PROCESSES IN BIOLOGICAL VISION:
including,

ELECTROCHEMISTRY OF THE NEURON

This material is excerpted from the full β-version of the text. The final printed version will be more concise due to further editing and economical constraints. A Table of Contents is at the end of this paper.

James T. Fulton
Vision Concepts

jtfulton@neuronresearch.net

April 30, 2017 Copyright 2008 James T. Fulton
19 Reading, listening and Other Complex Operations related to object recognition

[xxx move parts of earlier discussions re the operation (but not the mechanisms ) of interp and percept extraction to this chapter ]

[xxx review the multiple chapters of Gazzaniga (2004) related to reading and listening.

This chapter will address the most complex operations associated with vision. These include the mechanism for moving “targets” into the line of vision, the mechanism for tracking moving targets, and the mechanism for extracting interps and percepts from the images of the targets. Much of the material has a parallel within the auditory and other external sensory environments as well. However, different terminology is frequently encountered in these other modalities.

Reading is logically a sub-category of the object recognition task. However, it will be discussed first because it involves a more structured operating format that has been studied extensively. Thereby, reading provides considerable data related to the more complex generic task of object recognition.

Dehaene has written a book on reading for a general audience. It is quite useful for the non-researcher but contains little advanced material. Dehaene, although well known in European pedagogy, has presented a book from the perspective of a psychologist surveying the work of others. His first chapter with a subtitle of “The eye: a poor scanner” emphasizes his narrow perspective related to reading and totally overlooks the physiology of the human visual system. From this broader perspective, the eye exhibits a truly amazing optimization to satisfy a wide range of requirements associated with the environmental niche of the species. Dehaene does not delve into the details related to the physiology of reading and appears unaware of the role of tremor in vision. The term does not appear in his index.

The few numbers provided by Dehaene are ballpark. While presenting considerable material derived from PET, MRI and fMRI experiments, he is unable to say precisely what is going on in the areas of the brain highlighted, or what specific information is being processed. He provides little information related to ocular scanning during reading and does not address the role of tremor (or microsaccades) in reading.

While suggesting a 15 degree diameter region of the fovea, roughly 2.5 inches wide at 15 inches, is involved in reading (page 13), his examples on page 14 show a region of high acuity closer to the 1.2 degree diameter of this work (one-half inch wide. The word “read” at 15 inches extends over less than 2.0 degrees). His internal section title, “The Myth of Whole-Word Reading,” page222, addresses a false dichotomy. It also illustrates the chasm between the psychology community and the neuroscience community (based more firmly on physiology). He considers only individual letter-based reading and whole word-based reading when ocular studies clearly show that syllable reading is highly significant for words in context of more than about 5 letters. He does address the subject of syllables and morphemes on page 22 from an awkward perspective. While his semantics are highly readable, his discussions would be well served by a variety of block diagrams and flow diagrams that could be subjected to falsification. Such falsification would lead to new alternate diagrams that would move the subject forward. His chapter 2, regarding the performance of “Mr. C” provides useful information that helps delineate the CNS engines related to reading and reporting by the subject on what was read. This material, and other material from his book, will be incorporated into the diagrams of Section xxx. of this work describing the top level operation of the CNS and Chapter 18 of “Processes in Biological Vision” describing abnormalities of the visual modality of the CNS. It is clear that Mr. C suffered damage to the input circuits of the stage 4 engines assigned to the “association of information” extracted by the specific engines of the sensory modalities. It is not clear that he suffered damage to the later portions of the stage 4 engines at the junction of the “left ventral temporal-occipital” lobes of the CNS.

---

1Released April 30, 2017
Dehaene has provided a new book in 2014 covering the role of consciousness related to the brain. The book is aimed at the popular press but does provide a variety of useful definitions that will be considered in Section 18.1.1 of ‘The Neuron and Neural System’ by this author.

This chapter will rely heavily upon Sections 7.3 – 7.7 of Chapter 7 and Chapter 15 to provide the physiological and neurological model that the mechanisms of reading and object recognition depend upon. The figures in Section 7.4.1 are particularly pertinent to the following reading discussion. They focus on the direct paths between the thalamus and the parietal lobe (or cerebellum) and the minimum length paths within the precision optical system that make reading a unique capability found (so far) only in humans. In this respect, it is important to distinguish between recognizing simple pictographs as usually used in primate research and the level of fine detail associated with perceiving a specific word among other similarly spelled words.

The description of the horopter in Figure 7.4.1-14 describes the spatial limits of human visual acuity in object space. These limits proscribe the ability of the visual system to interpret scenes without version (motion) of the eyes relative to the scene.

This chapter will also rely upon Section 7.5 to set the stage for understanding the reading mechanism in humans. It is clearly the culmination of the animal ability to interpret global bucolic scenes followed by interpretation of more specific scene elements with the goal of identification of specific objects and finally the ability to interpret small highly structured scenes (and sequences of such scenes) into very complex abstract concepts independent of any residual structure in any domain.

To understand the above mechanisms and the level of performance they can achieve, advanced concepts of Signal Detection Theory must be addressed. The text by Green & Swets, although employing the Calculus in very sophisticated mathematics, is an important source of background for this chapter. Of particular importance is the discussion about the 1 of M, or M-ary, symbol detection technique and the signal-to-noise performance achievable with this technique. This technique is key to understanding the operation of all of the sensory modalities and not just vision. Another area Green & Swets addresses is that of the critical band. It will be shown here (xxx or in Section xxx) that the critical band applies to the spatial filtering (rather than frequency filtering) associated with the correlation mechanism within the TRN. This technique is also used generally in the sensory portion of the neurological system.

After discussing reading and listening (fundamentally one dimensional or “linear” signal and information extraction procedures, the Chapter will proceed to examine the object (and particularly face) recognition procedures employed by the neural system.

19.1 Overview of the general process of perception and cognition

When discussing perception and cognition within the neural system of biology, it is important to note that it involves two distinct pathways. The first pathway relates to the perception of the overall visuotopic scene available to the visual modality and will be labeled the awareness pathway. It will be associated with the concept of the where task, where things of interest are in the visual field. The second pathway is related to the high performance region of the visual modality typified by the small area centered on the line (point) of fixation and called the foveola. This 1.2 degree diameter foveola is primarily associated with identifying what individual elements of interest in the wider scene are, the what task. It will be described more fully under the label, the analytical pathway. Earlier works focused on object recognition associate the awareness pathway with “early” visual processing and the analytical pathway with “late” visual processing.

Garcia-Perez & Sierra-vasquez employed the terms early and late in their modeling activity after noting in 1995, “the present state of visual modeling is rather unsatisfactory.” They assumed the visual modality could perform Fourier transformations without any confirmation or citations. Peli, in a paper within the larger volume he edited (page 63) identifies a group of nonlinearities that make simulations based on linear...
filters awkward but proceeds to assume a spatial frequency environment within the human visual modality that cannot be justified based on the more recent literature. While his work does apply to optimizing displays to be examined by human eyes, it does not reflect the processing performed within the visual modality.

Lamme & Roelfsema provided a significant paper in 2000. Its introduction closely follows the framework of this work as it relates to the various perspectives of the visual process developed here based on the flow diagrams developed in Section 19.10.4 and in Section 4.6.3 of “The Neuron and Neural System.” They introduce three dichotomies in their introduction; 1) spatial and object vision, 2) attentive and pre-attentive vision and 3) the differences between conscious and non-conscious visual processing. While not using the term “declarative” as used here with respect to both memory and the description of the data stored in such memory, they assert humans “experience vision mostly through our conscious percepts.” Their figure 1b develops the star-network character of the neural system within the CNS. However, their figure 1c is oversimplified; there is no requirement that signals from V1 must pass through V2 to reach V4. They note these and other challenges in their Box 2. They also note the presence of other paths, such as via the brachia of the superior colliculus (the PGN) and the pulvinar as developed in this work.

In 2005, Epstein & MacEvoy have suggested an additional aspect of the scene recognition versus object recognition tasks. They note the ability of a person to perceive a previously perceived environment (such as a room with furnishings) as a whole almost instantaneously. It is obvious this early perception occurs before the brain has had time to analyze the information acquired during the initial (early) gaze on the scene and long before any precision analyses has occurred via the analytical (late) pathway of vision. Their observation and discussion appear to provide an answer to the question, “what purpose do the other areas of the occipital lobe, V2, V3, etc., serve?” Based on the histology, traffic analysis of the CNS, along with voltage evoked potential (VEP) data collected over many years, it appears these areas perform a low resolution but fast interpretation of the visuotopic scene as a whole based on previous observations of the similar scene (even from a significantly different aspect angle). Section 15.2.5, developed earlier, is consistent with this interpretation. It will be developed in greater detail in Section 19.10.5.2.

Epstein & MacEvoy proceed to elaborate on the differences between scenes and objects from their perspective.

[xxx add more background material ]

Shipp has recognized the major importance of the thalamus. Unfortunately, his global view did not recognize the shell covering the pulvinar as the critical TRN. Shipp proposed that the pulvinar itself acted as a “remote hub for coordinating spatial activity within multiple cortical visual maps.” Chalfin et al. give additional citations supporting this view and suggest, “At a minimum the lateral posterior pulvinar complex thus serves the brain to identify and focus attention on a particular object or task.”

[xxx mention underwood and bouma ]

Bouma has provided a good tutorial review of the processes of centering objects of attention on the foveola and the gross motions associated with reading and other tasks as of 1978. His work is discussed in Section 19.8.6. Underwood has presented considerable material on the operational aspects of reading and cognition of scenes in general (See Sections 7.5.3 and 19.8.2).

The major role of both declaratory and non-declaratory memory in the object recognition is not appreciated by most academics and behavioral laboratory researchers. It is critically important that the protocols used in behavioral research control for memory utilization.

---

19.1.1 General perception of a scene—role of the awareness channel

The awareness pathway of biological vision consists of the full retina, the lateral geniculate nuclei of the midbrain and multiple stage 4 information extraction engines of the occipital lobe of the cerebrum (or cerebral cortex). It will generally be labeled the LGN/occipital pathway. The awareness pathway employs aggregated sets of stage 1 sensory neurons processed by the stage 2 signal processing neurons to provide a high sensitivity to changes occurring anywhere in the visuotopic scene imaged on the retinas. The resolution of the awareness channel is much less than that suggested by the density of sensory neurons due to the aggregation process. The actual resolution is limited by the number of distinct neurons passing through the optic nerves from each ocular to the stage 4 information extraction engines of the CNS.

The stage 2 signal processing neurons are aggregated using a winner-take-all strategy to insure that a distinct change in stimulation applied to a single sensory neuron will be relayed to the CNS as a significant change at the where location within the field of view. This will cause the CNS to focus its attention on this location within the field of view in order to perform the what task discussed below.

19.1.2 Detailed perception of a scene—role of the analytical channel

The analytical pathway of biological vision consists of the small foveola (circular region of only 1.2 degrees diameter centered on the point of fixation), the perigeniculate nuclei of the midbrain and the pulvinar of the midbrain. The pulvinar appears to rely upon its close association with the visual portion of the cerebellum for long term declarative memory storage purposes. Thus, this pathway will generally be labeled the PGN/pulvinar pathway with the understanding that a more appropriate label might be the PGN/pulvinar/cerebellum pathway.

The foveola consists of only about 23,000 sensory neurons in an array with a diameter of about 175 pixels. The sensory neurons are packed into a hexagonal array with a seventh receptor in the center of each hexagon. Among some in the histological community this has been identified as 7-connectivity.

The primary task of the PGN/pulvinar pathway is to extract information about and appropriately label elements within the field of view of the foveola following a series of saccades based on the where information extracted by the LGN/occipital lobe pathway above. The resulting labeled information is deposited in the stage 4 saliency map that can be interrogated by the stage 5 cognitive engines in order to determine a plan of action if required. The PGN/pulvinar relies heavily on the learning cycle described below and the memories stored previously in the cerebellum to accomplish its mission.

19.1.3 Detailed perception of a scene—role of the Precision Optical Servomechanism

The role of the ocular nuclei, controlling the pointing of the eyes, has not been widely studied and placed in the proper context in order to recognize their important function (particularly at the micro-saccade level of activity). The ocular nuclei are a part of the Precision Optical Servomechanism (POS, Section 7.3) of the visual modality. It is responsible for two major tasks. The first is the conscious pointing of the eyes in the direction specified by the stage 5 engines of cognition. The second task is the more important one from a functional perspective. It is the non-conscious pointing of the eyes to areas of interest determined by the alarm mode of the awareness pathway as well as the more routine semi-coarse pointing of the eyes to elements of the visuotopic scene identified by the awareness pathway, and finally the very fine pointing of the eyes involved in extracting the what information concerning individual elements within the scene projected on the foveola. These last motions are generally below the instrumentation level typically used in the behavioral performance laboratory of typical university psychology departments. They are at the micro-saccade and tremor level of eye motion (Figure 7.3.2-1).

19.1.4 Important and specific terminology in perception

It is critically important to define a set of terms applying to steps in the perceptual processes independent of the sensory modality involved. Tentatively, and specifically in this chapter, this work will use the following sequence of steps in the context of the vision modality. Note carefully, Figure 19.1.4-1 shows the prescribed set of terms can be implemented or arrived at via multiple paths.

Discovery, step 1: discovery first occurs prior to any perception in the signals generated by the sensory neurons. The signals are initially passed up the signal pathways associated with alarm and awareness. Typically the alarm mode is not activated unless the signals from the stage 1 circuits indicate an overload
of the circuits or other alarm situation.

**Recognition**, step 2,

**Identification**, step 3;

**Reporting**, step 4;

It is important to note, experiments can be arranged so that a report is issued (frequently a key press) before the complete recognition and identification of an event occurs.

### 19.1.5 A The routine versus learning states of operation (old 15.4.1)

[xxx may want to abbreviate this section and move part concerning reading to follow 15.6.6 & 15.6.7]

Each of the couples defined in Section 15.6.5 [xxx out of order] can perform in conjunction with two different state diagrams. In the routine case, any signal sensed and reported to the old brain will be interpreted and reported to the appropriate portion of the neo-cortex as a percept according to the routine state diagram for that modality. However, whenever a signal is sensed that cannot be interpreted by the appropriate couple, a supervisory signal will be sent to the TRN. The TRN will in turn change the operating mode to follow the “learning” state diagram. This state diagram is more involved and employs more elements of the CNS. The resulting procedure is slower in execution but it results in a new interp being stored in the appropriate memory element, and being available for future interpretation via the routine state diagram.

#### 19.1.5.1 The generic aspects of learning and routine operation EDIT

[xxx edit out ref to cerebellum specifically]

Based on this work, a critically important role for the pulvinar can be defined. That role is one of a major memory function (a very large lookup table). Its primary responsibility is to provide a translation service between low level and high level signaling languages. The role is very similar to that used in modern computers where a translation table is required between machine language (ones and zeros) and the stilted English of higher level languages like C++. Such a table is frequently required to be bi-directional. This is usually achieved by using two separate sections of the memory to perform the translations in different directions. In the visual system, the pulvinar has the additional task, in conjunction with other elements, of learning the required translation routines initially.

To accomplish both the learning and the translation of the appropriate signals, the pulvinar can be considered a special type of memory device called a write once, read many times (or read only) memory unit. At the technical level, such a man made device is called a PROM, a programmable read only device. The use of such a device can greatly reduce the computational requirement associated with routine tasks.

In the case of vision, the main role of the pulvinar is proposed to be the conversion of a mid-level signaling language generated by the two-dimensional correlator of the pretectum (expressing interps) into a higher level language (expressing percepts) that is then transmitted to the cerebrum for high level cognition. This role is suggested by Noback when he observes on page 146, “there are approximately three times as many
6 Processes in Biological Vision

cerebellar afferent fibers as cerebellar efferent fibers.”

In its role in oculo-motor and skeletal-motor operations, it plays the opposite role. It converts high level language signals from the cerebrum into mid level language signals that can be processed by the various motor neuron nuclei into the low level signals required to operate the various muscles. The conversion from a high level to a mid level language introduces a greater degree of detail into the commands. This level of detail results in both smoother and more highly coordinated movements.

The question remains, how does the pulvinar become programmed? A proposed method is illustrated by the state diagram of Figure 19.1.5-1. The diagram illustrates two main paths. The upper path is used when the pulvinar fails to recognize a signaling sequence sent to it. This initiates the complex sequence required to learn the meaning of the sequence. The second is the simpler case where the pulvinar recognizes the signal sequence, based on prior learning, and issues the appropriate signals. These commands are typically condensed and higher level in the case of vision, and expanded lower level commands in the case of motor operations. Only the vision case is illustrated.

Recent research is suggesting that the role of the parietal lobe in this figure (Brodmann area 7) should probably be replaced by the association areas of the cortex (in the vicinity of Brodmann areas 21, 22, 37 and 39) and the cerebellum as the primary location of long term declaratory memory connected with vision.

It is noteworthy that if the pulvinar is damaged following learning, the system can revert to the direct mode of connection between the pretectum and the cerebrum used during initial learning. The system will operate slowly, but it will operate. This mode is characterized by inadequate coordination in skeletal-motor activities. These motions are frequently tentative, not unlike those of a very young child.

**Figure 19.1.5-1 Candidate state diagram describing the operation of the pulvinar and thalamic reticular nucleus.** Moving left to right, the information density rises as the data volume decreases. The information loses all spatial correlation to the scene when converted to an interp. A “no” decision by the pulvinar is probably hierarchal. For efficient mode control, a “yes” decision is probably sent to the control point as well. See text.

The pathways shown in this figure are part of the analytical mode of vision. Physically, they utilize the Pulvinar pathway and do not involve areas 17 through 22 of the cerebrum.

**19.1.5.1.1 “Experienced” operation of the pulvinar in vision**

When knowledge of a pattern is available from previous experience, the pulvinar can convert a mid-level input into a high level output very quickly, on the order of the cycle time of the correlator (approximately
In situations, such as reading, where a fixed symbol set is used, performance of the pulvinar is routine and few errors occur. Little effort is spent on recognition of the syllable stream and the available effort can be spent on analyzing the ideas associated with a paragraph. However, if the symbol set is more diverse or unusual, the pulvinar may issue a “no recognition” signal to the control point. There may also be levels of “no.” It may recognize the object as a female, as an older female, as a grandmother; but will it label the scene as “my grandmother?” This type of situation is the source of the comment, “I did not recognize you initially.”

19.1.5.1.2 “Learning mode” of operation of the pulvinar

If the pulvinar issues an emphatic “no recognition,” the visual system reverts to the learning mode. This mode is slower and requires greater mental concentration. In this case, the control point must send the mid level data from the correlator to the cerebrum for cognition leading to a definition (in mid-level language) of the object (an interp) before sending a high level definition of the object (a percept) on to the higher cognitive centers for action. Both the interp and the percept must also be sent back to the control point for entry into the proper memory location of the pulvinar. In the figure, the learning loop is shown returning to the pulvinar directly for convenience.

Common examples of these learning cases include;

• the printing of “Jesus” in white on a black background and without a border. Cognition is required the first time to tell the pulvinar this is the same as a similar word printed black on white and with a white surround.

• encountering old German text where the f’s and the s’s are interchanged. Cognition is required the first time to tell the pulvinar to interpret the syllable by interchanging its translation of the individual characters.

If the pulvinar issues a conditional “no recognition” signal, the learning mode will also be entered in an attempt to more succinctly define the object under examination. This type of situation may be as simple as examining a word to determine if it is misspelled or a new word with a totally new meaning.

19.2 Perception and cognition of a general scene

19.2.1 Frequency of ocular motions during perception

Section 7.3.2 defined a series of ocular motions by angular size and defined the temporal frequency bandwidth related to those motions. Here, it is useful to consider the frequency of occurrence of these different size motions.

The question requires a brief discussion of the activity of the “average healthy human subject.” The frequency of saccadic operations varies greatly with the type of activity at a given time. Using the classifications given in Figure 7.3.2-1, there is a significant difference in saccadic activity related to each of the defined terms.

The microsaccades (tremor) occurs continuously in humans. It is only suppressed by very severe coma or death. (Historically, observation of the microsaccades has been used to diagnose death. The frequency of microsaccades peak near 30 Hertz and occur continuously. (thus, microsaccades occur at 30 movements per second all of the time). The bandwidth value of 120 shown in the figure corresponds to the fourth harmonic of the fundamental frequency of 30 Hertz. These higher frequency components are used to form the sawtooth shape of the tremor waveform.

At the other extreme, large saccades may not occur for up to a minute when “staring into space.” More frequently, they occur at a rate of a 5-25 per minute when driving, watching a live sports event, etc.

The rate of small saccades and minisaccades varies tremendously with the specific task. When reading, a saccade of about 15 degrees (a small saccade) occurs at the end of each line of text. This gives a rate of about one small saccades per second. Minisaccades occur as the eye moves from syllable to syllable along the line. These occur about every 200 milliseconds for a rate of about 5 minisaccades per second during reading.

19.2.2 Planimetry of eye motion supporting interpretation of detail
Figure 19.2.3-1 provides a figure from Padgham & Saunders. They interpret the important features of the figure as the points of fixation. This work takes an entirely different perspective. The important features are the periods of tremor as they outline the actual features of primary importance to the interpretation process. The individual tremor segments correspond to the principle edges in the viewed image, from the perspective of the two-dimensional correlator of the TRN within the POS. While these may not be the primary features envisioned by the investigator, they are those selected by the TRN control circuitry of the subject. They estimate the frequency of the tremor at 50 Hz. The small excursions, technically minisaccades or flicks, corresponding to a few minutes of arc, occur one or twice a second at a velocity of hundreds of degrees per second. Finally, there is a slow drift following a saccade and before the beginning of the tremor. This latter drift appears to represent the signal delay within the closed servomechanism of the POS.

19.3 Perception and cognition in reading

Ramus has provided a chapter on the basis of reading acquisition. While including the word neural in the

---

Reading & Recognition 19- 9

It is generally recognized that writing appeared in various primitive forms some 10,000 to 12,000 years ago. It was first reduced to a form recognizable today, and separate from simple wall paintings, about 3000 B.C.

Research into the origins of writing is extensive and only a few features will be discussed here. Coulmas has provided a very readable history and ontogeny of writing. Without initially defining writing, he lists its principle characteristics (page 17).

1. It consists of artificial graphical marks on a durable surface.
2. Its purpose is to communicate something.
3. This purpose is achieved by virtue of the marks’ conventional relation to language.

All indications are that writing, as opposed to an artistic rendition, first appeared in Mesopotamia as an aid to commerce.

The point of interest here is that writing is an artifact of human activity that is totally dependent on two features of the neural system. First, the ability of the visual modality to perceive the subtle differences between the symbols, and generally groups of symbols, used. Second, the ability of the higher neural centers to associate these percepts with previously considered (and essentially memorized) ideas.

19.3.2 Higher Levels of Perception–Introduction to reading & verbalization

Although the computational aspects of higher levels of perception are beyond the scope of this work, it is possible to draw some general guidelines for the structural aspects of this level of perception within the brain based on the fact that the retina is generally regarded as an extension of the brain surface. The general conclusions would consist of:

+ the mechanisms of the brain are based on electrolytic chemical principles which allow the brain to achieve its remarkable thermal efficiency by utilizing non-dissipative,
10 Processes in Biological Vision

reversible thermodynamic processes.

+ the “conductors” of the brain are non-metallic, electrolytic paths which exhibit charge transport velocities on the order of 10 meters per second (versus $10^8$ meters per second for metallic conductors).

+ the principle of feedback is utilized frequently in its internal form. External feedback paths around an amplifier are used infrequently if at all.

+ the structural foundation of the brain is the activa, the liquid crystalline based electrolytic transistor.

+ each neuron contains at least one activa at the junction of the dendritic and the axonal structures, which may or may not be located within the soma; the soma is primarily a housekeeping apparatus.

+ the neurons of the brain are interconnected primarily by “gap junctions” which are actually the active regions of additional electrolytic semiconductor devices, activas.

+ the activa of the brain are used primarily in analog signal processing, except where long distance communication is required, on the order of two millimeters or more.

+ the analog activa is easily converted into an action potential generator by the addition of a significant size capacitor to either the dendritic or axonal terminal of the device and the implementation of an larger impedance in the poda lead.

+ the differential amplifier configuration found in the photoreceptor cells of the eye are easily adapted to a two stable state device that could be used for long term storage in the brain.

19.3.2.1 Modes of vision using nomenclature of psychology

Koch & Ullman have provided an analysis of the modes of vision from a largely psychological perspective. They review the literature in terms of a two-stage model. The first stage is described as a “preattentive” mode. The second is an “attentive” mode. They then concentrate on further analyzing the attentive mode, under the label selective visual attention. They suggest this mode can be further divided into three operational stages. First, a set of elementary features is computed in parallel across the visual field and is represented in a set of topographical maps. Second, a “winner-take-all” (WTA) network localizes and selects the most active unit in the above saliency map. Third, an additional network directs this localized unit to a central representation unit, presumably for detailed analysis.

The Koch & Ullman approach is quite compatible with this work. However, they do not appear to address the alarm mode or differentiate between the analytical mode operating in response to an alarm or to volition. Their preattentive mode conforms well to the awareness mode of this work. The discuss the “pop-out effect and other aspects of conspicuousness. The pop-out situation appears to be a clear case of autonomous alarm mode operation, although not necessarily due to a high order of danger. In the second case, searching for a feature having several attributes, the subject is clearly commanding a detailed search. This procedure is clearly compatible with the volition mode of this work. In both cases, a central engine is making choices. These can be considered “winner-take-all” with respect to the whole organism or winner-take-all with respect to only certain signal processing sequences. In this work, that engine is the thalamic reticular nucleus. Their interpretation appears to be largely sequential. It does not emphasize the parallel operation of both the awareness (preattentive) mode and the analytic mode (whether invoked by the alarm, awareness or volition mode). They do recognize a relationship between their attentive mode and the fovea but do not detail it. They define the non-retinotopic nature of the information following selection by the WTA mechanism and the subsequent attentive processing. Their attentive processing mechanism appears to conform to the combined activity of the perigeniculate nucleus and pulvinar (and potentially the

---

*Human Neurobiol.*, vol. 4, pp 219-227
cerebellum) in this work.

Their discussion contains several concepts that can be associated with the introduction of reading as well as general scene element interpretation.

The literature of scene analysis and reading found in the psychological literature is quite large. It is necessarily largely conceptually founded and behavior based. The modeling of the image perception mechanisms remain at a very elementary level in 2014. The models described in Galmar & Hsiao\textsuperscript{16}, as examples, do not incorporate any blocks directly associated with the physiological elements of the neural system.

19.3.2.2 Operating modes extended to reading

While the investigations by Koch & Ullman were generic to signal processing in vision, Olshausen, et. al. have focused more specifically on the analysis of symbols\textsuperscript{17}. Their model was aimed primarily at the need to accommodate variations in scale, orientation and position in arriving at an object oriented reference framework. They do focus on the use of an associative memory as part of the processing model. Their considerations are mathematical and conceptual but do address how some of their conceptual circuits might be implemented biologically. However, their analysis does not incorporate a dual path visual system. All of their discussion centers on the retino-geniculo-striate pathway to area 17. They do focus on the generally held view that the pulvinar is a major player in controlling the routing of signals. At the next level of detail, this work assigns this role more specifically to the adjacent thalamic reticular nucleus. They summarize their model as a “zero-th order” model. It does not attempt to explain the mechanism for recognizing a symbol. They do attempt to size the spatial dimensions of the analytical (attentive) mode. However, the numbers are of “ball park” precision. They take the “window of attention” to occupy an array of only 30 by 30 elements. This appears far from the capability of the foveola and the associated analytical mode of this work. Such a small area does not appear compatible with the area perceived by the visual system during one gaze in reading.

19.3.2.2.1 Tasks involved in reading

The task of reading involves so many individual procedural steps and so many alternate paths that it can only be drawn at a coarse level of detail here. The diagram is taken from a review of the material by Kennedy in Chapter 7, and by Rayner, Reichle & Pollatsek in Chapter 11 of Underwood. The material of Murray in Chapter 8 is also of significance but suffers from a lack of specifics concerning the geometry and performance of the fovea and foveola.

19.3.2.3 Operating modes extended to verbalization

Pre-1980 literature generally spoke of verbalization as being centered in the left cerebral hemisphere. As an example, Trevarthen & Sperry discussed the ability to verbalize with respect to perception limited to the left or right visual field\textsuperscript{18}. More recent MRI and PET experiments have shown the subject to be far more complex.

19.4 Smooth and Salutatory Pursuit EMPTY

19.5 Interp & Percept formation within the diencephalon EXPAND

In this work, an interp is taken to be associated with a low level piece of information, such as a single syllable in a written sentence. When combined with other syllables (and possibly) multiple adjacent words, the result is a percept (an understandable feature of a text, scene or musical rendition). The combination of multiple percepts leads to a more complete perception related to the scene or other feature of the external environment. Such total percepts (frequently including the individual percepts) are stored in the saliency map of stage 4 for perusal by the stage 5 cognition elements of the neural system.


12 Processes in Biological Vision

19.6 Reserved
19.7 Reserved

19.8 Reading and the interpretation of fine detail

In the following introductory discussion the functions of reading and the analysis of fine detail will be considered one. However, there is a significant difference between analyzing fine detail and analyzing the typical detail in a scene. Because reading involves nearly every facet of the visual system, it has not been described in one of the previous Chapters that focus on specific functional areas. Therefore, this section will be subdivided into four portions to include the scope of the full subject. Section 19.8.1 will assemble a variety of relevant subjects in a background section. Section 19.8.2 will present an overview scenario of the total reading process. Section 19.8.5 will discuss the environment, mechanics and mechanisms involved in reading. Section 19.8.6 will provide some performance descriptors for specific reading scenarios. There are a series of clinically significant irregularities related to reading. These are discussed in Section 18.9.

Reading is the ultimate mechanism and capability provided by the analytical mode of visual system operation. The process of reading is probably the most sophisticated function in vision, if not in the entire neural system. It involves all of the features of the POS operating in the most sophisticated and choreographed manner. Furthermore, reading clearly involves distinct training, operating and senescent phases. Therefore both the initial introduction to reading and the subsequent time scale become significant factors in discussing and evaluating reading performance. The interconnection of cognition related to the awareness mode of vision and the cognition related to the reading process is highly dependent on the training, memory capability and innate intelligence of the subject. This situation complicates the description of the capability level reached by a subject, and of similar subjects at a given time.

19.8.1 Background

This material will highlight a difference between the performance of selected primates (other than man) and man. As found by many experimentalists, the general premise is that the other primates mature intellectually and visually along a course similar to humans up to an equivalent age of three or four. This capability is clearly evident where chimpanzees and other primates are taught to respond to verbal and graphical materials by selecting images on a television screen, or (reversing the process) expressing their desires by selecting an image on a television screen. Rare animals may even be able to express more complex thought patterns involving a series of images on a television screen. However, these images are more pictures than writing. Change the font of the words on the screen or alternate the colors of the letters in the words and the experiment quickly comes to an end. Humans have little difficulty with such changes beyond the age of six or seven (on average).

The overall capabilities to be developed and discussed in this section as well as the individual mechanisms involved exhibit a wide range within a given human population. In nearly every case, this material will attempt to describe a typical subject unless it is clearly indicated that a special case is being discussed.

Crystal has provided some conceptual block diagrams of the reading process as suggested by a linguist19. He differentiates between reading by eye and reading by ear. He also describes a variety of dyslexic and aphasic conditions. That work focuses on letter recognition rather than the more advanced syllable recognition.

19.8.1.1 Glossary

To develop a comprehensive discussion of reading, it is necessary to introduce a variety of terms from a wide assortment of disciplines from biology, computers and the education process. Many of these terms will be defined in greater detail in the following sections.

- Comprehensions: High level thoughts reflecting a high degree of cognitive understanding. Generally stored in the highly associative saliency map and only subject to change as a result of new cognitive

processes.

**Controller**—A type of neural engine specialized to control the flow of data into and out of memories located throughout the neural system.

**Default mode**—In the case of reading, the preprogrammed mode assumed in the absence of feedback from a controller indicating a failure in one of the cognitive steps following initial image scanning.

**Engine**—A general purpose analog computational facility found in perfusion throughout the brain. They are frequently programmed as special purpose machines based on PROMS. These PROMS are initially programmed through learning and experience.

**Interp**—A vector description of the symbol(s) imaged on the foveola issued by the pretectum following a nominal 50 ms scanning (at the microsaccades level) and cross correlation process.

**Percept**—The vector created by the assembly of individual interps generated by the pretectum in response to the scanning and cross correlation process performed by the Precision Optical System.

**PROMS**—See Write once–read repeatedly memory.

**Saliency engine**—A higher level mechanism in perception that provides the addresses of pieces of perceived information that must be considered as a group to achieve cognition. The various vectors, relating to the perceived information, when grouped act as individual bytes in the overall cognition vector describing the recognized event.

**Saliency Map**—The general database employed by the cortex as a archive of all sensory information collected during the life of the subject. Highly associative. Not generally re-writable except under instruction from the controller following significant cognitive processes calling for such a change.

**Shift register**—A type of memory that is designed to accept short values (consisting of several characters) in part of its memory (defined by an address) and then accept subsequent values at other memory locations with the general intent of assembling a longer, more comprehensive, value from the sequentially received individual values. The comprehensive value can be accessed by addressing only the initial address.

**Write once–read repeatedly memory**—A type of memory, commonly called a programmable read-only memory (PROM) in computers, that once data is written to a specific address in the memory, it cannot be changed. However, the address can be accessed repeatedly to recover a copy of the stored data.

**Vector**—A coded multidimensional signal incorporating multiple values derived from multiple sensory or cognitive sources (possibly over a time interval).

**19.8.2 Overview scenario of the reading process REWRITE**

Several sections addressed previously are important to an initial understanding of how the brain manipulates information related to these subjects. Section 15.6.3.6 addresses the role of the pulvinar in creating percepts of a scene. These percepts depend on the ability of the two-dimensional correlator within the pretectum to generate the initial interps of the elements of the scene. The correlator design and organization are discussed in Sections 15.3 & 15.4. A number of consequences associated with improper operation of the POS, the control point, the pulvinar and the cerebrum are highlighted in Section 18.3, 18.7 & 18.9.

Reading should be differentiated from the observation of logos, and other large graphics that may contain letters drawn from a local language. Logos are meant to recall a previous image. They are not designed to convey new information.

Reading is a highly stylized method of acquiring information from a written message. It invariably involves a step and repeat process wherein the eyes move progressively along a line of symbols, pausing for 50–200 msec. before proceeding via a small saccades. During the pause, the eyes perform a complex pattern of microsaccades (only poorly categorized in the literature under the heading of tremor). These individual motions are used to convert the spatial pattern of the message into a temporal pattern that can be
14 Processes in Biological Vision

manipulated by the neural portion of the visual system. The control of these mini and microsaccades emanates from the Precision Optical System. The POS is a group of high performance servomechanisms that control the motions of the eyes in response to the requirements of the portion of the midbrain tasked with extracting intellectual meaning from the symbology projected onto the foveola of the human eye. This portion is called the pretectum and it is a part of the thalamus. The pretectum attempts to extract the fundamental meaning of small groups of interconnected lines we generally call syllables. When the pretectum is able to do this successfully, the POS follows a preprogrammed sequence of step and repeat motions (sometimes called a one-dimensional Levy flight in the recent psychophysical literature). These motions present a new image to the pretectum every 50-200 msec as required. The pretectum extracts the meaning of each small group of lines in the form of an initial (or basic) interp. These individual interps are assembled sequentially until they represent a higher level interp, typically a word. Multiple interps are then assembled into a percept, typically a short sentence or a phrase. The process continues to build until the subject comprehends the entire meaning of the written message.

The gross steps in reading are accomplished serially, but in a continuous pipeline fashion, and can be defined as;

- sensing, the acquisition of material imaged on the foveola,
- Initial cognition (based largely on prior experience) resulting in an interp,
- Second level cognition (based largely on prior experience) resulting in a percept,
- Third level cognition (based to a significant extent on prior experience) resulting in a theme,
- Higher cognition leading to a comprehension of a message or idea.

Much of the material comprehended is stored in “permanent” memory.

These terms will be defined in detail below and in the Glossary.

19.8.2.1 State of knowledge concerning the circuits of reading

The current state of knowledge concerning the process of reading is spotty. The architecture of the process can be reasonably understood through comparison to man-made circuits of similar function and through reverse engineering (typically involving diagnosis resulting from trauma, disease or intentional lesions). The signaling capabilities of the individual neurons are also well understood. However, the coding used to transmit the messages, and the nature of the messages is totally unknown at this time. As illustrated below, our knowledge of the more gross movements of the eyes in reading are well documented. Our knowledge of the more critical microsaccades of the eyes has not progressed since Yarbus and Ditchburn in the 1970's.

While a large amount of dissection, at the neuron anatomy level, has provided a large amount of detailed information about specific signal paths, the overall signaling patterns have been difficult to define on a global basis. This is changing rapidly with the introduction of nuclear magnetic and nuclear emission imaging techniques, particularly the functional MRI (fMRI) which is fast enough to represent individual areas of gross neural activities simultaneously over a significant area of the neural system. Unfortunately, these techniques are still classed as a tool in traffic analysis from a communications perspective. They do not describe the precise commissure path between engines, the number of parallel neuron circuits in the path, or anything about the messages or the coding of the messages along these paths. To answer these questions will require an increase in the sophistication of recent electrophysiological techniques. Such techniques are available in the available test equipment base. They involve synchronously recording multiple neural paths, each at finite time intervals during variations in the stimulus applied to the subject. Such techniques are readily applicable to the simple neural commands of the oculomotor systems now. By advancing back along the neural circuits of the oculomotor sections of the POS, additional understanding of the codes and messages can be obtained. Hopefully, this will eventually lead to the interception and decoding of neural signals along selected simple paths that relate to various levels of comprehension.

19.8.2.2 Setup procedures prior to reading

While it may seem obvious, the material to be read must be placed to present imagery within a specific size range (if it is to be read as opposed to just being recognized as a graphic). The smallest character to
be analyzed must normally represent an angular size of at least 40 seconds of arc. If the size exceeds five
times that height, the full capability of the analytical process to read efficiently will not be realized. The
material should be lighted to within the photopic range. Ideally, it should be illuminated with a daylight
source (nominally 7053 Kelvin) to insure maximum performance by all of the photoreceptors in the
foveola. The material should be of the highest contrast, black on white being preferred.

If serious reading is to be performed, the subject should prepare to read by isolating himself from other
distractions that might instigate needless activity in the alarm mode of vision. Such alarm signals will
override and interrupt the reading process.

19.8.2.2.1 Achieving focus

The auto-focus capability of the eye is dependent on the POS. The POS in turn is dependent on detail in
the image to achieve focus. This detail should consist of at least one high contrast edge in the field of the
foveola. As Campbell & Greene noted, a sine wave does not provide a good signal for auto-focus
purposes. The focusing operation typically requires 50-100 ms. See Section 18.2.4. It will become clear
that the visual system does not refocus following each saccade. It relies upon short-term memory of the
scene to provide ques to proper focus, only stopping to refocus following abrupt movements of the scene
relative to the eyes.

The capability of the memory system is largely unexplored relative to reading. It appears that the visual
system relies heavily on memory to provide an estimate of the focus condition required to return to any
line of fixation previously achieved. Thus, changing from reading to looking across the room appears to
rely on memory for an initial estimate of the focus condition required to focus on the far wall.

19.8.2.3 Typical sequence of minisaccades during reading

The following material is reproduced here from a fuller account of reading dynamics presented in Section
7.5.3.

Becker, et. al. have also provided empirical material on the reading mechanism as of 1999. Both
Underwood and Becker, et. al. concentrate on observable eye movements (at the minisaccade level) as
the key to our understanding of reading. This work takes an entirely different view. It treats observable
eye movement as merely a mechanism for imaging individual scene features, symbols and character groups
onto the foveola where the actual process of perception is initiated. This perception involves eye
movement at a level not normally observed by the clinician or academician, the microsaccades or tremor
level.

Based on the conceptualization developed in this work, reading can be defined as the
act of assembling a sequence of perceptions acquired through the sequential
analysis of individual symbols or character groups and interpreting these
perceptions in accordance with a set of syntactical rules. In this definition, symbols
include hieroglyphics and other glyphs. The initial interpretation of each symbol or
symbol group by the POS results in the generation of an individual “interp.” When a
series of interps are combined, the resulting interpretation will be called a percept.

Within this work, the foveola and the fovea are defined based on the morphological characteristics of the
retina. The foveola is defined as 1.18° in diameter, the fovea is defined as 6.26° in diameter in object
space and the parafovea is defined as beyond 6.26° from the fixation point. Therefore the fovea of
Kennedy conforms closely to the foveola of this work and his parafovea will be assumed to equal the fovea
of this work.

Kennedy also states “In reading, each word is inspected by an initial fixation at a particular position

20Campbell, F. & Green, G. (1965) Optical and retinal factors affecting visual resolution J Physiol vol 181, pp
576-593
Cogn Comput vol1, pp 64–76
resulting from an ‘entry saccade’ of a given size, launched from a particular location in another word.” The conditions, variations and significance of this entry saccade are discussed in some detail on page 152-153. The relationship of such saccades on the empirical model of Rayner, Reichle and Pollatsek in Chapter 11 is also reviewed. The referral to a “location in another word” is about as close to the discussion of characters in a word that is achieved in the overall Underwood text. More material is presented on regressive saccades, from one word back to a previously examined word, than is given to the examination of the characters within a word.

