VISUAL PERCEPTION

Theory and Practice

Terry Caelli

PERGAMON INTERNATIONAL LIBRARY

Visual Perception Theory and Practice

by

TERRY CAELLI

The University of Newcastle,
New South Wales, Australia

PERGAMON PRESS

OXFORD · NEW YORK · TORONTO · SYDNEY · PARIS · FRANKFURT

U.K.	Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 OBW, England
U.S.A.	Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.
CANADA	**Pergamon Press Canada Ltd., Suite 104, 150 Consumers Rd., Willowdale, Ontario M2J 1P9, Canada**
AUSTRALIA	Pergamon Press (Aust.) Pty. Ltd., P.O. Box 544, Potts Point, N.S.W. 2011, Australia
FRANCE	Pergamon Press SARL, 24 rue des Ecoles, 75240 Paris, Cedex 05, France
FEDERAL REPUBLIC OF GERMANY	Pergamon Press GmbH, 6242 Kronberg/Taunus, Hammerweg 6, Federal Republic of Germany

First edition 1981

British Library Cataloguing in Publication Data

Caelli, Terry
An introduction to modern approaches in visual perception.—(Pergamon international library).
1. Visual perception—Physiological aspects
I. Title
612′.84 QP475 80-40167

ISBN 0 08 024420 3 Hardcover
ISBN 0 08 024419 X Flexicover

Printed and bound in Great Britain by
William Clowes (Beccles) Limited, Beccles and London

PERGAMON INTERNATIONAL LIBRARY
of Science, Technology, Engineering and Social Studies
The 1000-volume original paperback library in aid of education industrial training and the enjoyment of leisure
Publisher: Robert Maxwell, M.C.

Visual Perception
Theory and Practice

Other Titles of Interest

NAWRATZKI
Impaired Vision in Childhood

BUNGE
The Mind–Body Problem

To four lovely girls:

Kate

Luisa

Emma

Lia

Preface

THE aim of this book is to familiarize the reader with recent technologies being used in vision research, and to review current findings and theories of visual information processing which use such technologies. For this reason the book has two parts: Part I—which briefly exposes the technologies as such; and, Part II—a review of more central areas of vision research including spatial vision, motion perception, and colour.

One of the greatest problems with working in a multi-discipline area, as vision research is now, is that of communication—in particular the understanding of various languages which are drawn from one discipline and applied to a common issue. We have all had the experience of wanting to understand a particular process but have just not become familiarized enough with the technology to adequately comprehend the process. In most cases we have so little time that it is not possible to start from our first-year undergraduate days and work upwards. Often, when we nobly decide to do just this, we end up flipping through some esoteric book precisely on the subject—in search of some easy-to-read statement about the specific issue that drove us to such a reference.

This is particularly true in areas where both humanities and science graduates are involved—like vision. In one way the scientist has little introduction to the historical and philosophical traditions behind the subject, while the humanities researcher just does not understand the intricacies and assumptions underlying specific procedures. This latter position generates ritual replications of past paradigms—at best justified by the argument to authority. The former error can generate meaningless technical exercises, which have little to do with solving more central problems in vision.

This book is aimed at overcoming the technical problem. It will not satisfy the dedicated mathematician, engineer, etc., and I hope it is not presuming too much for the not so technically minded.

Acknowledgements

THIS book is the product of research in visual perception over the past 5 years during which time the author has realized just how important objective definition and communality in technological skills are to the unification of research areas. It is simply not sufficient to find colleagues conducting research in an area where one employs a language completely not understood by another. The point is that many excellent ideas and models exist in the literature which are just not understood by many other researchers in the same area. This book is aimed at helping to solve this problem.

I am greatly indebted to all my fellow researchers in visual perception for helping me in my striving to think better thoughts in this most fascinating subject and, simply, for their creative contribution to vision research. In particular I thank Peter Dodwell, Bela Julesz, and William Hoffman for sharing their various thoughts with me over the years, and John Keats and David Finlay for that academic company most essential for all.

Contents

CHAPTER 1

Introduction: Languages, Processes, and Perception

IF WE consult any one of a variety of journals specializing in vision research we shall find that the area is most diverse covering issues from retinal biochemistry to visual memory and our perceptions of depth. In addition, we should find on closer examination that there are many different approaches to precisely the same problem—particularly in visual perception. Although the development of microelectrode recordings has revolutionized some aspects of vision research over the past 20 years, even such "objective" measures of cortical encoding of visual images have not unified vast areas of visual perception research.

