

27. Physiology

Academic and Research Staff

Prof. J.Y. Lettvin, Dr. J. Gardner, Dr. S. Jhaveri, Dr. L.A. Kamensky, Dr. D. Perlman, Dr. G.M. Plotkin, Dr. S.A. Raymond, Dr. S. Wiesner, G. Geiger

Graduate Students

L.R. Carley, A. Grant, B. Howland, A. Medina–Puerta, G. Pratt

27.1 Nervous Signals in the Neuropil of Tectum

Bell Laboratories, Inc.

Ortho Instruments

Jerome Y. Lettvin, Edward R. Gruberg²², Eric Prenowitz

In 1959¹ we reported not only on the retinal operations in frog's eye as recorded in optic nerve but also on what seemed the same signals recorded extracellularly in tectal neuropil. We assumed, and thereafter the subsequent literature from other laboratories took for granted, that this sort of tectal record represented the invasion of the terminal bush of an optic nerve fiber. The proliferation of branches in the bush, so we thought, might multiply the local signal current of the fiber by the number of branches and, so, bring the invading signal well above noise level. This hypothesis seemed to explain why we could not record from the fibers in passage but only where they terminated.

But in 1982 and early 1983, Drs. Edward Gruberg and Jerome Lettvin discovered some material that provided a very different account. The tectal neurons have two major known inputs, one from the eye, the other from nucleus isthmi, an ipsilateral slave nucleus to the tectum in that its sole input is from tectal cells, and its output, ipsilaterally, is back to the same cells. This arrangement, by the way, is ubiquitous in land vertebrates from frog through reptiles and birds to mammals, as has been shown by Dr. Harvey Karten.² From a variety of experiments by others as well as by us, the ipsilateral n. isthmi fibers are inhibitory to tectal cells. What we discovered was this: Direct stimulation of n. isthmi produced no obvious extracellular signals in the ipsilateral neuropil, while stimulation of optic nerve produced fine signals. The terminals of n. isthmi fibers alternate with optic nerve fibers as specific layers in the depth of the neuropil. It was, therefore, disconcerting to find no electrical sign of their activity in extracellular records.

From n. isthmi, a small fraction of the fibers go contralaterally to the opposite tectum, as was described in earlier papers by us. These fibers mediate the crossed information used in the frog's

²²Temple University

binocular vision. Like the direct optic nerve fibers to the opposite tectum, they are also excitatory, and electrical stimulation of these crossed fibers produces excellent extracellular signals in the superficial neuropil of the opposite tectum.

It became obvious to us that the sharp, discrete nerve spikes we were recording in tectal neuropil may not be the electrical signals on afferent fibers in the neuropil, but be the subsynaptic responses to the excitatory afferents. This hypothesis is not easily checked except by happy accident. The extracellular electrical spikes in tectum, seen through low resistance metal microelectrodes (of our new design, $\sim 20\text{ k}\Omega$ at 1 KHz for a $5\ \mu$ tip diameter) are fairly often not simple signals but can comprise up to five distinct phases and, so, be well individuated from other signals. These complex transients are of constant amplitude and shape if the electrode tip is not displaced. We found two cases where the overlapping receptive fields of the direct optic fiber of one eye and the relayed fiber from the opposite n. isthmi, representing the same part of the visual field in the other eye, evoked exactly the same characteristic complex spike at the same point in the superficial neuropil. This convinced us that there was a common signal generator activated by both fibers, and could only be the common area on the dendrite on which both ended.

The importance of this finding is great. To explain the significance: It is possible, if our finding is proper, to work out electrophysiologically the topography of endings on dendrites of cells in the central nervous system. The self-same kinds of spikes, appearing in neuropil where the cell bodies are remote, are found everywhere in the brain and spinal cord. They have been dismissed as fibers of passage, etc. Intracellular records only give a distant and smeared integrated view of all the synaptic activity in all the dendrites of the impaled cell. This method locates precisely active excitatory endings of known provenance. By their nature, inhibitory endings, which evoke shunts to the V_{K^+} or V_{Cl^-} across cell membrane, cannot produce the same current-generating responses as excitatory endings. Thus we now have the tool for telling whether a particular input to a neuron is excitatory.

References

1. J.Y. Lettvin, H.R. Maturana, W.S. McCulloch, and W.H. Pitts, "What the Frog's Eye Tells the Frog's Brain," Proc. I.R.E., 47:1940-1951, 1959.
2. H.J. Karten and R.O. Kuljis, "The Frog and the Peptide Swamp," to be published.

27.2 Sensing of Texture by Retinal Ganglion Cells

Bell Laboratories, Inc.

Ortho Instruments

Jerome Y. Lettvin, Arthur Grant

In 1959¹ we reported a variety of ganglion cells from the frog retina that became known as the "bug" detector. This element is the most frequent sort to be found in the retina and constitutes well over 50% of the population. However, it has a small cell body, and its axon is approximately $0.2\ \mu$ in

diameter at most, so that it is difficult to detect by its electrical signal. Initially, we found this type in optic nerve, using electrodes that, for some reason, have never become very popular. (They are a modified form of the Dowben–Rose probe that uses platinum–black at the tip. The real component of the input resistance of such a tip in contact with physiological saline solution is $\sim 20\text{ K}\Omega$ at 1 KHz for $\sim 4\ \mu$ diameter, so that the noise is extraordinarily low. But they require individual preparation — they cannot be batch–produced.) This retinal element, type II, however, is readily detected in tectal neuropil by the responses described in the previous section.

The literature since 1959 has more or less rejected our initial description, not because the experimental results could not be repeated, but because there was a simpler description to be had by the responses to less complex stimuli. That is to say, the details that we found so interesting were to be considered ancillary, epiphenomenal.

Our original description, while incomplete, had the misfortune of being improperly named as "net convexity" detector, and this name, rather than the substance of the account, became the object of attack and the grounds for denying the complexity of image processing that we found. It is worthwhile, therefore, to recall the original findings so as to set the basis for the new work to be reported here.

A type II ganglion cell in frog retina has an intrinsic receptive field of $\sim 5^\circ$ in visual angle. That is, on a screen, 15" away from the eye, the area to which it is directly sensitive is about 1" in diameter and is fairly sharply defined. The sensitivity is indicated by the production of pulse trains in the fiber. This area is directly surrounded by an annular region whose outer border is vague and variable; it is called the "surround". Visual events in the surround diminish the responses made to other events in the central receptive field. This surround annulus can be over 2" thick; the bad definition of the outer border simply reflects the lessening of the inhibitory influence of a stimulus by its distance from the border of the central receptive field, hereafter call simply "center RF", so that one can talk of center RF/surround RF relations.

The center RF is located by moving a small black spot $\sim 15'$ of arc in diameter over a blank white screen. The position of its center (to be called "focus") is estimated, and the spot is moved radially toward it from about 2° – 3° away. The border of the center RF is defined by the occurrence of a pulse train where the spot crosses the boundary between surround RF and center RF. The outline of center RF is somewhat circular or oval, but occasionally is cardioid. Movement of the black spot centrifugally from the focus of the center RF produces a smaller response (see below).

Once the eye of the frog is fixed against any rotation, the center RF of a type II neuron is well determined on a fixed screen and can then be studied for the relation between visual events and the pulse trains given by the neuron.

Two important insensitivities are found immediately. First, no steady state of the region covered by

the center RF and surround RF excites any nervous response in the fiber. The background white screen can be supplanted by a large color photograph of a complex scene, and the neuron is indifferent. Second, no uniform change of illumination on the screen excites a response. The general lighting in the room can be switched on and off to the complete indifference of the neuron.