Kennedy discusses the term prompt, with the word gaze in parenthesis following it, as the sum of all fixations prior to the first excursion outside of its boundary (page 157). In the Abstract to the chapter, he says the time to process one of two possible foveal “prompt” words was examined using measure of gaze and fixation duration.

**Figure 19.8.2-1** presents a modified, semi-standard figure from Liversedge, Paterson & Pickering in Underwood. It should not be inferred that most of each time interval shown relates to the latency before the next saccade. These saccades may be a part of a planned saccade sequence that does not require a significant latency between the end of analysis within one gaze and the beginning of a saccade to the next gaze location. Thus, a more specific set of subdivisions of the term latency is probably called for here. The times do define the maximum length of time required for the visual system to analyze the structure of the symbols within the foveola adequately for the subject to ascertain their semantic content (probably via a lookup table). The above authors did not address the size of the type (subtense of the height of the characters) used in their experiments but they did say the tracking data was quantized every millisecond.

**Figure 19.8.2-1 Hypothetical eye movement record showing the time in milliseconds spent in a gaze between saccades. The text being read is at the top of the page. The zig in the line is indicative of a regressive saccade. Total reading time is over two seconds for this single incongruous sentence. The process is highlighted by the regressive scan to analyze further the use of the word invited. From Liversedge, Paterson & Pickering, 1998.**

Inhoff & Radach, writing in Chapter 2 of Underwood, reported on eye movements when viewing long strings of printed characters. Their data provides good information on the precise nature of the small saccades related to eye movement during reading but little information at the microsaccade level or on the nature of the saccades used to perceive word meaning in the context of reading. Their text introduces a number of potential experimental variables (flexibility of the eyeball leading to transient movements of the lens group–cornea and/or lens) but does not provide a foundation for overcoming or controlling them. Their figure 1 shows a number of minisaccades at the 0.1-0.2 character width level in between saccades of one and seven character widths. However, the noise level of their equipment was not specified or shown and they may have excluded microsaccades from their analyses. Their discussion includes presentation of an interesting dichotomy (citing Deubel & Bridgeman). First, that the eye is an imager and that small “post saccadic movements will smear the retinal image.” Second, that the post saccadic motions are relatively small and principled, and the reader may be able to extract useful information during that

---

Radach & McConkie prepared Chapter 4 in Underwood. It discusses the determination of fixation positions in words during reading. One of their conclusions is that, “In all cases (where there are spaces between the character groups), eye movement control during reading appears to be word-based.” This control appears to involve two distinct mechanisms, a selection mechanism and a performance mechanism. It is proposed that the selection mechanism determines how one word rather than another is selected as the target of a saccade, and the performance mechanism determines where the eyes actually land given the above selection. The discussion centers on the general likelihood that the saccade is aimed at the center of the selected word. In the data presented, the imagery was presented (a German translation of the initial text in *Gulliver’s Travels*) in page format with five to seven double-spaced lines of up to 72 ASCII characters each on a 15 inch VGA monitor. At 80 cm viewing distance, each letter corresponded to approximately 0.25° of visual angle (15 minutes of arc or three times the 20/20 character definition).

One of the conclusions of the above authors is that “where the eyes go with respect to selected saccade target words, is the result of low-level visuo-oculomotor control factors, almost completely unaffected by higher cognitive processes.” This is probably true for relatively familiar words not calling for regressive saccades or multiple saccades within one word. The above authors go on to caveat their statement. One of their caveats is “in the case of regressive inter-word saccades, the saccade parameters we have looked at suggest a control mode different from the low-level default routines.” The intra-word characteristics of the reading process were not discussed in their chapter. Here again, the word inter-word appears in the index to Underwood but the expression intra-word does not. Heller & Radach made note of two important facts. They noted the work of Dunn-Rankin in 1978 that showed that the initial fixation point on words was not at their center but at positions left of center. They also noted the work of Rayner & Pollatsek in 1981 that showed that the final decision concerning the direction and magnitude of the next saccade was made during a given fixation interval.

Rayner, Reichle & Pollatsek presented Chapter 11 in Underwood. They discuss the effect of limiting the length of time available within a gaze to analyze the text characters. They show that if a given gaze (the traditional fixation interval by name, even if it involves tremor) is interrupted before 50-70 ms have elapsed, the reading process itself is interrupted and comprehension suffers or is lost. They also suggest that a preview of a word, while it is imaged in the area outside the foveola, has a positive impact on reducing the time required to interpret it when it is moved into the foveola.

The above authors review several conceptual models of the eye movement control system required to implement reading. A process model by Morrison is summarized. This model is still conceptual but includes the concept of some preanalysis of a word before it enters the foveola or before it is brought to the point of fixation within the foveola. A logic is provided that controls the length of the gaze and/or the series of interim fixations associated with each word. The model explains two aspects of the eye movement phenomenon in the reading process: (1) the fact that there are fixations that are much shorter than the nominal 195-200 ms saccade latency in simple oculomotor tasks and (2) the occurrence of unusual landing positions, such as between words. A competing model by O’Regan, that they describe as a strategy-tactics model, is also summarized along with a critic of its features.

Finally, the authors summarize their proposed E-Z Reader Model. They describe it as similar to Morrison’s process model except more refined through the implementation of two additional facets. First, it decouples the signal to shift attention from the signal to program a saccade. Second, it is better specified in that it has been implemented as a computer simulation program. They refer to the fifth generation of this program as the program under current discussion. They describe the simulation with a schematic representing five basic processes:

1. A familiarity check on a word.
2. Completion of lexical access.
3. An early, labile stage of saccadic programming, which can be cancelled by subsequent saccadic programming.
4. A later, non-labile, stage of saccadic programming.
5. The actual saccadic eye movement.

They define the first two steps as products of a single cognitive process which occurs during a preview while the word is still in the fovea (not the foveola) and occurs before the movement of the word to the

---

25 Becker, et. al. (1999) Op Cit. pg. 341
center of the foveola. They develop the fact that completion of the familiarity check depends on two additional factors. These factors slow the rate of processing as the eccentricity of the word, relative to the point of fixation becomes larger. This factor was added to recognize the rapid falloff of the resolution (acuity?) of the eye with eccentricity.

They also discuss the operational distinction between an interword and an intraword saccade. They suggest the basic eye movement strategy conforms to the “dumb” default strategy: That strategy is to plan to fixate each word from more than one viewing location unless the word’s familiarity indicates that a refixation is unnecessary. The dual fixation strategy could obviously be useful in words that frequently have unusual or multiple suffixes (syllables).

The dumb default strategy emphasizes the importance of having a large vocabulary in the field encompassing the text being read. It also suggests that the average reading rate is dependent on the frequency of occurrence of words found in the vocabulary.

Some authors have argued there is no “magic moment” of word identification, that identification (of individual words or total thought) only comes with a growing amount of data collected on a continuous basis. This appears to be a question of semantics between authors since the motion of the eyes is clearly not continuous with time.

19.8.2.3.1 The familiarity default procedure in reading

The above “dumb default” strategy suggests a major change in the operating mode of the POS during reading. As soon as a word is recognized cognitively (the magic moment?), the POS initiates a saccade and proceeds to the next fixation point. This suggests that the visual system operates in a manner similar to the “auto complete” feature of an INTERNET browser. The browser compares the initial key strokes of an entry with its short term memory and suggests the appropriate completion of the typed entry. If the suggestion is wrong, the typist is free to enter an alternative. An equivalent scenario can be defined for the visual system. After analyzing only a few symbols, initially in the fovea area surrounding the foveola, the system may believe with high probability that it knows what the entire symbol group means. In that case, it will instruct the POS to proceed to the next symbol group. If the following symbol group is recognized but it does not fit logically into a recognized syntax with the first group, the POS may be instructed to perform a regression saccade in order to review the previous symbol group for an alternate suffix or other difference from the assumed meaning. This procedure is illustrated in Figure 19.8.2-2. The system attempts to interpret a short sentence. It can methodically perceive and interpret each character group as in A or it can adopt a more aggressive approach and make a guess based on the likelyhood that the character group “dres” is part of the longer word “dressed.” This results in the sequence shown in B and some time is saved as one gaze (fixation) is eliminated from the initial saccades sequence. If however, the assumption was made that “dres” was part of the word “dresses,” the same procedure can be followed until the second or third character group in “yesterday” is reached. At this point in the interpretation, a context conflict is recognized through comparison of the initial concept file with the saliency map of the individual. As a result, a regression saccade is called for back to the word that was actually “dressed” and not “dresses.”

![Figure 19.8.2-2 The procedure of perceiving and interpreting a sentence showing three alternatives.](image)
It is instructive to consider the time line of the above activity to determine if it suggests the course of the signals moving through the visual system. Certain time lines may suggest activity limited to the POS, to the POS in conjunction with the pulvinar or to the POS in conjunction with either area 7 or the posterior areas of the cerebrum. Figure 19.8.2-3, from Becker, et. al. addresses this subject directly. There is a statistically shorter latency associated with the regressive saccade than with the progressive scan. Apparently, as soon as the system notes an inconsistency in the proposed syntax, it cancels the analysis of that character group and calls for a regressive saccade to reestablish a viable baseline. These experiments were carried out with considerable statistical precision and the original source should be reviewed before proceeding.

Figure 19.8.2-3 Distribution of fixation duration for two subjects in a standard reading experiment. Top, cumulative distribution. Middle, fixation duration following syntax failure at fixation after a "dumb default." Bottom, distribution of fixations before normal progression to next fixation. See text.
19.8.2.3.2 Typical time sequence in reading a paragraph

Figure 19.8.2-4 presents a record of a subject reading a paragraph. The complexity of the line is suggested by the number of steps in the descending traces.

19.8.2.3.3 Difference between good and poor readers

Gibson has discussed attentional strategies in reading. He has reproduced a figure from Tinker that sheds some light on the difference between good and poor readers. The number of minisaccades per sentence appears to be much higher in poorer readers. The dwell times between minisaccades appear similar.

19.8.2.4 The fundamental phenomenon and mechanisms of reading

The above synopsis of much detailed work reported in Underwood and elsewhere leads directly to a discussion of how the semantic meaning of a character group within a word is determined during the individual fixations (or gazes). It is likely that the reader initially scans textual material as a scene and evaluates its general characteristics, concentrating on paragraph arrangement, figure inserts, indentation etc. The system then calls for an entry saccade to position the line of fixation on the first character group of interest.

[Figure 19.8.1-4] demonstrated the importance of considering the size of the text imaged onto the foveola and fovea. The preliminary conclusion was drawn that less than five (nominally three) characters of nine point type were usually imaged on the foveola at one time and this number quantified the amount of information that could be analyzed by the visual system without additional saccadic activity. It also showed that more characters could be imaged across the fovea. However, the resolution of the optics and the different signal processing employed in the foveal region suggests that these characters could not be analyzed with sufficient precision. However, the structural accouterments of the character groups (capitalization, illumination—including capitalized acronyms, hyphens, etc.), and particularly the space between them could be determined using the awareness channels of vision.

The data on progressive intraword and regressive scanning suggests the system is prepared to make estimates of the meaning of an initial character group based on only the first few characters and then repeat the estimation procedure on subsequent character groups until the cumulative estimate fails on the basis of syntax or the analysis and perception of the word is completed successfully. The failure calls for a regressive scan to repeat the interpretation process. This regression may involve several character groups or several words. Success calls for the implementation of a progressive scan to the next word. The empirically measured location of adjacent gazes also suggests that the system is prepared to ignore many if not all two and three letter character groups when they are recognized in the fovea and before they are fixated upon. There is an exception to this policy where the system is expecting such groups such as in highly technical text.

19.8.2.4.1 The mechanism of detailed character group analysis

As a result of the above procedural steps, the critical remaining task is to specify how the visual system recognizes the individual character groups during an individual gaze. Figure 19.8.2-5 describes the scene.

---

imaged on the foveola of the retina for a word imaged at the scale of the 20/20 line on a Snellen Eye Chart. The letters are 3.125 minutes of arc high to reflect the immersed optical system used in vision. This size corresponds to 5 minute of arc characters in object space divided by the index of refraction of the vitreous humor of the eye. The characters are shown in a bold serif font resulting from the use of a serif font in image space, such as Times New Roman, and passing the image through a resolution limiting optical system such as that of the human eye.

There appears to be little appreciation in the Ophthalmology community for the difference in performance of the visual system when viewing characters with and without serifs. This is demonstrated by the back to back samples shown without relevant discussion in Miller & Newman. The discussion of the minimum angle of resolution (MAR) parameter is useful, as is the history of the Snellen Eye Chart.

Each photoreceptor of the retina is shown at its correct nominal size of 2 microns diameter. However, it is shown in a square packed array for ease of discussion when they are actually packed in a hexagonal array. At this resolution, a five equally spaced character group representing a word spans about 43% of the foveola. As discussed earlier, this size image corresponds to about 1.5 point type at 15 inches and most people cannot resolve the individual characters at this scale and the spacing shown. Only when the characters are well separated and in block form can they be perceived under these conditions. The fact that they can be perceived is suggested by the fact that the microsaccades of tremor are sufficiently large as to cause the photoreceptors of the retina to scan across the edges of the individual characters. The degree of

scanning is sufficient to determine the contrast edges associated with each character, even with the hexagonal packing of the actual photoreceptors. However, the degree of scanning is or is becoming marginal. As a result, the question arises as to whether the scanning is adequate to generate an outline of the character as in the letter P, or whether only a stick character version, such as the R, is available for perception by the brain. This question becomes moot as the size of the characters is increased. Using nine point type at 15 inches, the individual strokes of a character are about the width of twelve photoreceptors in a line. It is clear that the combination of the photoreceptor size and the scanning mechanism can resolve both edges of a stroke of a character this large. The hexagonal packing of the photoreceptors can only provide additional detail.

Because the complete initial character group falls within the foveola during a single gaze, the parallel processing capability of the neural system connecting the foveola to the brain is able to analyze all of the characters simultaneously (assuming adequate resolution of the space between characters). Hence, the brain is able to assemble an estimate of the three individual characters and to reach a preliminary interpretation of the meaning of this syllable-like group.

Note however, the eyes are not stationary during this nominal 220 ms gaze. They are continually moving with amplitudes of micro-arc-seconds and frequency components in the 100 Hz region. Based on the findings of Rayner, et. al., this gaze interval will be subdivided into individual “glimpse” intervals of about 50 ms. The assumption being that the pretectum is performing its cross correlation calculation based on the edge crossings that occur during the glimpse interval. It then issues an interp based on this calculation (after performing a correlation with the interp values stored previously in the pulvinar).

Having made an estimate of the meaning of the first character group, it has two choices. It can make an assumption as to the meaning of the entire group of characters in the word, and execute a “dumb default” to the next word or it can make a progressive saccade to the next character group within the overall word. In the latter case, it proceeds to build an estimate of the meaning of the entire word essentially one syllable at a time. In the former case, it takes a gamble that is proven good or bad based on the syntax of the previous and following words in the line of text. If the estimate appears to fail, a regressive scan is performed to reestablish a running estimate of the meaning of the multi-word sentence. The more technical the text material, the more likely the estimate is to fail and the more frequently regressive scans will be observed in the experiments.

What is not known at this time is the scanning pattern used by the POS to scan each multiple character group. It appears clear from Shakhnovich and Yarbus that the horizontal and vertical components of each microsaccade can be implemented individually within times of less than 10 (probably less than 5) milliseconds. Experiments are needed to determine the statistical relationship between these two components for a variety of simple symbols and symbol groups as well as conventional alphanumeric character groups.

Recently, Deering has developed a computational model of the spherical retina within the eye of remarkable realism. Figure 19.8.2-6 shows the remarkable similarity between his representation of the human retina with a Sloan Eye Chart imaged onto it. The Deering model is able to introduce tremor, microsaccades and other motions associated with the visual process. It only contains photoreceptors of the type generally described as cones. Although he notes that the putative rods outnumber the cones by 16:1, neither they, or the spaces to accommodate them, are found in his model (Section 3.2.2.3).

---

24 Processes in Biological Vision

19.8.2.4.2 Typical sequences of microsaccades

Only gross information is available at this time concerning the nature of the microsaccades associated with reading. Section 7.5.3 summarizes what is known. The greatest lack at the current time is the phase relationship between the microsaccades associated with the motion of the eyes in elevation and angle. Currently available instrumentation can provide this information. What is needed is a bio-engineering institution, associated with a strong electronics instrumentation organization, that can look into this subject. Merely producing the Lissajous figures associated with these movements as a function of time (in about 10 millisecond steps) would produce a goldmine of information about vision.

19.8.2.4.3 Suggestions for future research in reading

Figure 19.8.2-6 Sloan acuity chart 720 x 720 pixels imaged on the cone. The upper half of the chart uses only 543 nm light. The lower half uses broadband light. Compare the upper 20/20 representation to the previous figure. From Deering, 2005.
While excellent insight into the reading process has been achieved during the previous exploratory research, it is appropriate to consider ways to raise the precision of the empirical process. One method of improving the statistical relevance of the data would be to remove the variable of word length from the sample text material in a given experiment. This change would allow more precise conclusions to be drawn with regard to the effect of word length on the statistical properties of the observed saccades. It would also allow more precise determination of the nature of the interword saccades. If the words were randomized, the element of syntax could be removed from the perceptual process, although the element of vocabulary would remain. Ultimately, these experiments may lead to a better understanding of the role of short particles of speech in the interpretation process. When combined with the current database, they may also lead to a better understanding of the process as a whole.

19.8.2.5 The Strategy Employed in Reading

The strategy employed in reading does not differ significantly from that for general scenes. However, to aid in communications, additional sets of rules have been adopted related to spelling and syntax that are not required for general scene interpretation.

When presented a scene containing text, the observer initially examines the complete scene as discussed above. It notes areas of the scene (objects) that appear to contain a particularly structured texture. If this object appears upon further examination to be text, the visual system performs an "entry saccade" that brings the line of sight to the expected entry point of the text (the upper left hand corner for most western languages).

After locating the first word (and character group of that word) the analytical procedure must initially determine the language of the text and the font of the character set used. It then continues based on the set of style, syntax and spelling rules for that language.

Reading involves the interactive process of creating an initial concept file just as when examining a natural scene. However, there are two significant variations. The stylistic rules dictate the direction in which the text is to be scanned. The anticipated rules concerning the shape of individual objects are replaced by a set of syntax rules unique to the specific language.

The syntactic rules include a stop symbol (the period) that defines the end of the presentation of a particular concept. By this means, a sentence is formed that is similar to an object within a general scene. The concepts associated with multiple sentences can be grouped into a broader concept. This grouping is labeled a paragraph in pedagogy.

19.8.2.5.1 Examples of the reading process

The following simple examples illustrate the process of perception and interpretation.

CHECKING THE SPELLING ON THE FLY

The following example shows how making a determination of the precise spelling in the first character group of the third word, and checking for that form against the saliency map, changes the concept contained in the initial file created during the reading process.

- The boy performs well on cue.
- The boy preforms the kit parts.

MODIFYING THE CONCEPT

Each of the three letter, and one two letter, groups in the first line is the target of an individual minisaccade. The character-group as a whole is then scanned by a series of horizontal and vertical microsaccades to determine its meaning. In the second line, one word is too long to be perceived and interpreted completely (at the geometric scale of the text). It is treated as a two-group (syllable)word. The first minisaccade brings the character-group "aro" onto the foveola and the next minisaccade brings the term "und" onto the foveola.

- The boy ran to the box
- The boy ran around the box

As a result of the saccade sequence, and the continual checking of the spelling of each character-group,
these two samples are interpreted slightly differently and result in distinctly different concepts. No conflict has arisen in the perception and interpretation of these simple sentences.

**SIMPLE CONCEPTUAL CONFLICT**

This example is obvious. Following the second saccade following the entry saccade, the initial file contains a conflict. Boxes do not run.

```
The boy ran to the box
The box ran to the boy
```

**SHARPENING THE CONCEPT WITH SUFFIXES**

This example merely illustrates how additional precision can be added to the concept using a suffix or an adverb.

```
The boyish girl ran to the box
The boy ran obliquely to the box
```

**MORE COMPLEX CONCEPTUAL CONFLICT**

In both of the above examples, the additional character-group within the two longer words would have caused an additional saccade to be added to the initial saccade sequence list. This would slow the reading process. However, the system has adopted a methodology to avoid this problem a high percentage of the time. After interpreting the first character group in boy or in obliquely, it frequently makes an estimate of the impact of the second character group of the word and calls for an immediate saccade to the next word. In the following example, the error is of negligible importance.

```
Jim dresses well
Jim dressed well
```

However, proceeding in this manner can result in trouble. In the following example a conceptual conflict arises in the first sentence several saccades later.

```
Jim dresses well yesterday
Jim dressed well yesterday
```

When such a conflict is encountered, a *regression saccade* is called for. This saccade takes the image projected on the foveola back to a point where resolution of the conflict is possible. The word dressed must be distinguished from the word dresses by using two saccades.

The frequency of multiple case specific suffixes contributes to the difficulty of a novice trying to read Russian.

**19.8.3 Extension of the model of the visual servosystem**

The above review of the operation of the visual system during analysis of a bucolic scene, a scene containing local changes and of reading suggests further sophistication of the visual servomechanism subsystem. The subsystem is so complex as to require a state diagram as an aid to understanding the physical structure of the system. The appropriate state diagrams and more detailed circuit diagrams will be presented in Section 15.2.5. What is clear from Sections 7.3 through 7.5 is that the servomechanisms of vision, centered on the POS, are highly sophisticated. The overall system involves multiple loops. The outer loop includes the midbrain and the cortex, with initial entry at area 19. The inner loop may include the cortex, particularly area 7, or it may be embodied within the midbrain and nearby specialized ganglia. It is most likely that the outer loop passes through the LGN on the way to area 19 while the inner loop passes through the Pretectum. The system is necessarily a sampled data system since all of the signals from the retina are converted to pulse signals in the signal projection stage between the retina and the midbrain as well as for transmission over the association fibers of the brain.

As shown above, the servomechanisms operate in a variety of modes during vision. Switching between these modes appears to be under the direct control of the Pretectum. However, it is safe to assume that the Pretectum frequently looks to the cortex for advice of a cognitive nature and it is always alert for both
stereo-optic signals and alarm signals from the computational capabilities of the LGN.

Besides using sampled data servomechanism techniques, the visual system is able to incorporate a large amount of computational capability within each servo loop. This makes the application of simple analog servo techniques to the understanding of the visual system a questionable procedure. Only under the most circumscribed experimental conditions is it appropriate to use analog concepts in interpreting the visual process. The servo loops of vision incorporate significant lumped constant time delays due to the nature of the signal projection function. While analog concepts have been used in the past to estimate the high frequency capability of the visual system, they have been grossly in error. Even the low frequency region of the overall signal passband cannot be described as a low pass network. Since the system exhibits a zero at zero frequency, even the characteristic of the servo passbands between zero and five Hertz must be described as a bandpass network. In parallel with this bandpass network is the separate high frequency portion of the overall characteristic associated with tremor. This region extends from somewhere above five Hertz, probably above 30 Hertz into the region above 120 Hertz.

Major limitation of the inner servo loop of the visual system are the finite lumped constant delays related to the projection stage. These are located between the retina and the Pretectum, and between the Pretectum and the oculomotor muscles. These two delays sum to a nominal 6.6 ms. With even minimal time delay in the signal processing portion of the loop, it is effectively limited to a 160 Hz maximum frequency response. In terms of these delays, the response of the servo plant, consisting of the oculomotor muscles and the inertia of the eyes, in traveling through a rotation of only 20-40 seconds of arc is minimal.

19.8.4 Extension of the concepts leading to cognition

There is a lack of defined terms describing the process leading to cognition of symbolic material at the level required to understand the mechanisms involved. A specific problem is the lack of a term to describe the signal produced by the pretectum following the cross correlation of the image presented to the foveola in the process of reading. It appears that the pretectum is only able to handle syllables or short words in a single time frame of about 50-70 ms. The output produced by the pretectum is obviously not a complete thought, nor is it likely to be expressed as a complete thought. In this work, the term “interp” will be used as a noun to describe the vector created by the pretectum and the POS in response to stimulation of the foveola by a syllable, group of syllables or a word during a single 50-70 ms of a gaze interval. The pretectum presents to the higher cognitive centers this interp (in vector form) that can be incorporated into the saliency map. Alternately, it could present the vector to a short term saliency map used to collect time sequential vectors (interps) that are collected during a single gaze of approximately 220 ms. These interps would then be assembled into a complete perception of a message (percept) in vector form. The term percept is used in psychology to define the complete message. These individual interps will be accumulated within the POS or within Area 7 of the cerebral cortex. Multiple interps, when accumulated sequentially are presented to the higher cognitive centers as a percept. One or more percepts (possibly from multiple sensory systems) are then analyzed. This analysis leads to the cognition of a specific thought concerning a stimulus input.

The terms interp and percept are defined here out of expediency. No claim is made as to originality or precedent. They may be defined in a more sophisticated context in psychology. The definitions used here will eventually be correlated with those other terms.

19.8.4.1 Role of memory in reading

Ullman has provided some useful ideas related to reading and speaking language, particularly related to the roles of a “mental lexicon” and “mental grammar” along with thirteen pages of references31. Both of these subjects require a memory system like that discussed in Chapter 12 and 17 of “The Neuron and Neural System.” Ullman has also noted some inconsistencies in the interpretation of these terms in the psychology literature. The distinction to be made here is between those elements of language that are associated with declarative memory and either the perceptual or procedural portions of non-declarative memory. This work supports Ullman’s Declarative/Procedural model (DP) when the model is expanded to

28 Processes in Biological Vision

represent procedural memory as a subordinate portion of the broader non-declaratory memory category (Section 17.1.1) [xxx expand ]

19.8.4.2 Reading alone is achromatic

As developed in detail in Chapters 11 and 15, the signals of the analytical path connecting the foveola to the midbrain are not coded for chromatic information. All of the neural paths reaching the pretectum are treated equally, regardless of spectral representation, in the process of correlation discussed in Section 15.2.5. This leads to maximum spatial resolution in the overall process. It also leaves the outcome of the reading process achromatic. This is generally unimportant. If an idea of the color present at a relatively coarse level is desired, it is available from the same signals processed via the awareness mode of vision. It may be that such a coarse indication of color and color contrast is routinely attached to the signals passing through the circuitry of the analytical mode before they are stored in the Saliency Map. If so, the indication would be represented by another value attached to the vectors created during analysis.

While the reading process is achromatic in structure, presenting material in other than black and white causes a loss in signal-to-noise ratio within the data stream. This condition can result in a significant loss in legibility.

19.8.5 The mechanics and mechanisms of reading

As mentioned above, reading is the ultimate mechanism and capability provided by the analytical mode of visual system operation. In its full implementation, it is only currently found in homo sapien. This appears to be primarily due to the higher level of development of the Precision Optical System associated with the midbrain. It is this subsystem that implements most of the uniquely human capabilities of the human visual system. It is this subsystem that goes beyond the simple pattern recognition capabilities shared with all animals, and allows humans to interpret patterns of fine detail that other animals merely observe as “pretty” at best. This position is in agreement with Chomsky and Lenneberg who have argued that human language must be recognized as a species-specific ability based on unique bio-morphological structures.

The unique capability of reading is centered on the midbrain, and particularly the thalamus. This structure involves a large number of unique engines that perform a wide range of roles in vision. The roles of interest in reading center on the pretectum and a special engine labeled the control point in this work. These two engines form nodes that are shared with a variety of distinct servomechanisms and cognitive elements. These roles will be summarized below.

Traffic analysis, reported in Carpenter & Sutin show significant two-way connections between the cerebellum, the thalamus and the cerebrum. These connections form the foundation of the low level cognitive processes associated with reading. It will also be shown that most of the activity along this trunk originates at the TRN (the control point), proceeds to the pulvinar via the PGN, returns to the TRN and is then forwarded to the cerebrum, where additional high level cognition takes place.

19.8.5.1 Functional definitions needed to understand the reading process

Lacking clear terminology, this section will discuss two levels of cognition. Low level cognition is based largely on routine operations based heavily on previously learned relationships. It is characterized by the formation of individual and groups of interps, as well as the formation of individual and groups of percepts. An interp is the first-level output of the multi-dimensional correlator located within the pretectum. In reading, it corresponds to the internally coded representation of a syllable imaged on the foveola. A group of interps generally corresponds to a word. A percept is defined as the internal coded message corresponding to a series of words and constituting an idea. Such a group of words may be either a clause or a complete sentence.

Groups of percepts will be labeled ideas in this discussion. An idea is a complete thought generally equivalent to a complete sentence. For purposes of organization, the interps and percepts will be considered to be encoded in a low level “machine language” in the language of computers. They are assembled into groups of ideas that will be labeled themes. Ideas generally correlate with sentences while

themes are broadly equivalent to paragraphs. Ideas and themes will be considered to be “compiled” into a higher level language than the machine language of interps and percepts. These are the languages of the comprehended material. Prior to the first stage of comprehension, the information from the retina is passed to the pretectum in a form that will be described as raw data. There may also be a separate language used by the controllers, specialized neural engines designed to control the collection and dissemination of the collected and comprehended material. For purposes of discussion, this language will be called the control language. It is distinct from the machine and control languages used to handle acquired information. It is generally sent over separate, supervisory, neural channels connecting the various engines of the brain. These terms can be collected into an initial Table 19.8.5-1 for ease of reference and comparison. These terms may or may not be compatible with important documents in the literature. They are subject to change in the process of further definition and coordination.

### Table 19.8.5-1
**Summary of Terminology Describing Signal Formats**

<table>
<thead>
<tr>
<th>Term</th>
<th>Equivalent in Reading</th>
<th>Equivalent in a Graphic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interp</td>
<td>Syllable</td>
<td>Line or Point</td>
</tr>
<tr>
<td>Group of Interps</td>
<td>Word</td>
<td>Feature</td>
</tr>
<tr>
<td>Percept</td>
<td>Complex Word or clause</td>
<td>Recognizable object</td>
</tr>
<tr>
<td>Group of Percepts</td>
<td>Clause or sentence</td>
<td>Group of objects</td>
</tr>
<tr>
<td>Idea</td>
<td>Sentence</td>
<td>Image on foveola</td>
</tr>
<tr>
<td>Theme (or Scene)</td>
<td>Paragraph</td>
<td>Image within fovea</td>
</tr>
<tr>
<td>Act</td>
<td>An initial comprehension</td>
<td>Environ. within inst. field of view</td>
</tr>
<tr>
<td>Play</td>
<td>A complete comprehension</td>
<td>Total environment</td>
</tr>
</tbody>
</table>

Most of the mechanism involved in the cognitive aspects of reading are illustrated in Figure 19.8.4-1. This figure will also be used to discuss individual elements of the system in following paragraphs.

#### 19.8.5.2 Major functional steps in the reading process

Reading can be described as a highly stylized process involving a broad collection of neural engines found in the visual system, itself a major subsystem of the central nervous system. Its goal is to acquire information on any abstract subject via a highly stylized presentation of simple groupings of interconnected line elements. The stylistic aspects of the material include direction of sequential presentation of the groups, order of presentation of the groups, spacing of the groups, capitalizations, punctuations, and various methods of emphasis. On a different level, typesetters employ additional techniques to make the overall text pleasing to the eye. These include font design, variable spacing of individual characters within groups and various layout and pagination techniques.

The process of reading is dominated by a complex step and repeat process. The process is virtually automatic and the reader adopts a specific operating strategy that can be considered the default strategy. This default strategy is interrupted when the cognitive processes leading to comprehension are either interrupted or they do not achieve a sufficiently high performance threshold. When the interruption is due to a failure to recognize a simple grouping of line elements (generally representing a syllable), the system normally reverts to a learning mode. This mode involves a more detailed analysis of the group, in an attempt to learn its meaning, followed by an updating of the knowledge base of the individual with regard to that group. When an interruption is due to a failure to accept an idea or theme (generally because of failure of a syntax rule), a different strategy is used. First the control point orders the oculomotor system to return to a previous position and re-examine an interp that describes the tense of a word. In English, this interp appears at the end of a word (frequently a recognizable suffix different from the one assumed during the default reading procedure). In other languages, the interp may appear at other locations in a word or clause. Some languages use multiple suffixes. Any one of these may be the cause of the failure to accept a compiled idea or theme.

The performance level achieved in reading depends on a wide variety of disparate parameters. These range from the homeostasis of the individual to the light level of the material exhibiting the interconnected line element groups. They also include the familiarity of the subject with the language, idiom, semantics
and lexical form used in the material. The performance may also be influenced by the quality of both the physiological and neurological optics of the subject.

The default reading process can be tailored by the individual (largely unknowingly) during his early education. Techniques such as omitting the last syllables in long English words can lead to greater reading speed, but with a potential loss in precise comprehension. Similar methods of overlooking short prepositions can lead to higher reading speed but with a loss in precise comprehension. Comprehension of a multi-syllable word normally involves about 200 msec. This suggests that perusing a line of text for less than a full second invariably results in a loss in precise comprehension. “Speed reading” becomes a matter of trading off speed in completing a task versus the level of comprehension achieved. This tradeoff is directly related to the inherent information density of the material. If the material is particularly elementary, many of the individual words can be overlooked by a well educated subject. Only “10 Dow” in a sentence would be adequate to identify the residence of the British Prime Minister. Alternately the “Rose House” might not be enough to identify the residence of the Argentine President on Playa de Mayo near the obelisk in downtown Buenos Aires.

19.8.5.3 Memory is crucial to the reading process

Memory is an absolutely crucial element in the reading process. It is used in a number of discrete locations and for a variety of purposes. In general, the memory associated with reading can be divided into short term and long term types. These types can then be subdivided further with respect to their functional role. While the precise location of some of these memory types is not known, this situation is changing rapidly. The use of the new non-invasive in-vivo imaging techniques are beginning to provide site maps that may be used to further identify memory locations.

Short term memory is critically important in the assembly of information derived from an image into a serial signaling stream that can be used in cognition and comprehension by the brain. Long term memory is used for reference purposes. It is used at a high level, where recent themes can be compared to themes acquired long ago. It is also used at a very low level, as a lookup table in the conversion of raw image data into interps and percepts.

The method of controlling the flow of data and information into and out of these different memories is not understood at this time. However, the capabilities required are well within the capabilities of the typical feature extraction and command generation engines of the brain. It may be appropriate to define an engine specialized for the purpose of controlling the deposition of data in and the access to that data in a memory, a controller—as it is known in computers.

19.8.5.3.1 Types of short term memory

The short term memory used in the visual process can frequently be described as similar to the short term, shift register type memory of computers. It is used to hold specific values and to sequentially assemble short values into longer values.

It is common to speak of the content of a value in a shift register as a word. However, this will lead to great semantic difficulties here. In this discussion, a value will consist of a number of address slots filled by undefined characters that define a property of a feature. When multiple values are assembled into one larger value, the larger value will generally be considered a vector. Since the coding used to describe these characters, and to define properties to which they relate is still unknown, the discussion must be in a sense totally symbolic. The term word will be restricted to its use in human semantics.

The nature of short term memory in the visual system can be described with the aid of Figure 19.8.4-1. The elements along the diagonal line constitute the short term memory elements that can be described in terms of shift registers. They are typically filled during a time interval appropriate to their task (with some details below) and then refilled after the controller has called for the stored data to be forwarded to the next register. In the case of the initial interp assembler, this time interval will be taken as a nominal 50 msec. The time interval associated with the percept assembler will be taken as 200 msec. Both of these times assume the PROM located in the pulvinar has been able to provide a correct response to the individual interps and percepts. If not, the cycle times are extended while further analysis of the visual input from the retina is considered. This is the purpose of the oculomotor path shown. It causes the oculomotor portion of the POS to perform additional scans of the image and generate new or alternates information for processing into an initial interp by the multi-dimensional correlator.
Figure 19.8.4-1 The functional organization of the reading mechanisms including the types of memory employed. The interp assembler receives individual interps from the multi-dimensional correlator of the pretectum. The right hand two boxes are dotted to suggest the task can continue to higher levels of comprehension. The arrow from the idea assembler is also dotted because of its less defined status. The heavy lines are indicative of paths used when learning (adding new comprehensions) or correcting previous comprehensions. See text.

The pretectum contains a multi-dimensional correlator that “digests” individual character groups (or other symbolic groups and described here as a syllable) and creates an interp. This interp is essentially a description of the syllable in terms of a group of straight lines of finite length and specific arrangement. The nature of the code describing these lines is unknown. It is likely that the code group used to describe a specific syllable is very similar to the one used to describe a sound to be created in speech, in response to that syllable. If so, it simplifies the short term memory requirements associated with the overall neurological system. If so, it would also suggest that the same code group would be formed in response to an aural input describing the same sound. This scenario of code group formation is in concert with how children are taught to read and speak. They are essentially taught to assign a verbalization first to what they hear and then (generally a few years later) to the same syllable presented to them in text.

In the context of this discussion, the initial step in reading is to compare the interp created by the correlator to stored interps related to a specific sound. This is also the process used by young readers who speak the sound they have been taught corresponds to the syllable on the paper. As they progress, the young reader begins to verbalize sequences of syllables. In this context, the sequential grouping of interps (syllables) leads to a percept. A percept is a simple concept that is usually devoid of a purpose. The purpose is determined when several percepts are assembled into a theme. Such an assembly may incorporate a verb, in order to form a sentence. Alternately, it may only incorporate a series of words forming a clause. In either of these two cases, the group will be labeled an idea. As the ideas are assembled, they generate a theme. A theme includes a complete thought. In this context, an idea is correlated with a sentence and a theme is correlated with a paragraph. For purposes of illustration, the figure shows the themes being assembled further into an act and the acts into a play. In this analogy, a theme or paragraph could be labeled a scene in a play.

The remainder of the figure is to suggest that the subject is not aware of the individual interps or percepts.
In fact, his interest is not in the sentences. His cognitive processes are focused on the themes, acts and plays. This is a description of the cognitive processes involved in seeing. The subject is only conscious of the cognition related to the themes, acts and plays. He has no control over the lower level processes (except the ability to verbalize them in sequence).

19.8.5.3.2 Types of long term memory

19.8.5.4 The input to the cognitive portions of the reading mechanism

19.8.5.4.1 The Precision Optical System (aka Auxiliary Optical System)

19.8.5.5 The role of the control point in reading EMPTY

19.8.5.6 The role of the thalamus in reading EMPTY

19.8.5.7 The role of the cerebellum (and/or pulvinar) in reading EMPTY

19.8.5.8 The ultimate result of the cognitive portion of the reading mechanism

19.8.6 Performance in reading and analysis of fine detail EMPTY

Reading performance is highly dependent on the level and the quality of the light used to image the material. This is shown clearly in the description of the spatial contrast function as a function of signal to noise ratio. The criticality is stressed when the performance at a given signal to noise ratio is restricted to a small, square test stimulus.

19.8.6.1 Reserved

19.8.6.2 Exploratory research related to crowding

The psychology community has recently focused on the observable characteristics of reading related to the number of characters that can be viewed simultaneously during the process of reading. Most of the work has employed characters considerably larger than limiting resolution size characters. The work has not relied upon any model of the reading process and has generally relied upon only a few relationships that might be called first order laws. The so-called Bouma’s Law was developed in 1978 based on very limited experiments coarsely spread across the angular field of view33. The extent of the data collected parallels the data collected by Anstis relating to acuity as a function of field angle34. His data was limited to angles greater than 5 degrees from the point of fixation. Pelli et al. made an apparently cursory review of Bouma’s work in their recent paper exploring the psychology based concept of crowding in reading35. Their discussion on page 6 did not recognize the actual graph presented by Bouma in 1978 (figure 2).
While Bouma did not provide an equation, his graph shows a finite “resolvable detail” at zero eccentricity (related to Pelli’s s) and a clearly parabolic characteristic with eccentricity (extending to 50°). The finite resolvable detail at zero eccentricity is approximately 0.02° as expected. This value is consistent with a critical character-to-character spacing of nominally 0.08° for the 20/20 line on a Snellen Eye Chart.

The Journal of Vision, volume 7, issue 2, 2007, has provided a special issue related to the subject of crowding in relation to the rate of reading, and other variables.

Four papers have been reviewed in this area and are included in the files.


The experiments in these papers were short on protocol development. A variety of different fonts were used, some with serifs and some san serif, some with constant character spacing and some with variable spacing (although not noted in the text of the papers). The greatest shortcomings were in two areas: first, the lack of any model of how reading is accomplished and second, the use of large characters to determine the limits of the capabilities of the neural system to accomplish reading.

It is clear that considerable more work is needed to describe the relationship between the span of characters and the reading rate.

19.8.6.3 Reserved

19.8.7 Progress in machine vision related to reading

Recent progress in machine reading has been impressive, to the point that virtually every automated teller machine (ATM) incorporates check reading software that identifies and reads virtually every feature entered by handwriting or printing. A2iA appears to be the leader in the implementation of such technology. A wide range of papers have reported on variety of mathematical approaches to machine reading that extend well beyond the optical character readers of the past. Most involve some type of Markov Field analysis. A specialization in this area is a family of Boltzmann Machines designed to analyze two-dimensional scenes. The challenge of recognizing “holistic” character sets (words or multiple syllable groups within long words) appears to have been solved. Marti & Bunke have presented a variety of papers on their investigations36,37. Lavrenko et al. have also presented a useful paper38. These works and others will be discussed more completely in Section 19.10.5.

