* [160] used two four–luminance stimuli to perform their matching experiments. The simpler and more cohesive six–luminance stimulus employed here has a number of advantages. In addition to displaying all of the relevant luminances simultaneously within central vision, it leaves the luminance of only one patch to be manipulated. This setting has only one correct solution according to the subjects' internal model of transparency, since according to any algebraic transparency formulation requiring four luminances, the system's two

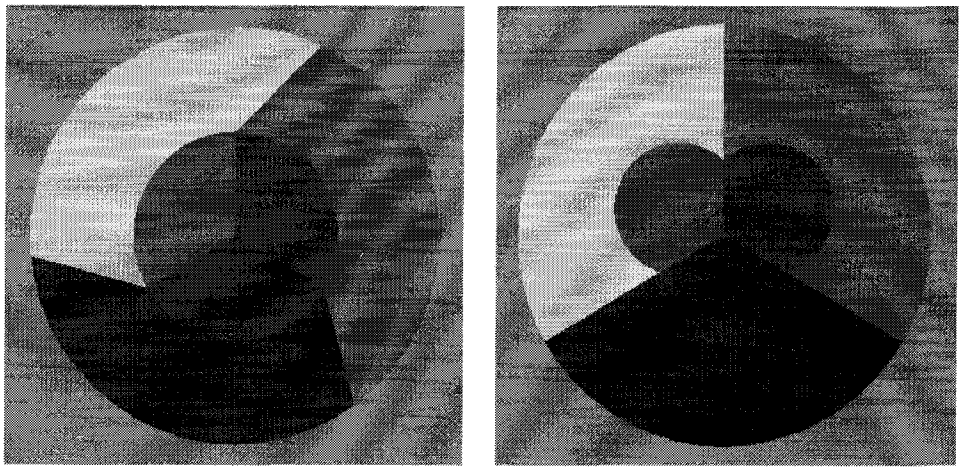


Figure 4.8: Examples of modifications to the six–luminance stimulus for investigating figural conditions in transparency. ,

unknown parameters can be uniquely identified, as evident from Eqs. 4.2. The results of Experiment 2 suggest that if there is a so–called ideal point that is preferred by an observer, it is flanked by a reasonably wide range which also engenders, albeit to a lesser extent, perceptually valid impressions of transparency.

Use of the six–luminance stimulus need not be limited to the study of luminance–based cues to transparency. There are figural conditions which must be satisfied in order to evoke compelling transparency percepts. Variants of the six–luminance stimulus, such as those in Figure 4.8, are currently being used to explore the role of X–junctions. Metelli [6] provided compelling examples of stimuli where the luminances were appropriately selected, which nevertheless did not produce perceptually transparent images. More recently, arguments have been made for the need for X- or T–junctions (the latter sometimes implying X–junctions [169]). Anderson [25], and previously Beck and Ivry [164], have suggested a series of heuristics for the required contrast relationships across collinear segments of such junctions. These rules define the order of intensities (from brightest to darkest) around an X–junction as a cue for transparency. These contrast polarities were satisfied as a consequence of the modified Metelli equations used in these studies, and their role was not

explored here. The six–luminance stimulus is ideally suited to test the importance of these various proposed figural conditions, as it is easily employed to provide a rigorous and quantitative assessment of perceived transparency, rather than merely a qualitative assessment.

How can these findings be placed in the context of models of transparency applied to the chromatic domain? D'Zmura *et al.* [170] have proposed a chromatic transparency model which describes the characteristics of filters based on the translation and convergence of vectors in colour–space. It is at present unknown how perceptual models of chromatic and achromatic transparency are related, and how the conditions necessary for chromatic transparency can be collapsed from a colour–space into the luminance dimension. The LE model implemented here, however, corresponds to a scalar version of the vector convergence condition in the study of D'Zmura *et al.* Westland and Ripamonti [171] have also suggested a model which aims to predict cases of perceived chromatic transparency. Their studies show that ratios of retinal cone excitations (across all cone types) between a background and its corresponding transparent layer are invariant over a wide range of cases of simulated physical transparency, and they cite this invariance as a possible cue or mechanism for the detection of chromatic and achromatic transparent overlays. They are, however, careful to point out, as Metelli [6] did originally, that physical transparency does not necessarily give rise to perceptual transparency, nor do all cases of perceived transparency correspond to a physically transparent surface. In other words, the so–called proximal–to–distal mapping of perceptual versus physical transparency is not one–to–one, but rather many–to–many. The real potential advantage of these newer models (such as Westland and Ripamonti's, and Singh and Anderson's in the achromatic domain) over conventional models is their use of contrast measures (cone excitation ratios, or Michelson contrasts), rather than differences of luminances or reflectances. Considering the wealth of literature indicating that the visual system is well–adapted to detecting changes in contrast, it would be interesting

to see whether the results of their simulations are supported by the type of psychophysical evidence provided by the present study.

Chapter 5 Preview

In the previous chapter a number of issues relating to the luminance conditions required for perceptual transparency were discussed. A six–luminance stimulus was used to compare different algebraic models, and it was found that the luminance episcotister model was one of the best at predicting subjects' settings. The same model is used in the next chapter to assess the effects of changes in figural conditions on transparency processing. Chapter 5 is an adaptation of the following in press manuscript:

R. Kasrai and F. A. A. Kingdom, "Achromatic Transparency and the Role of Local Contours," *Perception* (accepted February 2002).

Chapter 5

Achromatic transparency and the role of local contours

Abstract

This chapter investigates the role of contours and junctions in the perception of single–plane achromatic transparency. In order to measure the accuracy with which observers encode transparency, a six–luminance stimulus was employed whose figural properties could be easily manipulated. Accuracy was measured by requiring subjects to select (either by the method of adjustment or by using a forced–choice procedure) the luminance that best completed a simulated transparent filter. The X–junctions in the stimulus were destroyed or perturbed in three experiments. Simple occlusion of the junction (Experiment 1), and perturbation of the orientation of the contours of the filter as they pass through the junction (Experiment 3) resulted in small but significant reductions in performance. On the other hand, a sudden change in orientation of the background (material) contours (Experiment 2) resulted in a small but significant *enhancement* of overall performance compared with the control stimulus. In the forced–choice task, reversals in the polarity of contours (as defined by the brightness order of flanking regions) around the junction were shown to effect large changes in subjects' accuracy in processing transparency. The overall results show that X– and Ψ–junctions are indeed salient properties of transparent stimuli. The findings suggest that jagged contours with sudden changes in direction are more likely to be attributed to reflectance (material) changes than to changes due to a transparent filter (or

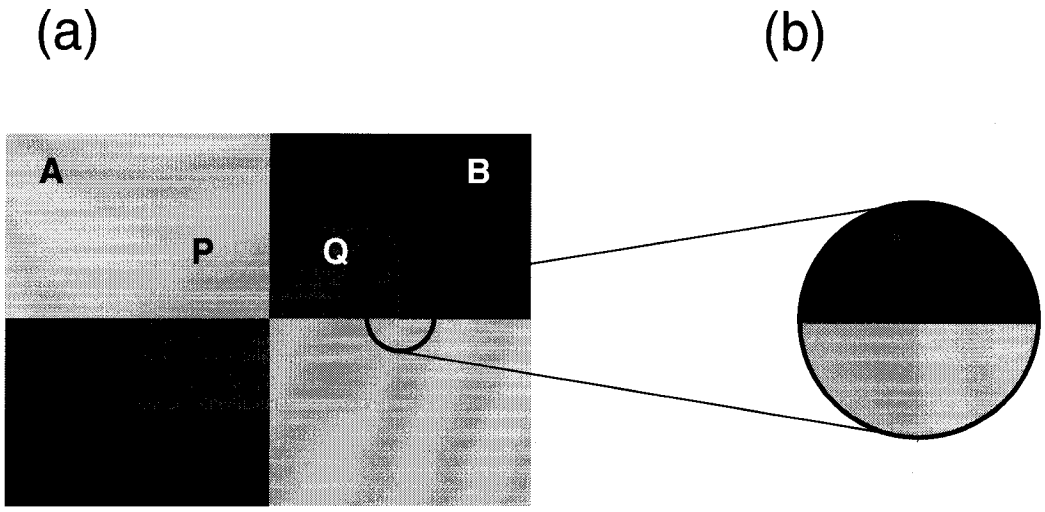


Figure 5.1: (a) Classical bipartite background overlaid with a transparent filter, and inset (b) X–junctions defined.

to illumination).

5.1 Introduction

Transparency, described phenomenally as 'seeing through' surfaces, has been the subject of much recent study. Interest in transparency is motivated by more general questions regarding scene analysis. Transparency can be thought of as belonging to a broader class of illumination effects, including shadows (a limiting case of transparency) and highlights, which cause abrupt changes in image luminance. While these changes in luminance may correspond in general to either material changes in surface property or to illumination effects, the mechanisms used by the visual system to disambiguate the two in the achromatic domain are as yet poorly understood. Specifically, how does the visual system distinguish between two contours which cross at a junction, when one contour is due to a material change and the other to a change in illumination?

Figure 5.1(a) shows a classical zero–disparity illusory- transparency stimulus consisting of a bicoloured background overlaid with a rectangular filter. The phenomenal segmenta-

tion of the homogeneous filter layer from the bipartite background is effortless and quite compelling. Kanizsa [22] referred to three types of conditions for the occurrence of this type of transparency: chromatic, topological, and figural. Collapsing the latter two, the conditions necessary for the perception of transparency can be broadly separated into intensity and figural parameters. Intensity conditions describe the luminance (or contrast) and chromatic relationships between parts of a putative transparent surface and its background. Most of the research in transparency has been devoted to the development of models which describe and predict the achromatic luminance and contrast relationships required to give rise to perceptual transparency [11, 20, 167, 160, 172, 4, 173, 174, 171, 175].

Figural conditions, on the other hand, determine the arrangement and orientation of contours that separate different parts of the image. As an illustrative example, take the two images adapted from Kanizsa [22] in Figure 5.2, which demonstrate the disparity in percept strength due to the different figural arrangements. Even though both images contain contours which cross (one caused by a change in background reflectance and the other by an illusory filter), the relative orientation of the contours is not identical. While comparative empirical analyses of different intensity models have already been reported [160, 175], a quantitative survey of the role of figural conditions remains incomplete. In this chapter the figural conditions necessary for single–plane achromatic transparency will be examined with a novel technique which has been successfully utilised to study intensity conditions.

5.1.1 Formal classification of junctions

To establish a common nomenclature, a brief overview of junction classification is presented (see also [176]). Junctions can be formally labelled as '3–junctions' and '4–junctions', depending on the number of contours which come together at the junction. Junctions of the type normally encountered in transparency stimuli depicted in Figure 5.1(a)

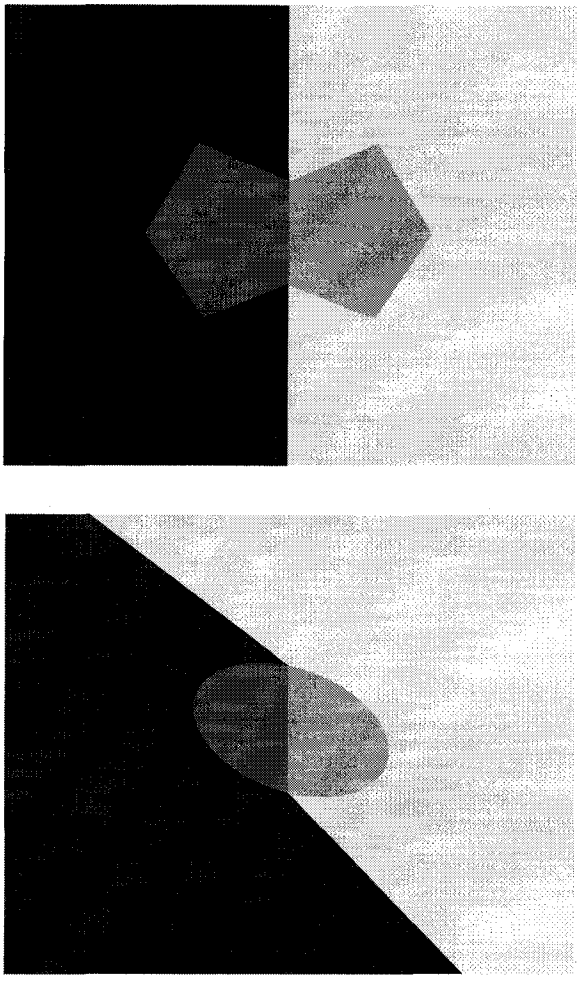


Figure 5.2: Discontinuity in direction of (*top*) the filter contour, vs. (*bottom*) the background (reflectance) contour (adapted from Kanizsa [22])

will have 4–junctions, as shown in the inset, Figure 5.1(b). If the 4–junction consists of two pairs of collinear (but not necessarily perpendicular) contours, it is called an X–junction, whereas if only two of the four contours are collinear, it is referred to as a Ψ–junction [see Figure 5.7(b)]. Similarly, 3–junctions consisting of two collinear contours are called T–junctions. In the context of these definitions, it has been suggested that T–junctions serve as a cue for occlusion of one surface by another [see Figure 5.3(d)], but that X–junctions trigger transparency perception.[1] But it is not true that all X–junctions automatically lead to perceptually valid transparency.

5.1.2 Intensity conditions

Although this chapter focuses on the role of junctions, it is necessary to consider the intensity relationships required for achromatic transparency in order to understand the particular techniques employed here to study figural conditions. Metelli [6] proposed the first quantitative model describing the reflectance relationships required for phenomenal transparency. Although others have since suggested alternative models which use contrast [8] or lightness [20] instead of physical reflectance, a more general model based on luminance, called the luminance episcotister (LE) model [4], is used here. The performance of human observers with stimuli containing illusory transparency has already been compared with the LE model and other models [20, 160, 8, 175]. The previous chapter demonstrated, using a new technique described in §5.1.4, that subjects' adjustments were not only very precise, but also very accurate compared to the LE model.