The only excitement of the fiber occurs when a phase boundary comes into being or is moved within the center RF. For simplicity, let us talk of two phases, black and white, where the black phase has ~ 0.12 the luminosity of the white. For a response to occur there are three conditions:

A. A single area within the center RF, up to and including the whole center in area must be sharply darkened, and the dark phases must have a sharp edge. The smallest area that produces a response in this way is about $2'$ of arc in diameter. The largest is slightly larger than the area of the center RF (which had been outlined by the radial moving spot method). The effect of blurring the edge of the dark spot compromises the response more than diminishing the contrast between the dark and light phases. The response is maximal at the time the black phase suddenly appears, then dies off in time. After about 20 sec, but often somewhat longer, the dark phase against the light background becomes a steady state system and is ineffective.

B. If a single black spot, fixed in area and shape and lying completely within the center RF, is moved in steps within the center RF, it excites the fiber with each step. The excitement is greatest when there is no centrifugal lightening by the movement. Thus a small black spot approaching the center of the center RF has a leading edge and a trailing edge. The response is large, except as the trailing edge moves predominantly away from the focus. Thus, for example, a black spot growing in size radially away from the focus gives a good response. So does a single black spot, part of whose border is co-extensive with the border of the center RF but which grows to occupy the whole of the center RF without trespassing onto the surround. All movements or growth of a black spot produce a good response so long as the spot is not in some sense "retreating", i.e., growing smaller or moving centripetally.

C. The moving dark phase or sharply appearing dark phase must be a single continuous spot, a single phase. This was the condition that excited enough disbelief that, apparently, it was never seriously checked or pursued. So, for example, three black spots, each capable of exciting a good response if alone, when rigidly spaced from each other and moving rotationally or translationally as a triad in the center RF, do not excite much of a response, if any. However, if the space between the dots is blackened so as to give a single black triangle, the response to its movement is again excellent. Furthermore, if yet a fourth spot is added and rigidly coupled, the neuron is insensitive to the tetrad. But if the fourth spot is decoupled from the other three and moved independently of them, it again excites the type II element.

A pair of rigidly coupled spots is not as exciting as a solid black bar of the same width and length as if the space between the pair had been filled. But the cell, nevertheless, does respond to a pair, if with

diminished vigor. Finally, a single black spot with a tortuous serrated edge, is almost as exciting as a black spot with a smooth edge.

It is evident from these three conditions that the type II element is sensitive in its center RF not only to phase boundary but to phase continuity. (This second aspect prophylactically rid us of Minsky and Papert's² later perceptron model).

Interaction of the center with the surround was also complex. For example, suppose against a blank white background a large black sheet of paper to be moved edge first into the surround RF and thence, jerkily, but with steady advance so as to intersect and finally pass through the center RF. There is absolutely no response. However, if the sheet is moved, corner first, into and through the receptive field in the same way, the response is strong. Several later researchers, notably Gaze and Jacobsen,³ felt that the growth of darkening from the rim of the center RF inward was excitatory purely as darkening if nothing occurred at the same time in the surround RF. In the case of the black sheet, edge first, the surround was darkened at the same time as the center was darkened, and, so, edge need not be involved — the interaction was simply a case of inhibition from the surround and on the basis only of diminution of flux. They conveniently ignored the case of the black sheet moved corner first. For some reason, the notion of shape detection in the retina was felt to be suspect. It needed, so the received wisdom held, the amenities of mammalian cortex. Edges possibly might, by some peculiar circuitry, be sensed by single retinal cells as was later found also in pigeon or rabbit, but certainly nothing as complex as shape or texture could be detected. Gaze and Jacobsen's³ experiment was thereafter cited as the example of how to reduce what seemed to be a complex system into an easily explained inhibitory interaction between center and surround on the basis of flux change alone.

With these comments as background, we can now describe our current work, emphasizing one specific aspect, that relating phase and texture. The primitive experimental setup is this: We locate a center RF of a type II neuron as described earlier. Then we cut a hole in the white screen slightly smaller than the center RF. The hole is backed with a wide, white cardboard flap that allows introduction of stimuli between it and the screen. Thus no stimuli ever appear in the surround. Five distinct stimuli are prepared. They are all circular paper discs, all of the same size, about half the diameter of the hole. Each is glued to a thin flat steel washer that can be moved over the face of the flap by a magnet behind the flap. The five stimuli are these:

1. A uniform black disc.
2. A white disc with several identical black spots on it. The area of the black spots is about the same as the area of the white ground of the disc.
3. Black disc with several white spots on it. Again, within the disc the areas of white and black are about equal.
4. A gray disc with half the luminance of a white disc.

5. A white disc with a black spot on it the size of those in stimulus 2.

While it is not technically correct to say it, stimuli 3 and 4 have the same first-order statistics and differ only in the second order.⁴ They differ also in another way. In stimulus 2, the continuous phase is white, in stimulus 3 it is black.

There is a good response to stimuli 1 and 4, and a good response also, though not as vigorous, to stimulus 5 when they are introduced into the center RF from behind the screen. There is no response to stimulus 2 but a fairly good and distinct response to stimulus 3.

If the hole is itself masked by a white sheet of paper and the stimulus set in phase behind it so that it suddenly appears when the sheet is sharply withdrawn, the same order of stimulus effectiveness is seen, and almost no response to stimuli 2.

These initial simple experiments are not as easy as they sound. The frog must be robustly healthy, immobilized in body and eye, the pupil must be constricted, and the retinal circulation good. Dilated pupil, poor circulation, the presence of barbituates or other anesthetic and any more curare than is just necessary for immobilization, all militate against successful study.

These early results show several things. Simple average darkening of a portion of the center RF is not a stimulus. Otherwise the responses to 2, 3 and 4 would be the same. Amount of edge is not pertinent, otherwise 2 and 3 would be equally excitatory. In distinguishing between the stimuli 2 and 3, there is the choice between saying the stimuli differ in second-order statistics or in which phase is continuous. There is, obviously, a weak relation between the two statements, except that we cannot see how continuity of phase can be inferred from the second-order statistics, while the other way round is fairly transparent although not simple. Furthermore, the textures, which we call microtextures to distinguish them from those which Julesz analyzed, are abstracted by one ganglion cell alone, not from an ensemble of such cells.

There has been one interesting by-product of this work. For a long time we have felt that something was wrong, extremely wrong, in using video screens as tools in studying the retina. It is evident enough for the frog that the envelope of dots in stimulus 2 is a far cry from the continuous single phase boundary of stimuli 1 and 4. Oddly enough the same microtexture prevents the perception of form in human peripheral vision. Form is not given by an envelope of dots if they are clearly discrete as pixils, but only by what are actually (rather than virtually) phase boundaries.

References

1. J.Y. Lettvin, H.R. Maturana, W.S. McCulloch, and W.H. Pitts, "What the Frog's Eye Tells the Frog's Brain," *Proc. I.R.E.*, 47:1940-1951, 1959.
2. M. Minsky and S. Papert, *Perceptrons* (M.I.T. Press 1969).
3. R.M. Keating and J.M. Gaze, "Observations on the "Surround" Properties of the Receptive Fields of Frog Retinal Ganglion Cells," *Quart. J. Exp. Physiol.*, 55, 129-142 (1970).
4. B. Julesz, "Experiments in the Visual Perception of Texture," *Sci. Am.* 232, 34-43 (1975).

27.3 Analogue Model of a Photoreceptor

Bell Laboratories, Inc.

Ortho Instruments

Jerome Y. Lettvin

Conceive a photosensitive pigment in a receptor that is several wave-lengths of light in width — a few microns. The pigment has a uniform spectral absorption. When all the pigment is in native state, it captures 2% of the photons entering the receptor in a flash of light.