19.9 Listening to speech

19.9.1 Historical description of Broca’s and Wernicke’s areas EMPTY

Boca’s area, BA 44-45. Wernicke’s area BA 22, 41-42 See Hearing chapter 8 and vision 18.9.1

38Lavrenko, V. Rath, T. & Mannmatha, R. (xxx) Holistic Word Recognition for Handwritten Historical Documents
http://pdf.aminer.org/000/347/156/likelihood_word_image_generation_model_for_word_recognition.pdf
19.9.2 More recent descriptions of brain areas

With the advent of non-invasive imaging of the body using functional Magnetic Resonance Imaging (fMRI) and Computer-Aided Tomography (CAT), the latter a form of X-ray imaging, describing the areas of the brain involved in almost any conceivable sensory (or even internal) stimulation has become a cottage industry. The major point to recognize is that even the 3.0 Tesla fMRI scanning machines have a resolution, defined by a voxel (on the order of a few mm on a side). A voxel of this size contains on the order of several million individual neurons. Thus, uncovering the details of neural operation is not likely using this technique. Only major traffic flow patterns are recognized at this resolution level.

Indefrey & Cutler have embarked on defining regions of the brain involved in speech recognition more precisely using Talairach coordinates. They focused on words, sentences, pseudo-words and tones. Their focus was on the surface of the temporal lobe. Their material needs to be reviewed with care as they do not review the role of the PGN, MGN or pulvinar in listening to speech.

The role of the diencephalon (PGN, LGN, MGN and pulvinar) has only recently (2010 or later) appeared in the brain imaging experiments reported in the literature.

19.10 Generic information extraction leading to object and face recognition

As a professor (a long term lead investigator in this field) has recently expressed to an associate, “the field of face recognition in cognitive psychology has not really progressed for two decades!!!” The reason is fundamental. The behavioral community has had neither a physiological framework, a functional neurological model or a neurological schematic to guide their psychophysical investigations. This work, and particularly Section 7.5 and the earlier sections of this chapter can provide such a set of guidelines.

The neural system does not have a strong computational capability. As discussed earlier (Section xxx ) it can not perform transcendental calculations directly (although it can obtain the necessary answers to such mathematical transforms through anatomical computation). It relies upon pattern matching techniques to achieve most of its information extraction capabilities, after an initial period of learning that involves storing specific patterns.

The key to understanding the mechanism of object (including face) recognition is to recognize the two-step mechanism of;
• coarse full field interpretation of the visual environment via the LGN/occipital couple with storage in the saliency map,
• high resolution sub-field (foveola field) interpretation via the PGN/pulvinar with the resulting interp(s) stored as an overlay on the full field saliency map (possibly in metadata form).
• comparison of the resulting saliency map of the same visual field with previous maps in long term storage.

This two-step procedure will be labeled the PEEP (Programmed Element Evaluation in Perception) procedure below. The procedure is highly dependent on the mechanism of normalization of the perspective of objects in the environment as part of the stage 4 interpretation and information extraction process.

Yarbus (Chapter 3) uses the simpler term, “fixation process” to describe only the motion of the eyes during the more complex PEEP procedure. He did not consider the neural examination of the image on the fovea that occurs during each individual fixation.

The successes in facial recognition made in the artificial intelligence field are applauded. However, these successes have been based on brute force algorithms with little in common with the functionality of the visual modality of the neural system.

It is appropriate to settle on some terminology before proceeding. Based on the earlier work in this chapter, it is suggested that the stage 4 and 5 elements of the neural system accomplish several functions in sequence.

---

• **Detection**— Typically, the becoming aware of a feature in visuotopic space that is unexpected, appearing at a different location than anticipated or moving at an unexpected rate. Occurs initially within the awareness and alarm pathways of vision.

• **Perception**— An initial interp is established by the stage 4 engines that can be associated with the perception of an object at a specific location in external inertial space. An awareness and alarm pathway function.

• **Identification**— This perception is normally processed further (by the analytical pathway) to establish an identification of the character of the object.

• **Recognition**— This identified interp is then compared with previously identified and stored interps in order to establish a recognition of the object within its inertial context. Recognition is an analytical pathway function.

**Pattern recognition**— A global term used to describe the mechanism of recognition of a pattern whether the pattern originates from alphabetical characters (reading), inanimate objects (object recognition) or animate objects (face or other feature recognition). Defined as early as 1973 by Ditchburn.

While others have used the same terms with different definitions\(^\text{40}\), that is to be expected in the absence of a standards setting organization such as found in the field of visual performance beyond the luminance and chrominance standards.

### 19.10.1 Historical background focused on object (face) recognition

Arbib & Hanson edited a large text at the dawn of the computer age and leading into the area of object recognition associated with robotics\(^\text{41}\). While it now appears primitive, contains little physiological information and uses archaic terminology, it includes an overview of 83 pages that describes the state of the art at that time in detail. They did stress the critical character of parallel computation compared to serial computation when treating imagery.

The academic literature of object recognition has focused primarily on face recognition primarily for historical reasons. Scientists sought to understand face recognition long before other objects of equal or greater spatial complexity became commonplace. No technical rationale has been offered for considering face recognition to play a larger role in object recognition than say military photo-interpretation, or understanding the artwork associated with microcircuit manufacturing.

Rayner et al. have provided an extensive bibliography associated with eye movements in visual cognition\(^\text{42}\).

It is also useful to define what prosopagnosia is and is not. **Figure 19.10.1-1** is misleading. It was prepared for the popular press using PhotoShop in a very crude modification. As will be shown below, sufferers of prosopagnosia typically see faces as normal but can not identify them. Bruce & Young express the disease as, “Prosopagnosic patients can, for instance, recognize faces as such even though they do not know whose face they are seeing (with three citations attached).” Other clinicians describe how their patients can describe facial details of a subject in great detail but cannot attach a name to the face. Caldara et al. have provided excellent information drawn from a constrained set of experiments\(^\text{43}\). They describe the ability of their patient, PS, as follows: “Her deficit truly appears to be restricted to the category of faces. With faces, PS is able to categorize a face as a face, . . .” However, “this is in stark contrast with her inability to recognize previously seen or familiar faces and to match unfamiliar faces.” Prosopagnosia frequently occurs as the result of traumatic injury but may also be present congenitally.

---


The perceptions the patient has are derived from the LGN/occipital couple pathway of the visual modality. Based on the above definitions, the subjects have achieved identification of a familiar face belonging to a person in front of them. However, they cannot assign a name to that face. It is the higher performance, PGN/pulvinar couple pathway that is not functioning properly. As a result the features of the face are not evaluated properly in order to achieve recognition. The related details will become clearer in the subsequent discussion. Because of the parallel signal pathway, a person with severe prosopagnosia can still identify and describe a face in detail.

Bruce & Young provided a paper in 1986 that reviewed possible frameworks addressing object, face and word recognition. The paper is primarily conceptual and begins with the assertion, “A photograph or other picture of a face will lead to the generation of a pictorial code.” Clearly a live subject standing in front of an observer will also generate what they call a pictorial code. They go on, “Nor is it simply equivalent to the viewer-centered information derived when a picture is viewed, since what we term a ‘pictorial’ code is at a more abstract level, at which information from successive fixations has been integrated.” The paper is worthy of careful study. They did not address the physiological or neurological aspects of object recognition. They did identify at least five distinct types of codes within their overall coding schema.

Valentine provided a paper attempting to develop a framework for investigating face recognition in 1991. He spoke in terms of a multi-dimensional space for describing facial recognition but in fact only described a two-dimensional space occupied by a group of vectors. Each vector was presumably associated with a relevant facial feature. He described, “Two specific models within this framework are identified: a norm-based coding model, in which faces are encoded as vectors from a population norm or prototype; and a purely exemplar-based model.” “The norm-based coding version and the exemplar-based version of the framework cannot be distinguished on the basis of the experiments reported, but it is argued that a multidimensional space provides a useful heuristic framework to investigate recognition of faces.”

Valentine presents an excellent overview of the face recognition problem in text form but his accompanying graphics are ony conceptual. He does define an “associated error estimate” when comparing one record of a face and a second record of the same or similar face. He also accurately describes the limitations of the state of the art at that time, “There are a number of important and as yet unanswered issues raised by the application of a multidimensional space framework to face recognition. Most notably, the dimensions of the space have not been identified, nor has the dimensionality of the space required. The determinants of similarity have also not been precisely specified-for example, the metric of the space has not been determined. In view of our currently imprecise knowledge of the dimensions underlying face perception, and therefore our impoverished understanding of similarity between faces, it would be premature to precisely specify the parameters of the framework.” He makes no assertion that his framework reflects any physiological or neurological model.

Figure 19.10.1-1 Misleading caricature of prosopagnosia. People with this condition perceive faces normally but are unable to assign a name to even familiar faces. See text. From BodyGeeks.com

---

44Bruce, V. & Young, A. (1986) Understanding face recognition Brit J Psychol vol 77, pp 305-327
Tanaka & Farah provided a paper in 1993 that was heavy on text but did provide a very few figures comparing similar features associated with different faces.

Nelson provided another paper heavy on text in 2001 while focused on infant and child development relative to face recognition. He focuses attention initially on the potential differences between face and object recognition. After reviewing many articles, his conclusion appears to remain unsatisfying. He then reviews what he describes as three discrete models. He concludes, “Overall, the bulk of the evidence suggests that the ability to recognize faces is one that is learned. Through exposure to faces, tissue in the inferotemporal cortex becomes specialized for face recognition, and, in theory, continued exposure to faces maintains this tissue until it becomes dedicated to face recognition (which is not to say this same tissue cannot be used for other related purposes, such as recognizing stimuli like greebles).” Bizarre computer-generated shapes are frequently called “greebles”. “Regardless, it appears that this specialization occurs rapidly, within the first months of life. With greater experience with faces, a form of perceptual learning takes place that further develops this tissue.” No discussion of the neural system was provided beyond conceptual generalities.

Sinha et al. have provided a review of what was known about human object recognition for the benefit of computer vision researchers in 2005. “In this paper, we present what we believe are 20 basic results, with direct implications for the design of computational systems. Each result is described briefly, illustrated, and appropriate pointers are provided to permit an in-depth study of any particular result. The paper provides a useful review for its intended purpose. Unfortunately, the paper does not address the physiology or neural circuitry actually used to achieve human object recognition. Their introduction begins,

“Notwithstanding the extensive research effort that has gone into computational face recognition algorithms, we are yet to see a system that can be deployed effectively in an unconstrained setting, with all of the attendant variability in imaging parameters such as sensor noise, viewing distance, and illumination. The only system that does seem to work well in the face of these challenges is the human visual system.”

Their Result 9 is that, “Vertical inversion dramatically reduces recognition performance. Upside down (‘inverted’) faces are harder to recognize than right-side up faces, despite the fact that the same information is present in both images.” They appear to overlook the fact that the “information” related to orientation is different and that recognition involves pattern comparison between the currently presented image and prior stored images. There are plenty of studies showing that within a matter of days at most, humans can adapt to wearing scene inverting glasses as their brains store new information from their recent saliency maps to compare with their current environment.

The concentration on whether a face is inverted or not in the literature is intriguing. Have the writers ever considered or measured the result of rotating a face only 90 degrees. this degree of rotation could be even more difficult to recognize than the inverted image.

Their result 18 appears particularly relevant, “However, it must be noted that the debate over faces being ‘special’ is far from over. It has been suggested that rather than being a true ‘face module’, the FFA may be responsible for performing either subordinate or ‘expert-level’ categorization of generic objects. There are results from both behavioral studies (Diamond & Carey, 1986; Gauthier & Tarr, 1997) and neuroimaging studies (Gauthier, Anderson, Tarr, Skudlarski, & Gore, 1997) that lend some support to this ‘perceptual expertise’ account.”

They close with, “Our presentation of results in this paper is driven by the goal of furthering cross-talk between the two disciplines. The observations included here constitute twenty brief vignettes into what is surely a most impressive and rather complex biological system. We hope that these vignettes will help in the ongoing computer vision initiatives to create face recognition systems that can match, and eventually exceed, the capabilities of their human counterparts.”

---

19.10.1.1 Mathematical concepts developed in connection with Neural Networks

The following discussion will incorporate considerable material from the rapidly developing field of (Artificial)Neural Networks, sometimes described by the acronym, ANN. A valuable resource in this area, although dated even in the third edition with respect to the neuron, is Haykin49. Figure 19.10.1-2 shows a commonly illustrated “neuron” as used in artificial neuron technology on the left. The right frame uses the same symbology to illustrate a real neuron of biology. There are several basic differences. The most important was noted by Haykin on page 10 (1999 edition),

“Unlike a synapse in the brain, the synaptic weight of an artificial neuron may lie in a range that includes negative as well as positive values.” This finding is a feature of all biological neurons, not just in the brain.

The second critically important difference involves the output of the basic neuron. Whereas, the ANN community assumes all neurons generate action potentials via the activation function shown, this is not the case for 95% of the neurons within a biological system. Only neurons employed as stage 3 encoders are arranged to perform the conventional summing and differencing operations of a stage 2 (and other stage neurons) as well as the action potential generation function exclusive to stage 3 neurons. These two facts are frequently overlooked by students and practitioners working in these two communities.

The normal inverting capability of the biological neuron is not recognized and bipolar weighting functions are used to compensate. Every neuron is assumed to generate action potentials via the activation function shown. Right; the biological neuron has two distinct input terminals, the non-inverting apical dendritic “tree” and the basal poditic “tree.” Each tree sums the signals applied to it and the differencing operation is achieved at the three-terminal Activa within the neuron. Stage 2 biological neurons do not generate action potentials unless they are hybridized with a stage 3 neuron.

The normal inverting capability of the biological neuron is not recognized and bipolar weighting functions

---

Figure 19.10.1-2 A graphical comparison of ANN compared to the biological neuron. Left; the conventional neuron of artificial neuron networks (ANN). The normal inverting capability of the biological neuron is not recognized and bipolar weighting functions are used to compensate. Every neuron is assumed to generate action potentials via the activation function shown. Right; the biological neuron has two distinct input terminals, the non-inverting apical dendritic “tree” and the basal poditic “tree.” Each tree sums the signals applied to it and the differencing operation is achieved at the three-terminal Activa within the neuron. Stage 2 biological neurons do not generate action potentials unless they are hybridized with a stage 3 neuron.
are used to compensate. Every neuron is assumed to generate action potentials via the activation function shown. Right: the biological neuron has two distinct input terminals, the non-inverting apical dendritic “tree” and the basal poditic “tree.” Each tree sums the signals applied to it and the differencing operation is achieved at the three-terminal Activa within the neuron. Stage 2 biological neurons do not generate action potentials unless they are hybridized with a stage 3 neuron.

Specifically, the weighting functions in figure 1-5 & 1.7 are limited to positive values in biological neurons as noted by Haykin. The negation operation is performed within the neuron. The apical dendrites of figure 1.2 provide the summation without negation of the sum while the basal dendrites (the podites) of the neuron perform summation followed by negation of the sum of all poditic inputs. The title should not limit the diagram to “pyramidal” cells. The figure is generic to the neuron portion (if present) of any biological cell. Other problems in the text relate to the limited physiological knowledge of Haykin with regard to the biological neural system. When discussing the auditory modality of the bat, he does not recognize the operation of the cochlea as a spectrometer. Second, he does not recognize that the outputs of the spectrometer are immediately demodulated, and only the envelope of the high frequency signals received by the auditory modality are processed within the neural system of the bat. Unfortunately, a large community has arisen based on the framework that Haykin espoused in his multiple editions of the book. Fortunately, most of the mathematical modeling of the neuron can be reinterpreted successfully as discussed earlier in this paragraph. Chapter 2 of the 2nd edition is a comprehensive discussion of potential learning machines. It provides a broad framework that can be applied to the neural network of any biological species. Overall, Haykin’s book is a tremendous asset.

A more introductory discussion of ANN ca 1996 was provided by Stergiou and Siganos as a class report. Their presentation follows the conventional ANN pedagogy and therefore suffers the same problems as discussed relative to Haykin’s work. They begin their 1996 discussion with, “Much is still unknown about how the brain trains itself to process information, so theories abound.” They go on, “However because our knowledge of neurones is incomplete and our computing power is limited, our models are necessarily gross idealisations of real networks of neurones.” These interpretations lead to needless complications in more recent work attempting to implement simulations of neuron networks in VLSI on silicon technology. Strangely, they do not cite Haykin.

Figure 19.10.1-3 provides a summary comparison of the elements used in an ANN of machine vision versus the biological neural network elements. Note the Node of Ranvier is not typically defined in ANN even though its role is critically important in the biological system. Similarly, the synapse is typically modeled without significant dependence on the features of the conceptually defined chemical synapse.

---

Chapter 2 of Haykin’s 2nd edition is a comprehensive discussion of potential learning machines. It provides a broad framework that can be applied to the neural network of any biological species. Overall,

<table>
<thead>
<tr>
<th>Artificial Neural Network</th>
<th>Biological Neural Network</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The Neuron</strong></td>
<td></td>
</tr>
<tr>
<td>A two-terminal device based on unknown operating principles</td>
<td>A three-terminal device incorporating a PNP junction type electrolyte semiconductor</td>
</tr>
<tr>
<td>Active element is an individual lemma typically associated with the axon (output) of the cell</td>
<td>Active element is a junction between two specialized lemma areas separated by a thin “base” region of water</td>
</tr>
<tr>
<td>Operation is described using three unsolved partial differential equations.</td>
<td>Operation is described using the Ebera-Moll model for any three-terminal low frequency semiconductor device</td>
</tr>
<tr>
<td>Topological relationship of dendritic terminal to Axon is undefined. Output &amp; input of same polarity</td>
<td>Dendritic terminal is a non-inverting input terminal</td>
</tr>
<tr>
<td>Axon output is assumed to be a binary positive monopulse in 100% of the cases (except in retina)</td>
<td>Axon output is typically a biphase analog waveform (96%) Axon output in stage 3 neurons is a positive going monopulse waveform (5%)</td>
</tr>
<tr>
<td><strong>The Synapse</strong></td>
<td></td>
</tr>
<tr>
<td>Generally described mathematically but relying upon the chemical theory of the neuron.</td>
<td>A three-terminal PNP junction device between an axon and a dendrite or podile wired as an “active diode”</td>
</tr>
<tr>
<td>Modeled as switch associated with a weighting that is a bipolar function</td>
<td>Operating as an analog device capable of transferring either analog or monopulse signals without inversion</td>
</tr>
<tr>
<td>Electrical properties are not usually modeled</td>
<td>Time constants and capacitances are negligibly short relative to the bandwidth of neural signals.</td>
</tr>
<tr>
<td>How weighting functions are established is not normally modeled</td>
<td>Impedance varies with size of PNP junction. Impedance does not vary within operating time scales</td>
</tr>
<tr>
<td><strong>Node of Ranvier</strong></td>
<td></td>
</tr>
<tr>
<td>Typically undefined in Artificial Neuron Networks</td>
<td>3-terminal PNP junction device used as monopulse signal regenerator in Stage 3 neurons. Generates monopulse output using internal feedback in podicle (base) circuit.</td>
</tr>
<tr>
<td><strong>Calculations</strong></td>
<td></td>
</tr>
<tr>
<td>Heavily dependent on transcendental calculations</td>
<td>No transcendental calculations</td>
</tr>
</tbody>
</table>

Figure 19.10.1-3 A tabular comparison of artificial and biological neural networks. The Node of Ranvier is normally not modeled since it plays no role in simulations using a digital computer. The synapse is not completely modeled because it is seldom a limiting circuit element. The neuron is typically modeled as a monopulse oscillator although that configuration plays a statistically minor role in the biological system.

Haykin’s book is a tremendous asset. Fukushima has described a “neocognitron” supported by some drawings of multiple layer Boltzman machines that are useful\(^5\). The work dates from 1980 with many expansions since then. Unfortunately, his descriptions of the “real neural system” are totally unrelated to the real biological neural system. His seven or eight hidden layers do not appear compatible with the histology of the visual cortex.

Figure 19.10.1-4 is a composite of three figures from Haykin (1.16, 1.17 & 1.18 of the 2nd Edition) explaining his concepts of two fundamental networks of Neural Network theory. Without his pedagogy related to these figures, no outsider can expect to understand the circuit diagrams of ANN. The left frame shows the basic concept of a recurrent network with the restriction that the output is returned, after a unit

delay, to all of the inputs except its own. The center frame shows the incorporation of this recurrent network into a signal processing channel consisting of two inputs and two outputs. The feedback and delay elements are not “seen” by the overall network and are designated as “hidden.” The right frame shows the shorthand used for a network incorporating the recurrent networks within the “hidden layer.” The goal is to reduce the proliferation of inputs into a lower-dimensional output. Treating the circuits of either the center or right frames as a black box and then performing conventional circuit analysis to determine the circuit within the black box would lead to identification of the hidden layer and all interconnections. With this information, the circuit could be reproduced using realization techniques. However, this procedure will generally fail in this situation for at least two reasons. Beyond more than a few inputs, the results of the analytical effort would probably be indeterminate. Second, these are dynamic circuits that are not amenable to analysis by static black box analysis techniques except under quiescent conditions.

Haykin addresses the subject of how to build prior information into the final network. “An important issue that has to be addressed, of course, is how to develop a specialized structure by building prior information into its design. Unfortunately, there are currently (1999) no well-defined rules for doing this; rather, we have some ad-hoc procedures that are known to yield useful results. In particular, we may use a combination of two techniques (LeCun et al.):

1. Restricting the network architecture through the use of local connections known as receptive fields.
2. Constraining the choice of synaptic weights through the use of weight-sharing.

These two techniques, particularly the latter one, have a profitable side benefit: the number of free parameters in the network is reduced significantly.”

By returning the signals uniformly (with equal weighting) across all signaling channels after a unit delay to the poditic (inverting) terminals of the neuron, a frame-to-frame differencing function is accomplished within the data stream. The overall circuit can be implemented in either analog or binary formats and either under synchronous as asynchronous control.

On page 24, Haykin asserts, “A major task for a neural network is to learn a model of the world (environment in which it is embedded and to maintain the model sufficiently consistent with the real world so as to achieve the specified goals of the application of interest.” He goes on, “The subject of knowledge
representation inside an artificial network is, however, very complicated. Nevertheless, there are four rules for knowledge representation that are of a general commonsense nature (Anderson, 1988).

**Rule 1.** Similar inputs from similar classes should usually produce similar representations inside the network, and should therefore be classified as belonging to the same category.

**Rule 2.** Items to be categorized as separate classes should be given widely different representations in the network. —The second rule is the exact opposite of Rule 1.—

**Rule 3.** If a particular feature is important, then there should be a large number of neurons involved in the representation of that item in the network.

**Rule 4.** Prior information and invariances should be built into the design of a neural network, thereby simplifying the network design by not having to learn them.

Rule 4 is particularly important because proper adherence to it results in a neural network with a **specialized (restricted) structure.**

Based on this material from Haykin, the general framework for object recognition within the biological vision and hearing modalities can be appreciated.

### 19.10.1.2 Recent conceptual approaches to understand face recognition

Gosselin & Schyns developed a concept for quantifying the performance, if not the underlying mechanisms, of face recognition in 2001. Their abstract begins, “Everyday, people flexibly perform different categorizations of common faces, objects and scenes. Intuition and scattered evidence suggest that these categorizations require the use of different visual information from the input. However, there is no unifying method, based on the categorization performance of subjects, that can isolate the information used. To this end, we developed Bubbles, a general technique that can assign the credit of human categorization performance to specific visual information.” The paper concludes with, “Note that the principles of Bubbles are not limited to faces but are also applicable to other object and scene categorizations. The technique is a human search for diagnostic features in any specified n-dimensional image generation space, even if the space is abstract.”

An unsupported, and possible misleading suggestion to future investigators, is the reliance of their procedure on spatial frequency segmentation followed by a search protocol based on a random walking small Gaussian aperture across the entire visual field. It has been shown in this work that the animal neural system does not employ spatial frequency analysis or any form of neural computation involving transcendental calculation (i.e. Fourier transforms or correlation functions). It does achieve the same results as these mathematical transforms in some cases by using spatial rearrangements (computational anatomy) among arrays of neural axons arriving at a given stage 4 engines (See **Section 8.4.2** in “Hearing: A 21st Century Paradigm” concerning the implementation of a Riemann transform by computational anatomy within the auditory modality of the neural system).

To avoid confusion, it must be noted that the only place signals are processed in the frequency domain in the hearing modality of the neural system is in stage 0 (the non-neural stage including the physical acoustics of the cochlea).

“Versatile categorizations are not restricted to faces. People can typically classify a given object as a car, at the basic-level, a vehicle at the superordinate level, and as a Porsche at the subordinate level, when they know this expert categorization.”

“From the outset, it is important to emphasize that the aim of this paper is to illustrate the fundamental principles of Bubbles in the context of simple, but In Bubbles, the observer determines the informative subset of a randomly, and sparsely sampled search space.”

They describe the images used in Bubbles, “Each face was partly revealed by a mid-grey mask punctured

---

*Vision Res* vol 41, pp 2261–2271
by a number of randomly located Gaussian windows (henceforth called ‘bubbles’) with standard deviation of 0.22’ of visual angle, see Fig. 3c for examples. We chose bubbles with a Gaussian shape because it is smooth and symmetrical (see Marr, 1982).

During the experiment, the number of bubbles per image was automatically adjusted, using an adaptive procedure, to reveal just enough face information to maintain a 75% correct categorization criterion (Bubbles is a self-calibrating technique). The size of the bubbles and the self-calibration are important aspects of the technique that we will discuss in the results section.”

“It is important to stress that subjects were not under any time pressure to respond and so could freely explore each stimulus. The experiment comprised a total of 512 trials (16 presentations of the 32 faces). A chinrest was used to maintain a constant viewing distance of 100 cm. Stimuli subtended 5.72×5.72’ of visual angle on the screen.” This is approximately the size of the nominal human fovea of this work.

**Figure 19.10.1-5** reproduces a modified figure 2 from the Gosselin & Schyns paper. Their algorithm employed a randomly located small Gaussian aperture that traveled (caricatured at several locations at upper left) over the 5.72 degree scene. Their algorithm highlighted the areas where a portion of the face can be seen (in the absence of the large circles). The highlighted areas are a good approximation of the 1.2 degree size of the image presented to the foveola (the large circles) for a line of fixation centered within each circle during a presumed sequence of saccades and gazes (sometimes labeled a Levy flight or a joconde–dance). The areas highlighted by their algorithm do match the diameter expected to be interrogated by the human foveola during a saccades/gaze sequence (the PEEP procedure defined below). The sensitivity pattern of the foveola however is formed by an area of constant sensitivity within each circle. The text labels correspond to the labels assigned during their first set of experiments. EXNEX refers to an expressionless face. The hair lines of all their images were standardized, as was the global orientation (full face presentation) and lighting, to avoid impacting the results. The subject was allowed to observe the face through the small randomly moving apertures. When they believed they could identify the face, based on previous training, they stopped the experiment and the number of distinct aperture views employed up to that point were noted. “With enough trials, a random search is exhaustive and all the search space is explored.”
After a very complicated numerical computation involving several sorts of their data, the algorithm develops what are called “diagnostic masks.” These masks are much larger than the size of the test apertures. The diagnostic masks do appear to approximate the size of the individual images formed on the foveola during the individual gazes in the normal operation of the human eyes. A set of simple circles have been added to the figures to represent the “diagnostic masks” as developed from human physiology later in this section. Only a few of these 1.2 degree circles equivalent to the size of the foveola are shown overlaying portions of their diagnostic masks.

McKone, Crookes & Kanwisher provided an early paper including the results of fMRI experiments that is
quite illuminating, although their figure appears to be subject to a lot of discussion. Their figure 1 highlights multiple locations within the human brain associated with face recognition. The original caption for this figure is quite complicated and uses some annotation different from that in the figure. The figure highlights the “matched operations” occurring on both sides of the brain (although singular labels are frequently used in the brain literature).

Figure 19.10.1-6 Areas associated with face recognition based on fMRI. Three face-selective regions are shown: the FFA in the fusiform gyrus along the ventral part of the brain, the OFA in the lateral occipital area and the fSTS in the posterior region of the superior temporal sulcus. For studies of face identification (rather than expression, etc), the FFA and OFA are of greatest interest. See text. From McKone et al., 2008.

Brewer published a short (typed) book on face recognition in 2011. It presents a very broad discussion of what is involved in face recognition in a context similar to this work. Pages 3 through 8 provide several clear summaries of other peoples work. Many phenomena are examined from the perspectives of different authors and conclusions are then drawn relative to these particular pieces of the larger question. A few experiments were performed using strangely altered faces of commonly known people (Heather Langenkamp). The discussion is from a social sciences perspective and little discussion of the neural system is included except anecdotal references to more detailed work.

McKone, E. Crookes, K. & Kanwisher, N. (as of 2008) The cognitive and neural development of face recognition in humans (awaiting publication)

Atkinson & Adolphs presented a review in 2011 on the neuropsychology of face perception\(^5\). It contained only one figure, a comparison of the face processing model of (a) Bruce & Young and (b) Haxby et al. involving conceptual block diagrams. They do suggest a more complex schematic than a simple feed-forward configuration but they do not develop any details relative to their proposals. They do provide a comprehensive review of fMRI, transcranial magnetic stimulation (TMS) studies and other non-invasive topological studies.

They do draw one conclusion supporting the hypothesis of this work, “We have seen that two patients with lesions encompassing much of lateral occipital cortex (one in the right hemisphere, one bilateral) with consequent lack of the normal face-selective fMRI response in this region nevertheless showed face-selective responses in spared regions of fusiform cortex and STS [three citations]. Both patients are able to discriminate faces from non-face objects, which is reflected in normal face-selectivity in right FFA of both patients despite the lateral occipital lesions, implying that the relevant information must be reaching FFA via a processing route that bypasses OFA.” The PGN/pulvinar couple pathway is the logical answer to this implied question. Without noting the limited resolution of the techniques they reported on, they do assert, “It also becomes critical to conduct group analyses, ideally on sample sizes in excess of 100 patients or so, such that deficits can be traced reliably to shared damage to a particular region, rather than attributed to the idiosyncracies of any individual lesion. One approach—the classic one—to address this problem, of course, has been to study patients with highly selective lesions. But such patients are exceedingly rare and still suffer from the fact that each is essentially unique. We need to know whether the results generalize, and we need to know this about neuroanatomical regions so specific that they are almost never lesioned selectively. Current approaches are using voxel-based lesion–symptom mapping with samples of several hundred to address these issues.”

With the advent of MRI and particularly fMRI, a number of laboratories and major research programs have arisen to study face (object) recognition. The program at the University of Louvain, Belgium is particularly well documented. The work of several of their investigators is discussed below. Gauthier is a current investigator into the association areas involved in face recognition\(^6\). She has focused on the fusiform face area (FFA) but has explored the possibility that it is involved in the broader task of object recognition. A paper apparently based on her 1997 PhD thesis provides a structure that will be discussed more fully in Section 19.10.3.5.

### 19.10.1.3 Features in presumed order of importance BRIEF

Standardizing the hair line in face recognition is a very poor idea. Features of the hair are some of the most important features in face recognition. The scalp line, and the facial hair (eye brows, eye lashes, mustache and beard) are critically important to face recognition. Many movie stars have not changed their hair features in 40 years because it is such an important part of their persona.

*Figure 19.10.1-7* shows an “awkward” photo illustrating the challenge of facial hair in face recognition. Except for the glasses on the father, the principle discriminator in these images is the hair line, the cruvature of the eye brows and the detailed shape of the mustaches.

---


\(^6\)http://en.wikipedia.org/wiki/Isabel_Gauthier
A comment on line by Pollik (04 May 2014 10:40am) in response to an article about computerized face recognition is suggestive of the actual task: 108%

97% is not enough for any system to rely on for security.

People don't just use the face...they use the face in conjunction with other information - voice, height, weight, signifying marks, knowledge displayed in conversation, mannerisms and body language, expectation of seeing a person in that place, smell, dress sense, gait.

The comment referenced an article in the Guardian/Observer citing the work of Taigman and associates. It included a low resolution image, Figure 19.10.1-8, indicating the brute force approach being pursued in the computer-vision community to achieve face recognition. [xxx need more text after full paper is issued. ]

19.10.1.4 Efforts to show upper or lower field preference in recognition

The psychological field has been attempting to determine the preferred area of the visual field for face recognition without recognizing the critical role of the foveola (generally by disbursing the test images about the fixation point at angles of 2 degrees minimum to the nearest edge of the imagery)58. The results are generally not sufficiently supportive when compared to results employing the 1.2 degree foveola in the

![Figure 19.10.1-8 Brute force approach to computer-vision face recognition REPLACE. Image is of low quality at this time, taken from a newspaper. From Taigman, 2014.](image)

![Figure 19.10.1-7 The significance of facial hair in real face recognition investigations. See text. From AwkwardFamilyPhotos.com, 2014.](image)

imaging experiments. They did report a minimum initiation delay of 200 ms (for the beginning of a full hand and arm motion) while employing 100 ms exposure times.

The Quek & Finkbeiner psychology protocol involved multiple independent functions from a physiological perspective. The protocol involved, learning (training leading to storing selected imagery in declaratory/working memory) before the experiment, priming (for 50 ms based on that learning) and testing for (100 ms). During the testing, it was this author’s experience that a major conscious saccade, on the order of 3-4 degrees, was required. That saccade was probably followed by a non-conscious PEEP procedure to actually extract the information required to recognize the character of the provided imagery. The subject was not asked to identify the image, only to categorize it.

19.10.2 Capability of eye-trackers in face recognition experiments

There are two major classifications of eye-trackers, those that are invasive and typically involve the use of contact lenses covering a portion of the ocular, and those that are non-invasive-using only remote test equipment to track natural features of the ocular. At the current time, with only a rare exception, only invasive eye-trackers provide adequate sensitivity to ocular motions to be useful in object (face) recognition experiments.

Several papers have appeared recently discussing the PEEP procedure, the serial combination of saccades and fixations employed in object recognition. Unfortunately, these papers focus on the statistics of the PEEP procedure rather than the time sequence, and on relatively large saccades and long fixations (saccadic pauses) related to version rather than information extraction. The nominal saccade was the full diameter of the fovea. This focus may be due to the high thresholds they employed to eliminate noise and blinking from their data. The Unema et al. paper contains a good review of the literature.

19.10.2.1 Capability of non-invasive eye-trackers

Figure 19.10.2-1 from Komogortsev et al. provides an overview of the current approaches to eye-tracking available in hardware from the industry. It is important to note the left three approaches employ velocity signals related to the saccades while the right two approaches employ position signals related to the gaze positions within the visual field. The lower frames of the figure illustrate how some of the techniques overlook many of the smaller saccades, a crucial shortcoming of these devices for purposes of understanding face recognition. Komogortsev clearly state their objective of achieving greater standardization of terminology and threshold criteria for eye-trackers in order to support greater understanding of, and hopefully higher performance, eye-trackers.

They open their Abstract with, “In an effort toward standardization, this paper evaluates the performance of five eye-movement classification algorithms in terms of their assessment of oculomotor fixation and saccadic behavior. The results indicate that performance of these five commonly used algorithms vary dramatically, even in the case of a simple stimulus-evoked task using a single, common threshold value.” They also note, “Recent interest has increased in the use of mathematical models to standardize classification of components relating to normal eye behavior in response to external stimuli or impairments relating to pathology or aging. However, the challenge continues to exist because analyses techniques used to track oculomotor movements continue to be highly variable and without universal standardization for system identification of specific eye behaviors. This ongoing problem has led to a preference in tedious manual techniques and reluctance to adopt automatic analysis systems with limited capacity for comparisons across settings.”

---

They describe the algorithms, “Every algorithm presented here can be described in the following general form. The input to an algorithm is provided as a sequence of the eye-gaze position tuples \((x, y, t)\) where \(x_e\) and \(y_e\) are horizontal and vertical coordinates of the eye position sample, respectively, and \(t\) is the time when the sample was taken. A threshold value is provided to allow classification of each eye position sample as a fixation or a saccade, according to the classification criteria implemented in the algorithm. Next, the “merge function” is employed to perform classification of consecutive eye position points as a part of fixation, and then, collapsed into a single fixation segment with center coordinates computed as a centroid of the fixation segment. Classified fixations are subsequently merged into larger fixation segments using criteria based on two parameters: length of the time interval between two fixation groups and the Euclidian distance between these groups.” The challenge is to set the thresholds used in these algorithms to avoid the loss of information while still suppressing spurious eyelid closures. This is a difficult task that has yet to be automated successfully. Because of the high thresholds typically employed, their figure 2 shows most reported saccade amplitudes are in the 8 degree or larger category and most reported fixations are of 0.5 second or higher duration. These values are of negligible value in object recognition experiments where tracking fixation intervals of 250 ms or less is critical and sub-degree minisaccades must be considered.

Their section E provides a discussion of the limitations of the current non-invasive eye-trackers. Until the industry is able to offer higher performance non-invasive eye-trackers, they suggest reliance upon more mechanical methods of data collection and analysis must be relied upon.
Komogortsev et al. provided a lot of graphical performance data but it is not clear how relevant it is to the face recognition task. In their 2012 paper, the Komogortsev team provided current data on a commercial eye tracker, EyeLink, when used in the I–VT mode. In their experiments, they used “A comparatively high classification threshold of 70°/s is employed to reduce the impact of trajectory noises at the beginning and the end of each saccade. Additional filtering discarded saccades with amplitudes of less than 4°/s, duration of less than 20 ms., and various trajectory artifacts that do not belong to normal saccades.” The limits of 4°/sec and duration of less than 20 ms effectively eliminated the saccades related to fine detail perception and to reading.

Kasneci et al. have provided a very well structured paper on saccades and fixation \(^64\). However, the paper does not address the analytical activities occurring during their intervals labeled fixations. A portion of their figure 3 is reproduced here in modified form as Figure 19.10.2-2. Note how the presumed fixations extend over 1/3° or more. The eye movements clearly indicate additional activity during the fixation intervals. The analytical mode of visual operations are highly active during this interval. The mini and micro saccades occurring during these intervals should not be interpreted as noise.

Fairbanks & Taylor have provided an analysis of human eye motions in the context of a log-normal random walk (a Levy flight pattern)\(^65\). However, their laboratory assembled equipment has significant limitations for object (face) recognition, Fairbanks & Taylor develop the performance of their test equipment in their figure 7.

```
[xxx condense this text or show the figure ]
```

“Figure 7(a) plots the horizontal (x) and vertical (y) locations of the gaze in units of screen pixels. Micro-saccades are expected to occur over an angular range of typically 0.5°. This angle translates to a distance of 15 pixels on the screen and, as expected, this approximately matches the typical width of the dwell regions observed in Figure 7(a). Given that the saccades and micro-saccades are produced by different physiological mechanism and serve different purposes within the visual system, it is expected that their scaling behaviors will be different. The scaling analysis of the spatial patterns will therefore be expected to reveal a crossover between the two processes – the saccades and micro-saccades at a size scale of approximately 15 screen pixels.

Figure 7(b) shows the corresponding temporal pattern by plotting the x position against time t. The periods of relative motionlessness are the dwell periods at a given location, during which time the eye is undergoing micro-saccades. The typical dwell time is approximately 0.4s. The time scale of the individual micro-saccades is expected to be approximately 10-20 milliseconds. We note that this is on the same order as the sampling rate of the eye-tracking equipment (16ms, 60 Hz). This measurement limitation would therefore impact on any studies of the micro-saccade.”

"The left - hand extreme of these plots matches the width of the screen: 1024 pixels (290mm), which corresponds to an angular motion of the eye of 29°. The right hand extreme corresponds to 1 pixel (0.3mm), which corresponds to an angular motion of 0.03 degrees.”

The performance levels discussed in Fairbanks & Taylor are far below those required for effective investigations into object recognition.

---


Putnam et al. provide data acquired using recent optical computer-aided (or coherence) tomography (OCT) techniques to track eye movements. The saccadic and gaze data is provided as an ancillary output of the stabilization servo loop used to support the primary investigations of the retina. Such eye movements are deleterious to the main goal of this equipment, the detailed examination of the structure of the retina using a scanning beam introduced through the pupil. The equipment is complex but some variants are now available in packaged form suitable for psychophysical experiments. More complex variants of the equipment may not be available in the market yet and may be required for object recognition experiments. See Section 18.8.3.5.3.

With the exception of the possibly available high performance OCT test equipment, the comments of Komogortsev et al. appear to remain relevant, the state of the art in commercially available non-invasive eye-trackers is far below that required for productive laboratory investigations into object recognition. An other exception may be in a tracker used by Juhasz et al. They used a pair of Forward Technologies Dual Purkinje eye trackers (Generation 5.5 and Generation 6), which sample eye position every millisecond and have a resolution of less than 10 minutes of arc. The eye trackers were interfaced with a personal computer. Sentences were displayed in white font on a black background using a Philips 21B582BH 24-in. monitor.

