EVIDENCE AND CONSEQUENCES OF FEATURE DETECTION IN

THE VISUAL PATTERN RECOGNITION OF READING

TAYLOR J. MEEK

LING407: LINGUISTICS AND READING

PORTLAND STATE UNIVERSITY

8 DECEMBER 2008

I. Introduction

Within the scope of visual pattern recognition for reading, the debate between feature and templatic theories has various supporting arguments for each (Crowder & Wagner, 1992). However, there is certain physiological and experimental evidence that suggests a feature-based system best represents the true nature of the recognition process. If this evidence holds fast, feature-based systems seem to be capable of adapting to novel situations and building up in complex ways that can be seen as precursors to a model capable of supporting efficient whole-word recognition.

At the basic physiological level, neurons in the retina have been identified which respond like feature detectors, and the structure and further assembly of those in higher-order animals hints at an adaptive means of building up a complicated feature detection and recognition process.

The nature of experiments on visual after-effects—where an extended stimulus tires certain detectors and causes a reaction to neutral stimuli—guides us to the locations of these various feature detectors. These after-effects also encourage a contrastive view of feature detection, in which each feature detector looks at a complementary pair of features and compares them to generate its output. The complexity and specificity of feature detectors is also acknowledged by research into after-effects, showing that higher-level feature detectors are capable of exacting responses to particular stimuli.

If features are indeed the underpinnings of pattern recognition, then whole-object recognition is the recognition of features in relation to each other, forming a grander schema for perception tasks to take on. Evidence supports this from experiments on physiological nystagmus, which is the rapid, unconscious, and uncontrollable jitters and self-corrections of the eye. The removal of nystagmus results in a surprising effect that gives us strong support for a part-based recognition model, but also raises some questions about meaning and top-down processes.

Having acknowledged the possibility of a feature-based system and the coalescent nature of its makeup, it is possible to look at object and whole-word recognition as a higher-level feature-detection process that may share common traits with the physiological systems it is built upon. In particular, the Pandemonium model can benefit from (and possibly be expanded by) these findings. The idea of a variable ranged activation of the neuron or basic feature detector encourages Pandemonium's theory of dynamically shouting demons. Even so, after-effects may pose a problem for studies showing that re-stimulation by a particular, or semantically related, previously stimulated item results in a faster response due to context-based lowering thresholds of activation (Crowder & Wagner, 1992).

II. Neurophysiology of Vision

Vision is first stimulated on the photoreceptors of the retina, which are specialized neurons that transmit chemical signals in response to light. Neurons all receive signals from other cells, responding by sending out a constant base signal to other neurons that can be increased by receiving excitatory signals and decreased by receiving inhibitory signals. Photoreceptor cells in the eye occur in two types: rods, which are highly sensitive to light but not color, and cones, which are color sensitive and concentrated in the fovea. (The fovea is the central area of the retina containing only cone cells and is responsible for the sharpest vision.) Cone cells are also separated into three types, each of which is most sensitive to a particular wavelength of light, approximating to red, green, and blue.

Photoreceptors then combine together by a series of other cells, including bipolar cells, which excite in response to one cell or inhibit in response to another. The signals from the bipolar cells then travel through a network of ganglion cells. (Ganglion cells are neurons with axons several inches long which carry the signals from the bipolar cells out of the eye to the brain, forming the optic nerve.) The visual cortex is the rear portion of the brain primarily responsible for more deeply processing signals from those cells.

III. Evidence About Feature Detectors

1. After-Effects of Difference Detectors and Complex Feature Detectors

There is a quirk of vision reported by Lindsay & Norman (1977) that gives us significant insight into the process of pattern recognition. If a subject looks at a pattern of red-colored squares for a minute or so, then immediately looks at a white surface, they then perceive the complementary (green) color in that same pattern in the absense of any stimulus that should cause such a response. This response is termed an *after-effect*, and is the result in the preceding example of the tiring of certain retinal cones responsible for color perception. Since the stimulus was exclusively red, this particular experiment would have the effect of tiring (or acclimating) only cones sensitive to red wavelengths. The resulting green after-effect arises from the interconnections of the bi-polar cells and cone cells.