Given how well it predicts human observers' responses to stimuli containing simulated transparency, the LE model (described below) was chosen as the intensity model for the stimuli generated in this study. Taking a stimulus with a simple bipartite background and

[1] At least two studies have shown empirically that under specific stimulus conditions, where the X–junctions are implied from T–junctions, the perception of transparency can be just as compelling [169, 25].

an overlying transparent filter [such as Figure 5.1(a)], the luminances of the transparent regions **P** and **Q**, are

$$P = tA + F, \quad \text{and similarly} \qquad Q = tB + F, \tag{5.1}$$

where t is the transmittance of the filter, A, B, P, and Q are the luminances of the two background and two layer regions, respectively, and F is the additive term due to the reflectance of the filter. The implication here is that the absolute value for the reflectance of the filter is inseparable from the illumination component of the filter, which in the general case is not necessarily the same as that of the background.

Contour polarity and ordinal rules

Just as algebraic models attempt to provide a fine–scale description of the optimal intensity relationships for transparency, ordinal rules governing the correct arrangement of intensities around the junction predict in a coarse fashion whether a stimulus contains a plausible or implausible transparent medium. These qualitative rules have been shown to give rise to consistent predictions regardless of the exact algebraic model used (cf., for example, [3] p. 213, [164] p. 588, and [25] p. 424).

A *polarity*, or relative direction, can be assigned to a contour depending on its flanking luminances, e.g. positive polarity if the region on the right is brighter than the one on the left. If the brightness order is reversed as the contour crosses through an X–junction, the contour is said to reverse polarity [Figure 5.3(b)]. If both pairs of contours in the X–junction reverse polarity, it is called a double–reversing junction [Figure 5.3(c)]. Non–reversing X–junctions [Figure 5.3(a)] result in bistable transparency percepts, where the depth order of the two surfaces is ambiguous; single–reversing junctions [Figure 5.3(b)] give rise to transparency with unique depth order; and double–reversing X–junctions do not evoke a transparent percept. Beck and Ivry [164] alluded to the same luminance–order

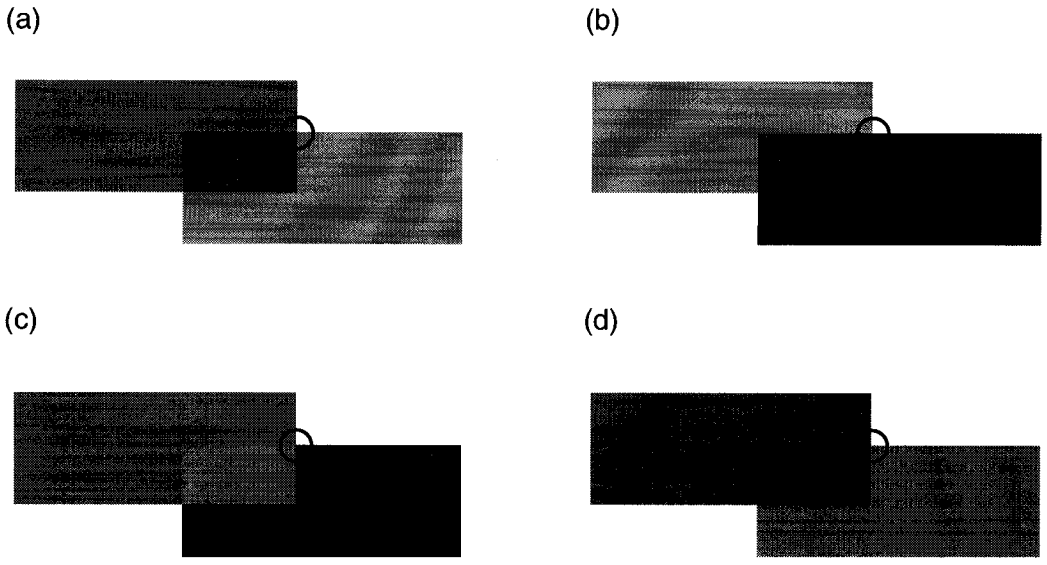


Figure 5.3: Example of (a) non–reversing, (b) single–reversing, and (c) double–reversing X–junctions; also, (d) T–junction.

constraints by suggesting that "the likelihood of perceptual transparency is increased when the direction of lightness changes across an x–junction is in the same direction" (p. 589), or in the words of Adelson and Anandan [177], p. 80), "an edge which is transparently occluded cannot reverse sign, while an edge which is in front may or may not reverse sign." Within the context of these broad qualitative rules, algebraic models simply place additional constraints on the solution–space of perceptually valid transparent stimuli. In the experiments described in this chapter the data were analysed in terms of both the coarse ordinal as well as the fine algebraic constraints in order to observe the effect of changes in figural conditions on the processing of transparency by the visual system.

5.1.3 Figural conditions

Metelli [6] outlined three figural conditions which follow from Gestalt principles: "figural unity of the transparent layer, continuity of the boundary line and adequate stratification" (p. 92). These conditions describe generally that the layer must: appear as a whole,

segmentable surface; not have any abrupt changes in the continuity of its boundary; and, appear to be between the observer and the background, such that "the underlying regions must appear to meet under the whole of the transparent layer" (p. 93). Kanizsa [22] also demonstrated that "each one of the two areas to be unified in the transparent surface [e.g. **P** and **Q** in Figure 5.1(a)] must be in contact with the homologous area [e.g. **A** and **B**, respectively] and *with only one of the two other areas*" (p. 158). Taken together, these define qualitative constraints on the nature of contours and junctions (see §5.1.1) which normally occur as a result of cast shadows, transparent media, or changes in material. Binocular X–junctions also occur in stimuli containing stereo disparity; however, a discussion of multiplane (stereoscopic) transparency is beyond the scope of this study.

That the existence of X–junctions is not a sufficient condition for perceptual transparency can easily be shown by observing Figure 5.1(b). The mere presence of an X–junction with four surrounding luminances does not result in a strong segmentation into layer and background without the added context of the rest of the figure. The questions that remain are whether X–junctions constitute a *necessary* condition for phenomenal transparency, and which characteristics of X–junctions change the saliency of a transparent percept. Adelson [178] has demonstrated a variant of the argyle illusion where two identical patches, one inside and one outside an illusory transparent strip, were shown to have a 70% difference in luminance. But when the X–junctions defining the illusory transparency were destroyed, the effect was found to be reduced to 20%. He concluded that "the sense of transparency tends to be reinforced by the X junctions" (p. 2044).

Overall, very few studies have endeavoured to quantify the specific role of junctions in phenomenal transparency. In this study, three stimuli were employed as figural manipulations to examine how destruction or perturbation of X–junctions affects the strength of a transparent percept, as measured by subjects' settings compared with the LE model.

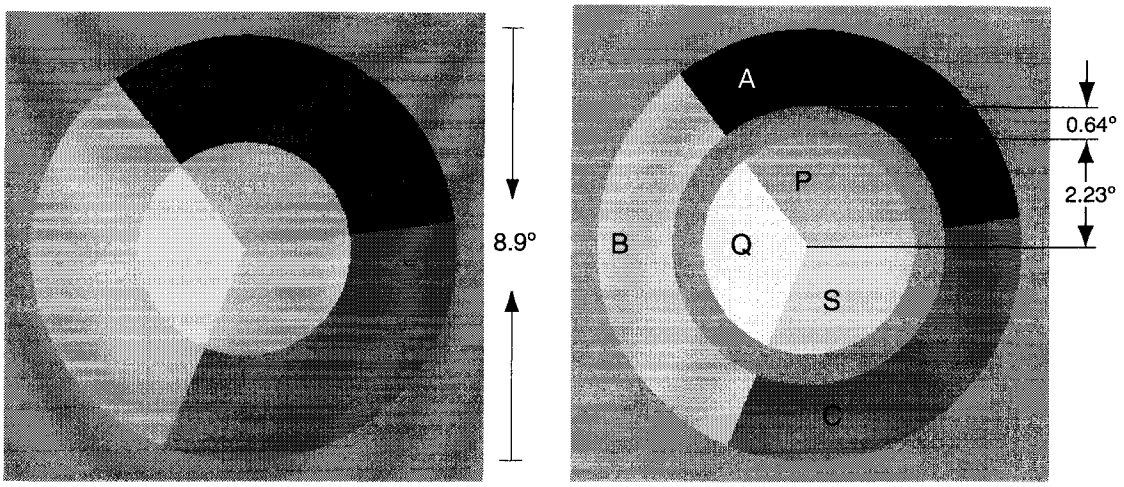


Figure 5.4: Circular six–luminance stimulus (*left*), and circular stimulus with occluding annulus and figural details (*right*).

5.1.4 Six–luminance stimulus

For this purpose a six–luminance stimulus described previously by Kasrai and Kingdom [175] was employed as shown in Figure 5.4. Six luminances are the minimum number required in a transparency stimulus to interrogate the visual system's internal model. Although four luminances (see for example Figure 5.1) are the minimum number required to obtain a transparent percept, fixing three of the luminances (e.g. A, B, and P) and manipulating the fourth luminance (Q) simply varies the reflectance or transmittance of the simulated filter [174].[2] On the other hand, in a six–luminance display such as in Figure 5.4, holding constant four of the luminances (two background and two foreground—A, B, P, and Q, respectively) fixes the transmittance and reflectance of the filter, whatever they may be in a particular subject's internal representation, while the fifth luminance (C) serves as a test background patch. The subject then adjusts the sixth luminance (S), which makes up one–third of the filter, to 'complete' the filter. It is worth noting that the geometry of the stimulus in its 'neutral' (concentric disks) configuration satisfies the conditions described

[2]The three–luminance stimulus described by Fuchs [21] which gives rise to a transparent percept under specific conditions is not considered here.

by Metelli and Kanizsa (see §5.1.3 above) in that the third (test) sector completes the inner transparent disk, and the reflectance contours of the background meet under the layer. As will be shown later, the six–luminance stimulus can also be easily modified to measure the efficacy of other figural conditions.

The qualitative effect of contour polarities on the six–luminance stimulus can be observed in Figure 5.5. The figure shows a range of six–luminance stimuli with the variable test patch gradually changing from dark to light (from top–left to bottom–right). According to the LE model, the image in the middle of the second row corresponds to the optimal transparent stimulus. The polarity of the contours is marked by arrows (pointing from darker to lighter) in three cases. Of these three exemplar stimuli it is only image **X**, in which the reflectance edges (white arrows in the angular direction) change polarity, that gives rise to a strictly nontransparent percept. In the cases of **Y** and **Z** the reflectance edges consistently have non–reversing polarity (note parallel white arrows). For the purposes of this chapter, stimuli similar to case **X** will hereafter be called polarity–reversing stimuli, referring strictly to the ordinal property of the reflectance edge.

As for the edge corresponding to the transparent filter, in this case the circular contour defined by the inner disk, recall that according to the contour polarity rules it may or may not change sign (note the parallel and anti–parallel black arrows). Also, when neither of the two intersecting contours reverses polarity, as in some of the junctions in stimuli **Y** and **Z** [just as in Figure 5.3(a)], the *local* depth–order information is metastable at the junction. Because of the unity of the figure, however, the depth–order ambiguity is resolved by the other single–reversing junction(s) in the stimulus.

In addition to the benefits derived from integrating the reference and test areas into a single display (as opposed to having two side–by–side four–luminance images as Gerbino *et al.* [160] did), the six–luminance stimulus affords certain advantages regarding stimulus

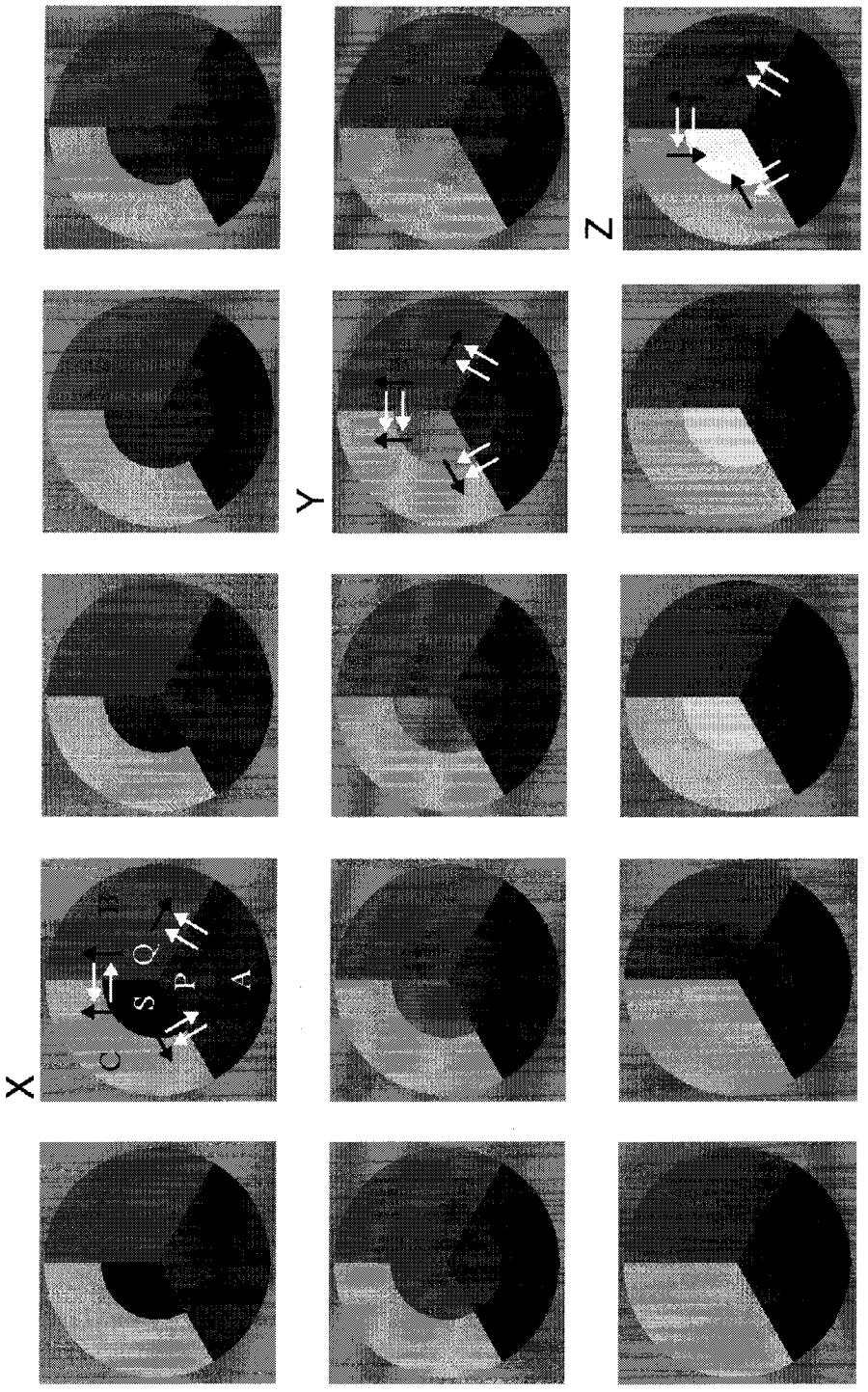


Figure 5.5: Example range of possible test patch settings. The image containing the correct luminance combination as predicted by the luminance episcotister model is in the center. Images labelled **X**, **Y**, and **Z** are sample stimuli, annotated with arrows illustrating contour polarities. White arrows indicate the polarity of reflectance edges while black arrows show the polarity of the filter's contours.

geometry. The stimulus contours can be easily manipulated to simulate different figural conditions, including those demonstrated in Figure 5.2. A brief report of some of these findings has been presented previously [179].