The eye-tracking laboratory at Oklahoma State University may be a good source for recent eye-tracker performance data. Komogortsev and colleagues have also presented a number of models of the oculomotor system in 2008 and 2012. However, they only model large saccades and truncate their performance at small angles and high frequencies. They do not incorporate the twitch function of the ocular muscles as developed in Section 7.3.4.

19.10.2.2 The capability of invasive eye-trackers

The most precise eye-tracking technology has always employed some form of contact lens applied over the ocular and modified to provide a signal in conjunction with external test equipment that allows recording at the tremor level of eye movement. Yarbus, working in the 1950's and 60's used a variety of contact lenses to provide excellent recordings (although limited by the recording technology of his day) of the eye tremor under a wide variety of conditions. Ditchburn worked in the same time period and provided similar but less extensive test results (See Section 7.3.3). The performance achieved by a wide variety of investigators using invasive techniques (modified contact lenses) is summarized in Section 7.3.3.5.3.

The 1996 paper by Kalesnykas & Hallett and associates provides a very detailed discussion of saccadic movements under a variety of conditions and recent performance data. The use of non-invasive eye tracking technology is more than adequate for supporting object recognition research and the use of contact lenses among potential test cohorts is much more common at this time.

19.10.2.3 New classes of eye-trackers

In the recent experiments in machine vision, a new hardware implementation has been introduced based on retroreflection from the eye of the subject. It notes the on-axis reflection of light through the pupil results in the red-eye effect so commonly observed in photography. The effect disappears under less retro-reflective conditions; the pupil appears dark. By taking the difference between two frames of data from a imaging sensory, the position of the eyes can be determined. Murmoto et al. have implemented a tracker


\[^{68}\] http://psychology.okstate.edu/component/content/article?id=103


of this type\textsuperscript{71} and have since found a commercial instrument exploiting this difference (the iniLabs. DVS128 spiking retina sensor\textsuperscript{72})

The Alcon division of Novartis, working with Google, has recently produced a new “active” contact lens that offers a new capability of achieving high accuracy micro-saccade and tremor tracking. It may be described as minimally invasive. Figure 19.10.2-3 shows a prototype of the accessory. Its utility for eye tracking has not been reported to date. It has been suggested one of its first uses will be in blood sugar tracking in diabetes.

**19.10.3 The PEEP procedure used in object recognition**

The previous discussions in Section 19.10 and in Section 7.5 of this work, have provided the framework for describing the procedure used in the vision modality of the neural system of at least the primates. It may also be used by the predatory birds, and many other animals where the database does not provide sufficient information about the use of a foveola within the fovea. The repeated sequence of gazes and saccades described in Section 19.8.2.3 for reading form the foundation for the following discussion formalizing the more general PEEP procedure of stage 4 information extraction via the foveola-PGN-pulvinar pathway of the vision modality.

This discussion will ignore the field of view outside of the fovea of the primates, a central area with an angular diameter of about 8.68 degrees (1.85 mm diameter at the retina) roughly centered on the point of fixation. The foveola will be taken as 1.18 angular degrees (0.35 mm diameter at the retina) more precisely centered on the point of fixation. The area outside the 8.68 degree circle is used primarily to support convergence of the two eyes and to sense forward motion of the head relative to its external environment.

[Figure 7.5.1-1] defines the names of the different ocular motions used in the following material. The amplitude of tremor is poorly documented but it appears to be on the order of 30 arc seconds, too small to be shown in the figure. The numbers found in the literature and summarized in Appendix L, the Standard Eye are given below.

Size of high frequency tremor--20-40 arc seconds in object field, 1 to 2 photoreceptors at the fovea
Reported frequency of tremor--30-90 Hertz (reports to 150 Hertz), nominal center frequency--50 Hz.
Baseline frequency of tremor--center frequency of ~90 Hz with sidebands extending from 40 to 130 Hz.
The horizontal and vertical components of oculomotor tremor are presumed to be in quadrature.

As shown by Padgham & Saunders in [Figure 19.2.3-1], the human eye achieves an accuracy of about 10 minutes of arc when it attempts a mini-saccade to a new absolute location defined by the test equipment. It can be assumed it can achieve similar accuracy when it attempts to move to a new location based on data reduction from the coarse saliency map. It can also be assumed the eye does not make an attempt to improve on the initial position achieved versus the desired location because the version system of the eyes operates open loop.

---


\textsuperscript{72}Lichtsteiner, P. Posch, C. and Delbruck, T. (2006) A 128x128 120 db 30 mw asynchronous vision sensor that responds to relative intensity change *IEEE ISSCC Digest of Technical Papers*, pp. 2060–2069
As discussed in Section 19.8.2, the scene recognition procedure begins with the recording of the entire scene at low resolution. Stage 4 information extraction, based primarily on spatially rapid changes in local signal intensity (not frequency), develops a set of locations within the field of the fovea that are used to point the line of fixation to these individual locations in sequence. [Figure 7.5.2-1] uses a caricature by Henderson to illustrate this procedure. The means of ordering of the sequence has not been determined but the resulting sequential saccades is believed to perform what is known as a Levy flight (statistically probably a log-normal random walk pattern). A more recent figure appears in Sheinberg & Logothetis (cited in Section 19.10.4.1.4)

Kalesnykas & Hallett have provided considerable information useful in describing in detail the individual fixations within a PEEP procedure. They did not concern themselves with the precise diameter of the foveola and at one point suggest it was about two degrees in diameter. They did note the so-called avascular foveola was about 0.7 degree radius (Polyak, 1941). They lacked any physiological model for their study except for the poorly defined foveola. In their introduction, they did note “The foveolar region has interesting characteristics” that they discuss primarily in text form. From these discussions, they define their test protocols.

Brady has written extensively on the steps involved in analyzing a scene, to include separating a gaze into individual glimpses before developing an overall perception of the scene. The contribution of Gosselin et al includes Figure 19.10.1-5 showing the fixated areas (gazes) imaged by the foveola during a PEEP sequence is discussed in detail below.

19.10.3.1 The signal manipulations used in object (face) recognition

The psychology community appears to stress a potential difference between object recognition and face recognition based primarily on anecdotal evidence or intuition. A comment frequently appears concerning the superior ability of the human to recognize faces over all other type and/or classes of objects without any scientific support or citation (Gauthier & Tarr, 2012, page 2). Many humans have achieved remarkable ability to recognize sonar signals (among sonar operators), specific sub-families of birds, sometimes down to the individual levels (experienced ornithologists) and a variety of musicologists who are able to label a complete musical piece after hearing only the first four or five notes. These people exhibit the results of considerable training that is virtually automatic with regard to face recognition in the course of day-to-day activity over periods of years.

This work treats pattern recognition as a global class where face recognition is only one among many subclasses. See [Figure 19.10.4-1] below.

[xxx modify this figure to stress the PGN-pulvinar couple ]

Figure 19.10.3-1, reproduced from [Figure xxx], shows the gross signal flow in caricature associated with the foveola and fovea operating in parallel based on this work and related findings presented in “The Neuron and Neural System.” In earlier parts of this work on vision, it was suggested the location of the saliency map, and potentially the final stage 4 information extraction engines were to be found in/on the posterior parietal lobe (potentially area 7). Subsequent review of the literature suggests these late stage 4 engines are more likely located in/on the posterior temporal lobe (in the vicinity of Brodmann areas 21, 22, 37 and 39). The figure calls out only area 37 for convenience.

References:

The papers of Boussaoud et al. have suggested in well supported documentation that the mapping of the visual modality into associative area 37 involves input from both the cerebrum (areas 17 through 21 as well as distinctly other paths that they do not identify. See Section 19.10.4.1.4. Figure 19.10.3-2 shows the scene in object space, also known as visuotopic space, from the perspective of the retina during the coarse and fine portions of the sequential gaze/saccade procedure. Visuotopic space is compared to retinotopic space, and the actual topological projections of these spaces onto selected areas of the brain in Section 15.6 of Chapter 15, part 2. http://neuronresearch.net/vision/pdf/15Higher2.pdf#page=18

The grid associated with the foveola is meant to show the much higher resolution of the foveola relative to the fovea (greater than 75:1 over most of the field, see Figure 7.4.5-1).

It is important to distinguish between attention and instantaneous gaze. In general, the conscious attribute, attention, does not change during the nonconscious PEEP procedure. In the current case, the entire view of the fovea represents the attention span in visuotopic space while the instantaneous gaze represents only the image space interrogated by the foveola.
Figure 19.10.3-2 The PEEP procedure sequence. The full field corresponds to 8.68 degrees in object, or visuotopic, space (shown as a square for convenience) upon initial exposure to a scene and encompassing the circular human fovea. After coarse processing by the LGN/occipital couple, an area of interest is determined and the line of fixation is established to the center of gaze one (G1). After a variable interval, a saccades (S1) moves the line of fixation to the center of gaze 2. After an additional variable interval, saccades S2 moves the line of fixation to gaze location G3, ... During the intervals at each gaze location (of 1.2 degrees diameter), fine information extraction by the PGN/pulvinar couple is accomplished.

Figure 19.10.3-3 shows an expansion of the well known figure 114 of Yarbus (1965) to caricature the first three steps in the proposed gaze/saccade procedure. The central frame shows the proposed initial gaze/saccades steps on the assumption the sequence starts at the mouth. The resulting pattern shows a strong similarity to the discussion in Section 19.10.1. During the pause at each point of fixation, the
PGN/pulvinar couple analyzes the orthogonal grid of photoreceptors within the nominal 175 cell diameter foveola. After creating a meta file describing the features of that part of the image, the meta file is transferred to the appropriate location of the saliency map representing that part of the external visual field in inertial space. After establishing a complete meta file for the scene in the saliency map, the saliency map and the file is available for stage 5 cognition and for long term storage in declarative memory (See Chapter 17 of “The Neurons and Neural System”).

http://neuronresearch.net/neuron/pdf/17Memory.pdf#page=7

The relevancy of this hypothesis is well supported by the laboratory and analytical work of Gosselin & Schyns discussed earlier.

Figure 19.10.3-4 from Yarbus (fg 112) also presented a scene strongly supportive of the current hypothesis. It is similar to that of Henderson but includes a number of people in the room. Rather than just show the points of fixation as in Henderson’s figure, Yarbus used circles representing the assumed foveola of 1.3 degree diameter centered on the points of fixation. The subject was allowed to study the frame at upper left for 35 seconds. The frame numbers indicate the gross order in which the individual groups of gazes were selected but not the time associated with each group or individual gaze.
In this figure, it should be noted that the scale of the picture may limit the neural system to recognition, but possibly not identification due to a lack of adequate angular resolution of the visual modality. Alternately, identification may be achieved more through the dress, silhouette, size and context (the location) of the subjects in the picture rather than through face recognition.

It is important to note, based on the right-most frame of the above figure of “the girl,” that the “attention” of the conscious (or cognitive) executive remains focused on the entire face within an inertial frame of reference. The conscious executive is totally unaware of the continual repointing of the line of fixation within this inertial frame of reference by the version subsystem under the control of the non-conscious executive (the TRN).

**Figure 19.10.3-5**, also shown in Figure 19.10.1-5, reproduces a modified figure 2 from a paper by Gosselin & Schyns. It provides an alternate representation of the PEEP procedure than that shown in the Yarbus example of a room. It is a more specific representation of the gazes associated with stage 4 information extraction by the foveola-PGN-pulvinar pathway of the neural system. Their algorithm employed a randomly located small Gaussian aperture that traveled (caricatured at several locations at upper left) over the 5.72 degree scene. Their algorithm highlighted the areas where a portion of the face can be seen (in the absence of the large

**Figure 19.10.3-4** Initial steps in object recognition prior to subsequent object identification from Yarbus. The upper left frame shows the unannotated picture. Subsequent frames were acquired in the sequence indicated. The circles represent the individual 1.3 degree diameter gazes brought to the foveola for purposes of recognition prior to identification. See text. From Yarbus, 1965.
circles). The highlighted areas are a good approximation of the 1.2 degree size of the image presented to the foveola (the large circles) for a line of fixation centered within each circle during a presumed PEEP sequence of saccades and gazes (sometimes labeled a Levy flight or a joconde–dance). The areas highlighted by their algorithm do match the diameter expected to be interrogated by the human foveola during a saccades/gaze sequence (the PEEP procedure). The sensitivity pattern of the foveola however is formed by an area of constant sensitivity within each circle. The text labels correspond to the labels assigned during their first set of experiments. EXNEX refers to an expressionless face. The hair lines of all their images were standardized, as was the global orientation (full face presentation) and lighting, to avoid impacting the results. The subject was allowed to observe the face through the small randomly moving apertures. When they believed they could identify the face, based on previous training, they stopped the experiment and the number of distinct aperture views employed up to that point were noted. “With enough trials, a random search is exhaustive and all the search space is explored.”

Figure 19.10.3-5 Figure 2 of Gosselin & Schyns overlaid with gaze circles and the apertures used in their Bubble algorithm experiments. Small circles in upper left frame; the size of the Gaussian apertures used to scan their facial images. The extent of the facial features shown were defined by the Bubbles algorithm. Each large circle; the expected size of the images examined by the foveola of the human eye during a single gaze. The algorithm does highlight areas of the size expected to be interrogated by the foveola. See text. Modified from Gosselin & Schyns, 2001.
After a very complicated numerical computation involving several sorts of their data, the algorithm develops what they called “diagnostic masks.” These masks are much larger than the size of the test apertures. The diagnostic masks do appear to approximate the size of the individual images formed on the foveola during the individual gazes in the normal PEEP procedures of the human eyes. A set of simple circles have been added to the figures to represent the “diagnostic masks” as developed from human physiology later in this section. Only a few of these 1.2 degree circles equivalent to the size of the foveola are shown overlaying portions of their diagnostic masks.

The figure shows sequential gazes of potentially short time duration. The glimpses shown in each frame constitute individual steps in the PEEP sequence normally used to interrogate a scene. During attention focused on an individual face, or while watching television, the dwell time of a specific gaze may be relatively long. [xxx how long? ]

19.10.3.2 The coarse information extraction by the LGN/occipital couple

The spatial resolution of the visual signals collected by the fovea (including the foveola) that is processed by the LGN/occipital lobe is an order of magnitude lower than that processed by the PGN/pulvinar (discussed in the next section). This resolution limit is determined by both the off-axis performance of the highly anamorphic optical capability of the stage 0 optical system. Because of this limited optical performance but the nominally constant spatial density of the photoreceptors in the retina, multiple photoreceptors of a particular spectral type are typically summed in the stage 2 signal processing circuits of the retina before they are forwarded to the LGN. After combining of the signals from each eye in the LGN (maintaining their spectral identity, and indicating their stereoscopic position in the z-direction as metadata, these low resolution composite signals are passed to the occipital lobe (area 17 or the visual cortex) along with some additional metadata for final stage 4 information extraction. As is becoming clearer during each calendar decade (particularly through use of fMRI techniques), area 17 is not limited to visual data alone. It may develop metadata from visual signals but may combine information from other sensory modalities closely related to the visual data before developing the final metadata.

One of the most intriguing current questions in the neurosciences is how the metadata developed by the stage 4 engines of information extraction are encoded (i.e. the metadata schema of the neural system. In exploring this area more deeply, it is strongly recommended that the investigator conform to the terminology suggested by the ISO for a particular field.

http://www.slideshare.net/OpenDataSupport/introduction-to-metadata-management  Slides 13 through 19 of this slide presentation are particularly relevant.

Both the LGN and area 17 of the occipital lobe are highly reticulated, an indication of the long dendrites and axons employed to form a nominally orthogonal information extraction matrix. At the gross histological level, these axons and dendrites are not rectilinear. However, their intersections are at the individual axon and dendrite level. See Section xxx.

After information extraction within area 17, it may be further processed either directly in areas 18 and 19. More likely, it is passed back to the switching and control portion of the TRN of the thalamus for redistribution to areas 18, 19, and elements of the associative cortex (such as area 37). After completing the associative processes, the information is placed in the saliency map probably found in holographic form near area 37. Part of the metadata returned to the switching and control portion of the TRN of the thalamus are the data describing the “busier” parts of the foveal scene warranting more detailed examination. This information is routed by the TRN to the version control system of the oculomotor subsystem.

19.10.3.2.1 The unordered saccades among primates

Kano et al. have performed a set of face observation experiments among the primates that define a set of
saccades for human, gorillas and orangutans. As noted below, their experiments were aimed at the conscious areas of interest (AOI) within a large angular field of view rather than the non-conscious points of fixation within each of these areas of interest. As such, the results reported by Kano et al. do not relate to object recognition as much as what might be considered a threat assessment related to the major portion of the visual field of view. This is consistent with the anecdotal remarks of Kano et al. at several points that when viewing faces, and probably particularly at the close range provided by these experiments, the apes and gorillas tend to focus more on the mouth area compared to the eyes than due humans.

Although they did not define the sequence in their individual saccade sets, they did make several assertions; “These results suggest that, although humans and great apes share a basic set of facial communicative skills, significant species differences characterize their habitual styles. . . .Both species were presented with pictures of whole bodies and faces, and they viewed the pictures freely. The species were strikingly similar in their patterns of face and eye scanning. For example, both species fixated on faces and eyes more frequently than on other parts of bodies and faces. However, several species differences were also identified. For example, chimpanzees viewed faces and eyes more briefly than did humans; typically, chimpanzees only glanced at eyes, whereas humans viewed both eyes (left and right eyes) alternately. Additionally, when presented with facial expressions, including conspicuous mouth actions, humans viewed the eyes rather than the mouth, and chimpanzees viewed the mouth rather than the eyes.”

Kano & Tomonaga have presented two related studies. The 2009 paper includes studies of non-face recognition as well. In the first paper, they “confirmed a non-random pattern in the first fixation on each scene/body feature by both chimpanzee and human participants when shown pictures of mammals.” They also provide large amounts of statistical information but not the order of the saccades. They did indicate, “In addition to the traditional index of looking behavior and looking duration, we also examined the sequence of fixations as participants viewed the pictures.” Images were presented for three seconds in this set. The 2010 paper includes some more complex scenes that indicate how the subject concentrates on both the faces and prominent edges in the surrounding scene. Images were presented for two seconds in this set. A few of their conclusions are, “Several similarities between humans and chimpanzees, such as the intense scanning of main facial features and the order in which each facial feature was inspected, are also consistent with the studies on monkeys (Guo et al. 2003; Gothard et al. 2004; Ghazanfar et al. 2006). Of special interest in this regard is the characteristic pattern of human face scanning involving sequential re-fixations on the eye region. This pattern is consistent with results of previous human studies.” “The direct causes of these characteristic scanning patterns in chimpanzees and humans remain unclear; further studies are necessary.”

In the Kano et al. studies, relatively large images were used, 18.5° high by 24.5° wide of visual field. They describe the sizes in the 2012 study slightly differently, “Facial pictures were converted to 768 x 1024 pixels (22 x 30 degrees at a typical 60-cm viewing distance) with a gray frame around the background (total 1280 x 1024 pixels). The internal parts of faces (eyes, nose, and mouth) were thus approximately 10–15 degrees in size at a typical 60-cm viewing distance.” As a result, their “fixations” corresponded to the conscious points of attention in this work rather than the nonconscious glimpses involved in actual object recognition. By using a coarse eye tracker (nominal accuracy of 0.5° visual angle), the experimenters focus was on the relatively coarse eye movements associated with the gazes of the PEEP procedure. Their statistical data is consistent with the expected gaze duration (measured in hundreds of milliseconds) but does not provide information about the length of the individual glimpses (typically in the 35 to 200 milliseconds range). In some of their figures, they did use the diameter of a circle centered coarsely on a feature, to indicate the relative length of time spent in one or more gazes at that feature.

Figure 19.10.3-6 rearranges their frames from figure 3 of their 2012 study for the purposes of this discussion. Some of the images were not reported for all three species. However, several trends appear

---


clear. Each of these species employs a similar if not identical conscious procedure when shown images at approximately the same scale of a HD television monitor. Each saccade set is concentrated along the vertical axis of the image with excursions to the eyes in a majority of cases, even when the image was not looking directly at the subject being tested.

The test subjects were not constrained with respect to time and the saccades order was not provided in the report. It appears that some of the subjects may have taken longer to examine some of the frames suggesting that they went back to reconfirm their evaluation of some of the glimpses previously analyzed.

Figure 19.10.3-7 reproduces Figure 6 of the 2012 paper shows state diagrams with the probability of saccade between states for each species when looking at conspecifics. The amount of attention given to peripheral features at the attention level is emphasized in these diagrams. Their figures 4 and 5 provide histograms of what features are attended to most specifically but do not indicate the order of attention.
19.10.3.3 The fine information extraction by the PGN/pulvinar couple

The PGN and the pulvinar are both highly reticulated structures. This reticulation is also suggestive of the long axons and dendrites used to form a nominally orthogonal information extraction matrix. At the gross histological level, these axons and dendrites are not rectilinear. However, their intersections are at the individual axon and dendrite level.

The signals generated by the sensory receptors of the retina destined for the PGN are provided with individual signal paths (long axons) connecting individual receptors with a counterpart location in one of the layers of the PGN. This is in contrast to the stage 2 signal processing of signals from the same receptors destined for the LGN. This distinct bundle of axons (xxx enter name of nerve here) forms the xxx of the optic nerve. The signals delivered to the PGN are used for two primary purposes; to develop a higher quality convergence signal for instructing the precision portion of the vergence system to obtain optimum focus on the object of interest falling on the foveola, and for combining the optimized signals into

---

**Figure 19.10.3-7 Typical scanning paths related to attention for facial pictures.** Probabilities of saccade destination are presented in percentages and scaled to size. If a human participant is currently looking at the right eye of a human face (top center), he/she would re-fixate the right eye with a probability of 2.6, the left eye with a probability of 15.6, the nose with a probability of 7.7, etc. (see their text for details). From Kano et al., 2012.
a single neural array of signals that can be passed to the pulvinar for information extraction.

It appears the array of signals passed to the pulvinar are not divided into left and right fields. The pulvinar receives one array of neurons representing signals from the entire 1.18 degree foveola field of view. This array is approximately 175 axons in diameter and contains approximately 23,000 axons. It is this array that must be processed by the stage 4 engines of the pulvinar to extract the information describing the foveola scene during an individual gaze (generally obtained during one Peep but conceivable confirmed during a second gaze).

Cooke et al. have confirmed that the limiting performance of the foveola in its analytical role is based on all of the sensory signals as a group without regard to their spectral sensitivities.79

19.10.3.4 The dimensionality of visual information stored in the saliency map

The question arises as to how many independent dimensions are employed within the saliency map to describe a complete scene as it exists at a specific time. The question is a complex one because of the signal processing performed in stage 2 and later stages. At each retina, two spatial dimensions and four spectral dimensions are readily defined in mammals (including humans). The pairing of the spectral channels prior to the optic nerve (at least for the signals destined for the LGN/occipital couple) tends to reduce this dimensionality. However, the combining of the data from the two retinas and the extraction of the third spatial dimension by stereoscopic processing tends to increase the dimensionality.

An additional question arises as how to handle the more detailed signals directed to the PGN/pulvinar couple. Each gaze produces additional signatures related to the specific object already identified via the three spatial dimensional and the combined spectral channels. As a minimum, an additional pair of spatial coordinates must be considered, even though these are later processed as an overlay on the map created from the LGN/occipital couple. It does not appear appropriate to expand the number of resolution elements (pixels) in the horizontal and vertical dimensions of the basic saliency map by a factor of about 75:1 to accommodate this higher resolution information. It is more likely, the additional information is incorporated as some form of metadata.

Furthermore, the extraction of abstract identity information by the PGN/pulvinar couple must probably be considered additional metadata consisting of one or more additional dimensions.

The minimum number of dimensions associated with the visual portion of the saliency map appears to be three spatial dimensions, three processed spectral dimensions (O-, P- & Q-), one or more dimensions of an as yet undetermined metadata schema and an unidentified time indicator.

19.10.3.5 Potential means of encoding visuotopic targets of interest EARLY

In this and the discussions below, the term visuotopic is used to mean the data under discussion remains within a subject-based coordinate system rather than an inertially-based system. It is probable that the visuotopic code used at the output of stage 4, information extraction, is compatible with a straight forward conformal transformation between these two coordinate systems. However, it may require a few more years of study to determine the character of, if not the actual, transform involved.

In the previous paragraph it was estimated that visual data alone probably involves more than six dimensions including a time marker and an unknown amount of metadata possible stored as a multi-dimensional array.

While there is no recognized need for an absolute or even precise time marker, memories must be stored within some relational time frame based on this investigators experience.

The more difficult question is how are all of the meta data associated with a complete visuotopic scene stored? This question is followed by the broader question of how is the meta data associated with the complete external environment in a 3D inertial space stored?

Looking first at a single visuotopic scene, such a scene typically contains a large amount of “background” imagery that is stored at least at a crude resolution level plus a variety of more important elements that have been examined in detail during one or more PEEP procedures. Many, if not most, elements of the background imagery is likely to have been examined by one or more PEEP procedures over the course of time (particularly for scenes visited frequently). If correct, the storage of such interpreted information about a scene in the normal inertial coordinate system of the neural system over a lifetime implies a nearly boundless memory capacity. Such a requirement strongly suggests some form of holographic data storage mechanism with each stored holistic scene consisting of a variety of individually encoded scene elements.

A holographic storage mechanism does not require a section of memory to be exclusively reserved for a single memory. It resembles the spread-spectrum technique of radio transmission where multiple messages can be sent simultaneously within a given frequency band with each message accessible by a separate and distinct “key.” This capability is easily demonstrated with an adequately prepared 2D visual hologram deposited on a transparent flat plate. When illuminated with an interrogation light beam of a specific wavelength, one 3D scene is recreated. When interrogated with a light beam of a different wavelength, a separate 3D scene is recreated. If both beams are employed simultaneously, both recreated 3D scenes can be observed simultaneously.

Adding messages in a hologram introduces a very small but measurable decrease in the signal to noise ratio associated with each message. In the case of neural memories, this may be related to the common expression that “my memory is fading as I get older.”

These individually encoded scene elements could have been collected at different times and merged into the previous stored record of the holistic scene. This presumption suggests the employment of a very general updatable database structure.

The above background would suggest the entire memory of an individual consists of a holographic memory map containing all of the sensory information collected over a lifetime via the “instantaneous” saliency maps. The information would be stored as a series of “holistic scenes” in a database and contain an unlimited series of individually encoded “holistic features.” It is conceivable that each holistic feature would contain a spatial address subordinated to the spatial address of the holistic scene (a hierarchical database).

The above potential scenario for information storage suggests that prosopagnosia is primarily a database addressing failure. Such a failure could be related to faces or any other class of perceived objects. It is not clear whether the failure occurs during record storage or record retrieval.

While the specific method of encoding the data from a two-dimensional fixation image (including any associated metadata) has not been uncovered to date, the methods used to interpret music may provide some clues. These methods are reviewed in Chapter 8 of “Hearing: A 21st Century Paradigm,”

http://neuronresearch.net/hearing/pdf/8Manipulation.pdf#page=52

19.10.3.6 The role of savants as related to the plasticity of the brain EMPTY

[xxx material has been moved to Section 18.9.2 ]
In 19.10.3.6, it will be nice if you could address the case of these two autistic savant that seems to have almost perfect visual memory who would encode not merely "holistic scenes".

19.10.3.7 Current state of knowledge about face recognition EARLY

Based on the hypothesis of this work and the literature (synopsized in the above discussion), the following points can be made;

1. The visual modality obtains imagery based on a scanning sensor and not a framing sensor such as a camera. In the absence of tremor, the visual system is blind.

2. The visual signal path consists of two nerves of about 1 million neurons each (the optic nerves) that are each segregated into two major channels associated with the complete visual field and with just the foveola
3. The two signal paths from the two eyes are merged separately in the pair of LGN (for the fovea) and the single PGN for the smaller foveola.

4. Following the merging of step 3, two major signal paths are maintained, that connecting each LGN to its appropriate hemisphere of the occipital lobe (forming the LGN/occipital couple pathway) and the single PGN connecting to its pulvinar (forming the PGN/pulvinar couple pathway).

5. The LGN/occipital couple focus on external environment awareness, extracting signals necessary to achieve coarse merging of the imagery from the two eyes extracting vergence information required to achieve coarse stereoscopic vision and supporting coarse object recognition (as opposed to identification).

6. The PGN/pulvinar couple focuses on extracting signals necessary to achieve fine merging of the imagery from the two eyes extracting vergence information required to achieve fine stereoscopic vision and supporting fine object analysis (that includes face recognition followed by identification).

7. The PGN/pulvinar couple is central to the gaze/saccades (PEEP) procedure.

8. The PEEP procedure is characterized by a perusal of the entire visual scene (at least within the fovea of the central 6.86 degree field of view) using what can be described as a log-normally distributed random walk (Levy flight) pattern.

9. Each gaze of the PEEP procedure is analyzed individually by comparison with prior stored foveola images stored primarily within declarative memory.

10. The nominally random ordered set of xxx resulting from the gaze perception/memory comparisons are used to create the ultimate perception of the overall scene (within the field of the fovea).

A challenge at this time is to discover how the stage 4 information extraction engines or stage 5 cognitive engines assemble the individual xxx resulting from the PEEP procedure followed by the xxx/memory comparison step into an unambiguous set of metadata describing the external scene.

[xxx need a term for the result of the individual gaze/memory correlation steps in the PEEP procedure.]

19.10.4 Establishing a multi-discipline object recognition framework

Rossion has stated a long-standing problem in neuroscience and psychology succinctly, “Understanding how the human brain builds a unified face percept is one of the greatest challenges of visual neuroscience because, in the early stages of visual processing, a face is represented by neurons with small receptive fields. These neurons provide information about local elements of the face, and these elements need to be combined for instance, by convergence of inputs to neurons in higher order areas with larger receptive fields, or by temporal synchronization of the activity of distributed populations of neurons in lower level areas (the ‘binding’ problem in vision; Treisman, 1996). Obviously, this question is not specific to faces and concerns a general problem for understanding human vision.”

This chapter has developed and/or described considerable material related to the overall object recognition task within the visual modality. To understand object recognition, generally following object identification (as to its character but not to the level of recognition of its individual personality), it is necessary to understand the modular character of the neural system of biology and specifically how the various servomechanisms and information extraction engines of the modality operate. The previous sections of this chapter have cited earlier sections of this work focused on these engines.
This section will summarize the hypothesis that the functional classification, “object recognition” includes a variety of sub-classes that include generic object recognition, minimally structured specific object (face) recognition and more highly structured object recognition (reading as an example). In the general case, object recognition is a multi-dimensional phenomenon and is ultimately a multi-sensory modality phenomenon.

Arguing whether face recognition is distinctly separate from object recognition is ephemeral and a major waste of printer’s ink in the psychological literature. This is particularly true when the debaters are unfamiliar with the physiology of the neural system underlying their discussion.

Faces, like dominoes, theater marquees, automobiles, etc. are subclasses of the class objects. The class, objects, is not limited to two-dimensional images; it is multi-dimensional. Any instantaneous perception of an object may include metadata. The metadata may be related to other information from the visual modality or to other sensory modalities. It may even include tags to other information attached to the perception in the process of comparing the current perception with previously stored perceptions of the same or very similar objects.

Figure 19.10.4-1 represents a simple test. Which of the tiles shown do not qualify as a legitimate tile in a conventional domino set? Which of the upper faces of the tiles actually represents a rudimentary “face” used in face recognition experiments with babies? Is the tile containing the rudimentary “face” still an object? How is a face distinguished from an object?

Obviously, the equilateral triangle for three in the upper tile face does not belong in the standard tile set. It represents a “face” used in face recognition research with babies (granted it is of reverse contrast, but standard domino sets come with both contrasts). The question is, are all of these tiles and tile faces still objects?

Arguing whether face recognition is distinctly different from object recognition based on simple experiments involving the inversion of two-dimensional facial images and the investigators intuition is also ephemeral. It may not have been noticed by such researchers that reading is functionally similar to face recognition and its success also depends on familiarity with the imagery involved. While intuition might suggest that reading material when the text is inverted is more difficult than reading non-inverted text. However, any experienced person working in marketing that can not read inverted material at a rapid rate is probably being overpaid. It is a distinct career advantage to be able to read a document (upside down to the marketing man) while the target client is reading the same document laid on his desk right-side up. Reading or identifying objects when inverted is dependent on the prior storage of the appropriate imagery in memory.

Any discussion of object recognition must explicitly address the question of prior training of the test subject in the context of the objects used in the experiments. This training may have occurred early in the test subjects life or been provided within the experimental protocol. The importance of training is highlighted in Table 1 of the Gauthier & Tarr paper discussed below. They proceeded to differentiate clearly between novices and experts in their experiments. They even introduce “Greebles,” synthetic images in an attempt to use training level as an experimental “control.”

There is a high likelihood that word serial/bit parallel signaling is used among the stage 4 engines of information extraction. Such signaling would allow a different interpretation of the multiprobe signal
19.10.4.1 Overview

In an important 1997 paper, Gauthier & Tarr developed a framework for discussing object recognition from the perspective of a psychologist that can be adopted (and possible modified) here. The paper appears to be drawn from the PhD dissertation of Gauthier. There are significant technical problems with the paper that will be discussed below. The discussion of classifications is useful but incomplete. Figure 19.10.4-2 provides a broader suggested framework compatible with the physiology of the neural system. Obviously only specific paths in this diagram can be expanded and annotated in one discussion as the discussion proceeds down the diagram. This diagram assumes the test subject is familiar with the presented objects and does not require training. Otherwise a different diagram incorporating a learning function must be provided.

All but the top and bottom lines of this figure relate to stage 4 information extraction engines. All but the top two and bottom lines of the figure relate to the PEEP procedure of vision. The line labeled 4th subordinate level is meant to indicate the collection of metadata from whatever elements are interrogated in 3rd subordinate level activities. The bottom line labeled “Extreme subordinate level” conforms to the label of Gauthier & Tarr. However, as seen in the next figure and discussion, the translation of the collected metadata into a holistic package of metadata and the retrieval of the short label describing the object (such as Max, Jim or Hassan) is actually accomplished in a morphologically separate set of stage 4b engines located within the cerebellum and employing a lookup table. The actual expression of the short label is by means of the peripheral portion of stage 6 incorporated into the peripheral nervous system (PNS) and the stage 7 (terminal neurons) controlling the activity of the musculo-skeletal system.

---

Before discussing this figure, it is useful to provide a block diagram describing the flow of signals and information within the visual modality of the afferent neural system along with the musculo-skeletal portion of the efferent neural system from figure 4.6.3-1 of “The Neuron and Neural System.” Figure 19.10.4-3 reproduces this figure. The left side shows the afferent stage 4 engines of the CNS. The right shows the efferent stage 6 engines of the CNS and the top hat shows the stage 5 engines of the cognitive
portions of the CNS. The boxes with suffixes “b” are important elements of the memory and the pattern comparison apparatus. Their role will become clearer in the discussions below. Only the visual and auditory modality engines are labeled at the lower left, although those associated with gustation and olfaction can be added. The separation of the stage 1 sensory signals into the LGN/occipital lobe pathway and the PGN/pulvinar lobe pathway is indicated. The parallel operation of the LGN/occipital pathway and the PGN/pulvinar pathway, both in conjunction with the version servo mechanism controlling ocular pointing, are key to the operation of the PEEP procedure discussed below. The alternate output paths from the lower row of boxes involve both the visual and aural pathways depending primarily on the level of familiarity of the individual with the material presented to the stage 1 sensory organs (in the visual case, the full spatial extent of the retina).

Failure of stage 4b engines of the cerebellum to recognize the metadata presented to it by the stage 4 engines as a result of the PEEP procedure, will result in the raw information, accompanied by the metadata being passed up the left-most path to the Stage 5 engines of cognition for their interpretation. These engines will pass the results of their review back to the 6b portion of the cerebellum via the stage 6 instruction pathways. Within the cerebellum, the 4b and 6b information will be linked to insure a coupling between them and stored as a recognized perception available for future comparison with new objects presented within visuotopic scenes (scenes represented in subject centered coordinates). This procedure is a critical element in the role of memory when discussing object (face) recognition experiments and models. The stage 5 engines of cognition do not need to spend time analyzing an object in a scene that has been analyzed before. The stage 4b engines of the cerebellum can retrieve the perception in metadata that has been created earlier when that same object is located within a new visuotopic scene.

Stage 6 and 7 of this figure can be expanded to show additional signal paths from the superior colliculus to other stage 7 elements of the neural system and the associated elements of the musculo-skeletal system than just the oculomotor system.
The afferent portion of the neural system is basically a change detection system. Upon any change in the external environment, described here as a new event, the appropriate stage 1 sensory receptors sense that.

Figure 19.10.4-3 Traffic flow to and from the prefrontal cortex related to the motor function. The afferent stage 4 activity is shown for reference. The afferent role of the diencephalon is shown in greater detail within the dashed box (along with the occipital and temporal lobes of the posterior association cortex), but does not include chemical and somato sensory inputs. The sensory signal outputs of the diencephalon can come from any of its engines. The dual roles of the cerebellum and basal ganglia are shown explicitly. Alt; alternate paths available under the supervisory control of the TRN. Compare to Fuster, 2008, pg 41.
change and proceed to transmit a signal to the central nervous system (CNS or the brain) both alerting the
CNS and providing a description of the change.

The stage 2 signal processing engines forward the alarm and the information signals over frequently
separate pathways, all of which employ labeled lines that do not require including address information in
the signal packages.

Which ever stage 1 sensory neurons are involved, the first task of the stage 4 information extraction
engines is to identify the objects involved in the new event with respect primarily to their location in the
environment. Following this identification, the second task of the stage 4 engines is to perform object
recognition for each of the identified objects. This procedure provides the information that must be placed
in the saliency map to allow stage 5 cognition (described as the extreme subordinate function at the bottom
of the figure) to completely understand the event. It then has the option of communicating that
understanding via the stage 6 engines of the efferent neural pathways and the skeletal-motor system.

The procedures involved vary in nomenclature depending on the modality involved. In this discussion, the
operations of the visual modality will be discussed in greater detail. However, a similar discussion can be
presented for each sensory modality.

For purposes of psychophysical experimentation, the objects identified at the organization level will be
sub-classified as animate or in-animate purely for convenience. They are processed identically. Gauthier
& Tarr describes the next or “basic level” as recognizing whether the object is a bird, dog, human or other.
This again is an arbitrary descriptor for the convenience of the researcher. The processes are identical
across the basic level.

Following a crude separation at the basic level, Gauthier & Tarr define a 1st subordinate level where the
breed, race, or other semantic grouping is defined depending on the type of animal. Here, a 2nd subordinate
level is defined for ease of discussion. This level describes what major features of the subject are
addressed in detail. The 3rd subordinate level describes what minor features of the subject are next
addressed. As defined here and confirmed by Gosselin & Schyns, these features are interrogated by the
PEEP procedure; the application of a higher resolution interrogation procedure to each of the most
important minor features of each major feature. This procedure generates metadata related to each minor
feature. The metadata associated with each minor feature is accumulated to provide metadata describing
the overall major feature of interest. This metadata is unique and can be compared to the previously stored
metadata related to this object. If the comparison indicates a sufficiently high confidence level perception,
the other cognitive engines can take further actions. One of those actions, by way of a lookup table, is to
prepare an instruction to the stage 6 command generation engines to prepare commands to the required
multiple muscles of the musculo-skeletal system that the object recognition task has been accomplished
and the object(s) are present in the new external environment as sensed during the event acquisition task.

Note that virtually all of the object recognition activity associated with the visual modality includes a
PEEP procedure (3rd and 4th subordinate levels) that has been well documented over a period of decades
during the 1960's and 1970's but has been essentially overlooked by the more recent psychophysical
community. The PEEP procedure is intimately associated with the version servomechanism of vision as
well as the stage 4 information extraction engines.

Some examples may be appropriate here. Nearly everyone in the developed world has acquired
memories adequate to recognize these people very quickly.
1. The Little Tramp was British actor Charlie Chaplin's most memorable on-screen character and
an icon in world cinema during the era of silent film. He was always identified with his body
contours and motion at the 2nd subordinate level.

2. Recognition of Mickey Mouse depends not only on his body contour but also selective features
within his frontal image, the pair of large buttons as one example.

3. Albert Einstein was recognized by his hair at the 3rd subordinate level. It was basically his
trademark for decades. As he grew older his face morphed to that of an old man, but his hair
combined with his semi-ruffled clothing remained his major identifier.

4. Lauren Bacall was always identified by her hairdo (visual modality) combined with her husky
voice (hearing modality) at the 3rd subordinate level. Her identification was further enhanced by
some of her famous quotations (at the semantic level of the hearing modality) such as, “If you want
me just whistle. You know how to whistle don't you? Just put your lips together and blow.”
Reading is primarily a linear operation that only utilizes a small portion of the information extraction capability of the stage 4 engines associated with the foveola and orthogonal to the direction of reading. Object recognition tends to utilize the horizontal and vertical information extraction mechanisms of stage 4 more equally than in reading. This appears to place a heavier work load on the object recognition task, particularly since the scenes involved in each peep of the foveola tend to be more highly structured than do those involved in alphabetic characters. On the other hand reading requires a higher level of procedural rigidity within the version system in order to read lines of text, perform reversions when meanings are not obtained in a routine manner and in moving down to the next line of characters as part of the return of the version position to the beginning of a line.