Physiological color vision in humans (and several other mammalian species) works by the comparison– through a specialized bipolar cell termed a *difference detector*–of the levels of activation for the three types of cone cells. Red and green cells are compared, and blue cells are compared to yellow (the comparison of red and green). In a neutral environment without any preceding stimulus, both red and green cells in that complementary arrangement would output the same *baseline signal*, and the difference detector that compares the two would emit its own baseline signal. With red stimulated, however, red cells begin exciting and outputting at a higher level than the baseline signal that the green cells output, which stimulates the difference detector into outputting above its baseline. Over time, the red cells will begin to tire from the red stimulus, and when that stimulus is removed and replaced by a neutral stimulus, the red cells emit a signal that is less than their baseline while green, which has not tired, emits its regular baseline signal. The result is that the difference detector is less activated by the tired red cells but not inhibited by the green cells so that it perceives green as having a higher signal than red, and so outputs a diminished signal itself, represented visually as green.

This effect, however, is dependant on the eye: When only the left eye is acclimated it will experience the effect, but the right eye will not, showing that this after-effect is physiological. The acclimation occurs only in the eye that is stimulated, but not at a cortical level that would transfer the effect to the other eye. On the other hand, certain after-effects do transfer between the eyes, which indicates a cognitive acclimation as opposed to a physiological one.

In another experiment, a subject is presented with a pair of stimuli–one of green horizontal bars on a black background, the other of red vertical bars on a black background–alternating every 5 seconds over the course of several (at least 5) minutes, and then presented with a stimulus (Fig. 1, right) of black bars on a white background in horizontal and vertical orientations. When the test stimulus is presented, the subject perceives the white horizontal bars as somewhat pink and the white vertical bars as somewhat green. Perception of diagonal lines is unaffected, but turning the head 45 degrees will result in the effects exhibiting relative to the perceived
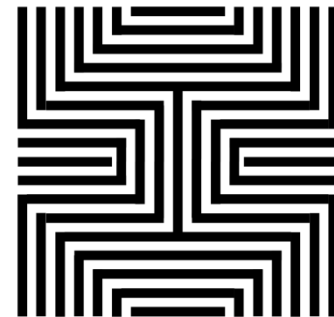
(Gibson & Harris 1968 in Lindsay & Norman 1977)
Fig. 1

orientation. This shows that at some level of processing, the detectors are sensitive not only to color but to lines in particular orientations. What's more, this effect also transfers between eyes: when the left eye is stimulated, the right eye will experience the after-effect. This not only shows that the detectors are more specialized than to just detect color, but also that they exist within the visual cortex as opposed to the retina.

Orientation of lines is not the only feature that affects acclimation of complex detectors. Observing water falling downward for a period of time followed by a neutral surface will result in an apparent after-effect of motion: an after-effect will appear, moving upwards in contrast to the original stimulus. This shows that the complementary difference detector nature of vision not only applies to the opponent process of color, but also to deeper cognitive processes, of which motion may be a member since some research (Lindsay & Norman, 1977, p. 49) has shown that motion after-effects transfer between eyes.

Lastly, these features can combine in complicated ways; subjects were presented with horizontal lines alternating moving vertically up and down. The direction of travel determined the color of the bars: when moved up they were presented in one color (i.e. green), and when moved down, presented in the complementary color (i.e. red). The testing stimulus following the acclimation period was a series of black-on-white bars moving upwards and downwards. The after-effect held according to direction of travel, showing the complementary color to each direction, which also shows that detectors in the cortex can be sensitive to color as well.

These after-effects give us evidence for three things:
1) Detectors function by comparing more than one input, calculating the difference, and outputting the relative difference of the two, irrespective of acclimation.
2) Fatigue transfer demonstrates the general site of detectors within the visual processing system.
3) Detectors are specialized for particular combinations of features describing a stimulus.

