Summary

Over the past year a special eye-movement monitoring and visual display system designed to further research on human visual information processing has been partially constructed and is nearing completion. The system can accurately measure eye-movements of less than 1' of arc and with the same precision control the movement of computer-generated visual targets relative to an observer's retina. A section of this report describes the nine subsystems of the visual apparatus and indicates their potential for experimentation. The theoretical reconstruction of the response profile to a visual target is complicated by the fact that under ordinary viewing conditions even during fixation the population of active elements in the visual pathway changes constantly. Using the special visual apparatus, it is possible to maintain a target on a fixed set of retinal receptors facilitating an orderly experimental analysis of visual response patterns. To direct this analysis toward predictive models, theoretical work has been focused on control sites for specific visual functions. At present models of visual transduction in human rods and cones have been devised and several properties numerically evaluated by computer, including temporal modulation transfer functions. Work is currently in progress on a model of visual acuity based upon the spatial modulation transfer function of neurons in the primary visual cortex.
Table of Contents

Summary ................................................. 1

I. Introduction .......................................... 5

II. Equipment ............................................ 7
   Eyetracker and Image Stabilization System

III. Models .............................................. 11
   1. Visual Transduction
   2. Visual Acuity

IV. Projections .......................................... 19

V. Bibliography ......................................... 21
I. Introduction

The human visual system is an action-oriented, information-sensing system whose active elements, although grouped into redundant subunits, consist for the most part of several billion interconnecting neurons. The problem of understanding a system of such staggering complexity has been ameliorated by recent scientific advances in the fields of neurobiology and psychophysics and by new technological developments in computation. With present neurobiological techniques it is now possible to determine in precise detail the neural mechanisms subserving perceptual processes and contingent behavioral action (Kandel, 1976). Typically, neurobiological experiments are carried out in non-human animals, but within reasonable limitations the results are applicable to man. In psychophysics, sophisticated methods of linear and quasi-linear analysis have been applied with conspicuous success to detection models for temporal and spatial parameters of visual targets (Kelly, 1972; Robson, 1966; Blakemore and Campbell, 1969). The availability of high-speed computers makes possible the implementation of mathematical models of visual processes too complex for standard analytic solutions. The project takes advantage of all three of these advances.

The approach is to generate mathematical models of visual processes consistent with neurophysiological results, than use psychophysical experiments to estimate parameters and validate the models. Previous quantitative models of human visual processes have been either simple summaries of empirical data of narrow scope with limited predictive value, or theoretical schemas neither in computable form nor realizable in terms
of known or plausible neurophysiological mechanisms (see Bibliography).
The novelty of our experimental procedure is to determine the performance characteristics of human vision under conditions of image stabilization.
Under conditions of ordinary viewing, even during fixation of a stationary target, the retinal stimulus is never constant due to eye-movements.
By initially removing the effects of eye-movements the problem of analyzing and synthesizing the response patterns in the visual pathway is considerably simplified. The system for image stabilization is described in the next section.
II. Equipment

Eyetracker and Image Stabilization System

In order to deconvolve experimentally the contributions of eye-movements from visual processing, we have designed and constructed in part a special optical system. By accurately measuring an observer's eye-movements over a target field, then optically compensating for these movements, the apparatus can effectively stabilize any target on the observer's retina. By the same feedback arrangement, it is possible to superimpose any arbitrary motion of a visual target relative to the observer's retina.

More specifically, the special optical system consists of subsystems as follows:

1. **Eye-Movement Monitor**: A modified version of the Stanford Research Institute Dual Purkinje Eyetracker (Cornsweet and Crane, 1973; Clark, 1975) measures monocular eye position to an accuracy of 1' of arc with a bandwidth of 150 Hz.

2. **Optical Deflector**: A custom-designed, high-speed dual optical scanner moves an image plane about the center of rotation of an observer's eye with a repeatability of 1' arc. When appropriately linked to the eye-movement monitor, the total lag time in following eye-movements is at most 1.5 msec, effectively stabilizing the visual target relative to the observer's retina.
3. **Image Stabilization Monitor**: From the difference in output signals between the eye-movement monitor and the optical deflector, a custom-modified video generator superimposes a marker indicating the observer's point of regard on a video camera image of the target field.

4. **Focus Stimulator**: A specially-designed lens produces changes in optical focus without changing target size or brightness so that the system has the capability of moving targets in three dimensions.

5. **Visual Display**: Visual targets can be presented by means of three display devices: a CRT, a Maxwellian view stimulator, and a Video monitor.

   (a) The CRT display can present two multiplexed channels of luminous bars or sinusoidally-varying luminance gratings whose extent, position, and contrast can be changed at rates up to 1000 Hz.

   (b) The Maxwellian optical system can provide up to three channels of controlled target or background illumination of varied spectral composition over a wide range of intensities up to $10^7$ for white light.

   (c) The Video display can present patterns of 512 by 512 elements with a 6-bit gray scale.
6. **Data Acquisition and Control Interface:** For each of the above devices, computer-control and data sampling are available in flexible form through an interconnection panel. The interface contains 16 TTL input lines, 16 TTL output lines, 16 multiplexed 12-bit A/D input lines, 8 12-bit D/A output lines, and a DMA channel for the video generator.