The pigment has four states. It is photosensitive only in the native state, A. When a molecule in state A captures a photon, it switches very rapidly to state B, the first intermediate product. A population of molecules in state B switches then to state C, the second intermediate, with a rate constant of β . In turn, a population in state C switches to the fully bleached state, D, with a rate constant of γ . And those in state D are returned to state A by an energetic process and with rate constant of δ . In respect to the rate constants, $\beta > \gamma > \delta$, as is required for stability of the state loop. The system is described by a simple ring, using state letters to signify the fractions of pigments in those states. φ is the flux entering the receptor, and Z is the capture fraction by the pigment when A = 1. Then

$$\frac{-dA}{dt} = A\varphi z - D\delta$$

$$\frac{-dB}{dt} = B\beta - A\varphi z$$

$$\frac{-dC}{dt} = C\gamma - B\beta$$

$$\frac{-dD}{dt} = D\delta - C\gamma$$

Under steady state of φ all these expressions equal 0.

We define two constants, $K_1 = \beta/\gamma$ and $K_2 = \beta/\delta$; and an attenuating constant

$$\frac{1}{K_3} < 1.$$

The state variables are then represented by conductances in the following circuit (Fig. 27-1), which is a primitive model of our present design:

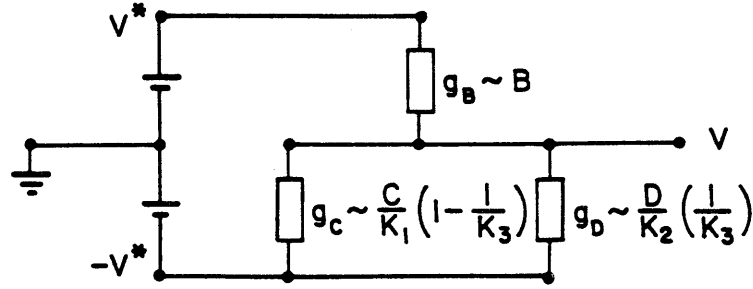


Figure 27-1

By Kirchoff's rule

$$\frac{V}{V^*} = \frac{g_B - (g_C + g_D)}{g_B + (g_C + g_D)} = S, \text{ the signal value } -1 \leq S \leq +1$$

and $S = 0$ at all steady states of φ .

Suppose the system has come to steady state and then a small step of $\Delta\varphi$ occurs.

At the time of the step

$$\Delta S = \frac{\Delta g_B}{2g_B} \quad \text{since } g_B = g_C + g_D \text{ just before the step.}$$

Thus $\Delta S_T \sim \frac{\Delta\varphi_T}{\varphi}$, which is the Weber-Fechner law (the subscript T signifies threshold).

If the step is maintained, S returns to ~ 0 with the approximate rate constant of β . When D is in steady state with C and B, for any step $\Delta\varphi$ away from the steady state φ_0 , at the instant of the step

$$\Delta S \approx \tanh \frac{1}{2} \ln \left(\frac{\Delta\varphi + 1}{\varphi} \right)$$

the operating characteristic. It is approximately what Norman and Werblen measured in actual cones.

And if the step is maintained, S returns to 0.

Note that ΔS is independent of φ_0 and depends only on $\frac{\varphi}{\varphi_0}$.

Note also that the saturating function S is almost linear with $n \frac{\Delta\varphi}{\varphi_0}$ between $\frac{1}{3} \leq \frac{\Delta\varphi}{\varphi_0} \leq 3$

which is the decade needed for handling reflectances. Note also the effects when $D > K_2 B$ and when $D < K_2 B$ and compare them with experience. This is not the final version, which takes account of the ohmic interconnectedness of cones, but it conveys the major ideas.

We will not go into how such a model accounts for several previously unaccountable phenomena and laws (e.g., the Rushton–Dowling law that for threshold to a flash under complete darkness

$$\frac{\log(\text{density of photons in a threshold flash})}{K} \sim D$$

where K can lie between 2 and 20, depending on the species of animal and the species of photoreceptor. The law holds for $.01 < D < .99$ as Dowling showed.) Instead we want to discuss what is involved in the strategy.

Since our eyes move about constantly in microsaccades and saccades to maintain the image, C takes a short-term running history of the fluxes encountered, and D takes a long-term running history. In brief, D does in time the averaging over many phases in many places of the scene because our eyes move about, while C does the averaging of the phases in the immediate region of the image. C is used for the normalization, the short-term adaption, and does not represent an independent measure or constraint. D supplies what was needed in the dimensional analysis without recourse to feedback from later processing.

This is a novel and, so far as we know, an unprecedented use of pigment intermediate breakdown products to supply the method for both the immediate local processing and the additional representation of the surround so as to account for the degrees of freedom in color vision. But the same theory also explains the Norman and Werblen "operating characteristic" measured on cones — a $\tanh \frac{1}{2} \ln \frac{\varphi}{\varphi_0}$ where φ_0 is the adapting light. The mechanism for such a process has not yet appeared or, to our knowledge, been suggested.

27.4 Enhancement of Form Perception Under Textural Masking

Bell Laboratories, Inc.

Ortho Instruments

Jerome Y. Lettvin, Gad Geiger

In eccentric vision, i.e. perception away from the fixation axis, there is a distinct interplay between form and texture. It occurs in foveal as in extrafoveal or peripheral vision and has been called lateral masking. Most easily demonstrated with letter and other shaped signs, it can be illustrated thus:

N X T E N E T

When the X is fixated with either or both eyes, the isolated N is quite visible; the N in TENET is not, although the two are equidistant from the X. This phenomenon was systematically investigated first by

Bouma.¹ Now there are over a score of relevant papers as given in the bibliography, which is far from exhaustive. To some extent the shape of a sign has much to do with its ability to mask or be masked, as does its boldness, contrast, distance from other signs and the sharing of direction between parts of neighboring signs. These relations have been examined extensively by Bouma and those who followed. But what remains as common in all cases is that the interaction between separate adjacent signs suppresses something related to their form so that the interior signs of a string are hard to identify. The most eccentric sign of an eccentric string is commonly the easiest to make out, and the least eccentric is the next easiest.

There has been a feeling, voiced again and again from Estes,² that such interaction, which makes eccentric vision less clear than would be expected from measurement of spatial resolving power, has its key in the nature of cortical receptive fields. In a sense this must be the case, the only reservation being whether those receptive fields have been well-enough described so as to accommodate the observations.

The reason for this doubt is that we have found an enhancing or unmasking interaction between the center of gaze and the eccentric field. Unmasking could not occur if the masking process was at so elementary a level that the information by which shaped signs are judged has already been lost. Interactions of this sort are transient as opposed to the masking which endures. They were sought and discovered in a simple way; the experiments to be described refined the observations so as to rule out various epiphenomenal causes.

Experiments

Two slide projectors were mounted behind and aimed at a translucent diffusing screen such as occurs in large film-readers. Inscribed on the screen was a fixation mark. Both projectors were equipped with current-driven shutters able to open or close in three milliseconds. The currents were governed by conventional gating operations whose timing could be set accurately. One projector displayed a test image on the screen for a period that could be adjusted up to 150 ms, but no more. Within this period the probability of one microsaccade is low — of two is negligible, and there is no time for a voluntary eye movement. From our point of view, the conditions for tachistoscropy are met, given the fixation of the eye up to and during the exposure.

The test image was followed by an "erasing" image, usually a square grid crossed by a grid of the diagonals. This erasing image always occurred with a delay after the test image was turned off. The delay from the onset of the test image never exceeded 250 ms. Erasure was important in establishing repeatability and reliability of the data. Otherwise there could be distinct variations that were related, we felt, to the use of after-images or to changes in adaptation.

The complex signs used in the test-image were bold, upper-case block letters of high contrast. They were about 35 minutes high in angular size from where the subject sat, and at most 30 minutes broad, and when presented in strings, were separated by 35 minutes between the centers so as to

provide clear spacing. These signs are called "complex". Another sort of sign, the "simple" one, consisted of bars with the same thickness of line as the letters but presented in different orientations.