5.2 General methods

5.2.1 Stimulus generation

Although the figural arrangement of the stimuli changed from experiment to experiment, all the stimuli used in these studies were composed of six luminances—three background and three layer luminances [see Figure 5.4(b)]. The three background luminances (A, B, C) were chosen randomly, along with random values for the transmittance t (between 0.2 and 0.8) and reflective component ($F \geq 0$) of the filter, such that the maximum luminance of the monitor ($\simeq$ 33 cd/m^2) was not exceeded. An additional constraint ensured that no two adjacent patches had similar luminances (i.e. were within 2 cd/m^2 of each other). Two of the three layer luminances (P and Q) were calculated according to the LE model (see Eq. 5.1). The third layer luminance (S) was the variable test patch. The luminance of the area surrounding the stimulus was fixed at an intensity of half the maximum luminance produced by the display (i.e., $\simeq$ 17 cd/m^2). Image pairs were viewed binocularly through a custom–built 8–mirror stereoscope with a principal–ray path length of 45 cm and a maximum aperture of 9.5 cm $\times$ 9.5 cm, for a maximum visual angle of 7.1°. Even though all parts of the image were in the plane of fixation, a stereoscope was used because these experiments were part of a larger project in which some stimuli were presented dichoptically [180] or stereoscopically [181].

5.2.2 Display

The display specifications were identical to those in the previous chapter (see §4.2.2).

5.2.3 Subjects

The subjects consisted of seven experienced psychophysical observers (RK, FK, PP, KW, CR, HW, and NP) with normal or corrected–to–normal vision. Apart from the two authors, the other subjects were kept naïve as to the purposes of the study.

5.2.4 Procedure

Method of adjustment

Two different procedures were employed in each of the following three experiments. The first procedure, the method of adjustment, consisted of subjects manually varying the luminance of the test patch **S**, using the computer mouse to drag a slider on the display, such that they perceived the transparent layer as a contiguous filter with uniform transmissive and reflective characteristics on the tripartite background. In Experiment 1 (see §5.3) each block contained both experimental and control conditions which were randomly intermixed, resulting in 60 adjustments in each condition. In Experiments 2 and 3, each subject made 120 settings over three sessions of 40 trials. The luminance S was set to a random value at the beginning of each trial. Figure 5.5 illustrates the range of possible settings of the test patch **S**, from 'black' to 'white', for a fixed set of the parameters A, B, C, t, and F. The subjects' settings of S for such stimuli were converted to residuals by subtracting the value predicted by the LE model described in Equation 5.1:

$$\begin{aligned} S_{\text{predicted}} &= \quad tC + F \\ \text{residual} &= \quad S_{\text{predicted}} - S_{\text{experimental}}. \end{aligned} \tag{5.2}$$

The root–mean–square (rms) of the residuals for each condition were taken as a measure of the subject's accuracy in processing transparency.

Two–alternative forced–choice

As a result of the unlimited stimulus presentation time in the adjustment task, subjects may have developed cognitive strategies which aided them in making adjustments. Therefore, a second technique with a short presentation time was also used in order to minimise the possiblity of such putative computational strategies. A two–alternative forced–choice (2AFC) technique was employed whereby each subject performed a total of 720 trials (360 for each of the experimental and control conditions) in blocks of 120 trials. The stimulus presentation time was 250 ms [182], with a 500 ms interstimulus interval. A small black fixation point was added in the centre of the disk. For the 'correct' stimulus of each forced–choice pair, the test luminance S was set according to the LE model (see Eq. 5.1). For the 'incorrect' stimulus, this model luminance S was multiplied by a factor chosen at random from the following list: 0.5, 0.7, 0.9, 1.1, 1.3, 1.5. Since both this factor and the model parameters (A, B, C, t, and F) were chosen randomly, it was ensured that the expected (i.e. correct) luminance S fell between P and Q so that polarity–reversals (such as in stimulus **X** in Figure 5.5) would occur in about half the trials. Clearly one would expect performance measured by the proportion–correct responses to be, in general, worse for factors close to 1.0 which result in similar test patch luminances (**S**) in both presentations, not only because subjects would not be able to discriminate as well between two very similar stimuli, but also because such small increments or decrements are unlikely to have changed the ordinal characteristics of the X–junction. The subjects' task was to select the stimulus with "best transparency", similarly to in the adjustment task. Audio feedback was provided for incorrect responses.

5.3 Experiment 1: Occlusion of the junction

In the first experiment, the aim was to assess the result of eliminating or occluding X–junctions. We used an annulus which was fixed at the mid–grey surround colour and straddled the contour of the simulated filter. Figures 5.4(a) and 5.4(b), respectively, show the circular stimulus without and with the occluding annulus. Figure 5.4(b) describes in detail the configuration of this stimulus.

For the adjustment task, the rms residuals as calculated from Eq. 5.2 are shown in the first column–pair of Table 5.1. Lower rms scores generally indicate better performance. Even though not every subject's results reach significance on an individual basis, there is a main effect of annulus ($F_{1,826} = 9.25, p < 0.01$) showing better performance overall (lower rms) in the non–occluded stimuli.

Turning now to the forced–choice results, for each subject, data from stimuli which had been perturbed from the LE model to produce contour polarities which do not induce transparency (as in case **X** in Figure 5.5) were grouped together *a posteriori*, regardless of the random perturbation factor. These were separated from those trials where the ordinal properties were consistent with unique transparency [as in **Y** and **Z** in Figure 5.5; see also Figure 5.3(b)]. Table 5.2 shows the 2AFC proportion–correct scores for all subjects. The results (first two column–pairs) thus show four proportion–correct scores for each subject: (with annulus versus without annulus) $\times$ (polarity–reversing versus non–polarity–reversing). These individual proportion–correct scores were pooled and averaged, and a binomial test of significance [183] was performed. Both the polarity–reversal ($p \ll 0.001$) and no–polarity–reversal ($p < 0.01$) conditions showed a significant difference between the annulus and no–annulus conditions.

These results, together with the rms values from the adjustment data show that the

	Annulus	(control)	Kink	(control)	Clover	(control)
RK	2.94 (0.34)	2.61 (0.29)	2.40 (0.30)	2.16 (0.28)	2.69 (0.33)	2.48 (0.31)
FK	4.83 (0.48)	2.95 (0.30)	4.09 (0.40)	3.89 (0.45)	4.86 (0.49)	3.10 (0.34)
PP	2.75 (0.34)	2.90 (0.36)	2.76 (0.32)	3.35 (0.39)	2.93 (0.41)	3.39 (0.46)
KW	2.82 (0.42)	2.65 (0.33)	2.38 (0.29)	2.93 (0.37)	2.85 (0.33)	2.51 (0.30)
CR	3.91 (0.55)	3.31 (0.38)	3.35 (0.42)	3.85 (0.47)	4.06 (0.46)	4.10 (0.46)
NP	3.44 (0.44)	2.54 (0.36)	2.24 (0.31)	3.07 (0.37)	3.22 (0.52)	2.58 (0.28)
HW	3.08 (0.35)	2.86 (0.33)	2.88 (0.38)	2.95 (0.34)	3.29 (0.42)	2.82 (0.36)
Mean	3.40	2.83	2.87	3.17	3.42	3.00
p	< 0.01		0.03		< 0.01	

Table 5.1: Root–mean–square (rms) residuals (cd/m^2) for all subjects. The second–to–last row shows the rms error for each condition, averaged across subjects. The standard error, based on repeated measures, is marked in parentheses next to each value. The last row shows the two–way ANOVA results.

simple occluding annulus shown in the stimuli in Figure 5.4 slightly, but significantly, reduces or degrades the ability to process transparency accurately.

It must be pointed out that the occluding annulus did not change the alignment of the contours that make up the X–junction, and was in fact concentric with the original layer–background and background–surround contours. An argument may thus be made that although the *local* X–junction had been destroyed in Experiment 1, according to the classical definition of X–junctions (see §5.1.3) the visual system may easily interpolate between the two pairs of collinear contours into a (non–illusory) X–junction.[3] Nonetheless the destruction of the local X–junction resulted in a sufficient disruption of the transparent percept to reduce performance on average in both adjustment and 2AFC tasks.

[3]Indeed, the observers in the this study did not report perceiving illusory contour completions across the annulus.

Expt.	Annulus				Kink				Clover			
Polarity	Rev		No Rev		Rev		No Rev		Rev		No Rev	
Condition	Exp	Cont	Exp	Cont	Exp	Cont	Exp	Cont	Exp	Cont	Exp	Cont
RK	0.80	0.97	0.71	0.67	0.93	0.95	0.68	0.69	0.91	0.98	0.66	0.65
(BE)	0.03	0.01	0.03	0.04	0.02	0.02	0.04	0.04	0.02	0.01	0.04	0.04
FK	0.80	0.86	0.55	0.70	0.81	0.84	0.65	0.62	0.57	0.76	0.55	0.63
(BE)	0.03	0.03	0.04	0.04	0.03	0.03	0.04	0.04	0.04	0.03	0.04	0.03
PP	0.66	0.83	0.56	0.60	0.78	0.71	0.68	0.64	0.85	0.78	0.59	0.61
(BE)	0.04	0.03	0.04	0.04	0.03	0.03	0.04	0.04	0.03	0.03	0.04	0.04
KW	0.89	0.89	0.60	0.60	0.96	0.92	0.64	0.68	0.93	0.95	0.66	0.63
(BE)	0.02	0.02	0.04	0.04	0.01	0.02	0.04	0.04	0.02	0.02	0.04	0.04
CR	0.66	0.84	0.57	0.62	0.82	0.91	0.64	0.69	0.79	0.93	0.62	0.66
(BE)	0.03	0.03	0.04	0.04	0.03	0.02	0.04	0.03	0.03	0.02	0.04	0.04
NP	0.56	0.89	0.58	0.68	0.78	0.88	0.67	0.68	0.91	0.88	0.59	0.60
(BE)	0.04	0.02	0.04	0.04	0.03	0.02	0.03	0.04	0.02	0.02	0.04	0.04
HW	0.72	0.79	0.60	0.65	0.80	0.86	0.68	0.56	0.77	0.78	0.56	0.66
(BE)	0.03	0.03	0.04	0.04	0.03	0.03	0.03	0.04	0.03	0.03	0.04	0.04
Mean	0.73	0.87	0.60	0.65	0.84	0.87	0.66	0.65	0.82	0.87	0.60	0.63
(BE)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
p	$\ll 0.001$		< 0.01		0.03		0.31		$\ll 0.001$		0.06	

Table 5.2: Proportion correct scores for all subjects. Each column–pair contains the score for the experimental condition ('Exp', on the left side) followed by the score for the corresponding control condition ('Cont', right side). The respective binary errors (BE) are shown below each score. The second–to–last row shows for each condition the average score across all subjects, and the last row shows the p values resulting from the ensemble binary tests for significance between experimental and control conditions.

In order to ensure that the performance differences observed in the presence of the annulus were due only to the occlusion of the X–junctions and not to the particular unified shape of the annulus, an additional control experiment was performed with a different stimulus, shown in Figure 5.6. Mid–grey disks with diameters equal to the width of the

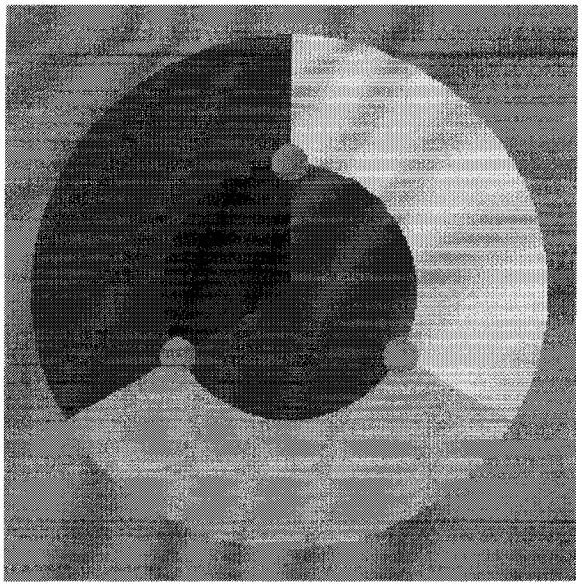


Figure 5.6: Stimulus for local junction–occlusion control experiment. Mid–grey disks are used to occlude the X–junctions.

annulus [see Figure 5.4(b)] were superimposed on the original six–luminance stimulus to occlude the three X–junctions locally. Only the 2AFC technique was used with the participation of four (RK, FK, NP, and KW) of the seven subjects. They performed 720 trials each over 12 blocks; the luminances in the stimuli were calculated as above and the two conditions (grey patches versus no occlusion) were interleaved randomly within the blocks.