2. Physiological and Cortical Feature Detectors

After-effects show us that detectors that are capable of responding to particular (simple or complex) features of a stimulus apparently exist in both the retina and visual cortex. Retinal dissections and stimulus/electrode-response experiments, into the physiological construction of these systems confirm this (Lindsay & Norman, 1977). Some feature detectors sit at the physiological level, closely tied to the photoreceptors in the eye. These basic detectors are inflexible to change and do not allow adaptability. For example, the eye of the frog contains detectors (Fig. 2) in the retina responsible for edge detection, moving contrast (edge) detection, dimness detection, and convex edge detection. These correspond to some basic behavioral patterns in frogs, and also shows their inadaptability.

Edge detectors respond most to a difference in contrast from one side to the other, for example the edge of a platform.

Fig. 2 - Edge, Convex Edge, Moving Contrast, and Dimness Detectors in the Frog

Moving contrast detectors look for the movement of these edges, irrespective of whether it is the animal moving or

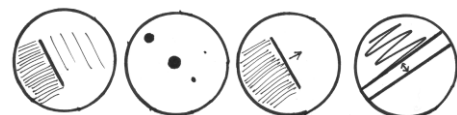the stimulus. *Dimness detectors* respond to changes in light, such as those from shadows of predators. Lastly, *convex edge detectors* respond to convex shapes (round objects) moving towards the focus of vision of the frog. These last detectors, in particular, are dedicated "bug detectors", responsible for the tongue-darting reaction of a frog to bugs entering their vision. Lindsay & Norman (1977) point out, however, that this particular detector is not very well suited to change; In an environment with many flying bugs, the frog will thrive, whereas in an environment with piles of dead, unmoving bugs, the frog will starve to death in the absence of the stimulus which enables it to hunt.

However, higher-order species have the ability to abstract their environment using powerful, yet simple, detectors to provide the same possible skills in an adaptive way. This begins at the level of the retina with *ganglion and bipolar cells* which in mammals, unlike frogs, comprise a series of detectors that form the building blocks of perception that are combined in neural processing to enable those species to respond to more complicated stimuli. Some of the abstract cells in these species include on-center/off-surround cells, which excite in response to light directly in the center of their aggregated photoreceptors and inhibit in response to light surrounding that spot. The opposing feature is also present in off-center/on-surround cells, inhibiting in the center and exciting in the surrounding cells. At this level, cells also exist that respond to light onset and extinction, color specificity, contrasts of light, and directional movement.

These basic features are transmitted to the visual cortex, which builds them into *simple cells* that form the same sort of detectors seen in the frog, such as edge, slit, and line detectors. They are all selective as to location, type, orientation/direction, and width. This means that a given feature detector will only activate when all of those constituent features match its criteria. From there, these feature detectors are further combined into complex cells which generalize the presence of a feature unrelated to its position, activating based on the presence of a feature of a particular orientation and width within a fairly wide field of vision.

Hypercomplex cells are built on top of the complex cells which respond to several more complex criteria:

▷ Termination/Size: A line not properly terminated within the responsive field of one of these cells will fail to activate the cell. It also must present enough stimulus (size, width) to activate. The termination may also only be specific to one endpoint, meaning that a detector may only respond if a particular line terminates at one end within the responsive field and continues beyond it from the other end.

▷ Movement Generality: Movement can occur in either direction along a given axis and still activate a cell.

▷ Angle: A given cell might activate most strongly from a right angle as opposed to one wider or narrower.

As these detectors build upon one another to create more specialized detectors, it takes more specific stimuli to evoke a response, but the location of that stimuli matters less. Eventually, this leads to highly specific detectors. For example, through neuron electrode experiments in monkeys, a particular detector cell was discovered that activated the most from an upturned monkey paw. Additionally, this same team found cells which responded to hemostat forceps and bottle brushes, which provides evidence that suggests these feature detectors are capable of learning to respond to new stimuli and adapting to the environment in the presence of enough novel input.