Initially, while the subject gazed at the fixation point, a string of three signs was flashed eccentrically $2^{\circ}40'$ away. The length of exposure and timing of erasure were adjusted until the subject reported correctly slightly less than 100% of the time, i.e., 80%–90%. This less-than-fully correct calling was quite stable and repeatable and often was retested at the end of a run. Thus, when the same strings were flashed at 8° eccentricity, which we used as a standard distance, the accuracy of report dropped to a fairly low level. This ensured that all tests were done below the threshold for recognition. Thus a base was provided against which enhancement of recognition could be tested for isolated signs or strings. Enhancement usually gave a 20%–50% improvement in the density of correct calls.

The subject used both eyes in the experiment, since we wanted to avoid comparing nasal and temporal fields in each eye.

Results

In the same image flash we began with two signs, one at the fixation point, one at 8° eccentricity. The signs were either identical or different. With disparate signs, identification of that at 8° was poor. But when the two signs were the same letter, identification was distinctly enhanced. This enhancement, however, applied only to complex signs. When simple signs were used, bars at the same orientation or at disparate orientations, no enhancement occurred.

With strings of three complex signs in a horizontal line, the center sign at 8° , there was distinct enhancement of any letter in the string when the identical letter was flashed simultaneously at the fixation point. When a letter not in the string was flashed at the fixation point, there was little or no enhancement, as if no fixation-point letter was flashed. And again, there was no enhancement with the same experiments done with simple signs instead of complex ones — the masking in the eccentric field stayed unchanged.

Similar experiments were done at greater eccentricities, and from them emerged two populations. The greater number of subjects showed neither resolution nor enhancement beyond 10° – 12° eccentricity. A small group, to our surprise, showed enhancement at 15° – 18° eccentricity. (The blind spot does not figure here, since the tests were done with binocular fixation.) This latter group were all characterized by significant reading difficulty. We mention this as an aside.

Finally, in admixtures of complex and simple signs in eccentric strings of three, enhancement always occurred when the identical complex signs lay at the fixation point and in the string, and never occurred when identical simple signs were in the same pair of positions.

Discussion

So far the work has addressed what may be called an interplay between texture and form. The interior sign in a string of eccentric complex signs has an odd quality when there is no enhancement. Something seems to be there — it has boundary in a way — but the spatial order is lacking by which that boundedness is given form. It is a textural part of the string, a vaguely prehended distribution, an innominate *chiaroscuro*, and the subjects do not guess wrong letters. They simply say they could see nothing clearly. When it stands out by enhancement in this transient display, it takes on an almost distinct form. We are driven to suppose, therefore, that there is a texture-breaking interaction between the fixation point and the eccentric string — an enhancing of unmasking influence.

If the masking was a primitive and information-destroying process, it would be hard to imagine how such enhancement was possible. The recognition is triggered only when the correct sign lies in the string. Otherwise there is no distinct percept at all. Therefore, the information that such and such a letter lies in the string must still be available in what appears, without enhancement, as formless — the masked center of the string.

We have as yet no way of accounting for this phenomenon, but its very existence calls into question those mechanisms that have been proposed to explain lateral masking.

While lateral masking is a robust and stable interaction, unmasking (enhancement) is transient and fragile. We needed tachistoscropy to show the interplay. The fragility of enhancement can be illustrated by the effects on it of microtexture. By microtexture we mean that the quality of the sign is not uniform; i.e., instead of being unrelievably white against black or black against white with sharp boundaries in either case, it is comprised of an assembly of dots or stripes, or what have you, for which we read the envelope as being the sign. With signs of this sort, such as are had from a dot matrix or TV screen, not only is masking most effective, but enhancement is singularly weak if present at all. This should occasion no surprise, because envelopes (textural boundaries) are not treated the same as phase boundaries in early visual processing. (We had shown this in the receptive fields of frog optic nerve fibers (Lettvin et al., 1959). The ease of using computer displays as stimuli has, to some extent, suppressed such observations in mammals.) However, enhancement is undisturbed by macro-texture or embedding texture, e.g., a cloud of noise texture (like that of a noisy TV screen but expanded to where the black spots are the size of letters). This cloud is inserted between the fixation point and the eccentric 8° letter and extends from the fixation point to the letter.

There are also indications that, taking the fixation point as the center of the system, masking in the tangential direction differs from masking in the radial direction, with consequent changes in the enhancement.

The enhancing interaction described here occurs up to a 10°–12° distance in the visual field for most of our subjects and over a markedly larger field for a distinct subclass of them. But it occurs only with

complex signs such as block letters. It does not occur with simple signs such as bars. That the interaction is expressed in the recognition of the eccentric signs raises the question of how local processing, whether in primary or secondary cortex, can be so connected over large distances to permit it to work in such a way, under tachistoscopic presentation, as to prevent or break lateral masking.

There are, of course, methods of building filters to reveal on the instant all examples of a particular sign in some integral transform (e.g., a hologram) of a two-dimensional cloud of different signs. If such sophisticated processing must be invoked, it invites a rethinking of the nature of receptive fields, particularly those found in the cortex.

References

1. H. Bouma, *Nature* 226, 177–178 (1970).
2. W.K. Estes, *Percept. & Psychophys.* 12, 278–286 (1982).

27.5 Physical Reasons Behind Caisson Disease

Bell Laboratories, Inc.

Ortho Instruments

National Institutes of Health (Grant 5 TO1 EY00090)

Jerome Y. Lettvin, Edward R. Gruberg²³, Robert M. Rose²⁴, George M. Plotkin

Most guesses about why diving animals do not get the bends have focused on possible mechanisms for preventing nitrogen from dissolving in tissue under high pressure. Examples are: the forcing of inspired air from alveoli into bronchi and trachea whence negligible gas exchange occurs; the shunting of circulation away from sensitive tissues; etc. On the whole, these measures, used to the extent needed to account for the absence of bends, are so incompatible with active life as not to be plausible under any number of compensatory hypotheses.

Recent work by Ridgway and Howard¹ lays to rest any need for such nonce engines. In their study of dolphins diving to about 100 meters over and over again, the dissolved N₂ in the muscle tissue rose to three times the partial pressure found in dolphins that remained at the surface. Despite rapid ascents that would certainly have given the bends to a human diver, the dolphins seemed to be quite comfortable.

It is not likely that dolphins simply do not complain about their blood boiling; their blood cannot froth, since no mammal, man included, could survive frequent gas emboli in heart or brain.

²³Temple University

²⁴Department of Mechanical Engineering

Therefore, the findings of Ridgway and Howard indicate that attention be directed not at the prevention of supersaturation but rather at the absence of the formation of bubbles under supersaturated conditions. The basic phenomena are evident to the thoughtful drinker of beer: Bubbles occur at specific nucleation sites such as scratches on the glass container, or solid particles adherent to the container or inadvertently introduced into the beer itself. The effervescence can be substantially reduced or eliminated by the careful use of appropriate containers, e.g., a smooth, clean, fire-polished beaker. Observations on a glass of beer near a radioactive source led directly to the invention of the bubble chamber as an experimental tool for high-energy physics. Bubble formation is nucleated by the particles as they decay along their paths and nowhere else. Another important example is the formation of CO bubbles in molten steel, which is necessary in steelmaking, and which is induced by mechanisms that nucleate the melt.