The data, analysed as above using a binomial test, reveal a significant difference in performance for the stimuli containing polarity reversals ($p < 0.001$), but no difference for the no–polarity–reversal stimuli ($p = 0.19$). Along with the results of the previous annulus experiment, we can conclude from the trends in this experiment that the occlusion of X–junctions does indeed hinder subjects' ability to encode transparency, be the occlusion limited locally or part of a global stimulus structure.

5.4 Experiment 2: Discontinuity of the background contour

In the next two experiments the significance of the alignment of the four contours was examined. Kanizsa's two examples ([22]; see also Figure 5.2) showed stimuli where the alignment of the contours was not preserved. In other words, either the contours corresponding to the transparency edge (filter–background), or those corresponding to the reflectance edge (separating the two background colours) have an undefined gradient at the Ψ–junction. Even though Kanizsa was not explicit in his categorisation of these two different conditions, he claimed in the former case that the figural conditions were "not favorable to the impression of transparency", whereas for the latter he states that "continuity of direction is not a necessary condition of phenomenal transparency" (p. 161).

Two variants of the original circular six–luminance stimulus were used to examine the two cases separately (in this experiment and the next), each with its own control condition.

The stimulus, shown in Figure 5.7, was the same as the original six–luminance stimulus, except that the reflectance contours of the background were kinked (or tilted) by 45°. In addition to the 120 adjustments on this stimulus, the subjects also made another series of adjustments on a control stimulus with 0° kink in the background contours. Even though this control stimulus is identical to the no–annulus stimulus of Experiment 1 [Figure 5.4(a)], these data were collected contiguously along with the 45° kink condition because of the time lag in data collection between Experiments 1 and 2. Note that whereas in Experiment 1 the experimental and control stimuli were interleaved randomly within the adjustment and 2AFC blocks, in Experiments 2 and 3 it was the experimental blocks which were presented in quasi–random order.

The residuals from the adjustment data were again compared to the expected settings

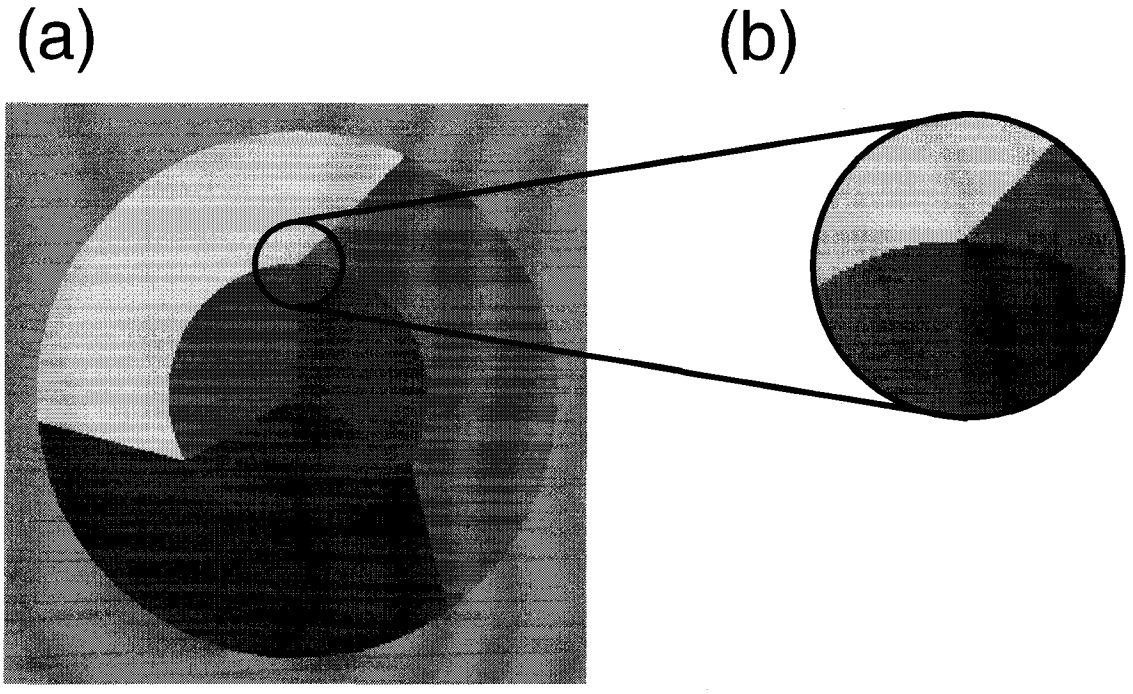


Figure 5.7: (a) Stimulus for Experiment 2 with 45° kink, and inset (b) showing Ψ–junction.

derived from the LE model. The subjects' results, similarly to Experiment 1, show no consistent *individual* differences between the kink and no–kink conditions. However, the pooled two–way ANOVA results show a significant effect of kink angle ($F_{1,1666} = 4.51, p = 0.03$) but, as shown by the subject means of Tables 5.1 and 5.2, this time with *better* performance in the kink condition.

The binomial test also revealed a significant overall difference in the pooled and averaged 2AFC proportion–correct data in the polarity–reversal condition ($p = 0.03$) but not in the no–polarity–reversal condition ($p = 0.31$). As with the adjustment results above, the subjects' overall performance was *better* (lower rms, and higher proportion–correct scores) with the experimental stimulus (45° kink) than with the control stimulus (no kink).

Figure 5.8: (a) Clover stimulus for Experiment 3, and (b) rotated clover stimulus as control condition

5.5 Experiment 3: Discontinuity of the layer contour

In this experiment we destroyed the *continuity* of the filter–background contour. For this, we used a clover–leaf pattern for the simulated transparent layer. The geometry of the layer consisted of three overlapping circles, each with a 1.4° radius, with centres about 2.2° away from the stimulus centre, producing a three–leaf clover pattern. In the experimental condition, the orientation of the clover pattern with respect to the tripartite background was such that a so–called minimum of curvature occurred at the Ψ–junction, as shown in Figure 5.8(a). In the control condition the clover pattern was rotated 60° with respect to the background, restoring the continuity of the filter boundary at the junction, as in Figure 5.8(b). The latter served as a natural control stimulus because apart from the contour geometry of the junction the individual sector areas and overall shape of the filter remained constant. Thus, any performance differences cannot be due to the shape or area of the overlay.

In both the adjustment and 2AFC experiments, subjects' overall performance was worse in the experimental condition compared with the control. A two–factor (subjects $\times$ clover orientation) ANOVA of the adjustment data shows an effect of the orientation of the transparency overlay ($F_{1,1666} = 8.69, p < 0.01$).[4] For the 2AFC data, the binomial test shows a significant difference only in the polarity–reversal data sets ($p \ll 0.001$), and not in the no–polarity–reversal data sets ($p = 0.06$).

5.6 Summary and general discussion

For ease of analysis, the averaged results are replotted in Figures 5.9 and 5.10. The plots show respectively the rms residuals for all adjustment conditions and percent–correct scores for all 2AFC conditions, averaged across all subjects. The results of the statistical analyses (see the bottom rows in Tables 5.1 and 5.2) generally correspond to what can be gleaned from the phenomenal experience associated with the stimuli, namely that under certain figural conditions the degradation of the transparency percept is matched by a reduction in the accuracy with which subjects process transparency. Performance is on average worse in relation to the LE model (see Eq. 5.1) in the stimuli where the X–junction is occluded (Experiment 1) or where an extremum of curvature of the layer–background contour exists at the Ψ–junction (Experiment 3). Conversely, a discontinuity in the direction of the contours which define a reflectance boundary (the background) at the Ψ–junction does not degrade performance (Experiment 2) but slightly improves it. These quantitative results are in agreement with Kanizsa's qualitative arguments demonstrated by the images in Figure 5.2.

[4]It should be noted that in this experiment, as in the previous two, there was also a significant effect due to differences between the subjects ($p \ll 0.01$). In addition, there were also effects of interaction between subjects (rows) and experimental condition (columns) in this experiment ($F_{6,1666} = 4.21, p < 0.001$) as well as in Experiment 1 ($F_{6,826} = 2.26, p = 0.04$). These effects are neither particularly surprising nor as interesting as the differences due to the changes in the stimulus condition.

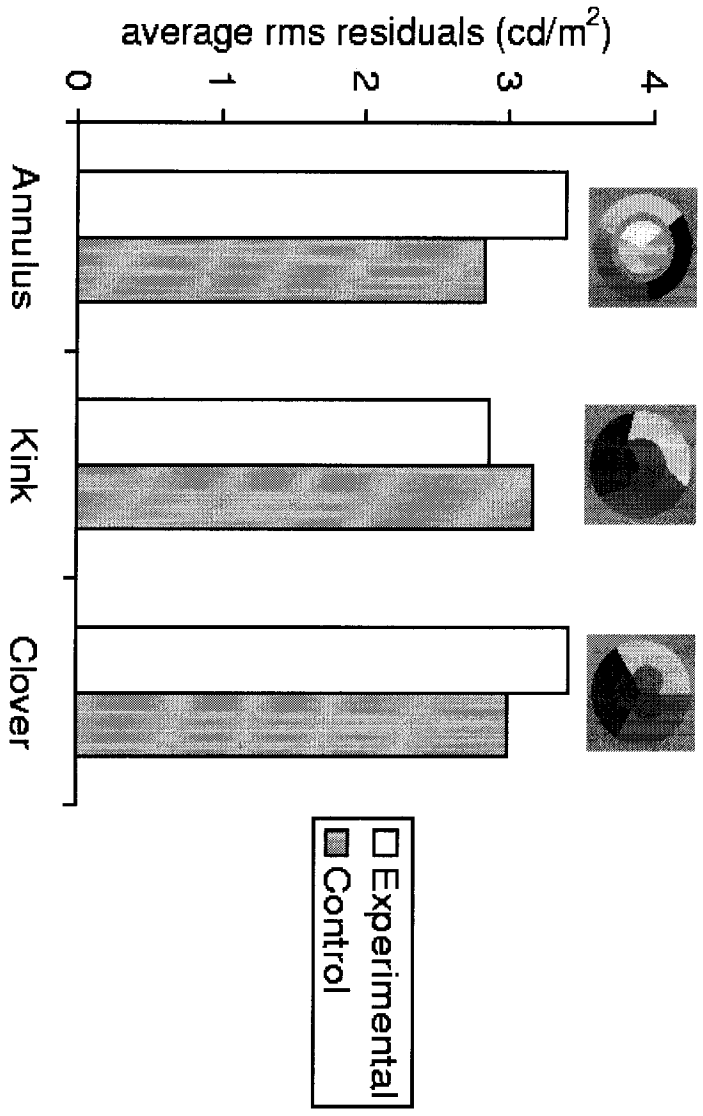


Figure 5.9: Adjustment data for all three experiments averaged across all subjects. See Table 5.1 for details.

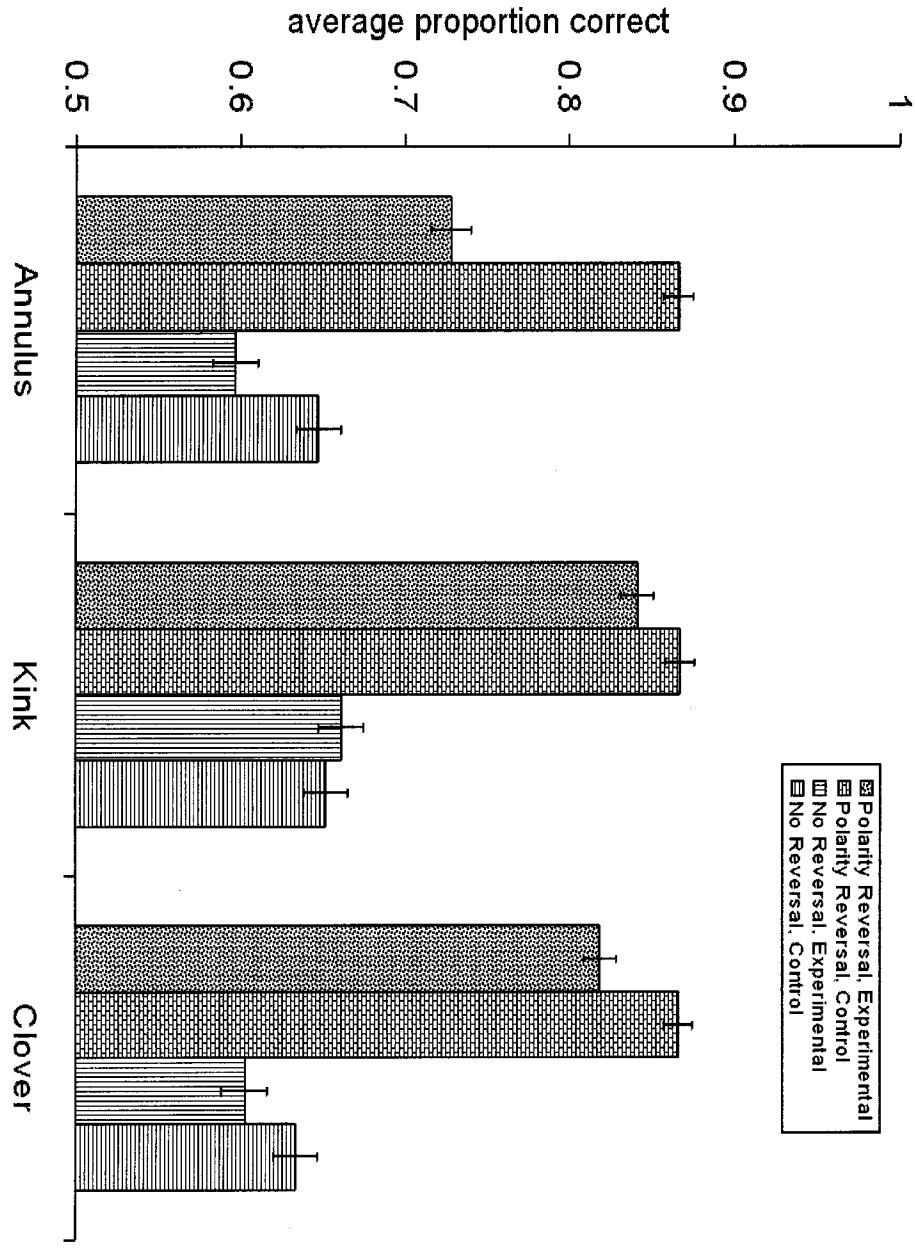


Figure 5.10: 2AFC data for all three experiments averaged across all subjects. Error bars are similarly subject–averaged binary errors (BE). See Table 5.2 for details.