Perhaps one of the greatest gaps in communication in the area of visual perception lies between those who use engineering/mathematical languages to describe events and processes—and those who do not. Those who do not use such technologies argue that the visual system could not enact such processes and that qualitative descriptions are more realistic. Those who use these methodologies argue that descriptions, which are not quantitative or explanations which do not involve an algorithm, are simply not useful or testable.

These issues raise the more fundamental problem as to the role of analogy in explanation and the criteria for metalanguages in a given system. I use "metalanguage" not in the sense of a predicate calculus but rather in the sense that many different explanations or models in visual perception make assumptions about things like the description of the stimulus, processing, and response relations, which all involve specific languages. For example, to describe a stimulus in terms of its amplitude and phase spectra generates different "explanations" than describing it (simply) geometrically. Some argue that this problem is not solvable but, rather, that a language is as powerful as the hypotheses and models or explanations it can generate. We shall see a clear example of this issue in Chapter 6 when dealing with current approaches to textures, contour extraction, and visual illusions.

It is also clear that our criteria for explanation are changing—possibly due to the influx of scientists from other disciplines. I remember speaking to a colleague concerning an experiment where the absolute threshold was found to increase over time for a particular visual stimulus. He confidently responded that the *explanation* for this was "adaptation". On further questioning I discovered that adaptation was defined in two ways: (a) by an increase in threshold, and (b) some vague notion of nerve cells getting "fatigued". Such "explanations" are becoming less popular.

There have been attempts to define generative languages to represent perceptual events—in accord with reasonable principles of neurophysiological function. The early "perceptrons" language of Minsky and Pappert (1968) is an excellent example. The problem with these languages has always been that they lack in application to the problems encountered in perceiving complex and more natural events and scenes. We shall deal with these various languages in Part II of this volume and, for the present, I shall deal with the "six-questions" of modern visual perception in the hope of illustrating the issues of language and processes underlying these specific areas of vision. These questions (I believe) are something like:

Q 1: To what extent are receptive field and response properties of individual cells determined by the complex neural connections in a specific area?
Q 2: What are the appropriate measures of neural activity?
Q 3: What features of an image are of specific interest to the visual system?
Q 4: What language(s) best describes these features and their detection processes?
Q 5: What relationships can be expected between individual cell activity and human psychophysical responses?
Q 6: What assumptions are necessary and sufficient in (5) to construct interpretable relationships?

I shall now deal with each question in some detail to illustrate the points made above, i.e. our implicit assumptions in research determine the experimental and theoretical outcomes as much as the explicit formulations.

Q. 1: *Single cells or networks.* One important aspect of vision research over the past 50 years is that the technology of electrophysiology has developed to such a stage that it is relatively simple to record electrical activity from individual cells along the visual pathways. From the discovery of lateral inhibition in *Limulus* (Hartline, 1949) to the feature extractors found in the frog's visual system (Lettvin *et al.*, 1959) and cortical feature extractors of Hubel and Wiesel (1962, 1968), the evidence clearly supports a correspondence between cell activity in specific cortical areas and stimulation of the visual field(s).

However, on further examination the interpretation problems become immense. Consider the initial conclusions of Hubel and Wiesel (even in 1968). They reported that cortical receptive fields, being driven by a contiguous collection of retinal ganglion cell fields, responded selectively to orientations of bars or slits of light. These "simple" cells were argued to lie in the same column structure of the cortex when close together in orientation selectivity. Hubel and Wiesel continued, of course, to postulate the complex and hypercomplex units, which, in turn, were driven by collections of simple cells as shown in Fig. 1.1.

Yet a series of recent experiments indicate that these results are, more properly, simple examples of a more general process in neurophysiological function. Results from intra-

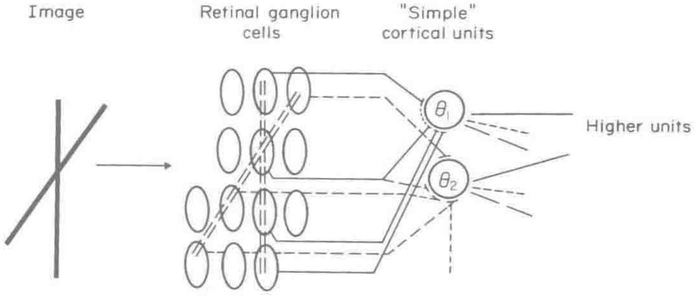

FIG. 1.1. The orientation selective receptive fields in cat's visual cortex discovered by Hubel and Wiesel (1962, 1968).

cellular recordings by Creutzfeldt *et al.* (1974) and others indicate that, with directional sensitivity, the cells' responses are not solely determined by the spatial receptive field arrangement proposed earlier by Hubel and Wiesel. Rather, such results indicate that response is a function of intracortical inhibitory connections spreading over large areas of the visual fields of the cat.