Ebullition of dissolved gases in liquids can be adequately described by the classical theory of Volmer,² Weber,³ Becker and Doring,⁴ as corrected by Lothe and Pound.⁵ The original theory was directed at the homogeneous nucleation of liquid drops from vapors, and later developed⁶ to deal with heterogeneous nucleation and expanded to deal with solids and liquids. The nucleation of bubbles by solid substrates in superheated liquids have been considered by Frenkel⁷ and Fisher.⁸ In general, the nucleation rate should be proportional to the expression

$$(\Delta G^*)^{1/2}(1 - \cos\theta^1)\exp(-\Delta G^*/KT) \quad (27.1)$$

where the activation energy G^* is given by

$$\Delta G^* = 16 \pi \sigma^3 \varphi(\theta^1)/3(P^* - P)^2 \quad (27.2)$$

and

$$\varphi(\theta^1) = \frac{(2 + \cos\theta^1)(1 - \cos\theta^1)^2}{4} \quad (27.3)$$

where σ is the surface energy, θ^1 is the complementary angle to the contact angle of the liquid to the substrate (i.e., π minus the contact angle), P is the imposed hydrostatic pressure and P^* is the partial pressure of the nitrogen gas inside the critical nucleus.

We suggest here that the difference in susceptibility between humans and, say, dolphins can be accounted for by the presence of more nucleation sites in humans and less chemical suppression of heterogeneous nucleation. With very few exceptions, heterogeneous nucleation is the rule in nature. Mammals provide numerous substrates for nucleation of nitrogen bubbles. For land mammals there are cartilaginous and calcific granules generated by the attrition of weight-bearing joints. There is a low but definite rate of lamellar fracture in cancellous centra of the vertebrae and the subchondral regions of the long bones. In most of us there are atheromatous plaques unevenly distributed in the

circulatory system which eventually calcify in arteriosclerosis. These "boiling chips" appear incipiently even in infants. There are huge numbers of other potential nucleation sites, e.g., calculi in gall bladder and urinary bladder; growths and scars on the edges of cardiac valves; etc. In fact, Eqs. (27-1) — (27-3) apply to flat substrates, and any re-entrant cavity will be a much more potent nucleation catalyst than a flat surface of the same tissue or material,⁸ so that any surface with fine folds or convolutions, e.g., villi, will be an excellent heterogeneous nucleant. In essence, every belch and crepitation vouches for our intolerance, not for the deep so much as for any rapid translation from it.

If, then, dolphins are to tolerate high supersaturations of dissolved nitrogen gas without fizzing, two approaches are possible. One is the elimination of potential nucleation sites by achieving an internal smoothness of high order; that is to say, that the circulatory systems of diving animals such as whales, dolphins and seals are "fire-polished" by evolution. On the other hand, heterogeneous nucleation can be suppressed by chemical inhibitors which reduce the catalytic potency of the substrates. This point of view is suggested by the fact that many fish survive extended supercooling until they are scratched or otherwise nucleated, and only then will suddenly freeze. They possess a potent "anti-freeze", a glycoprotein, which clearly is not in sufficient concentration to depress the freezing point significantly and, therefore, must be a suppressor of heterogeneous nucleation. The same point has been made abundantly clear in the formation of kidney stones and bladder stones, where supersaturations are attained in the normal human kidney⁹ only because heterogeneous nucleation has been suppressed. For the case of bubble formation, Eqs. (27-1) — (27-3) make it clear how such an inhibitor would work. A wetting agent would be highly effective, since complete wetting (or a zero contact angle) would take $\varphi(\theta)^1$ to unity and ΔG^* would be at its maximum value, i.e., the value appropriate to homogeneous nucleation of bubbles, and much higher supersaturations, up to those necessary for homogeneous nucleation, could be sustained. Alternatively, the number of nucleation sites may be so vastly increased in land-dwelling mammals that small quantities of nucleation suppressant may not suffice. In either case, it is possible to conceive of a chemical control for the "bends", if not on a chronic level at least for acute emergencies.

References

1. S.H. Ridgway and R. Howard, *Science* 206, 1182-1183 (1979).
2. M. Volmer, in *Kinetik der Phasenbildung*, Steinkopff, Dresden & Leipzig, 1939.
3. M. Volmer and A. Weber, *Phys. Chem.*, 119, 277-301 (1926).
4. R. Becker and W. Doring, *Ann. Physik.* (5) 24, 719-752 (1935).
5. J. Lothe and G.M. Pound, *J. Chem. Phys.* 36, 2080-2085 (1962).
6. J.P. Hirth, *Ann. N.Y. Acad. Sci.* 101 805-815 (1963), review.
7. J. Frenkel, *Kinetic Theory of Liquids*, Oxford U. Press, London (1946).
8. J.C. Fisher, *Appl. Phys.* 19, 1062-1067 (1948).
9. *Urolithiasis: Physical Aspects*, Nat'l. Acad. Sci. (1972).

27.6 Quantum Cryptography²⁵

Ortho Instruments

Stephen J. Wiesner

A class of codes is made possible by restrictions on measurement related to the uncertainty principle. Two concrete examples and some general results are given.

The uncertainty principle imposes restrictions on the capacity of certain types of communication channels. We will show that in compensation for this "quantum noise", quantum mechanics allows us novel forms of coding without analogue in communication channels adequately described by classical physics.

We will first give two concrete examples of conjugate coding and then proceed to a more abstract treatment.

Example One: A means for transmitting two messages, either, but not both of which, may be received.

The communication channel is a light pipe or guide down which polarized light is sent. Since the information will be conveyed by variations in the polarization, it is essential that the light, and that all polarizations of light, travel with the same velocity and attenuation.

The two messages are rendered into the form of two binary sequences. The transmitter then sends bursts of light at times that we will label T_1, T_2 , etc. The amplitude of the bursts is adjusted so that it is unlikely that more than one photon from each burst will be detected at the receiving end of the light pipe.

Before emitting the i th burst ($i = 1, 2, \dots$), the transmitter chooses one of the two messages in a random manner by flipping a coin or selecting a bit from a table of random numbers. If the first message is chosen, the i th burst is polarized either vertically or horizontally depending on whether the i th digit of the first binary sequence is a zero or a one. If the second message is chosen, the i th burst is polarized in either the right or left-hand circular sense depending on whether the i th digit of the second message is a zero or a one, Fig. 27-2. The receiver contains a quarter-wave plate and birefringent crystal, or some other analyzer, that separates orthogonally polarized components of the light wave into spatially separate beams. Following this is a pair of the best available photomultiplier tubes. If the first message is to be received, the analyzer is arranged so as to send vertically polarized photons to one phototube and horizontally polarized photons to the other. If the second message is to be received, the separation is made with respect to right and left-hand circular polarization.

²⁵As published in Association for Computing Machinery Special Interest Group on Automata and Computability Theory 15, 1, Spring 1983.

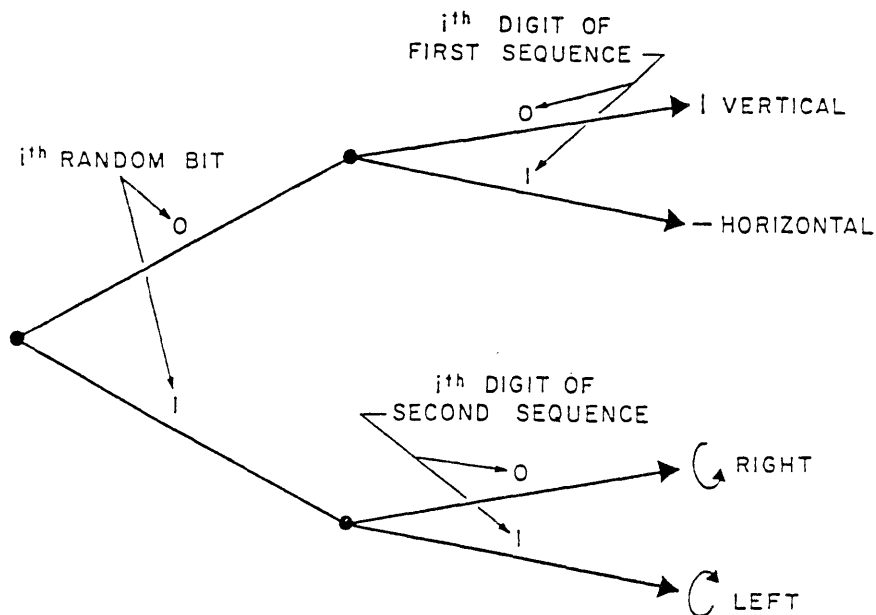


Figure 27-2: Polarization of the i th Burst

Now if the linear polarization of a photon is measured, all chance of measuring its circular polarization is lost. Thus, if the receiver is set to receive the first message, nothing at all is learned about the contents of the second message. Likewise, when the receiver is set to receive the second message, it destroys all information concerning the first message. If the receiver is set up to sort the photons with respect to some elliptical polarizations intermediate between linear and circular, less information about each message is recovered than when the receiver makes the best measurement for the reception of one message alone.