Singh and Hoffman (1998) have discussed in detail the case where the contour of the transparent filter (but not that of the background) has a direction discontinuity at the junction. They manipulated the sign of curvature, turning angle, and level of smoothing of the filter contour, and asked subjects to judge the change in the salience of perceptual transparency by rating on a 7–point scale the strength of a four–luminance transparency stimulus. They found that it was sufficient for the filter contour to have a strong maximum or minimum of curvature at the junction in order to have a loss of phenomenal transparency. In the study reported here, rather than rating the subjective saliency of the transparency, the subjects' *accuracy* in processing simulated transparency was measured. In addition, not only was the influence of the contour of the filter examined here, but also that of the reflectance contour. The results of the clover experiment (Experiment 3) are consistent with Singh and Hoffman's (1998) results insofar as the salience of the clover pattern as a transparent overlay was reduced compared with the condition with continuous contours through the junction.

5.6.1 Method of adjustment results

The results of the statistical analyses require a caveat, however. If the rms results are reanalysed in a leave–one–out fashion whereby the data from one of the five subjects are omitted in turn, the relative weight or influence of each subject's results can be determined. In the occluding–annulus condition (Experiment 1), leaving out CR or RK's results does not change the ANOVA results ($p \simeq 0.01$). Leaving out FK's data, on the other hand, raises the p–value to 0.11, indicating that the large difference between FK's annulus and no–annulus conditions single–handedly pushes the ensemble results into being statistically significant. Similarly, in the clvoer condition (Experiment 3), the p–value increases from $p < 0.01$ to $p = 0.31$ when FK's data are omitted. In light of this one may say that the

effect using the method of adjustment is subject to large interobserver variability.

5.6.2 Two–alternative forced–choice results

The 2AFC results are generally more consistent than the adjustment data, showing larger differences between experimental conditions (cf. RK and CR's 2AFC polarity–reversal data in all three conditions). A cursory inspection of the 2AFC data reveals that, even when subjects' results do not reach significance on an individual basis, the differences between the control and experiment data (in the annulus and clover experiments) across all subjects are generally in the same direction. Consequently the same leave–one–out examination of the results as in the adjustment data above was not performed. The differences between the adjustment and 2AFC results appear to confirm our apprehension regarding the use of cognitive strategies in the adjustment task. Since subjects had unlimited time to complete each adjustment, and there was no audio feedback possible for training purposes, it is difficult to know unequivocally the basis of their adjustment stratagem.

An additional set of analyses can be performed on the 2AFC results regarding the ordinal rules for '4–junctions' (see §5.1.2). With reference to Table 5.2, the effects of changes in contour polarity which are inconsistent with transparency perception can be seen by comparing each polarity–reversal column–pair with its corresponding no–polarity–reversal column–pair in the same condition. For example, considering the row of mean percent–correct scores (second–last row of cells) in the clover condition (Experiment 3), the scores for the polarity–reversing stimuli are dramatically reduced from 0.82 and 0.87 to 0.60 and 0.63. A binary test of the data for all three experiments showed a highly significant ($p \ll 0.001$) superiority in performance in the polarity–reversal cases compared to the ones with no polarity reversals, confirming what can be observed by visual inspection. Subjects thus performed on average worse on trials in which both stimuli had contour polarities

consistent with a transparent filter, compared with trials which contained an 'incorrect' stimulus with reversals in the polarity of the background contour. This indicates that the polarity of 4–junctions is an extermely salient feature which is used to distinguish plausible from implausible transparent media, as hypothesised.

5.6.3 Possible explanations

On the basis of these results it is possible to speculate as to the mechanisms underlying the perception of achromatic transparency. As mentioned in the introduction, transparencies and shadows can be more generally thought of as bringing about changes in illumination. It is conceivable, as Adelson and Somers [184] have suggested that the visual system considers jagged edges as more likely to be reflectance edges than illumination ones. Reflectance changes in an image define not only textures on surfaces, but also material changes between *different* surfaces, be they two surfaces in different depth planes or adjacent to one another. On the other hand, illumination contours are usually due to shading, changes in lighting level, or transparent filters. These latter contours are more likely to be either smoothly varying or blurred, or, when sharp, due to straight edges of objects which somehow occlude the illuminant. In other words, whereas reflectance contours are generally non–differentiable (i.e. have an undefined gradient) along the direction of the contour, illumination edges are more likely to be straight or smoothly varying.

This hypothesis may also serve to explain why there was a slight improvement of performance in the 45° kink condition compared with the control condition. Although in theory sudden kinks in the background (reflectance) contours are not inconsistent with transparency, it may be the case here that the control condition (0° kink) created a weak three–dimensional effect which was destroyed in the experimental condition. Even though there was no disparity in the images, some subjects reported perceiving the center of the

disk as the corner of a room or of a cube. This interpretation was less likely in the experimental condition, and this may have aided the subjects' performance.

Chapter 6

Conclusions

"But Marlow was not typical (if his propensity to spin yarns be excepted), and to him the meaning of an episode was not inside like a kernel but outside, enveloping the tale which brought it out only as a glow brings out a haze, in the likeness of one of these misty halos that sometimes are made visible by the spectral illumination of moonshine."

—Joseph Conrad, *Heart of Darkness*

THE science of optics in its early stages was motivated by interest in the relationship between the mind and the body, and in turn, in the interaction of the mind with the physical world. This dissertation continues in the direction of both of these traditional questions. Where is the boundary between Mind and Brain? How does the mind perceive its environment? If the eyes are windows onto the soul (or mind) then surely the same organ can be employed by the mind to view the world. As humankind's descriptive knowledge of the outside world increases, however, the mind's eye begins to gaze on the brain itself in the quest to understand itself. One may say that this in turn requires a specific understanding of the function of the mind's organ of sight—the visual system.

6.1 General summary

This work addresses the above questions using two related approaches. Whereas the approach of Chapter 3 made use of medical images as the stimuli in a medically relevant task, Chapters 4 and 5 used much simpler stimuli in tasks involving fewer variables. Neurosurgery and neurological procedures rely to an increasing extent on pre–operative images. Though the end–users of IGNS systems are clinicians with extensive training in neuroanatomy, their visual system is not free of the biases and shortcomings of the generic human visual system. As a result, it is worthwhile examining how humans interact with rich medical images when carrying out a particular task. This richness and complexity is defined by the very information that application designers are trying to convey to the end–user. This visual information, as discussed in Chapter 2, comes from a variety of sources and is typically spatially co–registered to form a multi–modal image often involving the use of transparency. Yet, during an image–guided procedure, rarely is all the information relevant at the same time. It becomes important then to convey only the most relevant data in the most understandable way. Work has already begun [144] on compiling a database of neurological procedures and their sub–parts in order to predict the types of images required according to a script. The question remains as to what makes an image most understandable.

6.1.1 Summary of findings

Using the approach taken in Chapter 3, having defined a specific clinically–relevant task, a series of experiments aimed to determine the combination of display variables which yielded the best performance by the subjects. It was found that in the task as defined, while stereo disparity was an important cue, the choice of surface rendering parameters (*viz.*

opacity and Gouraud–shading) did not influence subjects' performance. In addition, in a pilot experiment using both 2–D and 3–D mice to interact with and shape the surfaces of interest, it was found that subjects performed better using a 3–D tracking device compared to a standard (translational) 2–D mouse. While additional control experiments may be undertaken to clarify whether the performance differences were due to purely ergonomic factors (e.g. lack of limb support in the 3–D case) or to perceptual ones, the results indicate that three–dimensional interaction in this task does not confer an automatic advantage.

Following these findings, the second approach was to reduce the stimuli to their most basic elements. Perceptual transparency can be thought of as requiring two necessary conditions: intensity (or luminance) and form. Chapters 4 and 5 dealt with each of each conditions respectively, using a novel six–luminance stimulus simulating transparency.

The intensity condition describes the combination of background and layer luminances and the models which predict perceptual transparency. In Chapter 4 different models were tested based on the accuracy and precision of subjects' settings using the method of adjustment. The results indicated that subjects were both accurate and precise in their settings, which corresponded most closely to the luminance formulation of the episcotister model and the Michelson contrast model (see Equations 4.2 and 4.5).

The flexibility or range over which subjects perceive some degree of (not necessarily optimal) transparency was revealed by a two–alternative forced–choice technique. Since the 2AFC technique typically requires an *a priori* model which determines the 'correct' stimulus to be displayed during one of the two presentations of an experimental trial, a method had to be devised to analyse the data without assuming a particular model. As a result, the modified method of pair comparisons was developed in order to retrieve the probability density function describing the flexibility or range of acceptable transparency stimuli. The forced–choice results showed that while subjects can set their 'best trans-

parency' quite precisely, there is a wide range around this setting which evokes at least some degree of perceptual transparency. The limits of this range are constrained by the ordinal conditions described in §5.1.2.

Having validated the models describing the necessary and sufficient luminance relationships, in Chapter 5 the role of contour shape was examined. It has been suggested that when contours between the filter and background (a change in illumination) cross contours between the two background colours (a change in material), the resulting junctions play a significant role in transparency perception. In Chapter 5 the junctions in the six–luminance stimulus were either occluded or perturbed in four experiments.

The data suggest that the two types of contours are qualitatively different for the visual system. When the material (background) contour has a change in orientation at the X–junction, subjects' accuracy (as compared to the luminance episcotister model) improves, whereas orientation discontinuities of the illumination contour at the X–junction hamper accurate processing of transparency stimuli. Occlusion of the junction was also shown to degrade performance, indicating that the existence of X–junctions is a salient feature of simulated transparency. Even though these observed effects were not very large, nor necessarily consistent across all the subjects, the combined results were statistically significant. In the forced–choice studies involving an additional factor having to do with the polarity of contours, a highly significant superiority in the detection of 'incorrect' stimuli was shown. This finding not only confirms quantitatively what can be appreciated phenomenologically, but also what has been suggested in previous literature.

These results lead one to speculate as to the lack of improvement in subjects' performance in the segmentation task of Chapter 3 when the surface was rendered transparently. As discussed in §3.8, one may assume that the task was essentially one of contour matching. In light of the results of Chapter 5 one may suggest that the non–existence of

X–junctions may have contributed to the lack of difference in performance between the opaque and transparent conditions. Though the mesh–rendering condition contained numerous X–junctions, these junctions were all associated with a texture on the surface and not with crossings of background and layer contours. There is evidence to suggest that articulation[1] of the background surface, accompanied by the associated junctions on the edge of the filter facilitates transparency perception [174].

6.2 Suggestions for future work

In the context of medical image visualisation, it is as yet unclear whether transparency has an automatic beneficial effect. As Tiede [87] suggests, "there is no doubt that an object that is opaque in reality should be rendered as an opaque object to get the highest fidelity. Exceptions that justify transparent rendering are cases where we really want to look into an object that does not have distinguishable constituents or when we are not able to determine the true surfaces automatically." But if the goal is to display a structure as well as other objects it encompasses, what is the appropriate rendering method? Are judgments such as distance or colour (intensity differences) through transparent layers adversely affected? Is the human visual system capable of simultaneously distinguishing and processing more than one transparent layer? Clearly, further psychophysical research is required not only in the use of transparency in medical images, but also, in a general way, in displaying and compositing images from different sources.

Although this dissertation was motivated initially by applications of transparency to medical visualisation, psychophysical investigations of transparency of the kind described in Chapters 4 and 5 are interesting in their own right. It is quite plausible that, given the

[1] Articulation refers to luminance variations on a surface, such as a sine grating or a texture.

paucity of transparent layers in nature, the mechanism for detecting and encoding transparency makes use of pathways which already exist for colour constancy and the detection of shadows. This link has been suggested by the work of Westland and Ripamonti [171]. What is the relationship between the detection of shadows (a limiting case of transparency) and transparent layers? Work along these lines has already begun in examining the role of colour vision in disambiguating material and illumination changes in an image [185]. Similarly, while the studies of Chapters 4 and 5 examined specifically the visual system's ability to encode transparent layers, there are no known references in the visual psychophysics literature to studies of colour constancy behind transparent surfaces. To what degree is the visual system capable of discounting the characteristics of an overlying layer in order to maintain colour constancy on the background? The latter is the subject of ongoing work at the McGill Vision Research Group.