It can be shown, see “Hearing: A 21st Century Paradigm,” that music appreciation and in fact listening to speech involves the same type of linear neural operations (assembly of individual chords into bars and ultimately understanding of the entire composition) as involved in reading.

In the current discussion, it is the interaction among a great many pieces of meta data (derived from multiple applications of the PEEP procedure to objects identified within the presented event when discussing the visual modality) that must be combined and presented to the stage 5 cognitive engines to create a holistic representation of the new event. These levels of detail are not addressed in the recent papers by the Gauthier team and by Rossion. These contributions are discussed in more detail below.

The community has not been able to understand or model how the neural system is able to “normalize” the information accumulated by multiple applications of the PEEP procedure to allow object recognition regardless of differences in scale, orientation, lighting, etc. between the current and reference scenes. A recent review paper by Rolls provides background on this subject and his most recent proposals. It is important to recognize the limitations of the real neural system with regard to transcendental calculations when one begins developing “VisNets” and incorporating Gabor Functions into the calculations related to configurational, if not conformal, transforms.

Rolls has defined a computer program he names VisNet. Quoting Rolls, “VisNet can use temporal continuity in an associative synaptic learning rule with a short-term memory trace, and/or it can use spatial continuity in continuous spatial transformation learning which does not require a temporal trace. His abstract summarizes his model succinctly. However, it is yet to be shown that his assertion, “Then a computational approach to how invariant representations are formed in the brain is described that builds on the neurophysiology.” It is suggested that the expression, “might be,” should be inserted in place of the word “are” in this sentence. Alternately, his model should be considered a digital computer analog of a fundamentally analog neural system. The abstract describes how the model has been extended recently. “The approach has also been extended to account for how the visual system can select single objects in complex visual scenes, and how multiple objects can be represented in a scene.” This last extension appears to introduce procedures similar to the PEEP procedure defined in this work (although built on the discussions of earlier investigators).

Rolls admits in Section 7, “Brain mechanisms by which this computation could be setup are not part of the scheme, and the model does not incorporate learning in its architecture, so does not yet provide a biologically plausible model of invariant object recognition.”

Section 19.10.4.1.3 will discuss the Rolls paper in greater detail.
19.10.4.1.1 Recent contributions of Gauthier psychology team

The term holistic is used in many contexts. Richler, Palmeri & Gauthier have focused on these meanings\textsuperscript{84}. They note, "The term ‘holistic’ sometimes refers to a theoretical position regarding mechanisms. Different proposed mechanisms can be ‘holistic’ in different respects, yet the same term is used for all of them. Other times, ‘holistic’ refers to some measured behavior within specific tasks. The same term is applied to different measures, even though they may be capturing different things.” They do not draw a unique preference but encourage investigators to recognize the differences and define their perspective clearly. Their figure 1 stresses the lack of clear paths between hypothesized mechanisms and the observed operational measures. The figure itself employs multiple question marks and many terms (jargon) lacking simple unique meanings.

To date, the psychology community has not recognized the physiological operation of the visual system with regard to the PEEP procedure. When this procedure is recognized as a critical element in object recognition, one of the definitions they give appears particularly relevant, “An alternative “configural” view is that representations about individual face parts and their configuration are used to recognize faces, . . . Holistic processing then reflects explicit representation of spatial relations between features [with citations].” They conclude, "It remains unclear how these various proposed mechanisms and measures are related. It could be that all, some, or none of these holistic mechanisms explain any given combination of holistic measures.”

The sections in Richler, Palmeri & Gauthier titled “Multiple meanings” and “Holistic processing as a failure of selective attention” do not recognize the existence of the PEEP procedure. The volume of their text can be significantly condensed with recognition of the PEEP procedure as a fundamental mechanism of object recognition. Late in the latter section they note, “Behaviorally, holistic processing within the composite task is characterized as a failure of selective attention. But it is critical to recognize that meanings like ‘failures of selective attention,’ ‘sensitivity to configuration,’ or ‘the whole is greater than the sum of its parts’ do not map one-to-one onto possible mechanisms.” Their use of selective attention in this context, a term suggestive of and associated with consciousness, does not relate to the non-conscious PEEP procedure.

The 1997 paper by Gauthier & Tarr defined their experimental framework, protocols and many terms used when investigating object (face) recognition.

Gauthier & Tarr have also used the term configurational transformation in the colloquial sense rather than a mathematical sense. It should be differentiated from the mathematical expression, conformal transformation. In psychophysics, the term configurational transformation applies to spatial manipulations of elements of an image and not to a non-linear conformal transformation of the type used to achieve anatomical computations within the neural system. Gauthier & Tarr frequently compare the experimental results when subjects are presented with transformed configurations and/or studied configurations.

In their paper, they describe their “experts” versus “novices” very precisely with respect to their illustrations of greebles and their previously acquired exposure to greebles.

On the first page, they proceed to establish their framework, "Thus, if subjects are using independent part representations, there should be no difference in the diagnostic information available between the three conditions. Nonetheless, parts of faces were most readily recognized in the Studied configuration, less readily in a Transformed configuration, and most poorly in isolation, suggesting that the parts of faces are not represented independently (a so-called ‘holistic representation’). In contrast, none of the tests with control stimuli -- scrambled faces, inverted faces, or houses -- revealed any advantage for recognizing parts embedded in the intact configuration of the studied object."

"Whenever a particular effect, such as that just described, is obtained with faces and not control stimuli, the question arises as to whether this implicates a face specific mechanism. From our perspective it is prudent to consider specialized mechanisms only after the best possible control conditions have failed to replicate a given effect. In the case of faces, this means using nonface stimuli that adequately match many of the visual and categorical constraints found for faces. For instance, one of the most famous phenomena

\textsuperscript{84}Richler, J. Palmeri, T. & Gauthier, I. (2012) Meanings, mechanisms, and measures of holistic processing Frontiers Psychol pp 1-6
associated with faces, the inversion effect, in which there is a disproportionate cost for recognizing inverted faces (Farah et al., 1995; Yin, 1969), has been obtained with a homogeneous set of non-face objects (dogs of the same breed), but only for expert participants (Diamond & Carey, 1986). Similarly, Rhodes and McLean (1990) obtained the caricature advantage, that is, caricatures of faces are recognized more quickly than the actual faces, with bird experts who identified members of a highly homogeneous class of birds. Such demonstrations, however, do not necessarily rule out face-specific mechanisms in all phenomena associated with face recognition -- it is certainly possible that some of the effects which are considered to be face specific are mediated by a special mechanism. However, they do not identify any such candidate mechanism(s).

"One of the most salient characteristics of face recognition is that faces have similar features organized in similar configurations. Therefore, an adequate set of control stimuli should share this constraint. For this reason, sets of exemplars from a single visually homogeneous category such as species of birds or breeds of dogs have been used as control stimuli. However, it is not only the homogeneity of the subset of objects actually used in the experiment that matters -- for familiar classes of objects, the space of all known exemplars is also crucial. Thus, the apparent homogeneity of a control set may be insufficient if the larger class is not homogeneous (as in the case of houses or landscapes, Diamond & Carey, 1986). A second characteristic of face recognition is that faces are typically recognized at the exemplar-specific level. Thus, while we often recognize most objects at the basic level (e.g., 'chair or dog,' see Rosch et al., 1976), faces are generally recognized at the most extreme subordinate level (e.g., 'Jim or Max'). Consequently, it is important that control tasks addressing face-specific effects require the recognition of control stimuli at the subordinate level (e.g., distinguishing between several dogs of the same breed). A third characteristic of face recognition is that humans are highly expert at the very difficult task of discriminating between individual faces. Although expertise is difficult to define, it seems clear that it should be more than simply a practice effect in which performance improves with experience. One empirical definition that has been used and which we will adopt here is a qualitative shift in processing. Tanaka and Taylor (1991) found such a shift for bird experts who were as fast to recognize objects at the subordinate level ('robin') as they were at the basic level ('bird'). In contrast, non-experts are consistently faster on basic-level discriminations as compared to subordinate-level discriminations. Similarly, because humans are face experts, judgments of face identity (subordinate level) are as fast as judgments that are more categorical, for instance gender (Tanaka, personal communication). Therefore, because expertise interacts with the level of categorization, it is important that control tasks addressing face-specific effects use stimuli for which the participants are experts."

The above paragraph can be better understood if their second point is re-written to separate the recognition task of stage 4 (information extraction) or stage 5 (cognition) from the reporting task of stage 6 & 7. Their third point can also be strengthened if some citations could be provided demonstrating that humans are more expert at (human) face recognition than at all other types of object recognition. Similarly, it would be useful to cite a source supporting the statement that discriminating between individual faces is a "very difficult task" compared to recognizing, for example, an error in a mask used in microcircuit fabrication.

"In the present study we chose to investigate the nominally face-specific sensitivity to changes in configuration (Tanaka & Farah, 1993; Tanaka & Sengco, 1996). In prior studies, control stimuli for faces were houses, inverted faces, or scrambled faces. Given possible non-equivalence between these sets and normal faces, we used stimuli specifically constrained to be similar to faces along several dimensions, Greebles, as our control set."

The performance of their combined test equipment and protocol appear to fall well below that generally associated with the psychophysical experiments performed by non-psychologists (even recognizing the paradigm shift in technology since the days of the Trinitron display). "The experiment was performed on an Apple Macintosh LC 475 equipped with a Sony Trinitron 13" (10.4" x 7.8") color monitor with a resolution of 640 x 480 pixels (72 pixels per inch). The Greebles were all the same purple shade, with an overhead light and a stippled texture. Images were about 6.5 x 6.5 cm (2.5 x 2.5 inches) and presented in the display area subtending approximately 6.2 x 6.2° of visual angle. A simple calculation shows the vertical raster within their 6.5 cm (2.5 inch) working area was only about 25 pixels per degree of visual field. The number of pixels in the foveola field of 1.2 degrees is about 30 pixels. Their display had a coarseness of about six times that of the retina in the area of the foveola (or a resolution about six times lower than that of the eye of their observers). Recognizing that the sensory receptors are change detectors (edge detectors when combined with the perennial tremor) makes the resolution of the display even more important. Their results comparing novices and experts defined in their figure 2 might be statistically different if the resolution of their presented images had been better matched to, or significantly exceeded, the resolution of their observers. The relevant features of their greebles would be significantly more detectable.
76 Processes in Biological Vision

In their "Conclusions," Gauthier & Tarr make several points;

“The present study shows how extensive practice with previously-novel non-face objects can lead to some of the recognition effects typically associated with faces. We found that expertise training changed novices, who were presumably processing Greebles with their “default” object recognition system, into experts, who were not only faster and more accurate but displayed a greater sensitivity to configural changes. This effect of expertise acquisition on the part recognition paradigm can be compared to Stroop interference (Stroop, 1935). Robust interference is found in the Stroop task when subjects have to name the color of incongruously colored color terms. This interference is due to the automaticity of reading that has been acquired over years of practice. In a similar fashion, the acquisition of Greeble expertise leads to interference from information that experts have learned to process automatically. This is demonstrated by the fact that our experts cannot ignore this more global information, even when it would be more efficient to do so (e.g., in the Transformed condition). In contrast to the Stroop effect, not much is known about the learning process that leads to face or Greeble expertise, nor can our experiment illuminate the particular features that are used by experts.”

It is hoped that the material in this work can provide visibility with regard to how novel imagery is processed, compared to how previously experienced imagery (even if spatially convolved (but not transformed) is processed. See Section 19.13.1 regarding the Stroop Test. The state of familiarity with the subject matter is crucially important in separating the novice from the expert in any laboratory investigation of object recognition. Humans appear to be very quick learners with respect to the visual environment.

As part of a discussion relating to single-cell work with monkeys, Gauthier & Tan made a second point;

“K. Tanaka et al. (1996)\textsuperscript{85}, working with anesthetized monkeys, have recently investigated the minimal stimulus features necessary and sufficient to activate individual neurons in infero-temporal (IT) cortex. They have found that the critical features of these cells are moderately complex (e.g., an eight point star shaped pattern or a green square above a red circle) and may be thought of as an “alphabet” of features that could be combined to code complex objects. It is possible that the complex features for which IT cells appear to be selective are not fixed but can be modified as the result of structured experience such as expertise at subtler levels of discrimination.”

It should not be assumed that anything in the above paragraph leads to an exclusive neuron, Tanaka et al. only probed a few neurons at a time. Recent fMRI imagery has shown how dangerous it is to define a single specific neuron in a single neural engine as uniquely associated with a specific visuotopic image. This has recently been described as the Jennifer Aniston Effect in the popular press. It is more likely the identified individual neuron was associated with some of the metadata describing the imagery. This position is similar to that stated by Tanaka in a separate 1996 review\textsuperscript{86};

“Conclusions
Cells in area TE of the IT selectively respond to various moderately complex object features, and those that respond to similar features cluster in a columnar region elongated perpendicular to the cortical surface. Although cells within a column respond to similar features, their selectivity is not necessarily identical. The data of optical imaging in TE have suggested that the borders between neighboring columns are not discrete; there is a continuous mapping of complex feature space within a larger region containing several partially overlapped columns. This continuous mapping may be used for various computations, such as production of the image of the object at different viewing angles, illumination conditions, and articulation poses”

Gauthier and Tarr conclude. "In summary, we hypothesized that the putatively face-specific sensitivity to configural changes might be explained by a more general recognition mechanism fine-tuned by experience with homogeneous stimuli. The present results with Greebles provide some evidence that this is indeed the case -- experts showed greater sensitivity to a change in a studied Greeble configuration than did novices. These results suggest that expertise at discriminating between visually similar objects, such as Greebles or faces, produces the obtained sensitivity to configural transformations. More generally, we believe that such results illuminate the point that visual representations and mechanisms are not steady states and, as such, it

\textsuperscript{85}Tanaka, J. Kay, J. Grinnell, E. & Stansfield, B. (1996). Face recognition in young children: When the whole is greater than the sum of its parts. Unpublished manuscript, Oberlin College.
is essential to consider how they change with experience. As Johnson and Morton (1991) have argued in their work on infants’ face recognition, only a combination of both the cognitive and the biological perspectives can provide an answer to this fascinating question."

19.10.4.1.2 Recent contributions of the psychologist, Rossion

Rossion has presented an extensive paper related to the holistic question with discussions of many isolated phenomena87.

The paper has been severely criticized in some circles because of the vituperative words used to criticize the work of other in the opening paragraph of the “General Conclusion.” Apparently the rivalry has been a long standing one. Neither side has brought in neurophysiology to support their perspective and stated paradigms.

Not having been immersed in the Gestalt school of thought, the paper is difficult to comprehend. From the perspective of one trained in the physical sciences, the Gestalt approach can be considered ephemeral.

The long paper lacks an index. However, it focuses on three experiments.

Part 1: The composite effect and holistic face Perception
Part 2: The measure of an illusion
Part 3: The illusion of a measure
General conclusions

Rossion’s frequent referral to illusions is controversial unless used as jargon within his specialty. An illusion is characterized by a lack of understanding by the subject observing the illusion. Generally, an illusion should not be considered factual or a point of reference, but as an unsolved problem. As in the synonym, a magic trick is an illusion unless you understand how it is done. Then, it is merely a manipulation of the viewers mind. Many of the discussions would appear to be strengthened by the use of Greebles to avoid the intrinsic memories stored in the subjects declarative memory resulting from earlier exposure to similar, if not identical, images (common pictures of the President and various celebrities).

Rossion does provide a figure (page 49) adapted from Orban de Xivry et al. of 2008\textsuperscript{88}, it is reproduced here as Figure 19.10.4-4. The importance of this figure is that it demonstrates unequivocally that a well-known prosopagnosiac, PS, performs the expected PEEP procedure when scanning a face, even though she is unable to achieve complete recognition of the image (identify the image by name or assigned label). Her disease is apparently resident in the area(s) or engines of stage 4 assimilating the metadata from the normal PEEP procedure into a holistic perception recognizable by the cognitive engines of stage 5. The left frame showing a “normal observer” is significantly different than the typical observer of Yarbus, of Ditchburn and of others. In fact, it is significantly different that of figure 4 in the Orban et al. paper. The Orban et al. paper does contain considerable statistical data obtained under well defined circumstances. However, it does not discuss the physiology of the visual system or appear aware of the functional details (relative to the field of view of the foveola) of the PEEP procedure.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure.png}
\caption{A prosopagnosiac and normal scanning the same object. The left image is controversial and not of interest here. The right image provides strong confirmation that the well-known prosopagnosiac, PS, does in fact perform the PEEP procedure exactly as one would expect based on scan patterns of other normals. See text. From Orban et al, 2008.}
\end{figure}

\textbf{19.10.4.1.3 Recent contributions of the modeler, Rolls}

As noted in the Overview, Rolls has recently presented a 70 page review that is primarily a review of the work of he and his associates dating from the 1990's. The work is related to the task of achieving “invariant visual object and face recognition.” Note the presence of both object and face with equal weighting in the title of the article. Although a lengthy paragraph, the first paragraph of the introduction succinctly describes the goal of this area of study. The last paragraph of the introduction asserts, “This paper reviews this combined neurophysiological and computational neuroscience approach developed by the author which leads to a theory of invariant visual object recognition, and relates this approach to other research.” The tone of this sentence suggests the article is not a true review of the field, only the authors contribution.

\textsuperscript{88}Orban de Xivry, J-J. Ramon, M. Lefe, P. & Rossion, B. (2008) Reduced fixation on the upper area of personally familiar faces following acquired prosopagnosia \textit{J neuropsychol} vol 2, pp 245–268
Rolls’ discussion of the neural system is largely limited to his block level flow diagram limited to the surface of the cerebrum of the Macaque (his figure 2). The figure excludes the occipital lobe (V1 and V2) from his signal flow (except to suggest the introduction of signals from that area into the temporal lobe. The figure does not partition the temporal lobe to allow for its role in hearing, etc. More specifically, the block diagram does not incorporate any foveola or PGN/pulvinar elements into the analysis. Figure 1(right) starts at the level of the LGN, but again ignoring the parallel visual signal processing in the adjacent PGN. Section 2.2 describes the presumed convergence of sensory neurons based on receptive fields of neurons within V1, “in such a way that the receptive field sizes of neurons (for example, I near the fovea in V1) become larger by a factor of approximately 2.5 with each succeeding stage.” This description totally overlooks the much smaller receptive field of the individual sensory neurons of the foveola represented in the PGN/pulvinar (nominally 18.5 arcsecs or 0.3 arcmin, not one degree!!) It is an array of these individual sensory neurons (roughly 175 neurons in diameter) that support the stage 4 information extraction functions associated with the foveola that support object recognition within the neural system. No report has been found in the literature that has demonstrated that the array of individual sensory neurons of the foveola, presented to the PGN/pulvinar, has been presented to is represented in the LGN/occipital lobe (V1).

Rolls unknowingly undermines his own thesis, and supports that of this work, by noting that the receptive field of a neuron of the representation of the fovea at V1 is approximately one degree. That angle is larger than the total field of the foveola (1.2 degrees) and is about 180 times larger than the receptive field of the individual sensory neurons of the foveola-PGN-pulvinar pathway. It would make correlations designed to recognize small details within the foveola very difficult using only area V1 as a starting point. He does not provide any citation for his value.

One of Rolls related citations is Boussaoud et al. The first sentences in the introduction of that paper says; “Cytoarchitectonic area TE (Bonin and Bailey, ’47), located within the inferior temporal cortex, is known to play a crucial role in the ability to recognize and remember visual objects (for reviews, see Dean, ’82; Gross, ’73; Desimone and Ungerleider, ’89). Neurophysiological studies have found that TE neurons have complex stimulus selectivities and large, non topographically organized receptive fields, which almost always include the center of gaze and frequently extend into both visual hemifields (Gross et al., ’72; Desimone and Gross, ’79).” The last phrase identifies the input to the TE from the foveola-PGN-pulvinar pathway specifically. This is the only pathway including the center of gaze and sharing its visual representation with two widely separated locations (the points of fixation) of the two visual hemispheres.

In Section 2.3, Rolls does discuss some behavioral evidence that the sensory neurons nearest the point of fixation play a different roll than those sensory neurons farther removed from this point. Although the wording suggests a failure to recognize the mechanisms involved (and the distinct role of the foveola-PGN-pulvinar signal path), it does hint at the probability of some signal path not incorporated into his model. This alternate path would provide a distinct input into the associative engines of the neural system. Whether these engines are associated with his inferior temporal cortex (IT) rather than Brodmann’s area 37 is unclear. His IT area appears to be delineated more by areas 21, 22 & 39. However all of these areas are generally related to the association function used to integrate the inputs from all sensory modalities arriving along a multitude of paths.

Rolls does discuss a variety of diode matrices. Similar matrices can be and have been correlated effectively with the reticular areas of the brain. However, it is not clear that he relates the matrices with known reticulated areas. The very long sentence in his caption to his figure 3 highlights a problem with his model as formulated; “it would not be able to provide evidence about each of the stimuli (in a scene) separately.”

The paper incorporates a large number of descriptive networks at the conceptual level. It also reports on experiments performed by Rolls and various associates.

Section 5 describes many aspects of the VisNet model and its implementation. After considerable introductory discussion, it discusses the types of filters it considered and/or adopted between the neural representations of V1 and later representations. Rolls settles on two dimensional (2D) Gabor filters as useful in his model. Such filters can be easily implemented within the spatial domain of vision. However, taking their transform into a frequency domain is not possible within the neural system as presently understood. Such a transform would generally involve transcendental calculations.

The comments relating to the neural circuitry at the bottom right of page 19 of the paper (involving ganglion neurons of the retina and unspecified neuron types of the LGN) are not supported by any citations and appear frivolous or intuitive on the part of Rolls. A more expanded neural circuit model would show DC level removal is a major responsibility of the stage 1 sensory neurons, and particularly the adaptation circuit found within all sensory neurons regardless of sensory modality.

Section 5.1.5 introduces several potential performance measures that may find general use in neuroscience research. His initial results in testing his model (section 5.2.1) are interesting and suggest his use of the scanning pattern employed in this image information extraction routine is very similar to the normal tremor pattern used to scan the pattern imaged onto the foveola (not the coarser image of the fovea delivered to V1) of the visual modality. What he describes in words as a “Z” shaped scanning pattern would be more properly described as a two-way raster scan, left to right for three pixels on the top, right to left for three pixels in the middle and then left to right again for three pixels along the bottom.

Analysis of the rest of his section 5 requires additional effort to rationalize his terminology with that of the physiological community. It will be noted that his comment related to the conscious “spotlight of attention” in the last paragraph of section 5.8.2 does not refer to movement of the point of fixation used by the non-conscious scanning process defined in the PEEP procedure of this work.

Section 8.6 of the paper describes the relative performance of VisNet versus another complex digital computer emulation, HMAX but does not appear to compare its performance to that of humans. Section 8.2 provides information on the number of computational units (?nodes) used in both the VisNet model and a more complex model known as HMAX. “VisNet uses 128 x 128 neurons in each of 4 layers, i.e., 65,536 neurons”. HMAX is estimated by Rolls to be at least 100 times the number contained within the current implementation of VisNet. Even the complexity of HMAX is dwarfed by the complexity of the primate brain dedicated to visual modality information extraction and object recognition.

Section 9 describes several mechanistic approaches to object identification, but it is not claimed that any of these approaches are used by or approach the human’s ability to perform object recognition. Rolls concludes this section with, “The full view-invariant recognition of objects that occurs even when the objects share the same features, such as color, texture, etc. is an especially computationally demanding task which the primate visual system is able to perform with its highly developed temporal lobe cortical visual areas. The neurophysiological evidence and the neuronal network analyses described here and elsewhere (Rolls, 2008b) provide clear hypotheses about how the primate visual system may perform this task.” The last sentence is debatable. His use of the words “neurophysiological evidence” is based on a very superficial examination of the neurophysiology of the primate visual system. The words “provide clear hypotheses’ followed by “may perform” suggests a different and less assertive wording.

Rolls’ conclusion section appears appropriate for a review of computer or machine vision (MV) but it does not address the neural system of vision based on the detailed architecture, down to the circuit level, of the visual modality. He does not address the computational ability of the primate brain in any obvious way. There does not appear to be any investigation of the pattern matching capabilities of the brain (other than the introduction of some graphics from other investigators based on the identification of two putative types of neural cells, “S” cells and “C” cells). This is important because the various databases referred to all suggest that pattern matching is an important element in object recognition within the neural system of animals (and specifically primates and humans in ascending order). Rolls does point to an earlier paper of his that defined an alternate approach to a neural synapse. “Another approach to the implementation of invariant representations in the brain is the use of neurons with Sigma-Pi synapses. Sigma-Pi synapses effectively allow one input to a synapse to be multiplied or gated by a second input to the synapse (Rolls, 2008b). While such a synapse can be defined, it is a common feature of virtually any neuron within the neural system based on the electrolytic theory of the neuron presented in this work.

In section 7, “Further Approaches to Invariant Object Recognition,” Rolls describes the operation of the “S” and “C” cells of other investigators but does not describe them as synapses. In fact, he discusses them in a different context, “The function of each S cell layer is to build more complicated features from the inputs, and works by template matching. The function of each “C” cell layer is to provide some translation invariance over the features discovered in the preceding simple cell layer (as in Fukushima, 1980), . . .” This context includes the ideas of matrix processing by cells with long dendrites and/or axons that are compatible with the reticulation of multiple surface areas of the brain. The use of the term “template matching” implies pattern matching in general. It is proposed here that Rolls and the cited investigators are attempting to assign too many functions to individual neurons. Previously identified reticulated areas appear to employ thousands to millions of neurons in multi-dimensional configurations before passing their outputs on to other neural engines (typically consisting of millions of neurons) for further processing.
Rolls admits in Section 7, “Brain mechanisms by which this computation could be setup are not part of the scheme, and the model does not incorporate learning in its architecture, so does not yet provide a biologically plausible model of invariant object recognition.”

Until Rolls reinterprets his conceptual model along at least two different dimensions, i.e.;

1. reconfigures his conceptual model to include the role of the foveola-PGN-pulvinar signal pathway, and,
2. reorients his model to depend more heavily on pattern matching within the above signal pathway,

he will be unable to model the neural system. This does not say, he will not develop digital computer programming routines of considerable value in machine vision. He just will not be modeling the actual neural system of biological vision.

Reorienting his digital computer based model to more closely emulate the analog based computer of the brain will be more difficult. However, this can be accomplished using the associative computer techniques explored during an earlier time (the 1960’s and 1970’s (by Goodyear Tire and Rubber Co.–military research support organization—among others). The associative digital computer was explored during the transition period between the wide use of analog computers (both electronic and mechanical) and the predominant use of digital computers within military science and virtually all other fields.

19.10.4.1.4 Face recognition using MEG by Mosher

In a recent thesis (2008), Mosher was unable to support the position that the so-called M170 signal captured by a magnetoencephalograph (MEG) was exclusive to a particular face, to faces as a class or to a building as a class. See Section 15.6.7.

19.10.4.1.5 Integrating the work from other disciplines

Before the advent of fMRI, major laboratory investigations involving the lower primates were developing the “association areas” of the brain. Much of this work involved cytoarchitectural areas given names unrelated to the Brodmann labels applicable to the human cerebrum. Boussaoud et al. have provided an excellent summary of their work and surveyed the work of many others90. Since the mid-1990’s, the focus has changed to the investigation of the human cerebrum during tasks using fMRI. The results of these investigations have generally been related to the Brodmann areas although occasionally the more explicit Talairach coordinates are used. In many of the above cases, the data collected is expressed in terms of an expanded cerebrum in order to better interpret the data collected applicable to the sulci and the gyri.

A feature of the work using the lower primates was the early identification of the stage 3 signal propagation pathways to define a mylo-architecture, or myeloarchitectonics, of the cerebrum. Similar, more detailed and false color mapping of the stage 3 axon paths using modified fMRI techniques have caught the imagination of the public and have been widely published (see Chapter 1).

From the above work, it is becoming clear that much multiple modality merging is being accomplished in the areas of the posterior temporal lobe involving Brodmann areas 21, 22, 37 & 39. However it is not clear yet from the accumulated work whether these areas are the explicit site of the saliency map or only late stage 4 engines preparing information for the saliency map. The information present in these areas is recognized to incorporate metadata of multiple dimensions and to be only casually related to the visuotopic coordinates of the original imagery. Quoting Boussaoud et al., “The information provided by this direct pathway to TE is apparently sufficient for animals to make gross discriminations among three-dimensional objects differing along numerous dimensions, such as overall size, luminance, color, texture, etc., and to store coarse representations of these objects in memory. However, when tasks require fine or subtle judgements of shape, the essential contribution of TEO is revealed.” Based on the lower primates, the areas labeled TE and TEO were defined by (Bonin and Bailey91, ’47(xxx defin. TE and TEO ). Also quoting Boussaoud et al., “Terminology aside, subsequent studies have confirmed that lesions involving TEO cause a devastating impairment in the ability of monkeys to learn visual pattern discriminations.

---

which may be even larger than the impairment caused by lesions of TE (Cowey and Gross, '70; Kikuchi and Iwai, '80).” The Boussaoud et al. paper includes extensive graphics describing their findings. Their figure 13 illustrates the falling level of correlation of the information found in areas V4, TEO and TE versus the eccentricity of the source of imagery in visuotopic space (based only on the receptive field of the neuron interrogated within these areas). Their figure 15 is particularly interesting. However, the visuotopic meridians plotted there do not correlate well with the importance of TE and even TEO strongly suggesting the likelihood that these areas are receiving additional inputs from other areas of the CNS (as they suggest elsewhere in the paper). This work hypothesizes that these inputs are via the foveola-PGN-pulvinar pathway.

Jones & Powell, also working with monkeys, have provided an early study of the convergence of multiple sensory modalities in what they estimate would be Brodmann area 40 in humans.

Brewer et al. have extended the area of study by Jones & Powell based on more recent fMRI technology. They note, “The measurements are very stable across scans, and once instrumentation and protocols are established, data spanning large regions of cortex are relatively easy to obtain.” To give an idea of the complexity involved currently in using MRI to its full advantage, the following paragraph describes their MRI machine setup process.

Multislice fMRI was performed by the use of multishot (segmented) gradient-recalled echo-planar imaging (EPI). Volumes of 13 or 17 slices of 1 or 2 mm were collected, each with a field of view of 128 x 128 mm on a 256 x 256 matrix (0.5 x 0.5 mm in-plane resolution) and 2 mm slice thickness. The acquisition parameters were echo time, 20 msec; repetition time, 740 msec; flip angle, 50°; EPI zero phase, 8.192 msec or 40% of phase steps; pulse length, 3.0 msec; spectral width, 100 kHz; line acquisition time, 1.28 msec; number of segments, 8 or 16; segment acquisition time (MRI readout window width), 20.48 msec, repetition time between slices, 37.59 msec; and number of excitations per phase encode step, one.”

Figure 15 comparing the cytoarchitechtonics of the monkey and human (including area TEO) is particularly useful and is reproduced as Figure 19.10.4-5. Note the remote location of the fixation point, labeled the foveal point and indicated by the asterisks in the center and right portions of frame (a) relevant to the major area of the V1 mapping of the visuotopic space. The likelihood of a more extensive representation of the fovea at some other location within the CNS cannot be ignored based on this representation.

At least the monkey data was collected using a high-field (4.7 Tesla) MRI machine. “At present levels of sensitivity, human fMRI is useful for visualizing activity of the whole brain with millimeter resolution (Engel et al., 1997). In conventional monkey fMRI experiments, resolution can be at the submillimeter level (Logothetis et al., 2001), and for smaller volumes of interest, the use of implanted radio frequency coils permits voxel sizes as small as 0.012 μl (0.125 x 0.125 x 0.770 mm³; Logothetis et al., 2002). Using fMRI in a 4.7 T magnet and an anesthetized monkey, Logothetis et al. (1999) made preliminary measurements of the visual field organization in macaque.” The size indicated is a major improvement over the 2 x 2 x 2 mm usually reached and suggests measurements involving on the order of one million neurons [xxx check ].

The citation accompanying figure 15 notes the lack of a precise correspondence to a distinct fovea in the macaque (where they may be referring to the 1.2 degree fovea of this work), although this foveal representation may correspond to the extended fovea in TEO. They conclude with the assertion, “The angular and eccentricity maps in monkey and human brains are quite similar, apart from the overall scale. Common features include the large foveal representation at the confluence of the early visual areas, the concentric organization of the eccentricity map, the relative emphasis of foveal signals on the ventral (compared with dorsal) surface, and the distinct parafoveal band separating the foveal representations of V1-V3 from that of V3A. A notable difference is the presence of larger peripheral representations separating the early visual areas from the human MT complex and from the ventral foveal representation. The similarity of the fMRI maps between species provides a foundation for testing hypotheses about computational homologies based on stimulus–response measures.”

Sheinberg & Logothetis have contributed a significant paper on object recognition based on experiments with monkeys. Their Discussion opens with, “The purpose of this study was to examine the physiological properties of temporal cortical neurons during exploration of complex scenes. We found visual cells in the anterior regions of the temporal lobes with reliable and selective visual responses for visual objects that the monkeys had learned to recognize. These responses were similar whether the objects were flashed in isolation or found during search, suggesting that the observed activity is related to the process of noticing particular targets, independent of how they are found.” Notice the use of memory implied by the word “learned.”

Larochelle & Hinton have recently presented an extensive mathematical study of how the visual modality might perform foveola movement, including glimpse selection and prioritizing within the PEEP procedure. It is supported by a lecture presented at the NIPS convention in Vancouver in 2010 while

---


Larochelle was at the University of Montreal and a video simulation employing different parameters for the foveola than discussed above. The simulation sequentially numbers the gazes (glimpses) but does not provide their durations.

They define their foundation model, derived for use in machine vision applications, as follows, “To tackle these issues, we rely on a special type of restricted Boltzmann machine (RBM) with third order connections between visible units (the glimpses), hidden units (the accumulated features) and position-dependent units which gate the connections between the visible and hidden units.” They then say, “We start by describing the standard RBM model for classification, and then describe how we adapt it to the multi-fixation framework. Section 3.1 describes their RBM in a mathematicians context, “RBMs are undirected generative models which model the distribution of a visible vector $v$ of units using a hidden vector of binary units $h$. For a classification problem with $C$ classes, the visible layer is composed of an input vector $x$ and a target vector $y$, where the target vector follows the so-called “1 out of C” representation of the classification label, $l.” While this investigator can not vouch for their model, the results speak for themselves. The video suggests their mathematical implementation has successfully emulated the analog PEEP procedure faithfully, including selecting important elements from a bucolic scene or face and proceeding to follow a PEEP procedure to support stage 4 information extraction during each gaze (and glimpse) in the procedure. These activities follow a training (learning plus memory storage) activity.

Directed and undirected are used here in their graph theory manner. Directed infers a unidirectional signal path along the edge of a graphical representation whereas undirected infers at least a potential bidirectional signal path along such an edge. An edge is the equivalent of a connection between two nodes in the more common engineering notation.

In the neural system concept, an undirected path can generally be described as employing two separate and distinct paths.

The architectures currently being considered within the MV community will be addressed in Section 19.10.5.

It is suggested that the introductory remarks by Larochelle & Hinton concerning a variable resolution retina can be reworded to support two parallel signal paths between stage 1 sensory neurons and stage 4 information extraction engines, the high resolution foveola based path and a lower resolution fovea path (that does include information from the fovea but not at reduced resolution. See Section xxx. As noted earlier in this section, the foveola is circular with a diameter of about 175 sensory receptors. The fovea beyond the 1.2 degree diameter foveola extends to about 5.76 degrees from the point of fixation and contains millions of sensory receptors that are grouped to achieve higher sensitivity in areas where the resolution of the optics is degraded due to its anamorphic field. Such a configuration appears ideal for assuming a sparse field of sensory elements in the fovea when passing signals to an RBM. Figure 19.10.4-6 shows the current block diagram of the visual modality with the POS indicated on the left by a rotating circle.
Figure 19.10.4-6 A revised functional diagram of human vision ca 2002. The introduction of a restricted Boltzmann machine (RBM) may explain the purpose of the histologically identified and isolated “ganglion nucleus” at the entrance to the hatched area at lower left. Otherwise the RBM(s) could be provided within the engines labeled the LGN of the awareness path and the PGN of the analytical path, particularly in the lower right corner of each where the dashed line is shown. The PEEP procedure signals would then be passed through the thalamic reticular nucleus (TRN) to the ganglion nucleus.

The conclusion of Larochelle & Hinton appear prescient in the absence of reading this work, “Human vision is a sequential sampling process in which only a fraction of the optic array is ever processed at the highest resolution. Most computer vision work on object recognition ignores this fact and can be viewed as modeling tachistoscopic recognition of very small objects that lie entirely within the fovea (foveola of this work). We have focused on the other extreme, i.e. recognizing objects by using multiple task-specific fixations of a retina with few pixels, and obtained positive results. We believe that the intelligent choice of fixation points and the integration of multiple glimpses will be essential for making biologically inspired
vision systems work well on large images.”

The next question to resolve is where within the precision optical servomechanism (POS) of the visual modality is the implementation of one or more RBMs likely to occur. As noted in the caption, there appear to be two options at this level of detailing of the functional diagram. Inputs in relatively raw signal form could be received by an RBM located within the ganglion nucleus at the entrance to the hatched area at lower left. Alternately, an RBM could be located within the low resolution awareness path incorporating the LGN and the occipital lobe. Low resolution could be interpreted as a sparse data array by the RBM. In this case, the RBM would derive the pointing signals for the PEEP procedure and pass them to the ganglion nucleus through the thalamic reticular nucleus, TRN, for implementation by the musculatura in the hatched box at lower left. The signals generated by each subsequent gaze (glimpse) would then be passed to the PGN/pulvinar couple within the analytical loop for stage 4 precision information extraction. This latter approach appears more likely.

- - - -

Benjamin et al. have recently reported on their implementation of an initial emulation of the initial stages of the visual modality using a mixture of analog and digital silicon chip technology. They define their neuromorphic system as, “Neuromorphic systems realize the function of biological neural systems by emulating their structure. Their detailed description introduces a problem. Their modeling employs man-made MOS field effect technology (FET) transistors to emulate the Activa rather than the actual PNP bipolar junction transistor technology (BJT) involved throughout the biological neural system. Based on the definitions used in this work, the Benjamin team remain in the simulation, rather than the emulation, mode of modeling.

They did not develop their biological neural circuits beyond the conceptual stage and make the simple statement, “Axonal arbors, synapses, dendritic trees, and somas may be implemented in an analog or digital fashion: In the simplest fully analog implementation, these elements are emulated by a wire, a switched current-source, another wire, and a comparator, respectively. In fact, their elements are too simple. As an example, the typical myelinated axon arbor consists of a very high performance matched impedance coaxial cable rather than a wire (Section xxx).

Figure 19.10.4-7 reproduces figure 11 of Benjamin et al. Benjamin et al. simulated a circuit based on the conventional concept of the passive membrane of a neuron, the lemma, providing active amplification without any explanation of how this is achieved in biology. This circuit employs five active MOS devices to simulate the single BJT liquid crystalline device actually used in the sensory neuron of all sensory modalities of the neural system. This single high impedance BJT device, named the Activa, forms the adaptation amplifier of the stage 1 sensory neural circuit (Section xxx use figure from My Documents here). No output is shown within their figure but it can be assumed to be taken from M5 with further circuit modification. While active external feedback is used in the simulation, no external feedback is used within the actual biological adaptation amplifier, although the impedance in the emitter circuit can be considered an internal feedback element with regard to the adaptation function.

Their equations 12 through 14 are key to emulating a biological circuit in MOS. A BJT easily achieves the results sought by implementing these equations based on it typical common base or common emitter input characteristic.

---

Their circuit is highly stabilized by the external feedback implemented using several of the MOS devices. This stabilization is not required to emulate an intrinsically highly unstable first neural amplifier, the appropriately named adaptation amplifier.

In section B, they indicate how they realized their dimensionless model, “Before describing the dimensionless models’ circuit implementations, we illustrate how such models may be realized in the log domain [Frey99, 1993] by using MOS devices operating in the subthreshold regime to realize a passive membrane [see Eliasmith & Trujillo100]

The complexity of their simulation circuits is so high (compared to the biological equivalent) and the tables of parameters (mapping constants) so large, they will not be discussed here. Most of these parameters appear to be needed to achieve the non-linearities in their models that are intrinsic to analog PNP BJT devices.

Using MOS devices in the subthreshold regime is critically important in achieving the log-based operating environment used to emulate the actual biological system. However, a high impedance BJT device can do it at lower operating cost as noted in connection with the above figure. Only through the use of analog computing techniques with its significantly lower parts count will modelers be able to approach an emulation of 10 billion active devices and over 1000 billion synapses while consuming less than 22 Watts of power. The power consumption of the models summarized by Eliasmith & Trujillo in their Table 1 would approach the multi-megawatt level when scaled up from \(10^8\) neurons to the 10 billion neuron level. The IBM Blue Gene still involves a large room full of computers operating in a massively parallel configuration. Benjamin et al. note, “Simulating a human-scale cortex model (2 \(\times\) 10\(^{10}\) neurons), the Human Brain Project’s goal, is projected to require an exascale supercomputer (10\(^{18}\) flops) and as much

---

power as a quarter-million households (0.5 GW).”