7. **Laboratory Computer:** An Eclipse S/200 with 32 K words and 10 Mbyte disc is available to generate visual displays, control experiments, collect and analyze experimental data.
Figure 1. Theoretical impulse response functions for a human rod photoreceptor. For each energy level, given in relative units, the time-course of membrane hyperpolarization has been calculated from a mathematical model of visual transduction. With increasing pulse energy, rise-time decreases and peak amplitude saturates.
III. Models

The first year of the project has been devoted mainly to the generation of computer software and the construction of apparatus; however, some preliminary models have been devised for later experimental testing. Our strategy has been to focus attention on critical control points in visual processing. For example, extensive neurophysiological evidence supports the view that the only essential nonlinearity associated with intensity coding in vision and many other senses is confined to the periphery (Mountcastle, 1967; Stevens, 1970). Modelling the nonlinear transformation in the periphery sets the stage for the application of linear models for central processing.

1. Visual Transduction (with J. Daugman)

Electrical recordings from vertebrate retinas over the past decade have thoroughly discredited the photocell theory of visual transduction. It is now clear that vertebrate photoreceptors are active elements with complex response properties that to a considerable extent quantitatively account for the temporal and intensive discrimination capacity of the intact visual system (e.g., Boynton and Whitten, 1970; Kelly, Boynton, and Baron, 1976). Individual primate rod and cone photoreceptors have not been recorded for the prolonged periods necessary to obtain extensive quantitative results, but the initial findings suggest a close qualitative similarity to other vertebrate photoreceptors (DeMonasterio and Gouras, personal communication). So, in devising mathematical models for human rod and cone photoreceptors we have
been guided by the work of others on photoreceptors in non-mammalian vertebrates (Baylor, Hodgkin, and Lamb, 1974; Cervetto, Pisano, and Torres, 1977), taking into account the particular dimensions of human photoreceptors and the higher operating temperature (37.5°C).

Figure 1 shows the calculated response for a human rod photoreceptor to a series of light impulses of varying energy content. Two types of nonlinearity are evident in the response. First, a saturating nonlinearity for, as the energy is increased in equal ratio steps, the peak amplitude increases linearly at low levels, but approaches an asymptotic value at high levels. To a first approximation, the amplitude of the peak, V, can be described as a function of flash energy, E, by the following equation:

\[
\frac{V}{V_{\text{max}}} = \frac{E}{E + s_1}
\]

where \( V_{\text{max}} \) is the maximum attainable response, and \( s_1 \) is a semi-saturation constant. Equation (1) has been found to describe the amplitude of response to brief flashes in several species of vertebrates (Mansfield, 1976). The second type of nonlinearity found is a time-scale nonlinearity. As the energy increases, the rise-time remains constant at low levels, then decreases to an asymptotic value at high levels. To a first approximation, the rise-time of the peak, t, can be described as a function of flash energy, E, by the following equation:
Figure 2. Theoretical transfer functions for a human foveal cone. Amplitude sensitivity for three background levels of quantum flux is plotted as a function of temporal frequency of modulation based upon calculations from a model of visual transduction. Rather than being of constant shape as in a linear system, the shape of the transfer function is level-dependent.
\[
\frac{t}{t_{\text{min}}} = \left[ \frac{1^{1/3}}{1^{1/3} + s_2} \right]^{-1}
\]  

(2)

where \( t_{\text{min}} \) is the minimum attainable rise-time and \( s_2 \) is a constant.

Equation (2) has been found to describe the latency of response to brief flashes in a number of vertebrate species (Mansfield, 1973; Mansfield and Daugman, 1977). Despite these two nonlinearities found in the response characteristics of the model for human photoreceptors and those measured electrophysiologically in other vertebrates, a steady-state condition of constant illumination produces a linearization of photoreceptor response. Using the resulting small signal linearity prevailing at each level of light intensity, transfer functions can be calculated.

Figure 2 shows the amplitude component of transfer functions calculated for a human foveal cone for three light intensity levels. For clarity of presentation and for comparison with human psychophysical data for flicker thresholds, the functions are plotted using an ordinate in units of sensitivity (nanovolts/photon/cone-second). The shape of the family of curves resembles that obtained by Kelly (1971) for counterphase gratings which eliminate the familiar low-frequency attenuation by reducing lateral interactions in the proximal portions of the retina beyond the photoreceptors. Such a result is encouraging for it suggests that by obtaining the product of the transfer function for the appropriate portion of the proximal inhibitory network and the transfer function for the
appropriate type of photoreceptor, it will be possible to predict the form of the human detection threshold curve for a wide variety of spatio-temporal patterns.