In this way receptive fields of individual cells are proposed to be determined by the complex dendritic arborization processes of cortical cells, which, to some extent, have been theoretically examined by Leake and Anninos (1976) and anatomically studied by Valverde (1976). Recently, it has been proposed that the receptive field profiles of individual cortical cells represent the image profiles of two-dimensional spatial frequency filters of the image (see Spekreijse and van der Tweel, 1978: summary). As we shall see, although this approach has some advantages there are still problems and we shall deal with the issues in Chapters 6 and 7. At this stage it is sufficient to note this is a central problem of understanding how the visual system processes image information in the cortical areas: is the system fixed or dynamic?

Q.2: *Measures of neural activity*. Most microelectrode recordings are concerned with the measurement of spikes or the number of firings of the cell per second (pulses/second). However, it is not clear that this is a fundamental parameter of neural activity—certainly it is not clear that all information is transmitted via action potentials. For example, the firing probability is a function of the average potential and is not the same as the neuron's threshold function which depends on the membrane potential. That is, the firing probability is already a function of the average potential over the connecting cells (Sejnowski, 1976). Stein (1967) and Cowan (1971) have modelled the average firing rates by stochastic processes. Sejnowski (1976) summarizes these results by

$$\phi_a = N_a + \sum_b C_{ab} r_b,$$

where the average potential for cell a (ϕ_a) is determined by N_a, the average external input; r_b, the average firing rate of cell b; and C_{ab}, the neural connection matrix, where C_{aa} represents recurrent collaterals. The quantitative shape of C_{ab} is of fundamental importance to visual function and we shall continue with evidence for one form or another in Chapters 6 and 7. However, before these issues (in Q. 1 and Q. 2) are dealt with in more detail, the reader requires the technological background to filter theory and non-linear networks (Chapters 3 and 4).

Q. 3,4: *Feature specificity and language problems*. These constitute the central problems in vision research today—maybe for all times. When research is conducted in visual perception the experimenter constructs, tests hypotheses, and measures responses from perspectives, which are considered valid in the specific area. However, what is viable from one perspective is not from another. For example, if the experimenter commences a project on cortical feature selectivity by assuming that the visual field is Euclidean in nature, having a coordinate system (etc.), then specific stimuli as lines, edges, angles, and motions will be considered as fundamental—a reasonable conclusion. Yet an optical or electrical engineer, in regarding the visual processes as a system, would probably regard sine-wave gratings or simple waveforms as appropriate input stimuli—also reasonable.

However, results so far indicate that from individual cortical cells to gross psychophysical responses, the visual system responds selectively to all these various types of image parameters. We may well ask whether these results imply that the visual system functions

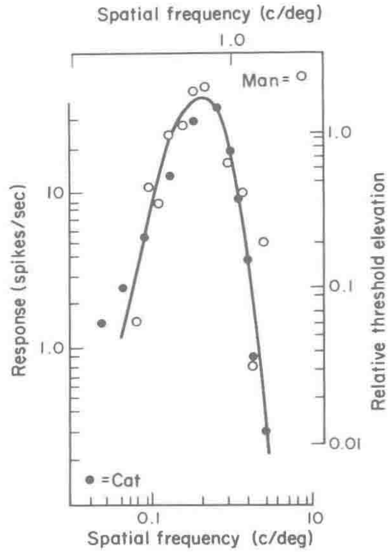

FIG. 1.2. Similarity between electrophysiological recordings from cat's cortical units (spikes/second) and human psychophysical responses (contrast sensitivity) for spatial frequency components of gratings. (From Campbell, 1976.)

on a coding strategy of which the above two code types must be examples. At this stage we have no answer to this equation—either in terms of neurophysiology or psychophysics—in cats or man!

As we shall see in Chapters 6 and 7, coding equivalences have to be established between various languages when a specific issue is being investigated. For example, what is the equivalent to orientation selectivity in the Fourier transform language? It is possible, by doing such, to arrive at what are epistemological criteria for one language as being more powerful than another.