In Section V(A), Benjamin et al. make the following opening assertion, “A neuron’s spike is dispatched from its array by a transmitter, communicated to its Neurocore’s parent and two children by a router, and delivered to the recipient by a receiver. All this digital circuitry is event driven – only active when a spike occurs− with its logic synthesized following Martin’s procedure for asynchronous circuits.” There are several major problems with this assertion. First, the biological circuits being discussed do not employ digital circuitry. The circuitry involves analog single shot monopulse oscillators. Second, the stage 3 signal projection neurons do not operate synchronously. They are asynchronous driven oscillators that are naturally quiescent except when driven. Contrary to the circuitry suggested by their figure 16, each “transmitter” is formed from a single BJT oscillator, each of the Node of Ranvier repeaters is formed from a single BJT oscillator, and each “receiver” is formed from a single BJT oscillator. There are no clocking circuits or latches or memory circuits used within the stage 3 signal projection circuits of biology.

All of the functions described in their Section V(B) for a “router” are actually accomplished by individual Nodes of Ranvier (NoR) in the biological system, with each NoR formed of a single active BJT device, the Activa, configured as a driven monopulse oscillator with one or more outputs. The NoR may project multiple pulse messages but it is not aware of the character (digital words, etc) of the pulse streams it generates. There are no word shift circuits used in stage 3 signal projection circuits of biology. While there is a ladder network commonly found in the physiology of both sensory and musculo-skeletal neural circuits, it is generally unidirectional and supports various reflex functions. It is much simpler than that shown in their Figure 17 and attributed to one of the authors, Merolla, and colleagues.

Benjamin et al. close their paper asserting the following, “Neurogrid uses a few watts to simulate a million neurons in real time whereas a personal computer uses a few hundred watts to simulate 2.5 million neurons 9000 times slower than real time. The human brain, with 80 000 times more neurons than Neurogrid, consumes only three times as much power. Achieving this level of energy efficiency while offering greater configurability and scale is the ultimate challenge neuromorphic engineers face.” Adopting an analog circuit configuration will be mandatory in order to reach that goal.

Eliasmith & Trujillo document recent attempts and claims to simulate the biological brain using digital computer techniques without meeting their two criteria, “The first is that such models need to be intimately tied to behavior. The second is that models, and more importantly their underlying methods, should provide mechanisms for varying the level of simulated detail.” While these may be important criteria from a modelers perspective, they are trivial compared to the primary criteria of actually emulating the quite simple circuits of the neural system that are used repeatedly ad infinitum. These circuits are all easily emulated in analog silicon transistor technology. An excellent example is the simple one Activa circuit used to generate, and regenerate action potentials throughout the stage 3 signal propagation engines of the neural system.

Eliasmith & Trujillo refer to two remarks of Markram. First, “Markram commented on the Spaun model. “It is not a brain model.”” Second, “In both cases, Markram is taking issue with what he sees as a lack of biological realism. Such disagreements about how to construct large-scale neural models are important. If we build a model that does not capture relevant aspects of the system, it will not be explanatory or predictive, and hence not useful. What kind of explanations and predictions do the current generation of models provide?” To be quite direct, any models built on an MOS technology, even operating in the subthreshold mode, do not provide any explanatory or predictive capabilities on which to build a true, or even reasonable emulation.

After discussing Markram, the remainder of the Eliasmith & trujillo paper resorts to philosophy that will not be addressed here.

19.10.4.1.6 The inappropriate models of the sensory neuron and retina of MV

During the 1980's, attempts were made to define the electrical circuit architecture of the biological retina and individual sensory neurons. These were generally the result of a cursory reading of the biological literature (which lacked any recognition of the electrolytic character of the neural system. The result was a highly conceptual understanding of the system. Neither the internal circuitry or the dynamic performance of the visual sensory neuron, or the topology of the retina were captured in the representations of those elements used in MV. This is true of the most commonly cited descriptions by Mead as well\textsuperscript{[101].} The

configurations illustrated in Ismail & Fiez in a widely used textbook (Chapter 7) are clear examples of the differences between the MV sensory neurons and retina topology and the actual biological configuration. Even describing “The main function of the synapse cell is to achieve a linear multiplication and to provide reliable storage of a weight value (pp 330-336)” is cause for major concern. The biological synapse is merely an “active diode” created by electrically bonding two terminals of a three terminal biological Activa. It can be formed with different cross sectional areas to affect its impedance in a network, but it exhibits no computational capability (Section xxx).

At the most fundamental level, the assumption that the sensory neurons were integrating type light sensors, rather than their actually being change detection type sensors (by both the biological and machine vision communities) has had the greatest impact on these two fields. It has made progress in signal interpretation related to stages 1 to 4 very difficult in the biological community. It has led the MV community to proceed down a path largely independent of any analogy to the biological system at the operational level.

The description of the bipolar cell as a three-terminal operational amplifier in MV defies its biological name, a morphological cell with two “poles,” a dendritic pole and an axon pole (Section xxx).

The complex hexagonal network of resistors that Ismail & Fiez incorporate into their MV circuitry is actually associated with the power supply (the limited permeability of the tissue substrate to the flow of nutrients to and from the cells) in a given confined area of biology (Section xxx).

Similarly, the “winner-take-all” circuits defined in conventional analog VLSI textbooks discussing MV are replaced by simple active diodes forming junctions with a common neurite in biological circuitry (Section xxx).

Because these implementations have become so ingrained in the MV literature, it is almost necessary to accept the MV representations of the sensory receptor and retina topology as permanent circuit representations in their own right within MV, but having nothing to do with the biological elements with the same names.

It is also important to note that the description of an analog circuit within the MOS and CMOS technologies of silicon generally refer to the use of MOS transistors in a non-saturating mode, and frequently a subthreshold voltage mode (Ismail & Fiez, page 363). The actual implementation only employs bipolar transistors in the implementation if the designation bipolar-MOS (BicMOS) is used.

19.10.4.1.7 Placing the hypothesis in the language of the psychologist

Gauthier & Tarr have presented an important paper describing a potential framework for understanding object recognition in a language understandable by psychologists and physiologists as well. Unfortunately, their test instrumentation introduced a significant spatial resolution limitation of the collected data.

The paper opens with, “Several researchers have proposed that configural information, i.e., the relations between parts, is especially important in the way faces are visually represented (Diamond & Carey, 1986; Farah, 1990; Rhodes, 1987; Sergent, 1988). If this is the case, face processing, as compared to the processing of non-face objects, should be particularly disrupted by changes in the configuration of parts.” They then describe an asymmetrical set of experiments attributed to Tanaka where the images of faces were treated differently than the way non-face objects were treated.

[xxx Define terms in next paragraph ]
Based on the physiology developed in this work, and in an attempt to foster uniformity in terminology, this work suggests four functional terms leading to identification of an imaged object followed by a response function.

**Discovery**– generally associated with objects appearing in the peripheral spatiotopic field prior to a gaze moving the image of the object to the foveola as part of a new gaze.

**Recognition**– a determination of the character of an object as part of a PEEP procedure involving

---

Vision Res vol 37(12), pp 1673-1682
90 Processes in Biological Vision

primarily the stage 4 engines. Occurs most readily on objects perceived before.

**Inner loop Learning**—associated with the recognition of an object for the first time and the storage of its metadata in the stage 6 cerebellum that allows future access by the stage 4 cerebellum.

**Identification**—a determination of the internal label used previously to relate to a recognized object.

**Response**—the action taken to indicate recognition and identification of an object to the outside world. Most often a verbalization, a highly stylized movement, such as a wave or a pushing of a specific button based on previous instruction, or swinging at a baseball or other object in sports.

19.10.5 The role of learning machines in vision

Section 19.10.5 on learning in machine vision can be set aside until cited from within Section 19.10.6 if desired. It is a superficial review of the field in order to familiarize the author with the architectural frameworks and progress being made in the field. Another short introduction to this subject appears in Section 15.9. The nomenclature employed in MV differs significantly from that of biological vision and the concept of the various neural elements is quite different (see comparisons in Section 19.10.1.1).

The ANN community has been studying a large number of architectures recently based on non-von Neumann machines. All of these architectures are useful in processing two dimensional imagery. The goal has been to discover an architecture that is capable of learning a pattern or class of patterns and then announcing when it has been provided with a similar pattern of the same class. The procedure is generally described as “reducing the dimensionality of the information related to the scene” where dimensionality refers to the content of the input scene and output message, not the spatial framework.

The MV community frequently suggests their implementations were inspired by the architecture of the biological neural system (typically the brain). These conceptualizations frequently suffer from a lack of understanding of the fundamental circuits of the brain and their limitations. As a result, they are neither emulation, simulations, and do not even operate in the same space-time domain of the neurological vision system. The MV community frequently assumes a long gaze interval, measured in seconds to tens of seconds, associated with a specific small angular view of the visual field. This interval is not compatible with the short gaze intervals associated with reading and the generic PEEP procedure. The community invariably assumes the visual modality employs integrating light photoreceptors of finite capture cross section to capture 2D information (by each eye). This is contrary to the actual use of change detectors and the mechanism of tremor to capture the 3D information in a scene (where later processing combines the information from both eyes before processing via the foveola-PGN-pulvinar and perifovea-LGN-occipital pathways.

On the other hand, a review of the architecture of these MV machines is highly informative as to what techniques may be, or likely are, employed in the equivalent vision and other modalities of the biological neural system within the analog portions of both pathways of stage 4.

Domingos has prepared a very informative video on machine learning and “big data”103.” It is based on his book of the same title. The lecture is important if one wants to understand where machine language and machine vision are going. Unfortunately, he has not realized the significance of bi-stratified neurons. He describes five current tribes active in exploring machine learning. He makes some critically important comments concerning the “Bayesians.” He notes that Bayesians feel they are the most abused tribe among those working in the field of machine learning. He discusses the relevance of the “prior” in considerable detail. He compares them to the analogizers.” Domingos describes the search for the “Grand Unified Algorithm of Machine Learning.” To achieve this algorithm, he introduces Markov Functions and Processes as Markov logic networks. Interestingly, he supports the idea that humans have not yet developed the fundamental ideas needed to achieve an adequate framework.

It will be made clear in Section 19.10.6 that none of the MV architectures are incorporated in toto into the biological system. This is due primarily to the availability of a stage 5 intelligence within the real time

---

information extraction and storage mechanisms. Where the MV community speaks of supervised and unsupervised “machines” for extracting information from 2D images, the visual modality includes a supervisory capability supported significantly by various memory systems.

Therefore, this section is provided primarily as a source of background in the technologies currently under investigation in MV.

There is a considerable similarity between the size of image processing arrays used in MV and the pixel size of the foveola array of human vision. The pixel size of the broader array supporting the fovea and/or full retina of each eye when represented at V1 of the occipital lobe is less well defined.

[xxx address texture of scene and short duration of glimpses snd or saliency map (short term memory)

The MV community has not recognized the fundamental character of the sensory neurons of biological vision (BV). BV employs change detectors versus integrating detectors (typically described as imaging when used in arrays). Forwarding change detector signals through the neural system conserves significant amounts of power compared to the signals associated with integrating detectors. The major power savings is in stage 3 where most of the NoR remain quiescent except when driven to regenerate an action potential.

They biological system exhibits a pseudo frame time of about 30 ms (programmable but not carefully investigated) based on the periodicity of the tremor waveform applied by the ocular nucleus to the eye muscles.

The MV machines are quite complex circuit-wise and are dynamic with respect to their state at a given time. They generally involve both feedforward and feedback circuits in complex arrangements. The complexity of these circuits makes it difficult to describe them semantically and the presence or absence of feedback frequently depends on the level of detail being discussed in a given sentence (see the Helmholtz machine discussion below)

To simplify speaking of the overall circuits involved, a shorthand is frequently used where a group of complex circuits within an overall circuit are described as “hidden.” The result is circuits with considerable implicit information not obvious to the novice. Haykin has discussed the character of these circuits in considerable detail.

**Figure 19.10.5-1** from Ghahramani, figure 6, provides a clear illustration of the hidden layers, $S$, in his typical hidden layer machine\(^{104}\). His methodology is quite clear from his abstract, “We provide a tutorial on learning and inference in hidden Markov models in the context of the recent literature on Bayesian networks. This perspective makes it possible to consider novel generalizations of hidden Markov models with multiple hidden state variables, multiscale representations, and mixed discrete and continuous variables. Although exact inference in these generalizations is usually intractable, one can use approximate inference algorithms such as Markov chain sampling and variational methods. We describe how such methods are applied to these generalized hidden Markov models. We conclude this review with a discussion of Bayesian methods for model selection in generalized hidden Markov Models.” The term “exact inference” is worthy of conjecture. His introduction further clarifies the goals of his tutorial.

While his mathematics becomes quite complex, he describes the nomenclature and operational significance of the directed links

---

92 Processes in Biological Vision

in the figure clearly in his section 5.2,

“In factorial HMMs, the state variables at one time step are assumed to be a priori independent given the state variables at the previous time step. This assumption can be relaxed in many ways by introducing coupling between the state variables in a single time step. One interesting way to couple the variables is to order them, such that \( S^{(n)} \) depends on \( S^{(n)} \) for \( 1 < n < m \). Furthermore, if all the state variables and the output also depend on an observable input variable, \( X_t \), we obtain the Bayesian network shown in the figure.

This architecture can be interpreted as a probabilistic decision tree with Markovian dynamics linking the decision variables. Consider how this model would generate data at the first time step, \( t = 1 \). Given input \( X_1 \), the top node \( S^{(1)} \) can take on \( K \) values. This stochastically partitions \( X \)-space into \( K \) decision regions. The next node down the hierarchy, \( S^{(2)} \), subdivides each of these regions into \( K \) subregions, and so on. The output \( Y_1 \) is generated from the input \( X_1 \) and the \( K \)-way decisions at each of the \( M \) hidden nodes. At the next time step, a similar procedure is used to generate data from the model, except that now each decision in the tree is dependent on the decision taken at that node in the previous time step.”

His section 6.3 provides a clear mathematical description of the “mean field” simplification for HMM and more advanced variants of the HMM model introduced below.

Ghahramani concludes with: “We have shown how HMMs are a kind of Bayesian network, and as such, the algorithms for learning and inference in HMMs can be derived from more general algorithms for Bayesian networks. It is possible to invent many generalizations of the HMM such as factorial HMMs, tree structured HMMs, and switching state-space models which, by using richer hidden representations, can model more interesting temporal relationships than HMMs. However having richer hidden state representations invariably leads to computational intractability in the algorithms for inferring the hidden state from observations.”

Hinton and most recently in collaboration with Larochelle as noted in Section 19.10.4.1.4, has been studying a class of computational frameworks defined as Boltzmann Machines. They are designed to analyze multi-dimensional data streams using pattern matching procedures.

[ copied all of the following to Section 15.6.6 This material needs to be refocused on just the applications of the RBM’s in reading and object recognition ]

- - -

The discovery of the Boltzmann machine, initially during the 1980's and the development of the RMB in its current form during the first decade of the 21st Century, provides for the first time the technology needed to understand the operation of the stage 4 engines found in areas 17 through 19 of the occipital lobe. It also provides strong confirmation of the position of this work that area 17 of the occipital lobe does not constitute the “primary visual cortex.” The function performed and the output provided by the occipital lobe, as part of the LGN/occipital couple inside the awareness pathway are not primary. Area 17 only forms an element of the PEEP procedure leading to meaningful information extraction primarily by the PGN/pulvinar couple included in the analytical path.

19.10.5.1 Overview related to BMs

Topology becomes very complicated in discussing learning machines. Figure 19.10.5-2 from Ismail & Fiez provides a partial framework that can be expanded upon below. Each node in this figure can contain all of the elements of a von Neumann Machine (defined below). As such, it is a microcomputer in its own right. These computers can share their information as suggested by the basic network (which you will note is a feed forward network). The basic network involves no feed back paths as shown. If recurrent signal paths are shown as in the case of the recurrent network, each node is capable of adjusting its algorithms as a result of these feedback signals (which will be thought of as learning below). The implementation of forward and backward signaling can be extended indefinitely. However, the design of such networks rapidly passes beyond the capability of humans. “Machines” are currently required to design these more complex machine implementations because of the number of permutations and weights associated with different paths that are possible. The multi-layer feedforward network shown exhibits an increase in “dimensionality” due to the fact it has more output than input terminals. This network also contains “hidden layers” that cannot be seen directly by the input signals or the output signals. Some authors label the nodes (von Neumann Machines by themselves) of these hidden layers as logic units. The massively
parallel signal processing that can be achieved using these topologies becomes obvious.

Hinton has provided an overview of such machines at the introductory level\textsuperscript{105}. He specifically develops the concept of a Restricted Boltzmann Machine in that paper.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{Figure19.10.5-2.png}
\caption{Three networks of interest in machine vision. Each of the nodes shown may contain all of the elements of a Turing Engine (or the more fully defined von Neumann Machine). Columns of nodes not visible to the input signals or output signals are generally described as “hidden layers.” Networks containing recurrent paths are generally capable of “learning” by dynamically adjusting their internal parameters. From Ismail & Fiez, 1994.}
\end{figure}

A major tutorial on Boltzmann and restricted Boltzmann machines is available online\textsuperscript{106}. There is a blog to help with sticky points\textsuperscript{107}. There is also a simpler introduction “for beginners\textsuperscript{108}.” This last introduction begins, “Restricted Boltzmann Machines essentially perform a binary version of factor

\begin{flushleft}
\textsuperscript{106}\texttt{http://deeplearning.net/tutorial/contents.html}
\textsuperscript{107}\texttt{http://stats.stackexchange.com/questions/48162/good-tutorial-for-restricted-boltzmann-machine-rbm}
\textsuperscript{108}\texttt{http://makandracards.com/mark/6817-bm-rbm-for-beginners}
\end{flushleft}
94 Processes in Biological Vision

analysis. (This is one way of thinking about RBMs; there are, of course, others, and lots of different ways to use RBMs, but I’ll adopt this approach for this post.) Instead of users rating a set of movies on a continuous scale, they simply tell you whether they like a movie or not, and the RBM will try to discover latent factors that can explain the activation of these movie choices.” It goes on;

“More technically, a Restricted Boltzmann Machine is a stochastic neural network (neural network meaning we have neuron-like units whose binary activations depend on the neighbors they’re connected to; stochastic meaning these activations have a probabilistic element) consisting of:

One layer of visible units (users’ movie preferences whose states we know and set);
One layer of hidden units (the latent factors we try to learn); and
A bias unit (whose state is always on, and is a way of adjusting for the different inherent popularities of each movie).

Furthermore, each visible unit is connected to all the hidden units (this connection is undirected, so each hidden unit is also connected to all the visible units), and the bias unit is connected to all the visible units and all the hidden units. To make learning easier, we restrict the network so that no visible unit is connected to any other visible unit and no hidden unit is connected to any other hidden unit.”

Hinton has provided a cookbook for implementing a restricted Boltzmann machine. It includes a clear description of what the RBM is designed to accomplish. Section 13 is key to the implementation of the machine in biological vision. Section 15-17 describe intermediate data that can be accessed in machine vision applications to determine how well the machine is performing against a given scene. Unfortunately, access to such information in the biological vision situation appears highly unlikely.

The Boltzmann machine is fundamentally a 2D pattern matching(or differencing) machine which is ideal for implementation by a biological neural system. Pattern matching within the CNS is its forte! In this case, the pattern matching should probably occur after the combining of left and right retinal images at the stage 4 LGN and/or PGN.

---

The pedigree of the RBM might be described by using simple definitions from Wikipedia;

**Hopfield network**

a form of recurrent artificial neural network invented by John Hopfield in 1982. Hopfield nets serve as content-addressable memory systems (an associative memory) with binary threshold nodes. They are guaranteed to converge to a local minimum, but convergence to a false pattern (wrong local minimum) rather than the stored pattern (expected local minimum) can occur.

**Stochastic process**
or sometimes random process (widely used) is a collection of random variables, representing the evolution of some system of random values over time.

**Directed acyclic graph (DAG)**

A directed graph formed by a collection of vertices and directed edges, each edge connecting one vertex to another, such that there is no way to start at some vertex v and follow a sequence of edges that eventually loops back to v again

**Deterministic Machines**

**Turing Engine (machine)**

Hypothetical device that manipulates symbols on a strip of tape according to a table of rules. A Turing machine can be adapted to simulate the logic of any computer algorithm, and is particularly useful in explaining the functions of a CPU inside a computer.(1936).

**Von Neumann Engine**

Defined the fundamental functional elements of a modern computer

---

such as a memory, a computational unit (CPU), a control unit and input & output mechanisms with greater specificity (1945).

**Stochastic machines**

Bayesian networks A probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). (Wikipedia) Networks and algorithms dependent on a priori estimates of the desired results.

Markov process a 'memoryless' stochastic process that can make predictions for the future of the process based solely on its present state. Its previous and future states are irrelevant to the prediction.

Hidden Markov Models A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be presented as the simplest dynamic Bayesian network.(Wikipedia).

Parametric HMM A machine architecture designed to extract parameters related to a signal stream rather than the explicit information concerning the signal stream.

**Machines based on learning (either supervised or unsupervised)**

Boltzmann Machine A programmable interconnected multi-plane read-write memory consisting of active nodes and incorporating internal logic allowing the programming to be self-directed. Hinton describes the nodes as neuron-like units that make stochastic decisions about whether to be on or off (but without defining the properties of a neuron).

Restricted Boltzman Machine A variant limiting the number of memory planes used as references

Mean field Boltzmann Machine A variant using "mean field" units that have deterministic, real-valued states between 0 and 1, as in an analog Hopfield net.

The Hopfield network forms the foundation for most of the machines described subsequent to its definition above.

Rabiner has provided an extensive tutorial on the use of hidden Markov models (HMM) in the implementation of speech recognition software during the 1980’s110. The work of that era depended on the provision of extensive lookup tables (his figure 11). Machine learning, whether supervised or unsupervised, was used only sparingly during that period. Since that time, commercialization of machines to extract information from streams of speech and music has been very successful. The functions involved in extracting information from the visual and auditory modalities can be expected to involve similar technology when the signals reach the pulvinar. It is likely that the MV and machine hearing (MH) approaches are evolving along parallel paths. Comparing the tasks and results associated with each machine modality may be of use in discovering the framework used within each modality of the biological system.

During the 1990's, only the Markov process or hidden Markov model were explored in detail by the mathematicians associated with machine hearing (MH)111,112 but it can be expected it has moved forward into machines incorporating supervised and unsupervised learning.

---

Wilson has provided a report on extracting information related to the parameters associated with gesturing in MV\textsuperscript{113}.

Forney has discussed the Viterbi algorithm that appears to be used toward the end of processing in both the PuLP and PuMP\textsuperscript{114}.

Muller et al. have explored HMM machines where color information was an important parameter\textsuperscript{115}. Objects were presented to the MV retina with random orientations and arbitrary scales. Their figure 8, reproduced below as [Figure 19.10.5-10] shows the ability of their engine to recognize isolated objects.

Great care must be used in interpreting the mathematical language Hinton and other use in developing the RBM concept because of the tendency to use words like “visible units,” “hidden units,” “visible layers,” etc when they are typically referring to the actual data planes and physical structures employed in a conceptual or physical representation of the mathematical concepts. The terms have no intrinsic relationship with the visuotopic scene represented by the data set used to “load” these physical structures when employed in the visual modality of the neural system.

The Hinton paper describes “Different types of Boltzmann machines, including one of particular relevance, a Higher-order Boltzmann machines.”

The stochastic dynamics and the learning rule can accommodate more complicated energy functions (Sejnowski, 1986). For example, the quadratic energy function in Eq. (4) can be replaced by an energy function whose typical term is the product of three terms.” Thus, the total input to the summation for index \(i\) that is the product of two terms in the update rule must be replaced by a three term product.

The Hinton paper also describes the RBM in “Relationships to other models”

**Markov random fields and Ising models**

Boltzmann machines are a type of Markov random field, but most Markov random fields have simple, local interaction weights which are designed by hand rather than being learned. Boltzmann machines are Ising models, but Ising models typically use random or hand-designed interaction weights.

**Graphical models**

The learning algorithm for Boltzmann machines was the first learning algorithm for undirected graphical models with hidden variables (Jordan 1998). When restricted Boltzmann machines are composed to learn a deep network, the top two layers of the resulting graphical model form an unrestricted Boltzmann machine, but the lower layers form a directed acyclic graph with directed connections from higher layers to lower layers (Hinton et al. 2006).

**Gibbs sampling**

The search procedure for Boltzmann machines is an early example of Gibbs sampling, a Markov chain Monte Carlo method which was invented independently (Geman & Geman, 1984) and was also inspired by simulated annealing in statistical physics.

**Conditional random fields**

Conditional random fields (Lafferty et al., 2001) can be viewed as simplified versions of higher-order, conditional Boltzmann machines in which the hidden units have been eliminated. This makes the learning problem convex, but removes the ability to learn new features.”

Hinton also notes the critical impact of the data structure and the number of reference planes used in a


\textsuperscript{114}Forney, Jr. G. (1973) The Viterbi Algorithm Proc IEEE vol 61(3), pp268-278

given instance on,

“The speed of learning:

Learning is typically very slow in Boltzmann machines with many hidden layers because large networks can take a long time to approach their equilibrium distribution, especially when the weights are large and the equilibrium distribution is highly multimodal, as it usually is when the visible units are unclamped. Even if samples from the equilibrium distribution can be obtained, the learning signal is very noisy because it is the difference of two sampled expectations. These difficulties can be overcome by restricting the connectivity, simplifying the learning algorithm, and learning one hidden layer at a time.”

Hinton provides a range of references in this paper.

- - - -

Another paper attempts to summarize the above insights and focus on the potential for misapplying an RBM in a given situation116. A limitation of particular interest is that “the likelihood function has local optima.” This may form the primary reason the PEEP procedure frequently revisits the same features in an intrinsically static image. The paper is a valuable source of citations ca. 2014.

- - - -

The Helmholtz machine

A Helmholtz machine (HM) has also been explored. It is particularly appropriate for applications requiring unsupervised learning. Quoting Haykin (1999), “The Helmholtz machine, first described in Dayan et al. (1995)117 and Hinton et al. (1995)118, provides another ingenious multilayer framework for achieving a similar objective without the use of Gibbs sampling.” The Dayan et al. describes the naming of this machine and its basic concept, “Following Helmholtz, we view the human perceptual system as a statistical inference engine whose function is to infer the probable causes of sensory input. We show that a device of this kind can learn how to perform these inferences without requiring a teacher to label each sensory input vector with its underlying causes. A recognition model is used to infer a probability distribution over the underlying causes from the sensory input, and a separate generative model, which is also learned, is used to train the recognition model.” It also describes both deterministic and stochastic variants of the Helmholtz Machine. The Hinton paper more clearly separates the recognition and generative phases of machine operation, using the labels “wake” and “sleep” respectively. In the biological case, these alternate labels are probably inappropriate. Both recognition and learning to recognize a new object both occur while the subject is totally awake. The Dayan et al. version of the Helmholtz model incorporates a restriction to limit the number of potential ways that a pattern can be recreated during the regenerative phase of operations.

Dayan et al. describe the restriction they apply to the HM just as they had in the BM to achieve the RBM configuration (their section 3).

“The key simplifying assumption is that the recognition distribution for a particular example $d$, $Q(\phi, d)$, is factorial (separable) in each layer. If there are $h$ stochastic binary units in a layer, the portion of the distribution $P(\theta, d)$ due to that layer is determined by $2^h - 1$ probabilities. However, $Q(\phi, d)$ makes the assumption that the actual activity of any one unit in layer is independent of the activities of all the other units in that layer, given the activities of all the units in the lower layer, $l-1$, so the recognition model needs only specify $h$ probabilities rather than $2^h - 1$. The independence assumption allows $F(d, \theta, \phi)$ to be evaluated efficiently, but this computational tractability is bought at a price, since the true posterior is unlikely to be factorial: the log probability of the data will be underestimated by an amount equal to the Kullback-Leibler divergence between the true posterior and the recognition distribution.

The generative model is taken to be factorial in the same way, although one should note that factorial generative models rarely have recognition distributions that are themselves exactly

116http://www.metacademy.org/graphs/concepts/restricted_boltzmann_machines
They also note, “It is essential that there are no feedback connections in the recognition model.” Based on these two restrictions, and more recent terminology, their HM should be labeled an RHM. They also incorporated several mean-field criteria, and some minimal preset biases for stability (see figure 4 caption), into their RHM. Their Discussion compares the operational results obtainable with a variety of different machines.

Figure 19.10.5-3 shows a typical Helmholtz machine. Quoting Haykin who uses the MV community concept of a synapse, “The Helmholtz machine uses two entirely different sets of synaptic connections, as illustrated in the figure for the case of a network with two layers of two-valued, stochastic neurons. The forward connections, shown as solid lines constitute the recognition model. The purpose of this model is to infer a probability distribution over the underlying causes of the input vector. The backward connections, shown as dashed lines, constitute the generative model. The purpose of this second model is to reconstruct an approximation to the original input vector from the underlying representations captured by the hidden layers of the network, thereby enabling it to operate in a self-supervised manner. Both the recognition and generative models operate in a strictly feedforward fashion, with no feedback; they interact with each other only via the learning procedure.” See Haykin (page 575), or more recent works, for further discussion of this network and the semantics of whether this network incorporates feedback or not. Dayan and also Hinton appear to describe the recognition and generative paths as top-down or bottom-up based primarily on the orientation of their diagram when they draw it on paper.

Figure 19.10.5-3 The architecture of a Helmholtz Machine. The feedforward paths are described by solid lines labeled Recognition. Dayan describes these paths as bottom-up. The feedback paths are described by dashed lines labeled Generative. Dayan describes these paths as top-down. See text. From Haykin, 1999.
19.10.5.1.1 Recent progress in reading of script by machine vision

Interpreting the recent progress in machine reading of handwriting is complicated by the separation of several distinct technological groups exploring this area.

Recent progress in machine reading has been impressive, to the point that virtually every automated teller machine (ATM) incorporates check reading software that identifies and reads virtually every feature entered by handwriting or printing. A2iA Corporation appears to be the leader in the implementation of such technology. A wide range of papers have reported on variety of mathematical approaches to machine reading. These now extend well beyond the optical character readers (OCR) of the past to Intelligent word readers (IWR). Most involve some type of Markov Field analysis. A specialization in this area is a family of Boltzmann Machines designed to analyze two-spatial-dimension scenes. The challenge of recognizing “holistic” character sets (words or multiple syllable groups within long words) appears to have been solved. Marti & Bunke have presented a variety of papers on their investigations including Marti’s PhD thesis of 2000. In the first paper, they assert, “We describe the three main components of the system, which are preprocessing, feature extraction and recognition.” Preprocessing includes a skew and slant correction step as well as horizontal scaling. They describe two approaches to the preprocessing step, “In the first, lines of text are segmented into individual words during a preprocessing phase, while the second method is totally segmentation free, i.e. the segmentation is integrated into, and obtained as a byproduct of, the recognition process.” In their section 4, Recognition, they discuss the viability of individual HMM’s for each word versus individual HMM’s for each character. In the third paper (appearing in a special edition of a journal dedicated to HMM), they describe the steps involved in using a simple HMM and “supervised” learning in text recognition. Figure 19.10.5-4 provides their flow diagram. They assert,

“In the presented system the following five preprocessing steps are carried out:
• Line Separation: The given page is divided into lines.
• Skew Correction: The text line is horizontally aligned. For this purpose the skew angle needs to be determined (see their Fig. 3, left part).
• Slant Correction: Applying a shear transformation, the writing’s slant is transformed into an upright position (see their Fig. 3, middle part).
• Line Positioning: The text line's total extent in vertical direction is normalized to a standard value. Moreover, applying a vertical scaling operation the location of the upper and lower baseline (see their Fig. 3, right part) is set to a standard position.
• Horizontal Scaling: The variations in the width of the writing are normalized.

Note that the order of these operations is not arbitrary. Because the nonlinear scaling in line positioning and horizontal scaling affect angles, skew and slant have to be corrected before positioning and scaling. Correcting the skew of the writing means a rotation of the image. This changes also the slant angle. Therefore the skew has to be corrected before the slant.”

These steps are also needed in biological vision and the precedent order is probably the same. Their paper develops each of these tasks in greater detail. In their section 5, they note, “Hidden Markov models (HMMs) are widely used in the field of pattern recognition. Their original application was in speech recognition. But because of the similarities between speech and cursive handwriting recognition, HMMs have become very popular in handwriting recognition as well.”

Figure 19.10.5-4 A flow chart for supervised script recognition using an HMM. Analysis on the left, supervised preparation of the data base on the right. From Marti & Bunke, 2001c.
The third paper also illustrates the concept, Figure 19.10.5-5, of using multiple HMMs to assemble a perceived sentence by concatenation of individual tasks. “In the recognition phase the character models are concatenated to words, and the words to sentences. Thus a recognition network is obtained.” As they note, to combine the character level task into the word level recognition process places a much larger learning task on the HMMs involved. Separating these tasks is more economical during the supervised learning activity. “In the network the best path is found with the Viterbi algorithm.”

Lavrenko et al. have also presented a useful paper122 focused on the holistic reading of groups of characters using the expression words (actually either words or multi-syllable groups as shown in Section 19.8). “A document is described using a Hidden Markov Model123 (HMM), where words to be recognized represent hidden states. The state transition probabilities are estimated from word bigram frequencies. Our observations are the feature representations of the word images in the document to be recognized. We use feature vectors of fixed length, using features ranging from coarse (e.g. word length) to more detailed descriptions (e.g. word profile).” “Recognition was performed with a Hidden Markov model with 14 states for each character. These Markov models were concatenated to produce word and line models.”

\[ \text{Figure 19.10.5-5 Concatenation of analysis cycles using HMM. From Marti & Bunke, 2001c.} \]

---


Hawkins has provided an excellent talk at the introductory level as to how RBMs can be implemented in machine vision\textsuperscript{124,125}. As appropriate for his focus on machine vision, he uses what might be considered a very condensed concept of how the visual modality actually works. He does not consider the physiology of the visual modality at all.

The field of machine vision is beginning to mature and it has drawn a very large contingent of young researchers into the research and a large amount of government money to support that research from many of the governments and large corporations around the world.

Hinton has provided an excellent video series on Coursera demonstrating how well an RBM can work to recognize simple text (even when in a variety of fonts)\textsuperscript{126}. There are many individual lectures by Hinton on YouTube.

\subsection*{19.10.5.1.2 The papers of Livingstone & Hubel}

The papers of Livingston & Hubel of 1984, 1987 and culminating in 1988\textsuperscript{127,128} provide a comprehensive set beginning to isolate feature extraction functions between the LGN, the V-labeled areas and the then largely unknown functions of the PGN and pulvinar in macaque monkeys. They do recognize the bifurcation of the optic nerve into bundles going to the LGN and what they label the superior colliculus (SC) instead of the PGN located along the Brachia of the superior colliculus (the neural path between the LGN and the SC). This work generally supports their early findings with change detection (including motion) the primary responsibility of V1. They were the first to describe the columnar structure of V1. They described the columns as blobs based on their cross section. They described these blobs as about 0.2 mm (200 microns) in diameter and with a center-to-center spacing of about 0.5 mm. The relationship between these blobs and the logic units of ANN’s requires discussion below. They also defined what they called interblobs. They asserted that they only observed blobs and interblobs within V1.

Figure 4 of their 1988 paper was meant to be comprehensive and steer their discussion. However, it is conceptual and contains a variety of poorly defined areas and question marks. It does suggest the continuation of the magnocellular path (the R–channel—brightness) of the optic nerve through the LGN to layer 4C\textalpha of area V1 and beyond. It also suggests the continuation of the parvocellular paths (O–, P–, Q–channels—chrominance) of the optic nerve through the LGN to layer 4C\textbeta of area V1 and beyond. These two paths relate to the full fovea. Based on current knowledge, it can be inferred that layer 4C of V1 is primarily concerned with change detection, primarily based on brightness information but potentially on some chrominance information. The R–channel represents the summation of the four spectral sensory channels of animal vision, ultraviolet, short, medium and long wavelengths. In discussing their figure 1, they did recognize that the R–channel data did include chromatic information in an undefined manner.

Since they did not appreciate that the brightness information presented to V1 is computed from the spectral signals, the remainder of their paper related to V1 becomes difficult to interpret or defend because of test protocol issues. Several footnotes in their 1988 paper and a concurrent paper\textsuperscript{129} discusses several inconsistencies between their results and those of others.
The luminance tests should be repeated under the condition that $P = Q = 0.0$. They did assert the neurons within a blob of V1 were not orientation selective. They also assert the neurons within an interblob of V1 were orientation selective. They show these signals being passed to the middle temporal lobe (MT) and their graphic suggests MT is distinct from V4. Their graphic suggests V4 is focused on color information. More recent data from follow-on investigations by other investigators may be available for review.

**Figure 19.10.5-6** from Wilson provides a clear representation of the various visual areas associated with the cytoarchitecture of the surface of the cerebrum.

![Figure 19.10.5-6 Designated areas of the cerebrum serving awareness vision. All of the cerebrum areas are within the two partial spheres of the occipital lobe and symmetrical about the longitudinal fissure. From Winslow, 1999.](image)

### 19.10.5.2 The RBM as a potential template for an engine(s) of the coarse awareness path

The concept of an RBM may well be the key to answering a long standing question; what is the actual role of the stage 4 engines of the occipital lobe of the cerebrum. While it has historically been assumed, based on very little histology but a lot of topographic mapping, that these areas are related to each other sequentially, as indicated by their numeric designations, this may not be the case.

As noted in the above citation to the metacademy.org presentation, RBM:s can be stacked. Areas V1, V2, V3 etc exhibit topographic maps of visuotopic space with what has been assumed to be lesser levels of precision as the numeric designation increases. The topographic maps of each region frequently abut the maps of an adjacent area in unexpected ways. These poorly understood ways can be taken to suggest a neurological continuity between the areas or, on the other hand, can suggest an independence and their independent implementation of multiple (potentially stacked) RBMs. The stacking could be within the thickness of the neural surface forming the cerebrum or they could be stacked functionally but be physically adjacent on the surface of the cerebrum.

It is critically important that future researchers appreciate that the biological vision modality is based on stage 1 *change detecting* sensory neurons. The signals passed from stage 1 through the early engines of
104 Processes in Biological Vision

stage 4 do not contain information about uniformly illuminated regions of the visuotopic scene. Where in stage 4, individual closed contours are recolorized (by filling in) is not currently obvious. It may be that the color of a given closed contour is only defined via the metadata for that scene element.

It is also critically important that future researchers recognize the importance of Riemann Transformations that are introduced between the visual field and the representations of that field at various locations within the neural system. Because of these transformations, the application of Boltzmann Machines (BM), whether restricted to Restricted Boltzmann Machines (RBM) or Support Vector Machines (SVM) can only be applied to small areas of the overall visual field.

To date, the use of the emulations based on various forms of BM have been restricted to small images (typically less than 200 by 200 pixels). These image sizes are most compatible with the foveola-PGN-pulvinar pathway in human vision and not the much broader field of view of the perifovea-LGN-occipital lobe pathway. They relate very well to the instantaneous size of the field of view of the high resolution foveola as reported in Section 19.8.2.3.

A majority, if not all of the work to date in information extraction through computer analysis suitable for use in machine vision offers little detailed insight into the biological neural system, except at the most gross philosophical level.

19.10.5.2.1 Potential implementation of an engine with hidden layers in V1

The area of the occipital lobe, labeled Brodmann area 17 is a very good candidate to be an implementation of an RBM designed primarily to perform two functions;

1. Discover the more important areas of a the complete visuotopic scene presented to it and

2. Preparing a high level instruction defining the sequence of saccades and gazes to be implemented by the precision optical servomechanism (Section xxx) in a subsequent PEEP procedure.

The PEEP procedure would perform the steps in providing individual gazes (glimpses) of the scene to the stage 1 photoreceptors (sensory neurons) of the foveola. The more localized, but very high resolution, image from each gaze would be encoded by the stage 2 engines and delivered to the analytical processing “service” described earlier as the analytical path within the visual modality (Section xxx).

---

In the following discussion, the term “sparse” will be used in two distinct contexts.

A sparse (thinned) subarray – An ordered subarray of a larger array designed to reduce the number of pixels that must be manipulated within the context of an RBM. Typically involves regular grouping of pixels into a lower resolution variant of the original array.

A sparse (thinly populated) subarray – A subarray (as above) where only a limited number of the nodes are at a different intensity or value from the majority of the nodes.

Based on these contexts, the stage 2 engines of the retina propagate a “sparse (thinned) subarray” of the complete array of sensory neurons to area 17 of the cerebrum via the LGN/occipital pathway, the forward propagation pathway of the awareness mode of operation in the visual modality. The signal structure propagated to the subarray delivered to area 17 consists of a “sparse (thinly populated)” set of data values. 

---

In this proposed implementation of an RBM in area 17, the topographic map would be of low resolution in order to provide rapid filtering of the entire visual field (or at least the complete foveal field of about 8.68 by 8.68 degrees; the nominal values used in this work). Rather than be described as a low resolution image, the image used by the RBM can be described as a sparse array in the language of mathematics.