Of course, even when the receiver is set for the first message, a full knowledge of the first sequence is not recovered. In fact, half the digits of the first sequence never even influence the transmitted signal and at the corresponding times, when the second message is being transmitted, the receiver output has an equal probability of being a zero or a one. This noise introduced by the coding scheme, as well as the noise due to the channel, the photon shot noise, and the photomultiplier noise, may be overcome if an error-correcting code of the usual sort is used in forming the binary sequences from the original messages. Care must be taken, for too much redundancy would allow both messages to be recovered by the alternate reception of one sequence and then the other.

There is no way that the receiver can recover the complete contents of more than one of the conjugately coded messages so long as it is confined to making measurements on one burst of photons at a time. In principle, there exist very complicated measurements that allow recovery of all the transmitted information. To see this, consider the transmission of two messages of finite length. The transmitter will produce a signal consisting of a finite number of bursts of polarized light, and the entire signal may be described by a single vector ψ in a large Hilbert space spanned by all possible

finite transmissions. If one of the messages is changed, a state corresponding to a different vector ψ' is produced. The change from ψ to ψ' could be detected unambiguously by a receiver of the type previously described, if set to receive the message that was changed. For this to be possible, ψ must be orthogonal to ψ' . It follows that the set $\{\psi\}$ of the vectors corresponding to all possible pairs of finite messages is ortho-normal and, therefore, there exists an Hermetian operator or a set of commuting Hermetian operators corresponding to a measurement or measurements that can distinguish all the possible signals.

There is an easy extension to the case of three messages, no two of which may be recovered. One simply transmits a third binary sequence using light in the two polarization states at 45° to vertical and horizontal. Extension to more than three messages is not straightforward.

The above system for sending two mutually exclusive messages could be built at the present time. Though it is possible in principle to beat the system and recover both messages, to do so would require measurements that are completely beyond the reach of present-day technology. The system, therefore, works in practice but not in principle. The next example is in the opposite category; it is foolproof in principle, but it probably could not be built at the present time.

Example Two: Money that it is physically impossible to counterfeit.

A piece of quantum money will contain a number of isolated two-state physical systems such as, for example, isolated nuclei of spin $1/2$. For each two-state system, let a and b represent a pair of ortho-normal base states and let $\alpha = 1/\sqrt{2}(a + b)$ and $\beta = 1/\sqrt{2}(a - b)$ represent another pair.

The two state systems must be well enough isolated from the rest of the universe so that if one of them is initially in the state a or α , there is little chance that a measurement made during the useful lifetime of the money will find it in the orthogonal states b or β , respectively. There is no device operating at present in which the "phase coherence" of a two-state system is preserved for longer than about a second; however, the continuing advance of cryogenic technique will surely change this.

Let us suppose, to be definite, that the money contains twenty isolated systems, S_i , $i = 1, 2, \dots, 20$. At the mint they create two random binary sequences of twenty digits each which we will call M_i and N_i , $i = 1, 2, \dots, 20$, $M_i = 0$ or 1 , $N_i = 0$ or 1 . Then the two-state systems are placed in one of the four states a , b , α or β in accordance with the scheme shown in Fig. 27-3.

The money is also given a serial number which is printed on it in the usual way, and the two binary sequences describing its initial state are kept on record at the mint and perhaps at a number of branch banks.

When the money is returned to the mint, a check is made to see if each isolated system is still in its initial state, or whether it has switched to the orthogonal state.

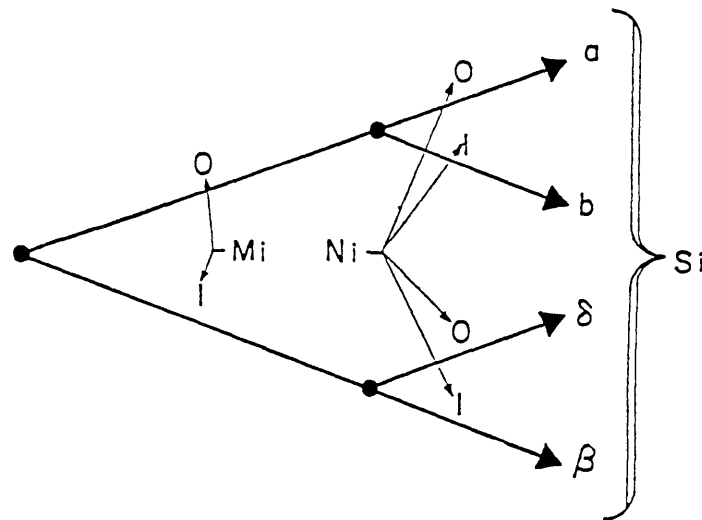


Figure 27-3

Now consider the problem of someone who would duplicate a piece of quantum money. He cannot recover N_i because, since he does not know M_i , he does not know what measurements to make on S_i . A measurement on a particular S_i that distinguishes α from b must necessarily destroy all chance of distinguishing α from β . Likewise, a measurement that distinguishes α from β destroys the chance of distinguishing a from b . Suppose a counterfeiter goes ahead anyway, makes some measurement on the S_i and produces money with the new S_i in the states found by his measurements. Then for each i , there is a 50% chance that he will make the wrong measurement, and in this event there is a 50% chance that a measurement at the mint will show S_i to be in the wrong state. Thus, there is a 1/4 chance of each digit being found wrong and the probability of the whole counterfeit coin passing inspection is only $(3/4)^{20} < 0.00317$.

Could there be some way of duplicating the money without learning the sequence N_i ? No, because if one copy can be made (so that there are two pieces of the money), then many copies can be made by making copies of copies. Now given an unlimited supply of systems in the same state, that state can be determined. Thus, the sequence N_i could be recovered. But this is impossible.

Conjugate Bases

If the momentum of a particle is known, then nothing is known about its position; in other words, it is equally likely to be found in all regions possessing a fixed volume V . Likewise, if the position is known, then nothing is known about its momentum. The same relation holds between all pairs of conjugate variables, and this suggests an extension of the idea of conjugation from variables to basis sets.

Let $\{a_i\}, i=1,2,\dots N$ and $\{b_i\}, i=1,\dots N$

be two ortho-normal bases for an N dimensional Hilbert space. We call such a pair conjugate if and only if $|(a_i, b_j)|^2 = 1/N$ for all i and j²⁶. Physically, if a system is in a state described by a_i , $i = 1, \dots, N$, then it must have an equal probability of being found in any of the states b_j , $j = 1, \dots, N$ and vice versa, if it is in a state b_j it must have an equal probability to be found in any a_i .

A collection of bases will be called conjugate if each pair of bases in the collection is conjugate. We can now present a definition.

A conjugate code is any communication scheme in which the physical systems used as signals are placed in states corresponding to elements of several conjugate basis of the Hilbert space describing the individual systems. Note that in the case where the sequence of signals has more than one element, the above definition does not require the vectors describing entire transmissions to be elements of conjugate base sets. This last condition was fulfilled in the second example but not in the first.