In addition, it seems unlikely that stereo disparity is unrelated to the detection of transparent layers. Layers, by definition, are on a plane independent of the background. In the absence of stereo disparity, and in the case of shadows, they are considered flush with the background, but as the Anderson [25] has suggested there may be a close link between transparent layers and junctions in the presence of disparity cues. Thus, interest in transparency is not limited to the few examples which occur in natural scenes, but also extends to a number of other phenomena with which transparency may interact.

Bibliography

[1] A. I. Sabra, ed., *The optics of Ibn Al–Haytham. Books I–III. On Direct Vision.* London: The Warburg Institute, 1989.

[2] M. Woo, J. Neider, and T. David, *OpenGL programming guide: the official guide to learning OpenGL, version 1.1.* Reading, MA: Addison–Wesley Developers Press, 1997.

[3] D. Kersten, *Computational models of visual processing*, ch. Transparency and the cooperative computation of scene attributes, pp. 209–228. Cambridge, MA: MIT Press, 1991.

[4] W. Gerbino, *Lightness, Brightness, and Transparency*, ch. Achromatic Transparency, pp. 215–255. Hillsdale, NJ: Lawrence Erlbaum Associates, 1994.

[5] M. Singh and B. L. Anderson, "Perceptual assignment of opacity to translucent surfaces: The role of image blur," *Perception*, vol. 31, no. 5, pp. 531–552, 2002.

[6] F. Metelli, "The perception of transparency," *Scientific American*, vol. 230, pp. 90–98, 1974.

[7] P. Whittle, *Lightness, Brightness, and Transparency*, ch. Contrast Brightness and ordinary seeing, pp. 111–157. Hillsdale, NJ: Lawrence Erlbaum Associates, 1994.

[8] M. Singh and B. L. Anderson, "Toward a perceptual theory of transparency (in press)," *Psychological Review.*

[9] L. da Vinci, *A treatise of painting.* London: Senex and Taylor, 1721.

[10] N. J. Wade, *A natural history of vision*. Cambridge, MA: MIT Press, 1998.

[11] F. Metelli, "An algebraic development of the theory of perceptual transparency," *Ergonomics*, vol. 13, pp. 59–66, 1970.

[12] A. Lejeune, ed., *L'optique de Claude Ptolémée. Deux stades de l'optique geométrique Grecque*. Louvain: Université de Louvain, 1956.

[13] P. van Musschenbroek, *Introductio ad philosophiam naturalem*, vol. II of *sect. 1820*. 1768.

[14] J. Plateau, *Dissertation sur quelques propriétés des impressions produites par la lumière sur l'organe de la vue*. 1829.

[15] W. H. F. Talbot, "Experiments on light," *The London and Edinburgh Philosophical Magazine and Journal of Science*, vol. 5, pp. 321–334, 1834.

[16] J. C. Maxwell, "Experiments on colour, as perceived by the eye, with remarks on colour–blindness," *Transactions of the Royal Society of Edinburgh*, vol. 21, no. 2, pp. 275–298, 1855.

[17] E. G. Boring, *Sensation and Perception in the History of Experimental Psychology*. New York, NY: D. Appleton–Century Company, Inc., 1941.

[18] J. A. F. Plateau, "Sur un principe de photométrie," *Bulletins de l'Académie Royale des Sciences et Belles–Lettres de Bruxelles*, vol. 2, pp. 52–59, 1835.

[19] H. L. F. von Helmholtz, *Handbuch der Physiologischen Optik*. New York, NY: Dover Publications, Inc., 1866/1962.

[20] J. Beck, K. Prazdny, and R. Ivry, "The perception of transparency with achromatic colors," *Perception & Psychophysics*, vol. 35, pp. 407–422, 1984.

[21] W. Fuchs, "Experimentelle untersuchungen über das simultane hintereinandersehen auf derselben sehrichtung," *Zeitschrift für Psychologie*, vol. 91, pp. 145–235, 1923.

[22] G. Kanizsa, *Organization in Vision.* New York, NY: Praeger Publishers, 1979.

[23] G. M. Heider, "New studies in transparency, form and colour," *Psychologische Forschung*, vol. 17, pp. 13–56, 1933.

[24] B. Tudor-Hart, "Studies in transparency, form, and colour," *Psychologische Forschung*, vol. 10, pp. 255–298, 1928.

[25] B. L. Anderson, "A theory of illusory lightness and transparency in monocular and binocular images: the role of contour junctions," *Perception*, vol. 26, pp. 419–453, 1997.

[26] A. R. Smith, "Painting tutorial notes," in *SIGGRAPH '79 course on Computer Animation Techniques*, ACM, 1979.

[27] B. A. Wallace, "Merging and transformation of raster images for cartoon animation," in *Proceedings of ACM SIGGRAPH '81*, vol. 15, pp. 253–262, ACM, 1981.

[28] T. Porter and T. Duff, "Compositing digital images," in *Proceedings of ACM SIGGRAPH '84*, vol. 18, pp. 253–259, ACM, 1984.

[29] K. Rehm, S. C. Strother, J. R. Anderson, K. A. Schaper, and D. A. Rottenberg, "Display of merged multimodality brain images using interleaved pixels with independent color scales," *Journal of Nuclear Medicine*, vol. 35, no. 11, pp. 1815–1821, 1994.

[30] T. Duff, "Compositing 3D rendered images," in *Proceedings of ACM SIGGRAPH '85*, vol. 19, pp. 41–44, 1985.

[31] M. Solaiyappan, *Handbook of Medical Image Processing and Analysis*, ch. Visualization Pathways in Biomedicine, pp. 659–684. San Diego, CA: Academic Press, 2000.

[32] M. W. Vannier and J. L. Marsh, "Three–dimensional imaging, surgical planning, and image–guided therapy," *Radiologic Clinics of North America*, vol. 34, pp. 545–563, May 1996.

[33] T. M. Peters, "Image–guided surgery: from x–rays to virtual reality," *Computer Methods in Biomechanics & Biomedical Engineering*, vol. 4, no. 1, pp. 27–57, 2000.

[34] J. Cox and R. C. Kirkpatrick, "The new photography with report of a case in which a bullet was photographed in the leg," *The Montreal Medical Journal*, vol. XXIV, p. 43ff, March 1896.

[35] N. L. Dorward, "Neuronavigation—the surgeon's sextant," *British Journal of Neurosurgery*, vol. 11, no. 2, pp. 101–103, 1997.

[36] H. Handels, J. Ehrhardt, W. Plötz, and S. J. Pöppl, "Three–dimensional planning and simulation of hip operations and computer–assisted construction of endoprostheses in bone tumor surgery," *Computer Aided Surgery*, vol. 6, pp. 65–76, 2001.

[37] A. Pernozzoli, C. Burghart, J. Brief, S. Haßfeld, J. Raczkowsky, J. Mühling, U. Rembold, and H. Wörn, "A real–time CORBA based system architecture for robot assisted craniofacial surgery," *Studies in Health Technology and Informatics*, vol. 70, pp. 253–255, 2000.

[38] S. Barré, C. Fernandez, P. Paume, and G. Subrenat, "Three–dimensional visualization system as an aid for facial surgical planning," in *Proceedings of SPIE Medical*

Imaging 2001: Visualization, Display, and Image–Guided Procedures (S. K. Mun, ed.), vol. 4319, (Bellingham, WA), pp. 252–263, SPIE, February 2001.

[39] T. Fortin, J. L. Coudert, G. Champleboux, P. Sautot, and S. Lavallée, "Computer–asssited dental implant surgery using computed tomography," *Journal of Image Guided Surgery*, vol. 1, pp. 53–58, 1995.

[40] Y. P. Starreveld, D. G. Gobbi, K. W. Finnis, and T. M. Peters, "Software components for medical image visualization and surgical planning," in *Proceedings of SPIE Medical Imaging 2001: Visualization, Display, and Image–Guided Procedures* (S. K. Mun, ed.), vol. 4319, pp. 546–556, SPIE, February 2001.

[41] H. K. Gumprecht, D. C. Widenka, and C. B. Lumenta, "BrainLab VectorVision neuronavigation system: technology and clinical experiences in 131 cases," *Neurosurgery*, vol. 44, no. 1, pp. 97–105, 1999.

[42] J. Wadley, N. Dorward, N. Kitchen, and D. Thomas, "Pre–operative planning and intra–operative guidance in modern neurosurgery: a review of 300 cases," *Annals of the Royal College of Surgeons of England*, vol. 81, pp. 217–225, 1999.

[43] B. L. K. Davey, R. M. Comeau, P. Munger, L. Pisani, D. Lacert, A. Olivier, and T. M. Peters, "Multimodality interactive stereoscopic image–guided neurosurgery," in *Proceedings of Visualization in Biomedical Computing* (R. A. Robb, ed.), vol. 2359, (Bellingham, WA), pp. 526–536, SPIE, 1994.

[44] J. Haase, "Image–guided neurosurgery/neuronavigation/the SurgiScope—reflexions on a theme," *Minimally Invasive Neurosurgery*, vol. 42, pp. 53–59, 1999.

[45] J. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[46] D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, "Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space," *Journal of Computer Assisted Tomography*, vol. 18, pp. 192–205, 1994.

[47] W. L. Nowinski, A. Fang, B. T. Nguyen, J. K. Raphel, L. Jagannathan, R. Raghavan, R. N. Bryan, and G. A. Miller, "Multiple brain atlas database and atlas–based neuroimaging system," *Computer Aided Surgery*, vol. 2, pp. 42–66, 1997.

[48] W. L. Nowinski, T. T. Yeo, and G. L. Yang, "Atlas–based system for functional neurosurgery," in *Proceedings of SPIE Medical Imaging: Image Display* (Y. Kim, ed.), vol. 3031, (Bellingham, WA), pp. 92–103, SPIE, 1997.

[49] P. St-Jean, A. F. Sadikot, D. L. Collins, D. Clonda, R. Kasrai, A. C. Evans, and T. M. Peters, "Automated atlas integration and interactive three–dimensional visualization tools for planning and guidance in functional neurosurgery," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 672–680, 1998.

[50] J. Koivukangas, *Stereotactic and Image Directed Surgery of Brain Tumours*, ch. Development of ultrasound guided brain tumour surgery, pp. 111–124. London, UK: Churchill Livingstone, 1993.

[51] R. M. Comeau, A. Fenster, and T. M. Peters, "Intra–operative ultrasound imaging in image–guided neurosurgery," *Radiographics*, vol. 18, pp. 1019–1027, July 1998.

[52] N. Hata, T. Dohi, H. Iseki, and K. Takakura, "Development of a frameless and armless stereotactic neuronavigation system with ultrasonographic registration," *Neurosurgery*, vol. 41, no. 3, pp. 608–614, 1997.

[53] O. Ganslandt, R. Fahlbusch, C. Nimsky, H. Kobler, M. Moeller, R. Steinmeier, J. Romstoeck, and J. Vieth, "Functional neuronavigation with magnetoencephalog-

raphy: outcome in 50 patients with lesions around the motor cortex," *Journal of Neurosurgery*, vol. 91, pp. 73–79, July 1999.

[54] F.-E. Roux, D. Ibarrola, M. Tremoulet, Y. Lazorthes, P. Henry, J.-C. Sol, and I. Berry, "Methodological and technical issues for integrating functional magnetic resonance imaging data in a neuronavigational system," *Neurosurgery*, vol. 49, no. 5, pp. 1145–1157, 2001.

[55] D. Croteau, L. Scarpace, D. Hearshen, J. Gutierrez, J. L. Fisher, J. P. Rock, and T. Mikkelsen, "Correlation between magnetic resonance spectroscopy imaging and image–guided biopsies: Semiquantitative and qualitative histopathological analyses of patients with untreated glioma," *Neurosurgery*, vol. 49, no. 4, pp. 823–829, 2001.

[56] M. C. Preul, R. Kasrai, R. Comeau, S. Narayanan, A. Olivier, R. Leblanc, and D. L. Arnold, "Using integrated multimodal imaging for neurosurgical guidance in the resection or biopsy of brain tumours," in *3rd International Conference on Functional Mapping of the Human Brain*, Academic Press, 1997.

[57] M. C. Preul, Z. Caramanos, R. Kasrai, R. Leblanc, and D. L. Arnold, *Advanced Neurosurgical Navigation*, ch. Proton Magnetic Resonance Spectroscopic Imaging in Neurosurgery: Applications for Brain Tumors. New York, USA: Thieme Medical Publishers, Inc., 1999.

[58] T. M. Peters, B. L. Davey, P. Munger, R. M. Comeau, A. C. Evans, and A. Olivier, "Three–dimensional multi–modal image–guidance for surgery," *IEEE Transactions on Medical Imaging*, vol. 15, pp. 121–128, 1996.

[59] J. Talairach, M. David, P. Tournoux, P. Corredor, and T. Kvasina, *Atlas d'anatomie stéréotaxique*. Paris: Masson, 1957.

[60] J. Talairach and P. Tournoux, *Co–planar stereotaxic atlas of the human brain.* Stuttgart: Georg Thieme Verlag, 1988.

[61] J. Talairach and P. Tournoux, *Referentially oriented cerebral MRI anatomy.* Stuttgart: Georg Thieme Verlag, 1993.

[62] G. Schaltenbrand and W. Wahren, *Introduction to stereotaxis with an atlas of the human brain.* Stuttgart: Georg Thieme Verlag, 1977.

[63] J. Favre, J. M. Taha, T. T. Nguyen, P. L. Gildenberg, and K. J. Burchiel, "Pallidotomy: A survey of current practice in North America," *Neurosurgery*, vol. 39, pp. 883–892, October 1999.

[64] W. Penfield and E. Boldrey, "Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation," *Brain*, 1937.

[65] P. A. Starr, J. L. Vitek, and R. A. E. Bakay, "Ablative surgery and deep brain stimulation for Parkinson's disease," *Neurosurgery*, vol. 43, no. 5, pp. 989–1015, 1998.