The multiple layer structure of the cerebrum in area 17 appears ideally matched to the configuration requirements of an RBM. This functional multilayer configuration supporting a sparse array appears ideally suited for detecting changes in the visuotopic field rapidly. The changes may be in terms of changes in brightness of a given pixel or small group of pixels (summed to form the sparse array), or of
such changes caused by motion of objects (or lighting changes) within small areas of the visuotopic scene. Such changes would be very compatible with the laboratory findings of Hubel and associates during the 1970's. A summary of Hubel's work appeared in 1988. To minimize the required number of data planes within V1, it is likely that this RBM is implemented to only extract information involving changes in value associated with elements of the sparse array. However, its operating logic may be sufficient to correlate such changes between adjacent elements of the sparse array in order to determine the direction of changes of distinct line or 2D objects in the visuotopic field.

If V1 is dedicated to extracting very simple changes in signal intensity among individual elements or small groups of adjacent elements in the visuotopic field of view, it is rational to presume that V2, V3 etc may implement RBMs designed to be selective for more complex patterns of change in signal intensity extracted from the visual field of the fovea.

The order of the RBM used in V1 is not currently known, however, it is clear that the signals from the merged imagery formed within the LGN clearly contain at least one additional parameter besides their angular position in the visual field. They clearly include depth information related to stereographic imaging. This combination of metadata would support at least a third order RBM as described by Larochelle & Hinton.

The output signals from V1, and possibly the higher numeric areas, would be forwarded back through the thalamic reticular nucleus (TRN) for purposes of prioritization before the forwarding of a single PEEP procedure instruction to the ganglion nucleus of the precision optical servomechanism (POS) shown in figure xxx.

Is this “forwarding back to the TRN” of extracted signals the “back propagation” of Hinton? How does Hinton use the term “geometrical transformation?” Is it the equivalent of anatomical computation used to describe the rearrangement of axon arrays to accomplish mathematical transformations without employing transcendental computations? It appears that the answers to these questions are positive.

To further reduce the processing load on the stage 4 engine of area 17 of the occipital lobe, it appears likely that the RBM associated with V1 does not employ a set of rules that require frequent updating. It may only use a largely static set of rules that are replicated across the extant of the subarray like computer display wallpaper, albeit as a continuum rather than individual tiles.

Esser et al. have now simulated the performance of V1 using their non-von Neumann neural network architecture developed under the TrueNorth Program sponsored by DARPA. The program employs a “crossbar” matrix very similar to the configuration of an A. T. & T. system 5 central office crossbar switch of the era before the introduction of all electronic central office switches. In this case the relays used to lock the long “dendrites” and “axons” in place are replaced by neurons.

Figure 19.10.5-7 shows their simulation using their “optical flow” algorithm, two frames of a file, and four of their single hidden layer corelets. Note how their simulation, and its tailored algorithm, provides an indication of change at every column (indicated by the arrowheads) simultaneously when implemented in the biological equivalent.

---

As noted in Section 3C, V1 can only accommodate a few layers of a potential RBM based on the TrueNorth architecture. Many investigators have also noted the considerable loss in speed associated with RBM architectures employing several hidden layers. These facts suggest that V1 is used primarily to ascertain changes related to objects in the visuotopic scene, primarily due to the motion but also sensing changes in light intensity associated with a given pixel. Subsequent neural networks, V2, V3, etc. would be implemented with serial FBM’s consisting of only one or two hidden layers as suggested by figure 3 in Hinton & Salakhutdinov [37]. This interpretation is in good agreement with the morphology of the human brain shown in Figure 15.2.5-6 from Chapter 15 [38] and the learning procedures in Chapter xxx [xxx] of the authors work.

In realistic operation, the visuotopic scene presented to V1 has had most of the global motions of the scene removed, only relative motions within a given gaze need be identified by V1. Thus the challenge of each V1 are two fold;
• to compile a list of locations within the total field that exhibit local temporal changes between the frames presented, and
• to compile a list of locations within the total field that exhibit rapid changes in signal amplitude between spatially nearby pixels of the most recent frame presented.

Without having an algorithm from either machine or biological vision to examine, the only available information is behavioral information from the psychophysical laboratory.

19.10.5.2.2 Potential introduction of engines with hidden layers in V2 through V5/MT

Traffic analysis and VEP measurements have developed several concepts of the role played by the Stage 4 engines labeled V2 (of the occipital lobe) and V3 and higher of the parietal lobe. However, few diagrams or schematics have appeared to support definitive discussions. Figure 19.10.5-8 presents a block diagram of the proposed operation of these engines for purposes of discussion. The figure is derived from Figures 15.2.5-6 and 19.10.4-3 of this work and figure 4.6.3-1 of “The Neurons and Neural System.” The first figure is derived from a paper by Gross, Bender & Rocha-Miranda132. The latter two figures are expansions of a Fuster paper133.

As noted by Maunsell & Newsome134 in an excellent source, many of the stage 4 engines areas (out of more than 20) identified shown in Figure 15.2.5-6 are found in the “association areas” related to the temporal lobe rather than the parietal lobe. Many of the areas are not located on the exterior surface of the cerebrum but on more interior gyri and sulci. Recent fMRI studies have made it clear that many of the stage 4 engines are not exclusive to the visual modality but may be considered “association areas” merging sensory data from many sensory modalities. These findings further complicate this figure and the figures noted above.

noted in lesion experiments, “Visual acuity declined sharply as the test stimulus approached the visual field location corresponding to the V1 lesion, and no threshold could be measured at its center. In contrast, lesions of area V2 caused no measurable decrease in acuity, nor was there any substantial effect on several measures of contrast sensitivity.”

“In this first controlled fixation study of the visual effects of V2 (as well as partial V3) lesions, we found no alteration in visual acuity and little or no change in contrast sensitivity. Conversely, a small lesion of area V1 caused an apparently complete loss of visual function in the corresponding region of the visual field, with no recovery over the 10 months of the experiment. However, two tasks chosen to reflect shape discrimination or grouping abilities were severely impaired by the V2 lesions.” They conclude, “These results suggest that lesions of area V2 may selectively interfere with grouping processes without disrupting basic visual capacities.” Later, “V2 is the first, and largest, cortical visual area that can be lesioned without removing most of the input to later stages. Because V1 sends direct projections to V3, MT, and V4, ablation of V2 leaves some routes to higher visual cortex intact. The effects of the V2 lesions were dramatically different than those of V1 lesions, causing no change in acuity, and little or no change in contrast sensitivity.”

The role of the higher numbered V areas remains in question. Merigan et al. offered a number of comments relating to whether the higher numbered V areas are in serial connection with V1. “The effects of V1 lesions have also been examined by determining their effects on the response properties of neurons in other cortical areas. Inactivation of V1 by cooling caused a near complete loss of response in V2 and V3 neurons, although about 30% of neurons in V3a continued to respond (Girard and Bullier, 1989; Girard et al., 1991a). The same investigators also found a complete loss of visual responses in area V4 when V1 was cooled (Girard et al., 1991b). These results are consistent with the notion that the studied cortical areas are strongly dependent on that major part of their input that comes from the geniculostriate pathway. However, a very different result was obtained for area MT. Rodman et al. (1989, 1990) found little change in the directionality, binocularity, receptive field size, or topography of macaque MT neurons after extensive V1 lesions, although the amplitude of responses as greatly reduced.” They noted the experiments of others have been inconsistent in this area.

- - - -

[xxx add more on higher designations after V2 & MT, including their location relative to the parietal or temporal lobes.

19.10.5.2.3 A major problem associated with V2 through V5/MT

The data collected by the psychophysical community is difficult to interpret with precision because of the investigator’s use of a wide variety of poorly defined stimuli. Using the terms white, red, green and blue are grossly inadequate. Test stimuli that are to be considered “luminance only” are generally obtained from a broad spectrum source that affects all of the chrominance channels (O–, P–, Q–) as well as the brightness channel (R–). If the goal is to evaluate the response of only the brightness channel of the system, care should be taken to employ two narrowband light sources with the coordinates 494 and 572 nm ± a few nm from the New Chromaticity diagram. These values assure the researcher that the signals in the chrominance channels are at null, P = Q = 0.0 and the brightness channel exhibits a finite value, R = a finite value. When performing experiments based on color stimuli, narrow band illumination should be used centering near 610 nm (red), 532 nm (green) and 437 nm (blue). It is preferable that LED or other monochromatic sources be used to establish these colors.

19.10.5.2.4 The “inclusive” evaluation of Felleman & Van Essen re: V1 to V5/MT

Felleman & Van Essen136 performed a comprehensive study of the sensitivity of the major stage 4 engines of the macaque monkey as summarized in Figure 19.10.5-9. The column labels do not purport any exclusivity regarding that feature. No information was provided as to the quality of the “white” light used in the data collection process (color temperature of the source or spectral content). The majority of their color sensitivity was based on an unusual criteria derived from a spectral sweep using a narrow bandwidth monochromator as follows;

---

“Figure 5 Tuning profiles for four parameters in a typical V3 neuron. . .C: color selectivity. Responses to 12 narrowband isoluminant color stimuli and isoluminant white moved through the receptive field in the preferred direction at the preferred speed. This cell gave approximately equal responses to each narrowband color tested and was classified as noncolor selective.”

“This cell responded well to monochromatic stimuli extending from 460 to 660 nm. A color index (CI), analogous to the aforementioned direction index, was used to quantify the response of the least effective monochromatic wavelength relative to the most effective wavelength (CI = 1 - worst/best). This cell had a color index of 0.34. There is a small dip in the tuning curve in the range of 500-560 nm, but we find this variation to be not statistically significant.”

“Similarly, a white index WI = 1 - white/best monochromatic) was used to quantify the response to isoluminant white relative to the most effective monochromatic wavelength.”

Figure 19.10.5-9 Percentage of cells responsive to various stimuli in six visual areas. Their criteria were carefully defined but unusual. Their color sensitivity presumed a preferred direction and speed. See text. From Felleman & Van Essen, 1987.

There tests were concentrated in areas more than two degrees from the point of fixation. Note the use of photopic nomenclature (isoluminant and equiluminant) rather than the more appropriate (but not critical at their level of investigation) radiant intensity units.
110 Processes in Biological Vision

Their figure 14 A shows a joint distribution diagram for their white index and color index measurements associated with V3. “Cells varied continuously in their selectivity for color and their responsivity to white. Dashed lines indicate criterion limits for color-selective and white suppressed cells.”

Their figure 17 compares the data underlying the above figure with that from a variety of other investigators. The scatter in the comparison data is quite high.

The lack of sufficiently tight control within their protocol may be surmised from their conclusion #5, “Contrary to previous reports, ~20% of the neurons in V3 were color selective in terms of showing a several fold greater response to the best monochromatic wavelength compared with the worst. Color-tuning curves of the subset of color selective cells had, on average, a full bandwidth at half maximum response of 80–100 nm.”

The criteria used in the above and similar studies of the 1990's do not provide significant information about the operational purpose of the engines generally labeled V1 through MT/V5 Van Essen (1985, page 315 in Peters & Jones) has probably provided the best data regarding their function but the data remains sketchy and possibly controversial. He did note the relatively large number of responses to objects moving through the visual field at velocities at a 32 degrees/sec rate (range 2 to 256 degrees/sec).

19.10.5.3 The RBM as a potential engine of the precision analytical path

[xxx edit to broaden from RBM to include mean field BM and potentially a Helmholtz Machine ]

The material in Section 19.8 has provided a large amount of data concerning how the pulvinar extracts information during reading. Tracing the path followed by young children learning to read is instructive. Whatever, the mechanism, they learn their ABC’s of the Roman alphabet by learning how a variety of straight strokes can be recognized as letters through repetition. The individual characters are initially block letters without serifs. Based on the simple electrical displays used widely, each character can be displayed using a combination of only seven strokes. With time, the child can recognize three letter words requiring a maximum of only 21 strokes. Then they progress to recognizing a maximum of about five letters before requiring a saccade. Later, they are presented with less precise characters representing the same character set. Machine vision research has shown this process can be simulated using some variant of a Boltzmann machine. Success has been achieved to date with typically only one hidden layer serving the foveola-PGN-pulvinar analytical pathway. The principle question appears to be, can the same level of performance be achieved for features scattered about the 1.2 degree field of the foveola as can be achieved with symbols restricted to a central horizontal/vertical or diagonal band of the visuotopic scene? The answer appears to be yes because of any reason to doubt the capability. A follow-on question is, can the subject learn to recognize more complex patterns within the field of the foveola as time goes on? The answer appears to be yes based on personal experience. However, for new complex patterns, a significant repetitive learning process may be required.

The pulvinar is tasked with extracting features from a glimpse that is presented to the foveola following a saccade from a list created in V1. No major ocular movement occurs during the glimpse. However, the tremor associated with the mammalian eyes continues during the glimpse interval. This interval appears to be terminated by the pulvinar with an instruction to the POS to proceed with the next programmed saccade in the PEEP procedure list. With the continuing tremor of the eyes, the elements within the nominally 175 pixel diameter glimpse that represent edges within the visuotopic scene are highlighted. It is the ensemble of edges that must be interpreted by the stage 4 circuitry within the pulvinar.

As noted above, the neural system employs change detection techniques in its sensory neurons in order to save power during stage 3 signal projection. Even in the presence of tremor, the sensory receptors will not produce a significant output signal while viewing a uniformly illuminated region of the visuotopic scene. As a result, the concepts employed in the “front end” of any machine designed within a MV context must be modified if it is to be used in BV.

The analyses of Marti & Bunke in the development of MV feature extraction topology appears to be relevant to the biological case as well (Section 19.10.5.1). They define three main modules:

• the preprocessing, where noise reduction and normalization take place,
• the feature extraction, where the image is transformed into a sequence of numerical feature vectors, and
• the recognizer, which converts these sequences of feature vectors into a minimal-dimensionality code.

In the case of analyzing text, a fourth part, the language model, provides linguistic knowledge about the context in which a word is likely to occur.
In the general case, it is likely that the pulvinar is tasked with several preprocessing steps before beginning feature extraction in order to compare the presented image with the previously stored images. Modifying their terminology in order to apply to a general 2D image, the following steps appear appropriate:

- Feature Separation: The glimpse is divided into areas of significant spatial element density.
- Skew Correction: The image is horizontally aligned. For this purpose the skew angle needs to be determined (see their Fig. 3, left part).
- Slant Correction: Applying a shear transformation, any slant associated with the image is transformed into an upright position (see their Fig. 3, middle part).
- Vertical Scaling: The image’s total extent in vertical direction is normalized to a standard value. This may be accomplished by setting the vertical extremes to upper and lower baselines.
- Horizontal Scaling: The variations in the width of the writing are normalized. This may be accomplished by setting the horizontal extremes to left and right baselines.

“Note that the order of these operations is not arbitrary. Because the nonlinear scaling in line positioning and horizontal scaling affect angles, skew and slant have to be corrected before positioning and scaling. Correcting the skew of the image means a rotation of the image. This changes also the slant angle. Therefore the skew has to be corrected before the slant.”

These steps are also needed in biological vision and the precedent order is probably the same. Their paper develops each of these tasks in greater detail.

The first lecture in the Coursera series on “Neural Networks for Machine Learning” by Hinton shows an RBM in operation, and essentially reading the numbers on a hand written check (as if read one character at a time). This can be considered the primitive mode of operation for the analytical pathway. In fact, as in the case of reading (discussed extensively in Section 19.8), the pulvinar is able to recognize syllables and up to about five letters during a single glimpse. Thus values written onto a check up to about 99.99 can be recognized within a single glimpse.

The pulvinar is a very large knot of neurons in the context of the neural system. It is almost certainly folded in a manner similar to the LGN as described in Section xxx. Limited data suggests the folds within the pulvinar may be accessible from both sides as found for neurons passing through, as well as terminating or originating in, the TRN. This capability allows investigators to consider Boltzmann Machines with multiple hidden layers to be employed profitably within the pulvinar. Koh has provided a recent diagram of the complete thalamus showing both the regions of the pulvinar (using Latin labels) and the thalamic reticular nucleus (reticulaire kern or TRN) shown in cutaway. Chalfin et al. have provided information on the size and the number of neurons within the visual portion of the human pulvinar generally labeled the lateral posterior pulvinar (PuLP). According to Chalfin et al., the PuLP contains a nominal 10 million neurons within a volume of 3.560 mm³ (values are for both PuLP’s combined. No comment was made concerning the “saddle” connecting the two bulk bodies). Their method of determining the volume of the brain and the volume of the pulvinar in a given species is not described in detail. By their calculations, the pulvinar occupies only 0.267% of the volume of the human brain. Their values are consistent with other investigators within their morphology community (Section 1.2.3). Their terminology differs from that used in this work.

Their estimate of the volume of the human brain as equivalent to a sphere of 5 cm diameter would suggest they included the space dedicated to commissure and volumes associated with space involved in the folding operation. They assert they measured the volume beginning at the pyramidal decussation in the medulla including olfactory bulbs. A preferred estimate would consider the area of an inflated and flattened brain (thereby eliminating folds) multiplied by a nominal thickness of the resulting sheet. Their description of their counting technique is also very brief. Indicating the use of a 750x light microscope without describing the instrument quality or the wavelength of light used leaves many questions. Many small neurons (or at least their neurites and axons) can not be resolved with a light microscope.

---


Fischl et al. have focused on inflating and flattening the human brain to determine its surface area and ultimate volume on a more histological basis. Section 10.2.1 in “The Neuron & Neural System” reviews the available data. The reported areas vary between 0.15 and 2.0 meters². As an example, an area of 1 square meter multiplied by a thickness of 6 mm would result in a total volume of the neural sheet of only 6,000 mm³, twenty times smaller than the value of Chalfin et al. and the morphological community. Using that figure and a PuLP volume of 3,560 mm³, the pulvinar would represent over 50% of the human brain. The available data from different disciplines does not converge on a viable values for the volume of the neural sheet forming the brain or its major subdivisions at this time (ca. 2014).

Opposite each PuLP within each pulvinar is a matching, similar sized, structure, the medial posterior pulvinar. It is generally labeled the PuMP. Each PuMP is associated with the auditory modality and is believed to perform the same tasks as the PuLP, extracting information from the stage 4 signals presented to it from the earlier auditory engines.

The saddle connecting the two sides of the pulvinar is voluminous and most of the literature suggests the PuLP’s do not operate separately in a bilateral manner. The same hypothesis may apply to the two PuMP’s. The saddle may contain more than just commissure. The foveola representation within the pulvinar does not exhibit the characteristic features found in the fovea representation at the two V1’s of the occipital lobes. No reports have been encountered describing the loss of perception related to only one half of the foveola for an example. However, the medical literature does contain reports of some patients not being able to draw a complete image observed by just their foveola. Such a failure might suggest failure in some part or parts of the foveola-PGN-PuLP pathway.

Diagrams of the pulvinar usually include the LGN and the equivalent Medial Geniculate Nucleus (MGN) serving the auditory modality. They usually do not incorporate the vision or auditory Perigeniculate Nuclei (PGN) that are located along an arm (brachia) extending to the superior and inferior colliculi respectively (Section xxx). While the LGN and MGN are located adjacent to the pulvinar, they do not interact with the pulvinar significantly. The PGN in particular does. Its output signals terminate in the pulvinar.

The purpose of the PGN, as identified in this work, is similar to that of the adjacent LGN. It is responsible for merging the signals received from the two retina in the overlapping areas of the foveola. The tasks of the PGN are to extract high precision stereoscopic control signals for delivery to the POS, establish depth data associated with significant elements within the glimpse being examined, and deliver the merged image plus the depth tags to the pulvinar for information extraction (the what). Extracting the depth data may be performed in conjunction with the information extraction function of the pulvinar.

19.10.5.3.1 Performance of 256 x 256 Boltzmann machine arrays

With the present state of the art, the machine vision community has been implementing a variety of Boltzmann machines in silicon based on a 256 x 256 pixel array. This array size is only marginally larger than the 175 diameter array size forming the foveola. The demonstration video provided by Hinton of his 256 x 256 array RBM shows it is quite capable of recognizing one character of an alphanumeric string of characters no matter what the font looks like, even hand drawn digits. The first question is can it recognize two characters placed side by side without an additional period of extended learning? If yes, can it recognize three characters simultaneously?

19.10.5.3.2 Performance of the 352 x 264 array of Muller et al.

Figure 19.10.5-10, from Muller, Eickeler & Rigoll, show the recognition paths using their feed forward part of a mean-field BM or a mean-field version of an HM. The glimpse is introduced at lower left as a result of the tremor generating signals describing both edges and textures within the individual signal paths. The upper path corresponds to their feature stream 1, extracting brightness only information. The lower path is condensed but represents feature streams 2, 3 & 4 extracting chrominance parameters. The subsampling frames show how the results of their signal processing are presented in an r, φ coordinate system centered on the center of gravity (COG) of the image. This COG should be analogous to the “where” coordinates of the glimpse extracted within V1. Their procedure is described as;
They describe the preprocessing calculations performed by their machine as “The feature stream 1 of the sequence, which is derived from the shape of the object, is calculated by applying the following steps to each image in the database (see upper part of Fig. 2). First the color image is transformed into a gray-scale image and thresholded afterwards in order to obtain a binarized image. Due to a fairly homogeneous background, this relatively simple method could be used. After binarizing the given image, the center of gravity (COG) as well as the maximum radius $R_{\text{max}}$ of the shape are calculated and stored for later use. Thereafter, a lowpass filter is applied for avoiding alias components in the image which is subsampled in the final step. In order to calculate the color information based feature stream, only the filtering and sampling steps are performed on the individual red, green and blue channels of the color image (see also lower part of the figure).

The subsampling step is performed in an adaptive way which results in scale invariance and which leads also to rotation invariance together with the modified HMMs. Samples are taken on polar coordinates $(r, \phi)$ where the origin is placed at the COG.”

Figure 19.10.5-11 shows the details of the subsampling “raster.” The mathematics supporting this raster are in their paper. The caption may be misleading. The process employed was actually a linear transformation from an rectilinear to a circular coordinate system. The number of total samples was unchanged.

Their figure 8, reproduced here as Figure 19.10.5-12 shows the ability of their machine to retrieve images matching the
spatial properties and color of their query sketches.
Figure 19.10.5-12 Query sketches and retrieved images. The labels under each frame show: the number of features being aligned by each of two internal HMM's and the estimated rotation angle between sketch and retrieved image. Note the colors match very well, e.g. the grips of the tongs (sketch 3) are red and so are the first three retrieved tongs. From Muller et al., 2001.
In brief, they assert;

“Thus, the features used in our experiments are samples taken on a polar subsampling grid, which are arranged in a feature sequence of length T. The components of the feature vectors are divided into two feature streams, where the rst stream represents the shape of the object and has a dimensionality of k. The second stream is calculated on the red, green and blue channels of the image and consists of k samples for each channel, which results in a total size of 3k (see also Figs. 2 and 3). The use of more than two streams would also be possible and could be utilized e.g. to further subdivide the color stream into chrominance and luminance streams (using a YIQ color-space), which would allow the user to control the influence of the brightness of an object. Another interesting configuration would be to integrate shape, color as well as texture streams into a Markov Model (note that texture classification has been performed applying HMMs in Ref. 1). The feature extraction is, as far as shape is concerned, related to the so-called signatures, which have been successfully combined with HMMs in order to build a recognition system for hand-drawn sketches. In Ref. 6, signatures are used in combination with one-dimensional Markov Random Fields (MRFs) for the simultaneous delineation and classification of shapes in the presence of clutter. The feature extraction presented in the present paper is more flexible than the signatures, due to the possible integration of color and the use of unconstrained shapes rather than being restricted to closed contour shapes. However, a combination of the MRF-based delineation process and the approach presented in this paper has the potential to result in a system which is able to retrieve color images in a cluttered environment, especially since MRFs and HMMs bear many similarities.”

Their terminology, YIQ, from television practice is equivalent to the color axes of the subtractive color process used in printing. They did not explain why they chose to suggest YIQ rather than the R–,P–, and Q– channels of this work. It may have been because they were working with printed query material. The axes sets are merely rotations of each other around the central point of the New Chromaticity Diagram (Section 15.3). This New Chromaticity Diagram is the theoretical underpinning of the CIE L*A*B* and L*U*V* color spaces. They actually used the RGB (without any Y channel) color coordinate space in their experiments, which are converted to R–,P– and Q– channels within the biological vision modality.

Their paper also describes how they created a rotation invariant HMM. The mathematics are more than can be included here.

“It can be seen in Fig. 8 that the system performs very well and that images can be retrieved by presenting simple sketches. Note that the user of this system does not need to know the rotation angle of the desired object and is also free to apply sketches with a skew. The angles given in Fig. 8 are good estimates of the true rotation angle of the retrieved objects with respect to the sketch. Although these values may be of no importance in this application, they provide the possibility for additional evaluations of the rotation invariant Hidden Markov modeling technique.”

When discussing the time required for their HMM to complete a cycle, they discuss a tying technique that offers further decreases in settling time which appears compatible with a biological application of the HMM technique. Their conclusion section is worthy of careful study.

19.10.5.3.3 Thoughts on resolving a historical paradox—occlusions of the target

The potential use of a Boltzmann machine in the biological process of object recognition appears to provide an answer to the historical question of how does the brain recognize only partially complete drawings of some object. Since these machines are based on prior learning, employ a large group of individual strokes to identify an object, and give a report based on the statistical ability of these machines to match a pattern, the absence of a particular stroke or contour in the presented object merely lowers the confidence level in the reported recognition.
Pessoa, Thompson & Noe have addressed the subject of illusions. They are philosophers who have offered their views based on intellectual and intuitive concepts lacking any basis in neurology. If the neural system incorporates some variants of Boltzmann machines, which involve a learning phase, understanding these illusions takes on a different character. Figure 19.10.5-13 shows two common illusions. Nearly every school child encountered the Kanizsa triangle and learned to recognize the incomplete triangle after a few encounters with it. The placard with “JESUS” in reverse contrast appeared in this author’s environment in the 1960’s. As a bumper sticker, it took me nearly a dozen casual encounters before I interpreted it appropriately. It then became instantly recognizable ever after. These appear to be cases of my object recognition capability learning the features of these images critical for identification.

Removing the solid black area of the Kanizsa triangle, which is a fill-in within the visual modality, leaves only a contour consisting of a circular portion and a triangular portion. When attempting to recognize this image, it is likely that the three triangular forms at equal distances from and facing each other is quickly recognized as an equilateral triangle. If this is correct, then the recognition engine reports this set of features as an equilateral triangle based on its previously memorized form (which had continuous lines connecting the triangular corners. Pessoa, Thompson & Noe note that the white area within the Kanizsa triangle appears lighter than the surrounding white area. This was the first time I was made aware of this feature. This feature may tell the researcher more about the fill-in logic used in the visual system where the analytical pathway attempts to fill a closed contour with a color that is the average of the color sensed in the inside of the contour based on the change detection signals generated by the stage 1 sensory neurons. The high contrast between the solid black areas at the corners and the very white inside of those corners may be used to define a fill-in that is even whiter than the surrounding white areas. These surrounding areas are based on the sensing of the box including the complete graphic. It is noted that if one focuses their attention on only one of the white-on-white areas forming portions of the Kanizsa triangle, the difference in whiteness is not perceived. The difference in whiteness is only perceived in the context of the feature recognition process. Pessoa, Thompson & Noe discuss this subject in their section 3.3.4, Boundary completion, using the designations edge-induced and line-induced. Within the visual modality, the edge-induced contrast is generally higher because the step change in contrast is more faithfully reproduced than a more bandwidth limited response associated with a line-induced transition.

The Jesus illusion is of a simpler type; it just requires the subject to establish the feature that the illusion is presented in a contrast opposite to that normally encountered. The feature, along with other illusions involving reversed contrast, suggest that normal or reverse contrast is a significant feature that is stored as a minimal dimensionality signature related to a gaze.

Pessoa, Thompson & Noe discuss “feature completion” (i.e., fill-in) only briefly in their section 3.3.5. They did not address the question of whether they were exploring the operation of the foveola or the full fovea and the differences in spatial resolution between these areas. They did make some remarks concerning the potential function of V2 with respect to contour completion based on the work of von der Heydt et al. (1984). “von der Heydt et al. (1984) tried to draw a clear distinction between the stimulus-response relationship, on the one hand, and perceived entities, on the other. For instance, they

Figure 19.10.5-13 Two common illusions. Top; The Kanizsa triangle. Bottom; the reverse printed Jesus illusion. See text.
used the term illusory contour stimuli, rather than illusory contour cells, and they borrowed the term anomalous contours from Kanizsa (1955; 1979) to define a stimulus property without reference to perception. ” **These comments are important** when interpreted in terms of the theory presented in this work. The visuotopic image presented to the retina includes large areas of constant lightness and chrominance that are not reported faithfully by the stage 1 change sensitive sensory neurons. The signals produced at the output of stage 2 relate fundamentally to the contour information within the visuotopic scene. The broad areas of constant luminance and chrominance are reconstructed (filled-in) during stage 4 information extraction.

Pessoa, Thompson & Noe also address fill-in with respect to signals recorded from the area of V1 associated with the blind spot of the retina (section 4.2). Subsequently, they address the representation of scotoma introduced into the retina of monkeys on various V# areas. (section 4.3).

Section 5.1 of Pessoa, Thompson & Noe explores the many ways representations and the mechanism of fill-in may arise, or relate, based primarily on philosophical grounds. They readdress the subject of fill-in in section 6 after some ancillary discussions. “Two lines of evidence point to an early neural mechanism for illusory contour completion: (1) neurophysiological data, and (2) psychophysical studies of the similarities between real and illusory contours.” Their section 7 includes a variety of discussions of empirical investigations surfacing discrete phenomena related to vision. Their paper is accompanied by an extensive set of largely philosophical and behavior oriented comments (48 pages).
19.10.5.4 An alternate school using alternate terminology–SVM as a BM

A separate group of mathematically inclined scholars have employed what look much like Boltzmann Machines but labeled Support Vector Machines (SVM) to explore information extraction applicable to machine vision systems. The SVM’s employ multiple layers of matrix oriented signal processing in both linear and non-linear modes. They also appear to be compatible with the ideas of supervised and unsupervised learning.

Yu et al\textsuperscript{140}. provided a review in 2002 of several approaches to visual information extraction using SVM’s. The assertion is made that these approaches are based on “biophysically plausible implementations of the maximum operation.” They defined a maximum operation (MAX) based on using the peak value achieved for each of the analog neurons of a matrix of neurons representing some features of a limited visual field.

They did not demonstrate that the neural circuits of biology could perform the complex and largely transcendental mathematics of their MAX operation.

Jhang et al\textsuperscript{141}. followed with another biologically-inspired model.

Izhikevich et al\textsuperscript{142}. provided a paper that over-emphasized the role of action potential neurons in 2004. However, their introduction is informative.

The cerebral cortex includes a population of billions of neurons, each making thousands of synaptic contacts with its neighbors. Given the complexity of the connectivity inherent in cortical anatomy, efforts to describe the pattern of electrical activity in exact detail within even a highly localized population of cortical neurons would be extremely difficult. Our understanding of cortical dynamics is also rendered difficult by the temporal properties intrinsic to such a biological system.

The paper is worthy of a data mining effort even though many of the values are taken from very old sources. They make no distinction between analog and pulse signaling when they note,

The conduction velocity for the myelinated axons was 1 m/s, in accord with the value experimentally measured for myelinated cortico-cortical fibers in awake adult rabbits (Swadlow, 1994). In contrast, the conduction velocity of non-myelinated local axonal collaterals was only 0.15 m/s, in accord with values reviewed by Waxman and Bennett (1972); corresponding delays could be as long as 10 ms.

They employ a very large scale computer model, “The total number of excitatory synaptic connections is 8,000,000 and of inhibitory synaptic connections is 500,000.” Their figure 2 contains only artistic representations of considerably simplified neural pulse waveforms.

They conclude with a hint of the self organizing capability of the BM as a general case,

“Despite the lack of coherent input in the model, neurons self-organized into groups and repeatedly generated patterns of activity with a millisecond spike-timing precision. The properties of the model allowed us to explore the spike-timing dynamics of neuronal groups, in particular their formation, activity-dependent persistence, competition, cooperation and responses of receptive fields to coherent input. We hypothesize that the appearance of such neuronal groups is a general property of re-entrantly connected spiking networks with axonal conduction delays and STDP (spike-timing dependent plasticity).”

Jhuang et al\textsuperscript{143} provided a paper that used the expression biologically inspired machine vision architecture. In fact, the paper offered little information on the neural physiology or circuitry they inspiring their architecture. They do cite an earlier paper by two of the authors, Giese & Poggio\textsuperscript{144}. Their paper also includes an un-credited figure that looks very similar to some of the work of Hinton using a Restricted Boltzmann Machine (RBM) architecture. This may be due to their using the same template provided by the National Institute of Science & Technology.

The third paragraph of their introduction is based more on the protocol of earlier investigators than it is on the architecture of the neural system. They note Hubel and associates use of optical bars moving in the visual field of a subject but do not recognize or appreciate the Riemann Transforms employed between the visual field and the representative field on the surface of the occipital lobe of the brain. (See Tootell material in \textbf{Section 15.5} & \textbf{15.6}, specifically \textbf{Section 15.6.5.4.2} [xxx confirm after editing of that material ]. In fact, straight lines in visual space are generally converted into other conical forms at the visual cortex associated with the occipital lobe in order to support truly biological neural information extraction.

They note, “Our approach is closely related to feedforward hierarchical template matching architectures that have been used for the recognition of objects in still images. These systems have been around for quiet some time now, starting with the work of Fukushima and of LeCun et al. . Here we follow the more recent framework using scale and position invariant C2 features that originated with the work of Riesenhuber & Poggio.”

Their section 1.2 describes the scope and limitations of their model (including a variety of citations.

“Typically, computer vision systems for the recognition of actions have fallen into two categories. One class of approaches relies on the tracking of object parts. While these approaches have been successful for the recognition of actions from articulated objects such as humans, they are not expected to be useful in the case of less articulated objects such as rodents. The other common class of approaches is based on the processing of spatio-temporal features, either global as in the case of low-resolution videos or local for higher resolution images. Our approach falls in the second class of approaches to action recognition. It extends an earlier neurobiological model of motion processing in the dorsal stream of the visual cortex by Giese & Poggio.”

In the next paragraph they note, “\textit{In particular, the model of Giese & Poggio is too simple to deal with real videos.} It lacks translation invariance and uses a limited handcrafted dictionary of features in intermediate stages.” As noted throughout this work, the use of videos to represent the information captured by the neural system, where the retinas are actually made up of change detectors rather than integrating sensors, is not warranted in biologically-inspired models. In the text, they note, “In order to convert the neuroscience model of Giese & Poggio into a real computer vision system, we altered it in two significant ways: We propose a new set of motion-sensitive units which is shown to perform significantly better and we describe new tuning functions and feature selection techniques which build on recent work on object recognition.” They do not indicate any use of Riemann Transforms to provide a more realistic representation of the captured imagery within the biological system.

They do provide citations relating to the KTH human action dataset (which they downsized to 160 by 120 pixels) and the TheWeizmann human action dataset (180 by 144 pixels). They also note the size of the UCSD Mice dataset is 240 by 180 pixels. The resulting image sizes are very compatible with the foveola-PGN-pulvinar pathway of human (biological) vision but do not relate to any biological features related to the signal processing within the occipital lobe of the human brain. These sizes relate to animals with a unique foveola and probably do not relate to many of the lower animals, including some monkeys (\textbf{Section xxx}).

Escobar & Kornprobst\textsuperscript{145} presented a paper in 2008 that opened with, “Action recognition (AR) in real scenes remains a challenging problem in computer vision. Until recently, most proposed approaches considered simplified sequence databases and relied on simplified assumptions or heuristics. Some

examples of these kind of approaches are [five citations], where one could find therein other references and further information.” They assert, “In this article, we follow the same objective as in Jhuang et al., which is to propose a bio-inspired model of motion processing for action recognition (AR) in real sequences.” In their introduction to Section 2 they note, “Several bio-inspired motion processing models have been proposed in the literature [five citations], those models were validated considering certain properties of primate visual systems, but none of them has been tested in a real application such as AR.” Their figure 1 provides a flow diagram through their architecture but can in no way be considered a model of the biological information extraction architecture. Their paper does not incorporate any recognition or description of the role of the lateral geniculate nucleus (LGN) or any other components of the thalamus in controlling/reformatting the information provided to area 17 of the occipital lobe (such as differential delay adjustment to aid in information extraction.

Their closing paragraph offers a useful conceptual and philosophical view of what lies ahead based on their biologically inspired approach to information extraction in machine vision. It offers little relative to the actual operation of the neural system.

Subsequently, Escobar & Kornprobst released another paper. They continued to refer to bio-inspired modeling but made several profound statements concerning how earlier papers did not appreciate the complexity of the visual modality with regard to the V1/MT cortical areas. They made a demonstrably erroneous statement early in their paper, “The primary visual cortex, designated as V1, is the first cortical area encountered in the visual stream processing. It is probably the most studied visual area but its complexity still makes it very hard to accurately model.” They associate the above claim with, “our goal is to propose new motion-based features for action recognition, inspired by visual system processing. To do so, we propose modeling the motion-processing pathway, focusing on cortical areas V1 and MT.” Thereby, they totally ignore the crucial processing steps encountered in the thalamus of the diencephalon and specifically the role of the thalamic reticular nucleus (TRN) in switching and supervision of neural signals.

They expand their concept of biological MT processing with,

“The activation of a MT neuron inside its classical receptive field can be modulated by the activation of a surround area. This has been widely studied in the neuroscience community but is usually ignored in most MT-like models. Interestingly, in Born (2007) the authors found two different types of MT cells:
(i) Cells purely integrative where only the response to the classical receptive field (CRF) stimulation is considered. These neurons without surround interactions strongly respond to wide-field motion.
(ii) Neurons with antagonistic surrounds modulate the response to CRF stimulation, normally suppressing it. These neurons are unresponsive to wide-field motion but they strongly respond to local motion contrasts.”

After further discussion of the above two points, they assert,

Considering all this information, and more specifically the work by Xiao et al., we propose to model four types of MT cells as shown in Fig. 2 (illustrating their concept of the four types): One basic type of cell only activated by its CRF, and three other types with inhibitory surrounds. The tuning direction of the surround is always the same as the CRFs, but their spatial geometry changes, from symmetric-isotropic to symmetric-anisotropic and asymmetric-anisotropic surround interactions.

If corroborated, these additional types of MT cells have major impact on earlier symmetrical models.

They provide an interesting Table 1, listing the size of their fovea (the foveola of this work) in pixels and a variety of other parameters relating to the KTH and Weizmann databases. The dimensions given are comparable to those employed within the foveola-PGN-pulvinar pathway of the diencephalon but are not comparable to any biologically recognized features of the neural circuitry of V1/MT. It is noted that they recognize the size of the individual photoreceptor cell is larger than the pixel used to describe the resolution of the retina. The cells are 3 to 10 times larger in diameter than the pixel. To achieve these ratios, the photoreceptors must be acting as edge detectors and not integrating detectors with a specific cross sectional area.

As in the earlier papers in this section, Escobar & Kornprobst never identify what specific areas of V1 or MT they are addressing on the cerebral cortex. Their results appear to be more related to the operations of the foveola-PGN-pulvinar than to V1 and MT.

Their results appear more useful in the machine vision domain than in establishing any relationships between the biological vision modality and the computational structures they develop. Their mathematical steps involve a variety of transcendental operations not known to be performable by the neural system. Their appendices appear to be irrelevant to the biological visual modality.

19.10.6 Plausible model of biological object recognition

The machine vision community has made great strides in image recognition in recent years. The results en bank provide an adequate database on which to propose the method of operation used within the analytical pathway (foveola-PGN-pulvinar) of human vision. However, any proposed new models must respect the currently documented use of Riemann Transforms and change detection sensitive detectors in the biological system. Confirmation of any proposed models is needed before further progress can be made. This section will;

- outline a current understanding of the problem,
- review the appropriate functional parameters of the biological vision modality, and
- suggest an initial baseline framework for achieving object recognition within the pulvinar, possibly supported by the cerebellum.

The generic problem is for the stage 4 information extraction engines to accept the array of signals originating at the stage 1 sensory neurons of the foveola and extract a very much lower “minimal dimensionality” descriptor of the information contained in that set of signals. The desired dimensionality can probably be represented by single word of nominally 16 parallel analog bytes, of probably less than 4 bits per byte (similar in information content to the 16 characters of a credit card account number). Such a sequence is frequently labeled a feature sequence \((a_1, \ldots, a_T)\) in MV. It is also desired that the order of the bytes be related to their significance in the visuotopic image presented to the foveola. The criteria related to significance may be determined by the software more than what would be considered significant by a human evaluator. Ideally, the word should be independent of the scale and skewness of the image presented to the foveola, and to the greatest extent possible independent of the orientation of the object in the external environment. The word could then be stored in declarative memory and placed in the saliency map for access by the stage 5 cognitive engines. The overall word may contain both brightness and chrominance information ala the parallel streams used by Muller et al. [xxx Muller et al. cited here ]

In the Muller et al. example, they used a feature sequence of five dimensions in the shape stream and a total of 15 dimensions in the three color streams. It may not be necessary to maintain and/or store five dimensions in each of the three color streams.