In addition to pairs of conjugate bases, there are triplets of conjugate bases. For example, in a two-dimensional system we have

$$\begin{aligned} & \{a, b\} & |a|^2 = |b|^2 = 1 & (a, b) = 0 \\ & \{1/\sqrt{2}(a+b), 1/\sqrt{2}(a-b)\} \\ & \{1/\sqrt{2}(a+ib), 1/\sqrt{2}(a-ib)\}. \end{aligned}$$

Three such bases were used in the scheme for sending three messages, no two of which can be received.

Are there sets bigger than triplets? The following theorem shows that there is no limit to the multiplicity of mutually conjugate basis sets.

Theorem: In an Hilbert space of dimension $2^{(N-1)/2}$, there exists sets of N mutually conjugate basis sets. **Proof:** Suppose the theorem to be true for $N \leq M$. Let $\{A^\alpha\}$, $\alpha = 1 \dots M$ be a set of mutually conjugate ortho-normal basis on an Hilbert space H of dim. $2^{(M-1)/2} \equiv D$

$$A^\alpha \{a_i^\alpha\} i = 1 \dots D$$

and

$$|(a_i^\alpha, a_j^\beta)|^2 = \frac{1}{D}$$

²⁶ (a_i, b_j) is the inner product $\langle a_i | b_j \rangle$ in the Dirac notation.

for all $\alpha \neq \beta$.

We can then construct $M + 1$ mutually conjugate bases on the space $H \otimes H \otimes \dots \otimes H = H^M$.²⁷ For the first M basis, we take a natural extension of the basis sets A^α . Call $A^{-\alpha}$ the basis set of H^M consisting of the vectors $a_1^\alpha \otimes a_2^\alpha \dots \otimes a_M^\alpha, i, j, \dots \dots D$.

Note that is $\alpha \neq \beta$,

$$\left| (a_1^\alpha \otimes \dots \otimes a_M^\alpha, a_1^\beta \otimes \dots \otimes a_M^\beta) \right|^2 = \left| (a_1^\alpha, a_1^\beta) \right|^2 \times \dots \times \left| (a_M^\alpha, a_M^\beta) \right|^2 = \left(\frac{1}{D} \right)^M$$

so these basis sets $\{A^{-\alpha}\}$ are mutually conjugate.

For the last basis, we take the vectors

$$V(q, \{P^\alpha\}) = \frac{1}{\sqrt{D}} \sum_{K=1}^D e^{2\pi i \frac{qK}{D}} \times a_{K^1} \otimes a_{P^2(K)}^2 \dots \otimes a_{P^M(K)}^M$$

here, $q = 1 \dots D$ and $\{P^\alpha\}, \alpha = 2, 3 \dots M$ is a set of cyclic permutations on the integers $1 \dots D$, (i.e., $P^\alpha(n) = n + J_\alpha \text{ Mod}(D)$ for some integer J_α .) Call this last basis V .

Since there are D cyclic permutations on D integers, there are D^{M-1} sets $\{P^\alpha\}$ and $D \times D^{M-1} = D^M$ vectors $V(q, \{P^\alpha\})$ in V ; as there should be.

The proof that V is ortho-normal is obvious.

So, actually, is the proof that V is conjugate to the other basis sets, but I give it since it is the heart of the matter. Fix α and let $W \equiv a_1^\alpha \otimes \dots \otimes a_M^\alpha$ be a typical vector of A_α . Then

$$\left| (W, V(q, \{P^\alpha\})) \right|^2 = \frac{1}{D} \left| \sum_{K=1}^D e^{2\pi i \frac{qK}{D}} \times (a_1^\alpha \otimes \dots \otimes a_K^1 \otimes \dots \otimes a_{P^M(K)}^M) \right|^2$$

The inner product will be zero unless $a_{P^M(K)}^\alpha$ equals the α th term of W . (Let P^1 be the identity.) This happens for just one value of K , call it k . Then $|(W, V(q, \{P^\alpha\}))|^2 = 1/D |a_1^\alpha \otimes \dots \otimes a_k^1 \otimes \dots \otimes a_{P^M(k)}^M|^2$ where the α th vector is the same on both sides of the inner product. As for the rest, $|(a_1^\alpha, a_{P^M(k)}^\alpha)|^2 = 1/D$

²⁷ \otimes is the tensor product. $H \otimes H'$ is defined as the space of all linear functions from H into H' .

27.7 New Eye Testing Chart

Bell Laboratories, Inc.

Ortho Instruments

Bradford Howland, Antonio Medina–Puerta

Snellen optotypes have changed very little since this eye testing letter chart was introduced in 1862. Snellen letters achieved a wide and rapid success and they are now used all around the world in spite of being based in an arbitrary and inaccurate principle.

A novel eye testing chart has been developed consisting of letters (or figures) made of alternatively black and white stripes (or dots) on a gray background. Any cross section of any letter has a Fourier transform with a zero frequency component equal to the luminance of the gray background. When these letters are out of focus (or equivalently, low-pass filtered), the image of the letters on the retina rapidly fades into the gray background, rendering the letters invisible rather than simply blurred as in a standard chart.

We have optimized the operation of the chart as follows. When sharply imaged the letters are highly visible. Then, when the image is defocussed, the letters will disappear as completely as possible. Such operation would clearly permit the optometrist to separate the lines of large letters which are visible and the lines of smaller letters which cannot be seen.

Snellen's "optotypes", introduced in 1862, (Snellen, 1862) achieved a wide and rapid success. Most of the distance test charts today are modelled on Snellen's original chart. One of the disadvantages of eye testing charts using letters as test objects is the well known fact that different letters are not equally legible, (Sheard, 1921), (Le Grand and Guillemot, 1951), (Coates, 1935), (Lebensohn, 1965), (Popp, 1964). Furthermore, the relative legibilities of the various letters of the alphabet differ from one type style to another. Hence it is not surprising that the related questions of the style of type and the selection of letters to be employed have been the subject of much discussion. (See for example, Bennett, 1965).

The optotypes proposed in this paper are free from this limitation and they offer two additional properties. Firstly, the letters fade completely in the background when their retinal image is out of focus making them invisible rather than just blurred as in a standard chart. Secondly, the relation of the size of these new letters to the size of Snellen's letters for the same visual acuity is not a linear one. Instead we have found that a six-to-one size variation of the new letters is equivalent to a twenty-to-one size range of Snellen letters.

A. Description and Operation of the Chart

This chart uses letters of different size arranged in lines as in a standard chart. The letters are formed from a stripe of uniform width, having an odd number of black and white lines; the stripe is

symmetric. This is the type of letter used in our experiments, and we shall concentrate on this particular case although similar results should be achieved with a chart made with letters with a different number of elements or even with letters made of black dots on white. The condition required is that the Fourier transform of the light reflectance of a cross section of a stripe approaches zero as ω approaches zero. This way letters are mostly represented by high spatial frequencies. Since refractive errors produce an out-of-focus image on the retina, which is equivalent to a low-pass filtered image, the lack of low frequency components of our letters thus renders them invisible when out of focus.

Decomposing the cross section light reflectance function of a five-element stripe into three π functions plus a constant enables us to calculate the following Fourier transform:

$$F(\omega) = \frac{-\sin A\omega + \sin [(1 - B)\omega] - G \sin \omega}{\omega/2} + G \delta(\omega) \quad (27.4)$$

To make this function approach zero as ω tends to zero, we take the limit of $F(\omega) - G \delta(\omega)$ as ω tends to zero and set it equal to zero:

$$\begin{aligned} -A + (1 - B) - G &= 0 \\ A + B &= (1 - G) \end{aligned}$$

This equation determines the level of grey, given the width of the stripes.

The choices of A,B and hence G are determined by three additional considerations. First, we wish to minimize the values of $F(\omega)$ for small ω . Secondly, we wish to avoid very small values of A and B, since these would correspond to line widths too small to reproduce by the photographic process used to create the letters. Finally, we wish to find values of A and B closely approximated by simple fractions, so that the letter stripe can be easily specified.