Q. 5, 6: *Individual cells and psychophysics.* One of the amazing findings of vision research is the agreement between individual cell recording results and gross psychophysical responses on identical stimulus material. For example, the human visual contrast sensitivity function is very similar to frequency selectivity responses of individual cells in the cat's visual cortex (Fig. 1.2). Similar equivalences occur between orientation selective curves from visual cortex cells and orientation specific masking effects as measured with humans.

Why do these similarities exist? Various arguments have been posed to answer the question from various rubrics: probability summation, linear systems, etc. However, what still seems amazing to the author is that the great complexities of visual cortex, decision-making areas, etc., seem to be suppressed in such a way that an individual cell can reflect the total activity of the human brain in making a decision. Pribram (Pribram *et al.*, 1974) developed a holographic theory of visual function based on this type of observation. This type of process runs counter to other highly interactive processes in known neurophysiological structures, e.g., inhibition.

This is not to deny that in many instances such a simple connection could exist. We (Caelli and Julesz, 1979) have recently discovered one clear indication that decisions concerning texture discrimination can be made on such an additive assumption. As will be reported in Chapter 6, we have found that texture discrimination can be predicted from the addition of dipole orientation statistics for each texture—discrimination being based on the amplitude difference between these distributions.

Perhaps the more common justification of the association between the "local" individual cell response and gross psychophysical represses is "probability summation". This simply states that the probability of perceiving a difference, or detecting a signal, is the (Euclidean) sum of the probabilities of detecting each signal component. This, as well as other assumptions, will be discussed in Chapters 6 and 7.

It is clear that such questions just cannot be answered without some understanding of the technologies involved, which brings us back to the aim of this book. In visual perception the choice of even the appropriate language to use in which to embed issues is not clear. With this in mind we now proceed Part I. the methodology.

REFERENCES

CAELLI, T. and JULESZ, B. (1979) Psychophysical evidence for global feature processing in visual texture discrimination, *J. Opt, Soc. Am.* **69**, (5) 675–678.

CAMPBELL, F.W. (1976) The transmission of spatial information through the visual system, *Scient. Am.* 95–193.

COWEN, J. (1971) Stochastic models of neuroelectric activity. In Rice, S., Light, J., and Freed, K. (eds.), *Proceedings of the International Union of Pure and Applied Physics Conference on Statistical Mechanics*, pp. 109–127, University of Chicago Press, Chicago.

CREUTZFELDT, O., KHUNT, U., and BENEVENTO, L. (1974) An intracellular analysis of visual cortical neurones to moving stimuli: responses in a co-operative neuronal network, *Expl Brain Res.* **21**, 251–274.

HARTLINE, H. (1949) Inhibition of activity of visual receptors by illuminating nearby retinal areas in the Lumulus eye, *Fed. Proc.* **3**, 69.

HUBEL, D. and WIESEL, T. (1962) Recaptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J. Physiol.* **160**, 106–154.

HUBEL, D. and WIESEL, T. (1968) Receptive fields and functional architecture of monkey striate cortex, *J. Physiol.* **195**, 215–243.

LEAKE, B. and ANNINOS, P. (1976) Effect of connectivity on the activity of neuronal net models, *J. Theor. Biol.* **58**, 337–363.

LETTVIN, J., MATURANA, H., MCCOLLOCH, W., and PITTS, W. (1959) What the frog's eye tells the frog's brain, *Proc. Inst. Radio Engrs* **47**, 1940–51.

MINSKY, M. and PAPERT, S. (1968) *Perceptrons*, MIT Press, Cambridge, Mass.

PRIBRAM, K., NUWER, M., and BARÓN, R. (1974) The holographic hypothesis of memory structure in brain function and perception. In *Contempory Developments in Mathematical Psychology* (Krantz, D., Atkinson, R., Luce, R., and Suppes, P., eds.), Freeman, San Fransisco.

SEJNOWSKI, T. (1976) On global properties of neuronal interactions. *Biol. Cybernetics*, **22**, 85–95.

SPEKREIJSE, H. and VAN DER TWEEL, L. (1978) *Spatial Contrast*, Akademie van Wetenschappen, Koninklijke Nederlandse.

STEIN, R. (1967) The frequency of nerve action potentials generated by applied currents, *Proc. Roy. Soc. B*, **167**, 64–86.

VALVERDE, F. (1976) Aspects of cortical organization related to the geometry of neurons with intra-cortical axons, *J. Neuroncytol*, **5**, 509–529.

PART I

Technology Relevant to Visual Perception