These *higher level detectors* show the foundations of some of the complicated relations of features that formulate larger objects and, in the context of written language, letters and words. What is left is a question of those relations: How do we perceive these unrelated features in our visual field as being related in the form of a whole object?

4. Nystagmus and Object Chunking

The eye has a unique characteristic about it that may provide evidence that we perceive objects in significant chunks, according to Lindsay & Norman (1977) following experiments by Riggs et al. (1953) and others. While fixated on a specific point, the eye can be observed moving in erratic and unconscious ways, though tiny; the smallest are as little as $\frac{1}{180}$ of a degree. These movements (Dodwell, 1970) appear (visually represented in Fig. 3) in several types:

1) Tiny and fast, only $\frac{1}{180}$ of a degree, 30-70 times per second.
2) Large oscillatory.
3) Slow drift, about $\frac{1}{20}$ to $\frac{1}{10}$ of a degree.
4) Rapid saccadic jerks, about $\frac{1}{12}$ of a degree, which correct the others.
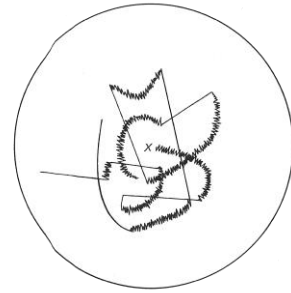
(Dodwell, 1970)

Fig. 3

The eye is constantly in a state of movement even while fixated upon a single point. This movement is referred to as *physiological nystagmus*. This contrasts to a *pathological nystagmus*, uncontrollable side-to-side eye movements, typically congenital or acquired from disease or damage, which can result in visual impairment (Kantor 2007). It is apparent from this that normal vision exists in the presence of these unconscious movements, but vision impairment appears when it becomes aberrant. What happens, then, when these normal movements cease, which is an impossible conscious task?

Since it is impossible to freeze the movements of the eye, it is necessary to match the movements to the image being presented. This was done in 1953 (Riggs, et al.) by affixing a mirror to a contact lens on the eye and projecting the image onto that mirror, which reflects it through an apparatus of mirrors that projects the image to be viewed onto a screen in front of the subject. The outcome after calibrating the system is that the eye is forced to view the same point of an image no matter where it gazes. What subjects reported seeing was that the stimuli in the images began to disappear, eventually leaving nothing but a homogeneous field. The images disappeared in a particularly interesting way (Fig. 4): Meaningful chunks would disappear and reappear all at once while the rest of the image would remain. The portion of the image under the fixation point lasted longest of all, as did more meaningful whole images (such as a face or word versus a field of erratic lines).
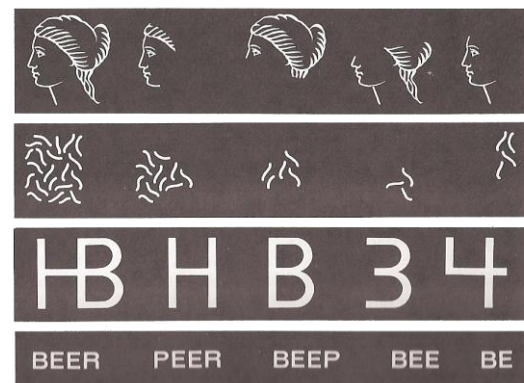
(Pritchard, 1961 in Lindsay & Norman, 1977)

Fig. 4

From this, a few things can be gathered. Firstly, an involuntary and small saccadic process is a necessary component to vision at a level lower than the saccades seen in reading. When the light of a stimulus no longer shifts around the retina, the image disappears completely in a

way analogous to acclimation in color-affects. If what are being stimulated are some sort of feature detectors, then the loss of active stimulation causes them to cease their reactions, and so contribute nothing to the meaning of the image. Following that, if those feature detectors are no longer receiving their physiological stimulation, it appears that for more meaningful images we are receiving some latent stimulation from somewhere else which causes the images to persist longer than the less meaningful images.