Values of A, B and G meeting all these requirements were as follows: $A = 1/3$, $B = 2/9$ and $G = 5/9$. Thus, the letter stripe consists of black and white lines having relative widths of 1,2,3,2,1 units, and having a total width of 9 units.

An alternative to the above would be to directly synthesize a letter stroke which has zero response for a group of spatial frequencies. Thus, for example, we might choose a flat frequency spectrum for frequencies higher than a predetermined value and zero for the rest and transform back into the spatial structure of the cross section of the letter. This structure will not be a composition of π functions, and since it is difficult to create stripes with gradual shading, we have not presently used this approach.

The operation of our chart is quite different from that of the Snellen chart. Snellen letters or any black letters on a white background which are imaged out of focus became illegible because the blurred letter does not resemble the sharp letter any more. Our letters in no way become illegible, but

they become invisible, as defocus progresses, much earlier than illegibility is possible. Fig. 27-4 illustrates this.

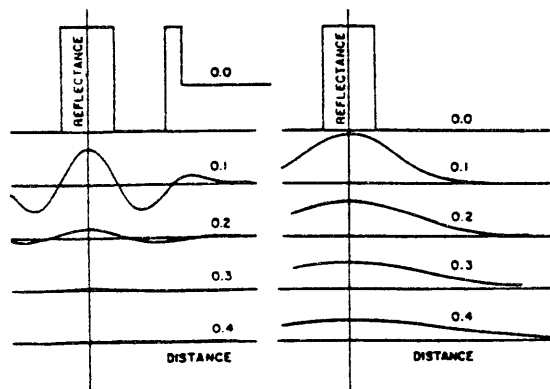


Figure 27-4: Result of Gaussian filter acting on single element and five-element letter stripes, with identical degrees of blur. The vertical line marks the center of the stripe

If we characterize the eye as a Gaussian low-pass filter, the curves on the right side of Fig. 27-4 represent the filtered image (in magnitude) on the retina of a single line for an increasingly narrow band of the filter. The curves on the left side show the image of a five-element stripe equivalently filtered. It is evident from the figure that the five-element stripe cross section becomes undetectable much faster than a single line.

B. Test of Operation with Subjects

Four subjects looked at the charts from a distance of 20 feet wearing different plus lenses on one eye (therefore converted to artificial myopes); the other eye was occluded. They were asked to read the lines in both charts. For each lens the line of smallest letters readable in both charts by the subject was recorded. This way lines of equal visibility were matched for every subject and later the values were averaged and plotted in Fig. 27-5. This figure shows the visual acuity necessary to read the letters of a determined height for the high frequency letters and for the Snellen letters. It is obvious from the plot that the new letters perform in a non-linear fashion, this means that there is no constant ratio between the size of the Snellen letters and the size of the high frequency letters for equal visibility. This result is a consequence of the peculiar spectra of the new letters and will be discussed further below.

This experiment, however, enables us tentatively to calibrate the new chart in the same way as the standard chart.

C. Discussions and Conclusions

To understand the results of comparative visibility tests of the Snellen and the new letters it is necessary to appreciate the differences in the operation of the letters. When the Snellen letters are rendered progressively more blurred, they gradually become unrecognizable as the blurred shapes of

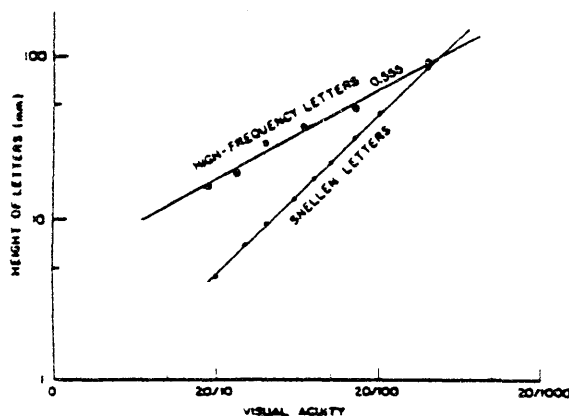


Figure 27-5: Plot of size of Snellen letters and the new letters as a function of visual acuity of subject

different letters become indistinguishable from one another. Even when the letter cannot be recognized, it is still present as a black blurred figure. With the new letters, the defocus acts to filter out the predominant frequency components of the stripe pattern, so as to render the image of the stripe indistinguishable from the surrounding grey area, i.e., the stripe is now invisible.

If one measures the predominant frequency component of the largest and the smallest of the new letters, corresponding to 20/200 and 20/10 vision, one finds spatial frequencies of 11 and 66 cycles/degree. It is known from the work of Fergus Campbell and associates (Campbell, 1968) that the eye's sensitivity to sine-wave gratings, i.e., its contrast sensitivity function varies markedly with spatial frequency, reaching a peak sensitivity at approximately 5 cycles/degree, and falling off rapidly for higher frequencies.

Thus, the smallest of our new letters, corresponding to 20/10 vision, and having 66 cycles/degree can withstand very little low-pass filtering before the predominate frequency component falls below the threshold of detectability of the retina. This, we believe, is the explanation as to why the sizes of the new letters are not proportional to the sizes of the Snellen letters of comparable visibility. Further work will be necessary to verify this hypothesis.

Although we have calibrated our new chart in the same way as the Snellen chart is calibrated, it should be noted that they do not measure exactly the same thing. The Snellen chart measures the highest frequency that the visual system is capable of detecting under the assumption that the transfer function of the subject's visual system is a "normal" transfer function, this is, the sensitivity to high frequencies progressively decreases. The Snellen chart cannot provide any additional information about the transfer function or contrast sensitivity function of the eye (Ginsburg, 1980). The new chart is made of letters of a narrow spatial frequency spectrum and, therefore, each line can be used to measure roughly the sensitivity of the human eye to that spatial frequency. Suppose, for example, that a subject could read lines 2 and 4, but not line 3 on the chart. It would suggest that his

contrast sensitivity function had a notch at that spatial frequency. Some pathological conditions of the eye have this peculiar characteristic (Walkstein et al., 1977, Regan et al., 1977), although Snellen acuity in these cases may be normal.

References

1. A.G. Bennett, *British Journal of Physiological Optics* 22, 238 (1965).
2. F.W. Campbell and J.G. Robson, "Applications of Fourier Analysis to the Visibility of Gratings," *J. Physiol.* 197, 551 (1968).
3. W.R. Coates, "Visual Acuity and Test Letters," *Trans. Institute of Ophthalmic Opticians* III, (1935).
4. A.P. Ginsburg, "Specifying Relevant Spatial Information for Image Evaluation of How We See Certain Objects," *Proc. S.I.D* 21, 3 (1980).
5. J.E. Lebensohn, "Visual Charts: Refraction," *I.O.C.* 5, 2 (1965).
6. Y. LeGrand and E. Guillemot, "Measurement of Visual Acuity with Blurred Tests," in *Opt. Devel.* 21, 2 (1951).
7. H.M. Popp, "Visual Discrimination of Alphabet Letters," *Reading Teacher*, January 1964.
8. C. Sheard, "Some Factors Affecting Visual Acuity," *Am. J. Physiol. Opt.* 2, (1968).
9. H. Snellen, "Letterproeven tot Bepaling der Gezigtscherpte," P.W. van der Weijer, Utrecht, (1862).
10. D. Regan, R. Silver and T.J. Murray, "Visual Acuity and Contrast Sensitivity in Multiple Sclerosis: Hidden Visual Loss," *Brain* 100, 563 (1977).
11. M. Walkstein, A. Atkin and I. Bodis-Wollner, "Grating Acuity in Two Sites with Tapetoretinal Degeneration," *Doc. Ophthal.* 12, 45 (1977).