2. Visual Acuity (with S. Ronner and G. Legge)

The threshold contrast for detecting a visual target with a particular spatial configuration is influenced by a number of factors, but under conditions where the nonlinear transformations in the retina can be set aside so that the response of the visual system can be considered linear, the cortical mosaic of neurons with spatially complex receptive fields functions as a major control site. At the level of the receptive field structure of striate cortical neurons, the spatial inhomogeneity and anisotropy of primate vision become clearly evident (Hubel and Wiesel, 1974; Mansfield, 1974). Fortunately, the inhomogeneity and anisotropy are orderly in nature. The average size of receptive fields is small in the foveal projection region of primary visual cortex, but increases approximately as the distance from the fovea. In addition, receptive fields preferentially sensitive to horizontal and vertical targets predominate in the foveal projection region, but not in the periphery. Were it not for the inhomogeneity and anisotropy a predictive model could be constructed from a single linear spatial filter using the Fourier transform of the unique point spread function.

As an instance of the detection problem we have considered the case of oriented line segments. To construct an appropriate spatial filter that is consistent with neurophysiological data,
Figure 3. Orientation selectivity in primate vision. The solid squares represent the threshold elevation in human observers following adaptation to a high contrast sine wave grating. The open circles represent the integrated impulse discharge elicited by a luminous narrow bar in a monkey visual neuron recorded in the foveal projection region of Area 17. The solid line fitted to the neural data represents the orientation response profile calculated for a theoretical spatial filter.
subunits, each of which has an identical spatial impulse response function of the form

\[ P(r) = e^{-r^2} - 0.04 e^{-r^2/25}, \]  

(3)

are combined in arrays and their outputs summed linearly. The response profile of such subunits resembles that found for sustained on-center off-surround mammalian visual neurons at the level of the retinal ganglion cells or lateral geniculate. For example, in the cat lateral geniculate the radius of the surround area is typically five times larger than the radius of the center area (Enroth-Cugell and Robson, 1966). When a line segment function is convoluted with the spatial filter at different relative orientations, an orientation response function is obtained such as that shown in Figure 3.

Figure 3 shows an orientation response function based upon nine subunits and possessing a half-maximal bandwidth of 48°. For purposes of comparison, two sets of data are shown: orientation sensitivity measures for a human observer derived psychophysically for a sine-wave grating target, and orientation response profile for a monkey visual neuron in the foveal projection region of Area 17. Both sets of data are in accord and are well fitted by the theoretical curve generated by the model. Since the bandwidth of the neuron is close to the mode of the population (Ronner, Legge, and Mansfield, 1976), the orientation sensitivity of the most numerous neurons may well be the determinant of threshold for the psychophysical task. By incorporating distribution parameters for
the spatial filters constructed from a cluster of subunits, it should be possible to account for inhomogeneity and anisotropy across the visual field and devise a more general model for the detection of visual targets.
IV. Projections

The understanding of vision in biological systems in terms of predictive mathematical models is a central problem from the scientific point of view whose solution is of great practical significance. How does the human visual system recognize patterns or reconstruct the visual world? Is a Fourier-like decomposition an intrinsic part of the process (Ginsburg, 1973)? At present our research has taken some initial steps toward answering these intriguing questions. Dynamic models for human rod and cone photoreceptors have been generated and the effects of cortical organization on the threshold detection of stationary, oriented targets examined. These models will serve as building blocks in the synthesis of more complex visual processes. The next step is to evaluate the models in psychophysical experiments and incorporate into more general models the effects of retinal adaptation and inhibitory spatial interactions. Such effects need to be studied in the absence of eye-movements. By stabilizing the images of visual targets on the observer's retina, it will be possible to obtain precise estimates of the parameters necessary to evaluate and extent the models.
V. Bibliography


temporal properties. Doctoral dissertation, Harvard University, 
Cambridge, Massachusetts.

spatial frequency channels or bar detecting units. Vision Res., 14, 
909-915.

Mansfield, R. J. W. (1973a). Brightness function: effect of area and 

Res., 13, 2219-2234.

In: H. R. Moskowitz, B. Scharf, and J. C. Stevens (Eds.), 
Sensation and Measurement - Papers in Honor of S. S. Stevens, 


Mansfield, R. J. W., and Daugman, J. G. (1977). Retinal mechanisms of 
visual latency. Submitted for publication.

mechanisms of color vision: spectral sensitivity and hue discrimination. 
In preparation.


Over the past year a special eye-movement monitoring and visual display system designed to further research on human visual information processing has been partially constructed and is nearing completion. The system can accurately measure eye-movements of less than 1' of arc and with the same...
precision control the movement of computer-generated visual targets relative
to an observer's retina. A section of this report describes the 9 subsystems
of the visual apparatus and indicates their potential for experimentation.

The theoretical reconstruction of the response profile to a visual target
is complicated by the fact that under ordinary viewing conditions even
during fixation the population of active elements in the visual pathway
changes constantly. Using the special visual apparatus, it is possible to
maintain a target on a fixed set of retinal receptors facilitating an
orderly experimental analysis of visual response patterns. To direct
this analysis toward predictive models, theoretical work has been focused
on control sites for specific visual functions. At present, models of
visual transduction in human rods and cones have been devised and several
properties numerically evaluated by computer, including temporal modulation
transfer functions. Work is currently in progress on a model of visual
acuity based on the spatial modulation transfer function of neurons in
the primary visual cortex.