[66] G. A. Krombach, U. Spetzger, V. Rohde, and J. M. Gilsbach, "Intraoperative localization of functional regions in the sensorimotor cortex by neuronavigation and cortical mapping," *Computer Aided Surgery*, vol. 3, pp. 64–73, 1998.

[67] K. Finnis, P. St-Jean, R. Kasrai, D. Clonda, and T. Peters, "Creation of a non–linearly warpable 3–D functional atlas for image–guided neurosurgery," *Medical Physics (Proceedings of the 44th Annual Meeting of the Canadian Organisation of Medical Physicists)*, vol. 25, p. 1094, June 1998.

[68] K. W. Finnis, Y. P. Starreveld, A. G. Parrent, and T. M. Peters, "A 3–dimensional database of deep brain functional anatomy, and its application to image–guided

neurosurgery," in *Medical Image Computing and Computer Aided Intervention—MICCAI 2000* (S. L. Delp, A. M. DiGioia, and B. Jaramaz, eds.), vol. 1935 of *Lecture Notes in Computer Science*, pp. 1–8, Springer, October 2000.

[69] P. Jannin, O. J. Fleig, E. Seigneuret, C. Grova, X. Morandi, and J. M. Scarabin, "A data fusion environment for multimodal and multi-informational neuro–navigation," *Computer Aided Surgery*, vol. 5, no. 1, pp. 1–10, 2000.

[70] P. Jannin, X. Morandi, O. J. Fleig, E. L. Rumeur, P. Toulouse, B. Gibaud, and J. M. Scarabin, "Integration of sulcal and functional imaging for multimodal neuronavigation," *Journal of Neurosurgery*, submitted January 2001.

[71] A. Olivier, M. Alonso-Vanegas, R. M. Comeau, and T. M. Peters, "Image–guided surgery of epilepsy," *Neurosurgical Clinics of North America*, vol. 7, pp. 229–243, 1996.

[72] A. Pommert, A. Schubert, M. Riemer, T. Schiemann, U. Tiede, and K. H. Höhne, "Symbolic modeling of human anatomy for visualization and simulation," in *Visualization in Biomedical Computing 1994* (R. A. Robb, ed.), vol. 2359, (Bellingham, WA), pp. 412–423, SPIE, 1994.

[73] M. A. Viergever, J. B. Maintz, and R. Stokking, "Integration of functional and anatomical brain images," *Biophysical Chemistry*, vol. 68, no. 1–3, pp. 207–219, 1997.

[74] G. E. Kraus *et al.*, "A technique utilizing positron emission tomography and magnetic resonance/computed tomography image fusion to aid in surgical navigation and tumor volume determination," *Journal of Image Guided Surgery*, vol. 1, pp. 300–307, 1995.

[75] D. J. Hawkes, D. L. G. Hill, E. D. Lehmann, G. P. Robinson, M. N. Maisey, and A. C. F. Colchester, *3D Imaging in Medicine*, ch. Medical imaging: Analysis of multimodality 2D/3D images, pp. 242–251. Berlin, Germany: Springer–Verlag, 1990.

[76] D. MacDonald, D. Avis, and A. C. Evans, "Multiple surface identification and matching in magnetic resonance," in *Proceedings of Visualization in Biomedical Computing '94* (R. A. Robb, ed.), vol. 2359, (Bellingham, WA), pp. 160–169, SPIE, 1994.

[77] P. Suetens, E. Bellon, D. Vandermeulen, M. Smet, G. Marchal, J. Nuyts, and L. Mortelmans, "Image segmentation: methods and applications in diagnostic radiology and nuclear medicine," *European Journal of Radiology*, vol. 17, pp. 14–21, 1993.

[78] E. Keppel, "Approximating complex surfaces by triangulation of contour lines," *IBM Journal of Research and Development*, pp. 2–11, January 1975.

[79] S. Obrador, "A simplified neurosurgical technique for approaching and damaging the region of the globus pallidus in Parkinson's disease," *Journal of Neurology Neurosurgery and Psychiatry*, vol. 20, pp. 47–49, 1957.

[80] E. R. Cosman, B. S. Nashold, and P. Bedenbaugh, "Stereotactic radiofrequency lesion making," *Applied Neurophysiology*, vol. 46, pp. 160–166, 1983.

[81] P. L. Choyke, P. Yim, H. Marcos, V. B. Ho, R. Mullick, and R. M. Summers, "Hepatic MR angiography: a multiobserver comparison of visualization methods," *American Journal of Roentgenology*, vol. 176, pp. 465–470, 2001.

[82] R. A. Drebin, L. Carpenter, and P. Hanrahan, "Volume rendering," *Computer Graphics (Proceedings of SIGGRAPH)*, vol. 22, no. 3, pp. 65–74, 1988.

[83] M. Levoy, "Efficient ray tracing of volume data," in *ACM Transactions on Graphics*, vol. 9, pp. 245–261, 1990.

[84] P. Lacroute and M. Levoy, "Fast volume rendering using a shear–warp factorization of the viewing transformation," in *Proceedings of the 21st International SIGGRAPH Conference*, pp. 451–458, ACM, 1994.

[85] O. Sommer, A. Dietz, R. Westermann, and T. Ertl, "An interactive visualization and navigation tool for medical volume data," *Computers and Graphics*, vol. 23, pp. 233–244, April 1999.

[86] M. Levoy, "Display of surfaces from volume data," *IEEE Computer Graphics and Applications*, vol. 8, no. 3, pp. 29–37, 1988.

[87] U. Tiede, K. H. Hoehne, M. Bomans, A. Pommert, M. Riemer, and G. Wiebecke, "Investigation of medical 3D–rendering algorithms," *IEEE Computer Graphics and Applications*, vol. 10, pp. 41–53, March 1990.

[88] J. John W. Keyes, "Editorial: Three–dimensional display of SPECT images: Advantages and problems," *Journal of Nuclear Medicine*, vol. 31, pp. 1428–1430, August 1990.

[89] L. M. Auer and D. P. Auer, "Virtual endoscopy for planning and simulation of minimally invasive neurosurgery," *Neurosurgery*, vol. 43, pp. 529–548, 1998.

[90] N. L. Dorward *et al.*, "Interactive image–guided neuroendoscopy: development and early clinical experience," *Minimally Invasive Neurosurgury*, vol. 41, pp. 31–34, 1998.

[91] J. M. McGregor, "Enhancing neurosurgical endoscopy with the use of virtual reality headgear," *Minimally Invasive Neurosurgery*, vol. 40, pp. 47–49, 1997.

[92] J. Burtscher, A. Dessl, H. Maurer, M. Seiwald, and S. Felber, "Virtual neuroendoscopy, a comparative magnetic resonance and anatomical study," *Minimally Invasive Neurosurgery*, vol. 42, pp. 113–117, 1999.

[93] S. L. Aquino and D. J. Vining, "Virtual bronchoscopy," *Clinics in Chest Medicine*, vol. 20, pp. 725–730, December 1999.

[94] J. Rodenwaldt, L. Kopka, R. Roedel, A. Margas, and E. Grabbe, "3D virtual endoscopy of the upper airway: optimization of the scan parameters in a cadaver phantom and clinical assessment," *Journal of Computer Assisted Tomography*, vol. 21, pp. 405–411, 1997.

[95] M. P. Fried, V. M. Maharir, H. Shinmoto, A. M. Alyassin, W. E. Lorensen, L. Hsu, and R. Kikinis, "Virtual laryngoscopy," *Annals of Otology, Rhinology and Laryngology*, vol. 108, pp. 221–226, 1999.

[96] S. Gilani, A. M. Norbash, H. Ringl, G. D. Rubin, S. Napel, and D. J. Terris, "Virtual endoscopy of the paranasal sinuses using perspective volume rendered helical sinus computed tomography," *Laryngoscope*, vol. 107, pp. 25–29, January 1997.

[97] M. D. Seemann, O. Seemann, H. Bonél, M. Suckfüll, K.-H. Englmeier, A. Naumann, C. M. Allen, and M. F. Reiser, "Evaluation of the middle and inner ear structures: comparison of hybrid rendering, virtual endoscopy and axial 2D source images," *European Radiology*, vol. 9, pp. 1851–1858, 1999.

[98] E. Neri, D. Caramella, M. Panconi, S. Berrettini, S. S. Franceschini, F. Forli, and C. Bartolozzi, "Virtual endoscopy of the middle ear," *European Radiology*, vol. 11, pp. 41–49, 2001.

[99] R. Frankenthaler, V. Moharir, R. Kikinis, P. van Kipshagen, F. Jolesz, C. Umans, and M. P. Fried, "Virtual otoscopy," *Otolaryngologic Clinics of North America*, vol. 31, no. 2, pp. 383–392, 1998.

[100] M. Wan, Q. Tang, A. Kaufman, Z. Liang, and M. Wax, "Volume rendering based interactive navigation within the human colon," in *Proceedings of Visualization '99* (D. Ebert, M. Gross, and B. Hamann, eds.), pp. 397–400, IEEE, October 1999.

[101] R. Frank, A. Stenzl, T. Frede, W. Recheis, R. Knapp, G. Bartsch, and D. zur Nedden, "Three–dimensional computed tomography of the reconstructed lower urinary tract: technique and findings," *European Radiology*, vol. 8, pp. 657–663, 1998.

[102] F. A. Jolesz, W. E. Lorensen, H. Shinmoto, H. Atsumi, S. Nakajima, P. Kavanaugh, P. Saiviroonporn, S. E. Seltzer, S. G. Silverman, M. Phillips, and R. Kikinis, "Interactive virtual endoscopy," *American Journal of Roentgenology*, vol. 169, pp. 1229–1235, 1997.

[103] B. Preim, W. Spindler, K. J. Oldhafer, and H. O. Peitgen, "3D distance measurements in medical visualizations," in *Proceedings of Interactive Medical Image Visualization and Analysis, Satellite Workshop of the Fourth International Conference on Medical Image Computing and Computer–Assisted Intervention (MICCAI 2001)* (S. D. Olabarriaga, W. J. Niessen, and F. Gerritsen, eds.), pp. 31–36, 2001.

[104] A. M. Chiu, D. Dey, M. Drangova, W. D. Boyd, and T. M. Peters, "Port placement simulation in a virtual 3–D thoracic model," in *Proceedings of Minimally Invasive Cardiac Surgery Symposium 2000*, (Key West, CA), 2000.

[105] K. Kreeger and A. Kaufman, "Mixing translucent polygons with volumes," in *Proceedings of Visualization '99* (D. Ebert, M. Gross, and B. Hamann, eds.), pp. 191–198, IEEE, October 1999.

[106] A. Perneczky and G. Fries, "Endoscope–assisted brain surgery: Part 1—evolution, basic concept, and current technique," *Neurosurgery*, vol. 42, pp. 219–224, February 1998.

[107] G. Fries and A. Perneczky, "Endoscope–assisted brain surgery: Part 2—analysis of 380 procedures," *Neurosurgery*, vol. 42, pp. 226–231, February 1998.

[108] D. Dey, D. G. Gobbi, P. J. Slomka, K. J. M. Surry, and T. M. Peters, "Automatic fusion of freehand endoscopic brain images to three–dimensional surfaces: creating stereoscopic panoramas," *IEEE Transactions on Medical Imaging*, vol. 21, no. 1, pp. 23–30, 2002.

[109] P. Milgram and F. A. Kishino, "A taxonomy of mixed reality visual displays," *IEICE Transactions on Information and Systems*, vol. E77–D, pp. 1321–1329, December 1994.

[110] W. Robinett, "Synthetic experience: A proposed taxonomy," *Presence*, vol. 1, no. 2, pp. 229–247, 1992.

[111] D. Zeltzer, "Autonomy, interaction and presence," *Presence*, vol. 1, no. 1, pp. 127–132, 1992.

[112] P. L. Gleason, R. Kikinis, D. Altobelli, W. Wells, E. Alexander, P. M. Black, and F. Jolesz, "Video registration and virtual reality for nonlinkage stereotactic surgery," *Stereotactic Functional Neurosurgery*, vol. 63, pp. 139–143, 1994.

[113] P. J. Edwards, D. J. Hawkes, D. L. Hill, D. Jewell, R. Spink, A. Strong, and M. J. Gleeson, "Augmentation of reality using an operating microscope for otolaryngology and neurosurgical guidance," *Journal of Image Guided Surgery*, vol. 1, pp. 172–178, 1995.

[114] L. J. Pisani, R. M. Comeau, B. L. K. Davey, and T. M. Peters, "Incorporation of stereoscopic video into an image–guided neurosurgery environment," in *Proceedings of IEEE Engineering in Medicine and Biology Society, 17th Annual Conference*, vol. 1, pp. 365–366, IEEE, 1997.

[115] W. E. L. Grimson, R. Kikinis, F. A. Jolesz, and P. M. Black, "Image–guided surgery," *Scientific American*, vol. 200, pp. 63–69, June 1999.

[116] Y. Sato, M. Nakamoto, Y. Tamaki, T. Sasama, I. Sakita, Y. Nakajima, M. Monden, and S. Tamura, "Image guidance of breast cancer surgery using 3–D ultrasound images and augmented reality visualization," *IEEE Transactions in Medical Imaging*, vol. 17, pp. 681–693, 1998.

[117] D. Drascic and P. Milgram, "Perceptual issues in augmented reality," in *Stereoscopic Displays and Virtual Reality Systems III*, vol. 2653, (Bellingham, WA), pp. 123–134, SPIE, February 1996.

[118] R. A. Drebin, D. Magid, D. D. Robertson, and E. K. Fishman, "Fidelity of three–dimensional CT imaging for detecting fracture gaps," *Journal of Computer Assisted Tomography*, vol. 13, pp. 487–489, May/June 1989.

[119] H. Rusinek, M. E. Noz, J. G. Q. Maguire, A. Kalvin, B. Haddad, D. Dean, and C. Cutting, "Quantitative and qualitative comparison of volumetric and surface rendering techniques," *IEEE Transactions on Nuclear Science*, vol. 38, pp. 659–662, April 1991.