Secondly, we get a very important clue about how perceived mental images are represented since they vanish in chunks at all. If we perceived the image of a word as a whole without regard to its individual features, we would expect it to disappear as a whole. But since we lose chunks but retain meaning (even if it is erroneous), it appears that our attention to an image relies on both features and the way those features interact to describe objects. When those features are no longer stimulated, they can no longer contribute to the image at large, causing whatever portion to which they contribute to fail and vanish. In much the same way that a pair of line detectors intersecting at a common point in a particular angle represent a vertex, a pair of vertices in a particular relation to one another may represent the terminated edge of a square.

When no representation in the mind of any known object can match the relations of the features stimulated, as in the case of a field of erratic lines, the only activation that occurs is that of those features. On the other hand, when some internal representation is located, it too is stimulated. Thus we see the field of erratic lines vanishing much more quickly than a face since the elements of the face have been called from memory, and so that memory representation is possibly cross-stimulating the higher level feature detectors that generalize that face even after the stimulus is gone.

Various theories of *feature classes* have been surmised, including one by Biederman (1995), which sets out to describe a variety of types of features, such as lines, vertices, parallelization, and symmetry. Biederman goes on to describe a feature class based on geometric shapes, *geons*, which are described in terms of the relations between the basic features, forming more specific but still generalizable forms (such as cylinders and cubes).

Objects are then represented as the relations between these geons, such that a coffee mug is the orientation of a narrow curved cylinder on the side (curved surface) of a squat cylinder, whereas a bucket is the same squat cylinder topped with the same narrow curved cylinder. Furthermore, it is the relations of those complex geometric shapes that produce scenes in the real world. This theory of geons appears to follow the previous idea of features building upon each other to construct more specific but generalized representations.

IV. Physiology, Features, and Pandemonium

The Pandemonium model describes a method of matching stimuli against internal representations of previously-encountered stimuli. Crowder & Wagner (1992) describe a model of letter matching, feeding the stimulus of a letter into a pool of feature demons who each examine the letter in parallel, and shout in relative loudness to their sureness of their feature being there. Cognitive demons representing previously-encountered stimuli listen for the set of feature demons that describe them, and shout in relative loudness to the sureness that the features out there are theirs. Finally, a decision demon listens for the loudest of the cognitive demons, ideally finding one to represent the stimuli with.

The research set out in previous sections contributes several things to the Pandemonium model. Firstly, Pandemonium is more accurate if a model of combinatory activation and inhibition similar to that in physiology is accounted for. Secondly, the built-up nature of feature detectors and the adaptiveness of higher level detectors shows that Pandemonium may be able to account for efficient whole-word recognition.

We saw that objects may be recognized as the relations between their features. However, if each cognitive demon activates based solely on the number of matching features it hears, ambiguity will occur. If the letter P is shown, (1) vertical line and (2) horizontal line features will both be stimulated as well as a curved line. In total, we could imagine 4 features being stimulated. Since the cognitive demon for P is looking for all of those features, it will shout loudly and likely win. However, if the letter F is presented, 3 of P's features will show up, and all 3 of F's

features will show up. If the cognitive demons for these two letters are shouting based on how many features they are matching, both will shout as loud as the other, and the decision demon will be unable to differentiate them. At the physiological level, difference detectors combine the inputs of their inhibitors and activators into a range of output, and we might expect the same characteristic to show up in deeper detectors. In this case, all cognitive demons have a standard maximum activation level; if all and only all of a demon's features are found, it will activate at the highest level possible of any demon, and no other demon will be capable of activating that high. Superflulous features will inhibit a demon, and missing features will fail to activate it as high as it can go.