[120] R. F. Smith, B. K. Rutt, and D. W. Holdsworth, "Anthropomorphic carotid bifurcation phantom for MRI applications," *Journal of Magnetic Resonance Imaging*, vol. 10, pp. 533–544, 1999.

[121] R. Fahrig, H. Nikolov, A. J. Fox, and D. W. Holdsworth, "A three–dimensional cerebrovascular flow phantom," *Medical Physics*, vol. 26, no. 8, pp. 1589–1599, 1999.

[122] R. E. Drzymala and S. Mutic, "Stereotactic imaging quality assurance using an anthropomorphic phantom," *Computer Aided Surgery*, vol. 4, pp. 248–255, 1999.

[123] A. Pommert *et al.*, "Simulation studies for quality assurance of 3D images from computed tomograms," in *The formation, handling and evaluation of medical images* (T. Pokropek and M. A. Viergever, eds.), NATO ASI Series F, Berlin, West Germany: Springer–Verlag, 1990.

[124] E. A. Benardete, M. A. Leonard, and H. L. Weiner, "Comparison of frameless stereotactic systems: Accuracy, precision, and applications," *Neurosurgery*, vol. 49, no. 6, pp. 1409–1416, 2001.

[125] R. Rohling, P. Munger, J. M. Hollerbach, and T. M. Peters, "Comparison of relative accuracy between a mechanical and an optical position tracker for image–guided neurosurgery," *Journal of Image Guided Surgery*, vol. 1, pp. 30–34, 1995.

[126] R. Khadem, C. C. Yeh, M. Sadeghi-Tehrani, M. R. Bax, J. A. Johnson, J. N. Welch, E. P. Wilkinson, and R. Shahidi, "Comparative tracking error analysis of five different optical tracking systems," *Computer Aided Surgery*, vol. 5, pp. 98–107, 2000.

[127] T. S. Paleologos, J. P. Wadley, N. D. Kitchen, and D. G. T. Thomas, "Clinical utility and cost–effectiveness of interactive image–guided craniotomy: Clinical comparison between conventional and image–guided meningioma surgery," *Neurosurgery*, vol. 47, no. 1, pp. 40–48, 2000.

[128] T. S. Paleologos, N. L. Dorward, J. P. Wadley, and D. G. T. Thomas, "Clinical validation of true frameless stereotactic biopsy: Analysis of the first 125 consecutive cases," *Neurosurgery*, vol. 49, no. 4, pp. 830–837, 2001.

[129] W. W. Choi, B. A. Green, and A. D. O. Levi, "Computer–assisted fluoroscopic targeting system for pedicle screw insertion," *Neurosurgery*, vol. 47, no. 4, pp. 872–878, 2000.

[130] T. C. Hartrum, T. C. Mallary, and J. W. Foley, "Evaluating user satisfaction of an interactive computer program," in *Proceedings of the IEEE National Aerospace and Elecronics Conference—NAECON '89*, pp. 508–514, IEEE, May 1989.

[131] M. W. Vannier, "Evaluation of 3D imaging," *Critical Reviews in Diagnostic Imaging*, vol. 41, no. 5, pp. 315–378, 2000.

[132] C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, vol. 8, pp. 283–298, 1978.

[133] C. E. Metz, "ROC methodology in radiologic imaging," *Investigative Radiology*, vol. 21, pp. 720–733, 1986.

[134] C. E. Metz, "Some practical issues of experimental design and data analysis in radiological ROC studies," *Investigative Radiology*, vol. 24, pp. 234–245, 1989.

[135] P. Hans, A. J. Grant, R. D. Laitt, R. T. Ramsden, A. Kassner, and A. Jackson, "Comparison of three–dimensional visualization techniques for depicting the scala vestibuli and scala tympani of the cochlea by using high–resolution MR imaging," *American Journal of Neuroradiology*, vol. 20, pp. 1197–1206, 1999.

[136] R. Stokking, K. J. Zuiderveld, H. E. H. Pol, P. P. van Rijk, and M. A. Viergever, "Normal fusion for 3D integrated visualization of SPECT and MR brain images," *Journal of Nuclear Medicine*, vol. 38, no. 4, pp. 624–629, 1997.

[137] P. Charland and T. M. Peters, "Optimal display conditions for quantitative analysis of stereoscopic cerebral angiograms," *IEEE Transaction on Medical Imaging*, vol. 15, pp. 648–656, 1996.

[138] F. Tendick, S. Bhoyrul, and L. W. Way, "Comparison of laparoscopic imaging systems and conditions using a knot–tying task," *Computer Aided Surgery*, vol. 2, pp. 24–33, 1997.

[139] P. A. Smith, F. F. Marshall, B. A. Urban, D. G. Heath, and E. K. Fishman, "Three–dimensional CT stereoscopic visualization of renal masses: impact on diagnosis and patient treatment," *American Journal of Roentgenology*, vol. 169, pp. 1331–1334, 1997.

[140] R. A. Robb, S. Aharon, and B. M. Cameron, "Patient–specific anatomic models from three dimensional medical image data for clinical applications in surgery and endoscopy," *Journal of Digital Imaging*, vol. 10, pp. 31–35, 1997.

[141] P. Kay, R. Robb, R. Myers, and B. King, "Creation and validation of patient–specific anatomical models for prostate surgery planning using virtual reality," in *Visualization in Biomedical Computing, 4th International Conference* (K. Höhne and R. Kikinis, eds.), (Berlin, Germany), pp. 547–552, Springer–Verlag, 1996.

[142] K. H. Höhne and A. Pommert, *Brain Mapping: The Methods*, ch. Volume Visualization, pp. 423–443. Toronto, Canada: Academic Press, Inc., 1996.

[143] A. Kaufman, M. Brady, B. Lorensen, F. Kitson, and H. Pfister, "Why is real–time volume rendering no longer a year away?," in *Visualization '98*, pp. 497–499, IEEE, October 1998.

[144] P. Jannin, M. Raimbault, X. Morandi, and B. Gibaud, "Modeling surgical procedures for multimodal image–guided neurosurgery," in *Medical Image Computing and Computer Aided Intervention—MICCAI 2001* (W. Niessen and M. Viergever, eds.), vol. 2208 of *Lecture Notes in Computer Science*, pp. 565–572, Springer, 2001.

[145] H. L. Kundel, "Visual cues in the interpretation of medical images," *Journal of Clinical Neurophysiology*, vol. 7, pp. 472–483, 1990.

[146] R. Shahidi, L. Clarke, R. D. Bucholz, H. Fuchs, R. Kikinis, R. A. Robb, and M. W. Vannier, "White paper: Challenges and opportunities in computer–assisted interventions," *Computer Aided Surgery*, vol. 6, pp. 176–181, 2001.

[147] H. Visarius, J. Gong, C. Scheer, S. Haralamb, and L. P. Nolte, "Man–machine interfaces in computer assisted surgery," *Computer Aided Surgery*, vol. 2, pp. 102–107, 1997.

[148] T. McInerney and D. Terzopoulos, *Handbook of Medical Imaging: Processing and Analysis*, ch. Deformable Models, pp. 127–146. San Diego, CA: Academic Press, 2000.

[149] H. M. Ladak, F. Mao, Y. Wang, D. B. Downey, D. A. Steinman, and A. Fenster, "Prostate boundary segmentation from 2D ultrasound images," *Medical Physics*, vol. 27, no. 8, pp. 1777–1788, 2001.

[150] "MNI display—3D and volume program." `www.bic.mni.mcgill.ca/software`.

[151] "Atamai, Inc., London, ON, Canada." `www.atamai.com`.

[152] D. G. Gobbi and T. M. Peters, "Generalized 3D nonlinear transformations for medical imaging: An object–oriented implementation in VTK," *Computerized Medical Imaging and Graphics*, submitted.

[153] F. L. Bookstein, "Principal warps: thin–plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[154] A. Hartov, S. D. Eisner, D. W. Roberts, K. D. Paulsen, L. A. Platenik, and M. I. Miga, "Error analysis for a free–hand three–dimensional ultrasound system for neuronavigation," *Neurosurgical Focus*, vol. 6, no. 3, 1999.

[155] J. Robert L. Galloway, R. J. Maciunas, and A. L. Failinger, "Factors affecting perceived tumor volumes in magnetic resonance imaging," *Annals of Biomedical Engineering*, vol. 21, no. 4, pp. 367–375, 1993.

[156] V. Chalana and Y. Kim, "A methodology for evaluation of boundary detection algorithms on medical images," *IEEE Transactions on Medical Imaging*, vol. 16, no. 5, pp. 642–652, 1997.

[157] J. D. Gill, H. M. Ladak, D. A. Steinman, and A. Fenster, "Accuracy and variability assessment of a semiautomatic technique for segmentation of the carotid arteries from three–dimensional ultrasound images," *Medical Physics*, vol. 27, no. 6, pp. 1333–1342, 2000.

[158] A. L. Gilchrist, "Perceived lightness depends on perceived spatial arrangement," *Science*, vol. 195, pp. 185–187, 1977.

[159] S. C. Masin, "Transparent surfaces and illuminated holes," *Perception*, vol. 24, pp. 761–770, 1995.

[160] W. Gerbino, C. I. Stultiens, J. M. Troost, and C. M. M. de Weert, "Transparent layer constancy," *Journal of Experimental Psychology—Human Perception and Performance*, vol. 16, pp. 3–20, 1990.

[161] W. Gerbino, "Colour constraints and optimal transparency," *Perception (suppl.)*, vol. 22, p. 2, 1993.

[162] N. A. Macmillan and C. D. Creelman, *Detection theory: A user's guide*. New York, NY: Cambridge University Press, 1991.

[163] J. P. Guilford, *Psychometric Methods*. New York, NY: McGraw–Hill, 1954.

[164] J. Beck and R. Ivry, "On the role of figural organization in perceptual transparency," *Perception & Psychophysics*, vol. 44, pp. 585–594, 1988.

[165] M. Singh and D. D. Hoffman, "Part boundaries alter the perception of transparency," *Psychological Science*, vol. 9, no. 5, pp. 370–378, 1998.

[166] D. G. Luenberger, *Linear and nonlinear programming*. Reading, MA: Addison–Wesley, 1984.

[167] F. Metelli, O. D. Pos, and A. Cavedon, "Balanced and unbalanced, complete and partial transparency," *Perception & Psychophysics*, vol. 38, pp. 354–366, 1985.

[168] J. Beck, "Additive and subtractive color mixture in color transparency," *Perception & Psychophysics*, vol. 23, pp. 265–267, 1978.

[169] T. Watanabe and P. Cavanagh, "Transparent surfaces defined by implicit X–junctions," *Vision Research*, vol. 33, pp. 2339–2346, 1993.

[170] M. D'Zmura, P. Colantoni, K. Knoblauch, and B. Laget, "Color transparency," *Perception*, vol. 26, pp. 471–492, 1997.

[171] S. Westland and C. Ripamonti, "Invariant cone–excitation ratios may predict transparency," *Journal of the Optical Society of America (A)*, vol. 17, pp. 255–264, 2000.

[172] S. C. Masin, "A weighted–average model of achromatic transparency," *Perception & Psychophysics*, vol. 49, pp. 563–571, 1991.

[173] S. C. Masin, "The luminance conditions of transparency," *Perception*, vol. 26, pp. 39–50, 1997.

[174] M. Singh and B. L. Anderson, "Toward a perceptual theory of transparency," *Investigative Ophthalmology & Visual Science*, vol. 41, no. 4, p. S218, 2000.

[175] R. Kasrai and F. A. A. Kingdom, "The precision, accuracy, and range of perceived achromatic transparency," *Journal of the Optical Society of America (A)*, vol. 18, no. 1, pp. 1–11, 2001.

[176] D. Todorović, "Lightness and junctions," *Perception*, vol. 26, pp. 379–394, 1997.

[177] E. H. Adelson and P. Anandan, "Ordinal characteristics of transparency," in *AAAI–90 Workshop on Qualitative Vision*, (Boston, MA), pp. 77–81, 1990.

[178] E. H. Adelson, "Perceptual organization and the judgment of brightness," *Science*, vol. 262, pp. 2042–2044, 1993.

[179] R. Kasrai and F. A. A. Kingdom, "Are X-junctions necessary for the accurate perception of transparency?," *Investigative Ophthalmology and Visual Science (suppl.)*, vol. 41, no. 4, p. S218, 2000.

[180] R. Kasrai, F. A. A. Kingdom, and T. M. Peters, "The dichoptic perception of achromatic transparency," in *Perception*, vol. 28 (suppl.), (Trieste, Italy), 22nd European Conference on Visual Perception, August 1999.

[181] R. Kasrai, F. A. A. Kingdom, and T. M. Peters, "The perception of transparency in medical images," in *Medical Image Computing and Computer–Assisted Intervention—MICCAI'99* (C. Taylor and A. Colchester, eds.), no. 1679 in Lecture Notes in Computer Science, (Cambridge, UK), pp. 726–733, Springer-Verlag, September 1999.

[182] S. C. Masin, "The time it takes to stratify two phenomenal surfaces," *Perception*, vol. 27, pp. 1417–1422, 1998.

[183] L. J. Bain and M. Engelhardt, *Introduction to Probability and Mathematics Statistics*. Boston, MA: PWS–KENT Publishing Co., second ed., 1991.

[184] E. H. Adelson and D. Somers, "Straightness, structure, and shadows," *Journal of Vision*, vol. 1, no. 3, p. 204a, 2001.

[185] F. Kingdom and R. Kasrai, "Discriminating material from illumination changes in complex scenes," in *54th Annual Spring Conference of the Society for Imaging Science and Technology*, (Montréal, Canada), pp. 5–8, Society for Imaging Science and Technology, April 2001.