Pandemonium can also adapt as the cognitive demons learn to interpret the various features associated with their particular patterns, such as the word THE, based on repeated stimulation. Studies of word versus letter recognition speeds (Crowder & Wagner, 1992) show that whole words are recognized just as quickly as, if not quicker than, their constituent letters. Studies of eyemovements and letter search tasks within words such as THE show that words can be recognized as whole units irrespective of their constituent letters. Fitting these facts into Pandemonium requires a very simple change: Enable the cognitive demons to form based on higher-level features comprised of lower-level features in relations to one another and you have enabled them to identify words alongside letters.

Since the maximal activation hypothesis carries over into this new hypothesis of cognitive demons seeking out higher-level features (i.e. letters or morpheme-like chunks), we would expect to see cognitive demons for words activating at their maximum when attention includes an entire word. Conversely, we would also expect to see that the cognitive demon for a constituent letter of an attention-given word would be inhibited by the presence of additional feature data from surrounding letters, and so would not activate as highly as the cognitive word demon. This would explain why it is difficult to locate the letter H in instances of the word THE in a body of text: the word THE has a specialized cognitive demon that shouts as loud as it can when it recognizes the whole pattern of THE, but the demon for the letter H clams up since the letters T and E make it second-guess itself.

After-effects pose a problem for studies showing that re-stimulation by a particular, or semantically related, previously stimulated item results in a faster response due to context-based lowering thresholds of activation. We would expect to see that previously stimulated items would tire from that stimulation, and so inhibit the ability to stimulate them or related items later. This is not the case as suggested by studies (Crowder & Wagner, 1992) which actually claim just the reverse, and the knowledge gleaned from physiological nystagmus might shine some light on this fact. Since more meaningful stimuli take longer to degrade in nystagmus-inhibition experiments, it is implied that what remains once the physiological stimuli no longer assist the task is the deeper representation of meaningful parts and whole objects that stimulates the recognition. Even after a stimulus is gone, its effects linger from the higher levels of processing in the mind, and activity from those levels may stimulate subsequent tasks by providing a stimulation of context from within the mind itself onto new stimuli.

V. Conclusion

Research into physiological feature detectors, after-effects, and physiological nystagmus provides promising data for refining the Pandemonium model. Dissections and cortical electrodes give evidence for the existence of adaptive and built-up feature detectors. After-effects show that detectors are comparative, specifically tasked, and localized to either the retina or visual cortex. Nystagmus shows that it is the relations between features that describe larger objects and that meaningfulness and top-down processing plays a critical role in all recognition tasks. All of these feed into the Pandemonium model, allowing it to account for the selection of precisely matched stimuli instead of stimuli with too many or too few matched features, for word superiority effects and adaptiveness in whole-word recognition, and for a possible link with context effects and top-down processing.

References

Biederman, I. (1995). Visual object recognition. In L. R. Gleitman, & D. N. Osherson (Eds.), *An invitation to cognitive science* (2nd ed., pp. 121). Cambridge, Mass.: MIT Press.

Crowder, R. G., & Wagner, R. K. (1992). *The psychology of reading :An introduction* (2nd ed.). New York: Oxford University Press.

Dodwell, P. C. (1970). *Visual pattern recognition.* New York: Holt, Rinehart and Winston.

Kantor, D. (2007). *Eye movements - uncontrollable.* Retrieved 12/6/2008, from http://www.nlm.nih.gov.proxy.lib.pdx.edu/medlineplus/ency/article/003037.htm

Lindsay, P. H., & Norman, D. A. (1977). *Human information processing :An introduction to psychology* (2d ed.). New York: Academic Press.

(Sidebar: It would be interesting to examine the results of context effects between media as well. If one word can help the recognition of another word, and if the image of an ethnically-identifiable face can help the recognition of a positive or negative word, can the stimulation by the written word speed recognition of the spoken word or vice versa? How would nystagmus-inhibition be influenced by the spoken word? Could it possibly help us determine whether context effects lie in the level of features or in the level of ideas?)