The initial conclusions are;
- None of the commonly named computing machines of MV are used in the biological system.
- Some of the techniques used in the MV machines are likely emulated in the biological system.
- The technique of parallel stream information processing is widely employed in the biological system.
- The separation of broad area gaze and local area glimpse information extraction is a built-in feature of BV systems.
- Some form of metadata encoding and transmission is a feature in BV systems.
- The biological system employs a stage 5 intelligence that must be supplied externally in MV applications.
- The biological system does not require the highly repetitious learning activity of MV.
- The BV system employs multiple asynchronous operating cycles with an average period of about 30 ms.
- The cerebellum appears to play an important role in support of the foveola-PGN-pulvinar pathway.

19.10.6.1 Neurological background specific to recognition task

Several features of the biological vision modality need to be highlighted before proceeding.

1. The stage 1 signals delivered directly to the stage 4 PGN are in RGB and partially blocked UV format. The signals presented to the pulvinar are dependent on ocular tremor to provide an analog representation of the central 1.2 degree diameter field (the acceptance range of the foveola) of the visuotopic scene centered on the point of fixation.
2. The signals from the foveola represent about 23,000 pixels forming a disc of radius 175 pixel diameter.
Maximum acuity in object recognition is achieved when all chrominance channels are activated, i.e., the scene is illuminated by broadband white light with a centroid of 494, 572 nm on the New Chromaticity Diagram (Section 15.3). The sensory neurons are packed into a hexagonal array with one sensory neuron in the center of the hexagon and six at the corners. The neurons of the array have different spectral sensitivities.

3. The PGN passes the foveola signals to the pulvinar with depth information as metadata.
4. The analog color signals are of notoriously low precision, probably not over 4 bits precision and already normalized with respect to individual sensory neurons and localized color information from the scene.
5. Brightness information (R–channel) if used at all is assembled in the PGN/pulvinar from the RGB signals.
6. The signals related to a single glimpse are presented to the analytical pathway until the pulvinar/cerebellum issues a signal that object recognition is complete.
7. Tremor within the precision optical system (POS) continues during the glimpse interval. The system requires at least 30 ms to create an initial representation of the visuotopic scene within a single glimpse. The tremor involves motion in both the vertical and horizontal directions and is only significant with respect to the foveola, but not the larger area fovea.
8. Gross angular rotation of the presented glimpse imagery can be flagged by the pulvinar and cause a reorientation of the imagery by movement of the head or body, via instructions generated in stage 5 and subsequent commands issued by stage 6. [xxx edit re cerebellum ]
9. It is proposed that the pulvinar has access to sufficient internal or cerebellum memory to store a large group of learned images.
10. The biological system necessarily employs unsupervised learning with respect to the stored images.
11. Color plays a significant role in biological vision.
12. Unlike in V1, there is no expected identification of areas of specific interest within a glimpse.

Based on these features, it can be deduced that the pulvinar and/or cerebellum employs an object recognition “machine.” The little available information concerning the internal topology of the pulvinar and cerebellum does not allow the determination of the complexity (in terms of hidden layers) of individual machines used within the pulvinar or cerebellum. If confirmed that axons can reach both sides of the neural tissue without difficulty, the use of multiple hidden layers within one machine of the pulvinar would be expected. More information is needed concerning the topology of the cerebellum.

19.10.6.2 Schematic of human learning–supervision or no supervision

The MV community has discussed methods for a machine to establish a reference image that can be compared to a test image. The generic method is for a machine to establish a minimal-dimensionality feature sequence \( \left( o_1, \ldots, o_T \right) \) extracted from an initial test image, store that sequence and then use that sequence to generate a representation of the original image. Two distinct situations are found in the architectures of MV machines. In the first situation, machines are designed to perform in two distinct modes, the first mode involves learning the minimal dimensionality feature sequence common to a large number of exposures to the same object from different perspectives. The second mode involves regenerating the best estimate of the original object from the final feature sequence. The other situation involves the initial presentation of an object to the machine which includes a set of similar objects already in memory. The machine then compares the object to each of the stored object and makes a judgement as to the highest probability match. It then identifies the selected stored image and presents it to a human supervisor for conformation of the choice. While the choice is normally presented graphically, a lookup table could allow it to enunciate the selection.

At that point, some external intelligence must be used to judge whether the regenerated image is an adequate representation of the original. This judgement may not be sufficiently precise. In that case, the protocol may be repeated, thereby collecting an additional feature sequence and a second judgement as to the quality of the regenerated image. By repeating this protocol and combining the feature sequences obtained over the interval of the experiments, a better quality of feature sequence is sought.

An alternate MV method is described as unsupervised. The protocol described above is followed but the regenerated image is subtracted from the originally presented image on a pixel by pixel basis and the resulting residue is minimized by repeating the protocol many times. If the machines architecture and instruction set are adequate to the task, the system will achieve a local minimum that adequately describes the test image.

In biological vision, a distinctly different protocol appears to be used as discussed in Section 12.5.1. A conceptual expansion of the relevant portions of the figure from that section is shown in Figure 19.10.6-1.
There is currently virtually no data regarding the steps in the biological object recognition task. All that is available are observations and interpretations of the assertions made by other laboratory investigators.

Similarly, there is very little in the academic literature on the role of the cerebellum, except for its role in smoothing the actions of the motor modality by expanding the command set issued by the premotor cortex before passing the expanded command set to the primary motor cortex. Words become important when discussing the cerebellum. Noback, and others, have indicated “the cerebellum plays no part in the appreciation of conscious sensations or in intelligence.” These are stage 5 cognition activities. Noback and others recognize the cerebellum has significant areas dedicated to the visual modality. In this scenario, these areas are dedicated to stage 4 information extraction activities, not cognitive activities. Chapter 15 of this work discusses the role of the cerebellum extensively.

As a starting point, it will be assumed the occipital lobe is only concerned with the “where” of an image and its features during a specific gaze. The object recognition task (the “what”) is assumed to be focused on the operations of the pulvinar. Not shown is the output of a saccades sequence by V1 of the occipital lobe that is passed to the precision optical system (POS) for presenting individual glimpses to the pulvinar.

---

Sheremata, Malcolm & Shomstein have recently made an interesting argument\textsuperscript{148,149}. They assert that...
“Visual short-term memory represents information in retinotopic, not spatiotopic, coordinates.” If supported, it would be significant, both in terms of both the physiology of the neural system and probably also with regard to the attention of the batter in baseball. It would suggest the reference frame of the information presented to the cerebellum, and possibly the association areas (BA 37/39) might be in retinotopic coordinates while the framework for the saliency map is almost certainly in spatiotopic coordinates. This situation would also suggest that the phenomenon of attention might also employ a retinotopic framework as far as vision is concerned.

This initial quiescent gaze is processed initially by the LGN and then V1 of the occipital lobe to identify features of interest in the visuotopic field. V1 delivers a coarse resolution interpretation of the visuotopic field to the visual portion of the saliency map as well as a saccades list to the POS to support operation of the foveola-PGN-pulvinar pathway of information extraction. Each saccade on the list is implemented sequentially in order to provide a glimpse to the high resolution foveola. The foveola of each eye transmits a high resolution representation of the glimpse to the PGN for merging into a single representation. It then passes the merged image to the pulvinar for preprocessing and feature extraction. The pulvinar performs an analysis resulting in a minimal-dimensionality representation of each glimpse as a whole. This representation consists of a feature sequence plus potentially other metadata. It passes this information to the cerebellum for comparison to its vast previously stored visual database. The cerebellum makes a preliminary review to locate similar minimal-dimensionality representations that might exhibit different angles of presentation and/or scale. If it finds a suitable match, it asserts that fact, thereby terminating the glimpse interval. It passes its “what” label for this object along as metadata to the association area for placement at the appropriate “where” location of the saliency map.

In this scenario, any de-rotation and scaling operations are accomplished by the cerebellum rather than the pulvinar. However, there is no data upon which to demonstrate this assumption.

If the cerebellum cannot determine a match between the presented representation and its memory files, it can either ask the prior system to reexamine the glimpse to acquire an alternate set of minimal-dimensionality feature sequence that it can re-attempt to match, or pass the submitted feature sequence along to the stage 5 cognition engines. The cognition engine may attempt to interpret the minimal-dimensionality feature sequence in the context of the full gaze and other sensory modality data available in the saliency map. If successful in identifying the object, it can assign a meaningful label to the feature sequence, and instruct the stage 6 efferent neural circuits to return the feature sequence and its newly associated label back to the cerebellum for long term storage. It may also instruct the stage 6 engines to take some action to indicate recognition of the object, to institute a new examination of the gaze or glimpse, or treat the whole glimpse or gaze as inconsequential.

The action of the stage 6 premotor cortex engines is dependent on the metadata (labeling, etc.) provided to it by the stage 5 cognition engines.

Major questions remain concerning the minimal-dimensionality feature sequence and metadata passed to the cerebellum;

• what is the character of the feature sequence passed to the cerebellum?
• does the feature sequence have other metadata attached to it?
• does the feature sequence present the most significant byte (MSB) first?

Given even preliminary answers to the above questions, how could or does the cerebellum search its visual memory adequately in less than one second?

To the extent the above scenario is correct, it suggests several specific situations.

• There is no reiterative procedure, involving a large number of cycles, employed in learning to recognize
an object. It recognizes the ability of the stage 5 cognitive engines to ignore one or more glimpses or to increase attention to the task.

- The pulvinar most likely performs a primary task seeking to create a minimal-dimensionality representation of each glimpse presented to it without attempting to pursue any de-rotation or re-scaling in the absence of any knowledge about the need to perform these tasks.
- The cerebellum is in a better position to perform any de-rotation and re-scaling needed to compare the presented minimal-dimensionality representation of the current glimpse and its stored representations.
- With failure of the cerebellum to recognize an object represented in the glimpse, the stage 5 cognition engines are in a better position to assign a label to the available minimal-dimensionality feature sequence based on the context provided by the saliency map or increase attention on the task by instructing the stage 4 engines to provide additional stage 4 analyses.

In short, it appears highly unlikely that the visual sensory modality spends more than a few seconds in the worst case on object recognition per se. This does not preclude the formation of a more precise record in declarative memory over a period of time involving multiple gazes containing the target object.

These proposed findings suggest the techniques of MV involving feed forward signaling are most likely relevant to the biological object recognition task except in the global relationships defined by the inner loop (and in some aspects the outer loop) defined above. The de-rotation and re-scaling techniques explored by Muller et al. in MV appear applicable to the similar biological task. These tasks can be considered post processing from the pulvinar perspective. They would appear to be an integral part of the processing within the cerebellum.

---

In the BV case, the foveola of the subject is presented an image that is stationary during the analysis of the individual glimpse except for the intrinsic tremor of the ocular system. For normal humans, this tremor is present from a very short interval after birth until death (absent any medical intervention). For an unrecognized object, the pulvinar establishes a minimal-dimensionality feature sequence that is stored in area 4b of the cerebellum. At the same time this feature sequence is passed to the stage 5 cognition engines where it is combined with other sensory information to establish a more complete description of the image (“Baby darling, this is your mommy”). As a result, the cognitive engine(s) forms a more complicated feature set, consisting of multiple feature sequences, that it instructs the stage 6 efferent neural pathways to return the complete feature set to area 6b of the cerebellum for declarative memory storage.

When again presented a similar glimpse, the pulvinar attempts to match the image with feature sequences stored as a part of a larger feature set within the cerebellum. If successful, it combines the new and old visual feature sets, along with a confidence level associated with the match, and forwards the complete feature set to the stage 5 cognitive engines. The stage 5 engines now have a description of the glimpse (“Baby darling, this is your mommy”) along with a confidence measure. The stage 5 engine can accept the identification and leave the prior memory in the cerebellum intact or instruct the stage 6 neurons to cause an update of the stored feature set within the cerebellum.

As a result of these procedures, the function of object recognition within BV appears to be unsupervised to an external observer. However, it is supervised from the perspective of the subject since the stage 5 cognitive engines constitute an intelligence nominally equal to that of the external observer.

### 19.10.6.3 Adaptation of the MV framework to biological vision

While the application of advanced probability principles from theoretical mathematics to MV can be very helpful, the understanding of the neural system adopted in MV research complicates any transition. As noted earlier, the assignment of bipolar weights to the synapses can be accounted for by reinterpreting all weights as positive with the potential for inversion being assigned to the neurites of the associated neuron. Reinterpreting the MV definition of the neuron is more difficult. The MV approach assumes all neurons generate action potentials, apparently on a monopulse basis with respect to each processing cycle. This interpretation introduces two problems:

- of the billions of neurons within the CNS, less than 5% generate action potentials,
- of the millions to tens of millions associated with a stage 4 engine dedicated to object recognition, only a few hundred at most generate action potentials, most of the outputs of one engine are passed to an adjacent engine in analog form,
- of the hundreds of stage 3 neurons projecting stage 4 signals to more distant stage 4 or stage 5 engines,
most create analog modulated multiple pulse groups with each processing cycle, and most importantly,
• the neural system is not a serial synchronous binary computer employing transcendental computations.

Based on the above facts, applying advanced probability principles to the CNS requires an entirely different philosophical framework than generally applied to image analysis within the MV context. It means reverting back to the basic interactions between the pixels of a Markov Random Field (or a Bayesian, sometimes Belief Field) and reinterpreting them in the context of a highly parallel asynchronous analog associative computer.

There is considerable histological evidence that the interdependencies between neurons extends beyond the nearest neighbor in both the neurons of the stage 1 retina and those of the stage 4 information extraction engines. Whether these patterns can be identified sufficiently to allow simplifying the necessary model of the functions involved by using mean field theory remains to be determined. Whether the other simplifications associated with the term “Restricted” as it is used in the Boltzmann and Helmholtz machines must also be determined. The patterns defined above appear to be roughly congruent with the logic units of MV machines and histologically with the neurons within a column of Livingstone & Hubel. The likely presence of “hidden” nodes and/or layers within these architectures makes the task of defining them functionally more difficult.

It appears most likely that a different “machine” architecture and label will be needed when modeling stage 4 information extraction in the CNS. It may also be necessary to establish machines of different architecture for use in the occipital regions and the thalamic regions of vision. While the pulvinar is likely to require significant support from the cerebellum in memory retrieval and deposition, it appears the engines of the occipital lobe are less dependent on memory in general and the cerebellum in particular (based on histological traffic analysis). Any machine representing the pulvinar will probably need to provide streaming (ala Muller et al.) to accommodate the color and brightness information. This may be less important in the occipital lobe where it appears multiple machines working in series may be found.

The idea of learning through generative action in MV may need to be considered differently in BV. While the system must learn without external human intervention, the stage 5 cognitive engines have a capability with human qualities. These engines also have access to other information available via the saliency map. [xxx keep going here with generative versus recognition ]

Bring in aspects of the corelet of Esser xxx et al. ]

19.10.6.4 An initial baseline using a mean field version of a BM or HM

The above environment leads to several plausible implementations of a Markov class of computational machine but several requirements that indicate the simple Markov machine would be inadequate. The requirement for unsupervised learning, requires some form of reiterative (recurrent) machine.

Either the Boltzmann machine or the Helmholtz machine appear to satisfy this requirement, but by different means. They both place a restriction on employing correlation among the individual nodes within a given hidden layer and the type of feedback employed. With these restrictions an acceptable machine should be re-labeled with a capitol letter R prefix (an RBM or an RHM). The restrictions reduce the ability of the machines to achieve ultimate probability of recognition. However, the biological system does not require ultimate precision. The pulvinar, in conjunction with the POS is capable of reprocessing an additional glimpse to confirm the accuracy of the first glimpse. The overall visual modality is also capable of repeating the complete gaze operation to confirm the reported recognition result.

The Helmholtz machine limits the recognition path to a feedforward path only. It also limits the generative path as a feedforward path only. Feedback is only allowed in the sense that the output layer is used as the input layer when operating in the generative mode, with the output of the generative mode provided to the input layer. The recognition and generative mode do not operate simultaneously. The use of the Helmholtz machine employs a topology placing a minimal requirement on the topology of the pulvinar, no intrinsic feedback within either the recognition or generative modes of operation. The analog character of the signals being processed require a machine employing the “mean field” concept. Based on these considerations, the initial baseline will consist of one or more mean field RHM’s within the pulvinar.

The title “mean field” was introduced to provide a simplified architecture machine. Rather than
accumulate a large number of individual binary signals from adjacent pixels from the same as well as adjacent layers, it was proposed that a single continuous variable with a value between 0.0 and 1.0 and representing the mean of the field could be introduced instead. One implicit result is that the integrating mechanism must be able to accept continuous analog information.

The work reported by Muller, Eickeler & Rigoll in 2001 appears to offer a “software” framework for the selected baseline machine. They employ a preprocessing step that provides resizing of the presented scene to match previously stored images as well as a de-rotation capability to provide a higher likelihood of achieving a satisfactory match. Their description of their software is more illustrative of a potential capability than prescriptive of an adequate capability. Their use of a single non-occluded object as their scene does not indicate a capability to select major features of the glimpse to the exclusion of less significant features, nor does it demonstrate an ability to handle occluded features. Their use of separate “streams” to handle brightness information and chrominance information offers excellent flexibility in optimizing the RHMs’s but may constitute technical overkill.

It appears that most animals only require a de-rotation process encompassing less than ±15 degrees. Beyond that value, the subject typically changes the position of its head or body to reduce the apparent rotation of the glimpse relative to its stored memories. Similarly, any tilt removal associated with reading need only remove individual character tilt of about ±15 degrees. Otherwise, the reading subroutines revert to a learning mode from the object recognition mode. A scale normalization is probably an important feature compensating for target images varying in size over a range of three up to at most five to one.

The physical motion of the eyes, when discussing the analytical mode are rectilinear and orthogonal due to the small angle approximation of trigonometry. However, the sensory array of the foveola is not rectilinear (Section 3.2 & Figure 3.2.3-3). In the language of Gonzalez & Woods150, the array must be treated as 7-connectivity when ignoring their individual spectral sensitivities. The neurons exhibit a lower level of connectivity and the retina exhibits a lower acuity, when only neurons of the same spectral sensitivity are of interest. The M–channel (green) maintains the highest acuity of the spectrally sensitive sensory neural arrays. Gonzalez & Woods provide a background for the transform tasks of interest in biological vision. However, as noted elsewhere, the biological system employs anatomical computations without employing transcendental mathematical computations. Anatomical computation involves the spatial rearrangement of the neural axons prior to their arrival at a given information extraction engine. The system does not employ any spatial to frequency domain transform, such as the Fourier transform or any correlation function.

As discussed in Section 12.5.1, the pulvinar is nearly adjacent to the cerebellum and it is very likely that the memory capacity of the cerebellum supports any memory requirements associated with the pulvinar in its role of object recognition. This memory is generally described as non-declaratory. For readily recognized scenes, the pulvinar can compare the scene to stored images from area 4b of the cerebellum in the figure of that section. If the pulvinar fails to recognize the object, it must rely on the signal path up through the stage 5 cognitive circuits and back down via the stage 6 neurons to the cerebellum to create a new properly labeled and permanent memory.

The individual chrominance channels are normalized by the sensory neurons themselves during signal capture. If needed, a brightness signal can be created by the pulvinar (or the PGN) and binarized if necessary. The Muller et al. report demonstrated skew removal related to the individual glimpse. It also illustrated the ability to emphasize the color or shape information by changing the weights used in the hidden layers.

19.10.6.5 Modifications to the baseline using a mean field version of a BM or HM

[xxx empty ]
130 Processes in Biological Vision

19.10.7 The minimum dimensionality required to index recognized objects

At the current time, nothing is known as to how the neural system encodes memories, and specifically visual memories. Many of the ANN machines do develop a “state diagram” related to their hidden layers after they complete a learning activity. This state diagram is then used as a template with which operational scenes are compared to accomplish recognition. Does such a state diagram adequately simulate, or even emulate, the state diagram stored within the biological neural system?

19.10.7.1 Rhetorical questions concerning object recognition

[xxx introduce “declaratory memory” as the type of interest in this section. ]

Both pedagogy and machine learning have established the critical importance of learning, as a critically important repetitive activity prior to implementing the object recognition task. This learning involves forming a memory related to the object of interest. While the operation of stage 4 and 5 engines of the biological system suggest their output employs word serial/bit parallel messaging, nothing is known about the code used to project or store these messages. Even the location of storage is unknown, along with whether the memories are stored in a recognizable form or holographically.

Several observations lead to a variety of open questions.

1. The minimal dimensionality associated with a recognized object is clearly more than just a noun describing the object. From personal experience, the minimal dimensionality appears similar to the result of the recognition of the thought involved in a complete paragraph in reading. The indexable record appears to identify the object as the result of a specific glimpse analysis within a traceable sequence of glimpses related to a specific gaze (including a minimal record of the epoch associated with the gaze).

2. The above considerations suggest the minimal dimensionality associated with a current glimpse and/or gaze involves a complex “word” using an unknown code. It is interesting to consider whether this word is an index to a general database of unknown character.

3. It appears any minimal dimensionality index includes the context of the recognition and some means of relating that context to the more general gaze, possibly what can be considered metadata.

Another question is whether the index recorded for future retrieval purposes is that of the most recent glimpse and/or gaze or whether it is the index of the ensemble of glimpses and gazes from both the current scene and the memory records used to achieve the recognition.

4. Reviewing the basics of database design could provide additional possibilities for identifying or at least hypothesizing the mechanism of memory creation and recall within the neural system.

The United States Patent and Trademark Office (USPTO) has been developing an index structure to satisfy its needs to retrieve graphic trademarks from its files. It has adopted a string of two digit numerical codes describing the major features of a trademark in order of precedence. As a simple example, the USPTO Design Code Search Manual begins with;

“The USPTO assigns all marks containing design figurative elements a 6-digit numerical code(s) for searching purposes. This manual indexes the categories, divisions, and sections that make up these codes. For example, a single five-pointed star would be coded in category 01 (celestial bodies, natural phenomena and geographical maps), division 01 (stars, comets) and section 03 (single star with five points), resulting in a complete design code of 01.01.03. The design search code manual also contains explanatory notes and specific guidelines that provide instructions for specific code sections, cross-reference notes that direct users to other code categories, sections and divisions, and notes describing elements that are included or excluded from a code section.” For complex trademarks, several six digit numerical codes may be combined.

The numerical groups are only partially defined at this time, 29 items in the first category. Category 02 includes the human body and its parts. A human nerve is coded as 24.17.02. Category 03 includes animals and animal body parts.

Can the biological code be thought of as a sequence of “syllables” in an unknown abstract code?
The visual modality appears to store gazes with glimpse tags attached.

Since reverse contrast features do not normally occur in the environment, it is unclear whether the normal visual system incorporates a contrast feature in the minimal dimensionality record committed to declaratory memory.

19.10.7.2 Current estimate of the minimal-dimensionality record of a scene

Based on the above review, analysis and discussion, a preliminary estimate of the signature prepared for storage in long term declarative memory can be assembled (ca. October, 2014) This signature appears to be based on;

1. a particular instance of a scene stored in low resolution form from Area V1 (the where element),
2. accompanied by indices related to each of the major elements within the scene (the what elements).
3. The where element appears to contain the location of the identified where elements in three coordinates.

   These coordinates may be in either the spherical system derived from the visuotopic scene by the ocular nucleus and the precision optical system (POS) servomechanism, or in inertial coordinates as calculated within the neural system.

   When stored in inertial coordinates, the indices of the low resolution record may be placed into the larger database recording the wider scope scene in which the visuotopic scene is only a minor part.

4. These indices are based on the previously stored records of these elements and not necessarily the latest example of the object found within a glimpse.

   The indices may relate to the updated stored record, as modified on the most recent element in the overall scene.

5. The low resolution where element is likely to be stored without color information in the predominant case. It is also likely that the what element indices are largely devoid of color information.

6. An epoch indicator (optional) .
7. Other as yet undefined metadata.

In cases where a new “what” element has been encountered (that cannot be found within the declarative memory database) and determined to be of significant value, the what element may be stored at foveola/pulvinar resolution along with its indices.

It is possible the resulting minimally-dimensional signature is similar to the code groups being defined by the US Patent and Trademark Office but not in decimal format. The biological code remains unidentified. It is interesting to consider whether the signature for a graphic is similar to the signature associated with a paragraph in reading.

19.10.8 Protocol for evaluating gaze content within PEEP procedure

PRELIMINARY

The prior section and earlier parts of this work have defined for the first time, the difference in the roles of the portions of the neural system allocated to the foveola (the PGN/pulvinar couple) and the larger fovea (the LGN/occipital loop). The PGN/pulvinar channel plays a heretofore undefined role in object (face) recognition that is separate from the role of the LGN and the occipital lobe (area 17 or the historically named “visual cortex”). It is appropriate to develop a laboratory protocol for evaluating the role of the foveola/PGN/pulvinar pathway in object recognition independent of the rest of the visual modality.

In any protocol related to PEEP, it is important to recognize two critical factors. Both the performance of the sensory receptors and any required initial major saccade are functions of the intensity of the stimulus. Any initial major saccade introduced into the experiment before reaching the point of initial gaze introduces a delay (see Figure 7.3.5-1). This delay is intrinsically a function of the delay introduced by the sensory neurons (see the delay term in the P/D equation, Section 7.2).

An experimental protocol is needed that would allow the psychophysical investigator to evaluate the
human subject’s ability to differentiate between similar scenes when the subject is restricted to only examining images provided within a 1.2 degree diameter centered on his point of fixation (the subject's nominal gaze).

Conceptually, the ideal would be to have the subject look down a tube with only a 1.2 degree image appearing at the end of the tube. An alternate arrangement would be to use a conventional video display with all but the center 1.2 degree field blacked out (or held at a fixed neutral gray value). The tube approach is preferred because it will not allow extraneous distractions that might generate one or more saccades. An instrument employing a “tube” approach is found in most optometry offices and is used to evaluate the stereoscopic ability of the subject related to the foveola. The image is not as dynamic as envisioned here. When the viewable field is restricted to 1.2 degrees diameter, the generation of mini-saccades is greatly restricted if not eliminated. Only micro-saccades and tremor will be present at significant levels of activity.

However instrumented for the subject, it is important that the display exhibit significantly more that 175 pixels per 1.2 degree diameter in order to not limit the acquired data due to the limited resolution of the test equipment.

A series of flicker comparison tests might be used where the subject is asked when he can notice the difference between two similar images. The duration of the exposure for each image needs to be considered further. It may also be important to vary the number of flicker cycles and/or the length of the overall flicker duration for a given pair of images. Huiskes, Thomee & Lew have published a review of flicker techniques including citations to a large number of flicker image pairs.

When preparing a new psychophysical protocol, the risk of false positives, such as reviewed by Marek should not be overlooked.

19.11 Utilization of extracted information by the Intellectual & Limbic Systems

Palermo & Rhodes have written a significant paper on the usage of information by the limbic system extracted by the stage 4 engines of the CNS. The subject is developed in Section 15.8.

19.12 A compendium of abnormal conditions associated with high level vision

Farah & Ratcliff have provided an excellent set of tutorial essays, ca. 1994, describing a wide variety of medical conditions lacking an organic cause. On page 17, Coway describes the result of removing the sensory portion of the superior colliculus (the perigeniculate nucleus) on the vision of monkey. In the same paragraph, he notes the significant volume of neurons connecting the PGN to the pulvinar and subsequently the pulvinar to the extrastriate visual areas (accompanied by a series of references). Parrett et al. discuss failures related to object recognition in Chapter 2 and also discuss the question of whether specific spatial orientations of an object are more readily recognized. Newcombe & Mehta address category specificity within the object recognition framework (including the possibility/probability of a specific category of faces in at least human cognition. Arguin & Bub develop material relating to problems in reading in chapter 6. Several chapters relate to “covert recognition” and “neglect” from a clinical perspective.

19.12.1 Behavioral data on reading problems–summarizing Dehaene

Dehaene has provided a vast amount of material related to reading from the perspective of a psychologist. He employs few cartoons, even fewer MRI images and no top level block or schematic diagrams related to the function of reading. He does provide some excellent figures summarizing the performance of the visual modality in reading.

His discussion of mirror image reading and writing in chapter 7 is quite illuminating. His discussion of patient A. H. provides unique information concerning the operation of the PGN/pulvinar couple, potentially the LGN/occipital couple, and potentially orthodromic engines along the signal path.

Dehaene goes into the development of alphabets in considerable detail, showing their most common structures and features they avoid. Whether the alphabets are developed to minimize reading difficulty or writing difficulty may be difficult to resolve.

19.12.1.1 Dyslexia occurs in three forms during development

Dyslexia appears in four distinct forms during human development,
1. the confusion among individual letters associated with their (primarily) left-right symmetry,
2. the confusion among words due to an interchange of letter order during analysis,
3. the transposition of short words in a series of short words and
4. the tendency to read in the opposite to conventional word order in a given language.

Webster’s Medical Dictionary provides a largely behavioral definition of dyslexia; “A variable often familial learning disability involving difficulties in acquiring and processing language that is typically manifested by a lack of proficiency in reading, spelling, and writing.” This section will be more concerned with the functional character of the condition.

The four listed problems are normally overcome through the loading of long term memory during the learning process. Failure to overcome any of these problems generally leads to a diagnosis of dyslexia. As Dehaene notes, the condition is so common in developing children, it cannot be considered a disease at that stage.

Dehaene, chapter 7, has provided a long scenario of child development and the subject of dyslexia. He shows that children generally go through short periods where they exhibit the above difficulties on a transient basis. Unfortunately, his caricatures in figures 7.3 & 7.5 should not be relied upon. they are based on archaic understandings of how the visual system operates. They fail to recognize the importance of the foveola in the analytical mode of vision and the PGN/pulvinar couple as the actual seat of written character and word recognition. Figure 7.5 contains major inconsistencies with the actual projection of the visuotopic field onto the surface of the two occipital lobes of the brain (see Section 15.2.5.7, particularly Figures 15.2.7-12 and –13). These figures demonstrate the grossly different representation of the fovea on the cerebrum. The letters A, B & C do not appear at all like they are represented by Dehaene. Furthermore, the fields of view projected onto the cerebrum are not circular as represented by Dehaene. They are parabolic. This difference is significant in the methods of information extraction available to the stage 4 engines.

Dehaene also discusses the results of MRI studies of the association areas of dyslexics in chapter 7. It can be assumed the BOLD form of MRI was used. While he noted the level of neural activity in the area of the junction between the temporal and occipital lobes, the procedure does not have the temporal or spatial resolution required to aid interpretation of neural signaling with the dyslexic.

In brief, Dehaene notes that children frequently encounter difficulty with the letter pairs b & d and p & q during their early development before they begin to address words in a more holistic manner. After growing through the left-right symmetry problem associated with these letters, they frequently encounter a similar problem with simple (one syllable) words such as the pair bad & dab. As they overcome this problem, they frequently encounter the letter order problem, confusing pairs like snail & nails. As they gain experience, actually through loading their memory, they advance to higher level problems related to word order. They frequently read one or more words in reverse order. Examples are house versus esouh and other reversals that lead to intellectual confusion. Throughout this period, the child may frequently switch from reading and/or writing from left-to-right to right-to-left.
Michelangelo is frequently described as using mirror writing in his notebooks as a means of writing in a secret code. It clearly is not a very secret code. Instead, it is more likely an exhibition of his encountering a neurological abnormality. This abnormality would have had little effect on his daily life.

Most of these difficulties are surmounted as the child begins to process longer words and particularly multiple syllable words. The longer character sequences seldom make sense and automatically cause re-scans of the relevant text as noted by many neuroscientists (Section 19.12.2.3). A new sequence of value is incorporated into permanent memory in order to avoid the problem the next time it is encountered. This is a basic learning methodology.

In chapter 6, Dehaene develops the important process of transitioning from decoding acoustic speech to reading and understanding graphic characters. On page 295, he notes the potential problem in children who have difficulty distinguishing between the spoken sounds for b & d and for p & q. He asserts, they “tend to be confusing, because they all refer to similar phonemes (all 'plosive') consonants. It is not surprising that this conjunction of difficulties poses specific problems for children. . .”

Thus, all children appear to encounter a variety of mirror symmetry problems as they learn to read more complex material. However, as the long term memory becomes trained in the subtleties of human written language, these problems are overcome. The problem is so prevalent, it is called developmental dyslexia. It is only the people that are unable to learn the subtle differences and continually to revert to slower methods of reading letter-by-letter that can be described as true dyslexics.

Dehaene discussed what appears to be a more serious form of dyslexia combined with a frequently transient error in motor modality response to a visual modality signal beginning on page 294. Subject A. H. exhibits clear dyslexic problems associated with reading but also frequently begins to respond to a desire to grasp a visual object by moving her hands in the opposite direction to the expected motion. His figure 7.8, reproduced as Figure 19.12.1-1, is illustrative of the overall group of problems suffered by A. H. The upper two frames were presented with discussion by McCloskey & Rapp in 2000156. The lower frame was discussed but not presented. A second 2000 paper discussed the lower frame157. Both papers relied upon the 1995 paper in a variety of aspects158. That paper noted in the conclusion, “In ongoing studies with A.H., we are examining her visual deficit at cognitive, neurological, and genetic levels. We are also studying the phonological deficit suggested by A.H.’s impaired phoneme discrimination performance, asking how (if at all) this deficit is related to the visual impairment. They offered no figure describing their neurological model.”

<table>
<thead>
<tr>
<th>Stimulus word</th>
<th>Reading response</th>
</tr>
</thead>
<tbody>
<tr>
<td>dog</td>
<td>bog</td>
</tr>
<tr>
<td>bone</td>
<td>done</td>
</tr>
<tr>
<td>pig</td>
<td>dig</td>
</tr>
<tr>
<td>star</td>
<td>tars</td>
</tr>
<tr>
<td>rib</td>
<td>rip</td>
</tr>
<tr>
<td>sun</td>
<td>nuns</td>
</tr>
<tr>
<td>skirt</td>
<td>skit</td>
</tr>
<tr>
<td>dust</td>
<td>dusk</td>
</tr>
</tbody>
</table>

Stimulus image:

- [Image of stimulus image]

Copy:

- [Image of copy]

**Figure 19.12.1-1** Behavioral aspects of mirror symmetry problems in A. H. ADD. From Dehaene, citing McCloskey & Rapp, 2000.
When discussing her excellent scores on a variety of standardized tests she took, the 2000 paper asserts, “Interestingly, A. H. mentioned during the interview that she occasionally failed to recognize familiar words when she encountered them out of context. She also stated that even for material that was not especially challenging, she sometimes had to read sentences or paragraphs more than once before they made much sense. These points take on some significance in light of the results we report.” The remarks are also significant when discussing the functional aspects of her neural system. She may have performed better in the word order tests if the words were longer and/or had more relevance to each other when they appeared in common usage. The 2000 paper presents extensive discussion of the details related to the performance of A. H.

In the upper frame, the subject clearly suffers from letter confusion and confusion of letter order within words associated with left-right mirror symmetry. However, in conjunction with the middle frame, it is clear she also encounters up-down symmetry problems related to the letters p & d and b & p. She may also have a font sensitive problem in resolving the difference between t & k when they are presented with serifs.

The central frame clearly shows the problems with both up-down and left-right symmetry. Since we do not know the vergence pattern of the eyes during the PEEP activities related to these scenes (Section 19.10), it is difficult to say whether these problems were limited to the PGN/pulvinar couple or also included a potential problem with the LGN/occipital couple (Section xxx). It is most likely the problem was centered on the PGN/pulvinar couple. Each of the objects shown fit well within the 1.2 degree diameter visual field of the foveola and PGN/pulvinar signaling channel.

The lower frame is indicative of a different signaling problem, in fact two distinctly different signaling problems. Again, the figure does not show the PEEP procedure used by the visual system. The PEEP procedure involves the fixation of the eyes designed to bring individual objects in the scene to the center of the foveola for sufficient time to extract information as to what the object is. These intervals are typically on the order of 200 ms and are not normally recorded in the experiments such as those performed in the typical psychology laboratory.

The 1995 McCloskey & Rapp paper illustrated the extent of A. H.’s misfunctioning both with respect to angular location in both azimuth and elevation, but also with respect to orientation of individual objects within a complex scene. They noted, “Although A.H. made both localization and orientation errors in direct copying (of a line drawing presented to her), we focus in this article on her processing of object location.” Their key results for localizing an object showed that “A.H. moved a computer mouse to indicate the location of an X presented for 250 ms on a video monitor. They described these as ballistic trajectories since there was no feedback as to whether she was actually approaching the correct position of the object. For stimuli at locations 6.7° left, right, up, and down from the screen's center, her error rates were 57%, 47%, 35%, and 36%, respectively (72 trials per location). All errors were confusions between left and right or between up and down (e.g., moving the mouse downward for a stimulus at the top of the screen). The mean overall error rate for 3 control subjects was 0.6%, and all of the subjects had error rates below 3% for each location.” They illustrated her actions in the lower frame of the figure. They interpreted their results as “This phenomenon suggests that A.H.’s deficit is not properly conceived as one in which visual stimulus information cannot be used to compute object location, and a random choice -more precisely, a choice unrelated to the stimulus - is made at some level of processing. Rather, it appears that (at least some of) A.H.'s localization errors occur when a visual process systematically computes an incorrect location representation from the stimulus (e.g., systematically computing a left location for stimuli presented on the right).” The next observation is critical to understanding her overall performance. “In contrast to her severely impaired performance in ballistic reaching, A.H. reached successfully for objects when allowed to change direction during movements. Under these conditions, she often initiated a movement in the wrong direction, but then quickly and smoothly self-corrected, reporting that she realized her hand was getting farther away from the object. This ability to self-correct may help to explain why neither A.H. nor other people were aware of her localization deficit prior to this study.” Self-correcting in this manner is descriptive of a different problem. She was not being asked to move a mouse to the location shown to her for a short period. In these later experiments she was being asked to bring her hand to the location of the object as if to grasp it while she observed the entire operation. This activity involves visual feedback during the course of the task (as suggested by the curved arrow at the bottom of Figure 19.10.4-3) [xxx wrong drawing, doesn’t have arrow] repeated here as Figure 19.12.1-2
Figure 19.12.1-2 Top Level Modes overlaying the Top Level Stages of vision ADD. This diagram is particularly useful in analyzing the performance of A. H. in terms of her long term dyslexia. See text.

[xxx discuss difference between the ballistic and the closed loop efforts to collocate with the object.]
McCloskey & Rapp made additional tests to demonstrate that the misfunctioning of A. H. was located within the visual modality. It did not appear during similar tests with her eyes shut and employing only her somatosensory modality of touch. They also demonstrated it was centered on localization and not the identification of an object.

McCloskey & Rapp discussed the “Nature and Implications of the visual deficit.” They concluded that their experiments were sensitive to the stimulus duration, motion related to the stimulus and contrast related to the stimulus. However, they did not seriously quantify these and other variables in their experiments. Their results are also highly sensitive to the luminance level employed in the experiments. It is important that these experiments be carried out within the photopic illuminance range if repeatability is to be achieved. Based on the data provided in Chapter 2, their experiments were actually carried out under low room illumination, about 100:1 lower light level (below sunset illumination conditions) than required for good reproducibility. Figure 19.12.1-3 reproduces their figure 4 with additional annotation based on the dynamic performance of the eye described in this chapter and chapter 7.

The data acquired during the first 50 ms or less is the result of the awareness mode of the visual modality. It is providing a relatively low error rate in terms of the angle and sign of object in the visuotopic field. The data results in a coarse signal (including meta data) being passed to the cerebellum (the stage 4b path) and also to the saliency map shown in [Figure 19.10.4-3]. After a first saccade, an initial report is prepared by the PGN/pulvinar couple of the analytical mode. This initial precision report and the

![Figure 19.12.1-3 Localization error rate as a function of exposure duration ADD for A. H. and 3 normal subjects. The data was collected under scotopic illumination conditions. The awareness mode involves the LGN/occipital couple and gives coarse location information. The initial and final reports involve the output of the PGN/pulvinar couple and suggest the source of the mislocation by A. H. See text. Annotated from McCloskey & Rapp, 1995.](image-url)
saliency map. In the case of A. H., these later reports are probably reasonably accurate with respect to signal intensity (the signal-to-noise ratio is low at the specified illumination level) but the sign of the angle is reversed (the major error). After about 250 ms, the eyes of both A. H. have performed additional saccades and proceeded to other points of interest in the visuotopic space as determined by the TRN (the nonconscious executive) regardless of the duration of the test exposure.

As noted above, A. H. will continue to perform in the above manner for open loop experiments. When, the experiments are changed to longer duration test exposures and instructions to reach for the object, the same performance will be obtained until the visual system is able to see the hand and object simultaneously. The initial performance will follow the same steps as described above until both the hand and object are perceived by the stage 5 cognition engines. These engines will now attempt to perform the requested hand actions, i.e., bring the hand and object together using feedback via the external environment. They will give instructions to the stage 6 command generation engines to drive the hand to the same location as the object (zero out any distance parameter between these two objects) regardless of any previous information about the location of the test object. Stop motion video of the hand will show it initially moving in the direction away from the object but then reversing course as the stage 5 engines become involved.

These actions of the awareness mode and analytical modes are consistent with what McCloskey & Rapp describe as the transient and sustained subsystems. Their suggestion that the parvocellular and magnacelluluar signaling paths of the optic nerve can be discounted as it is now known that the parvocellular neurons carry color information while the magnacellular neurons carry brightness information about the same objects in the visuotopic field. Their hypothesis first presented in their concluding remarks, “that A. H. has a deficit affecting the sustained visual subsystem. . .” is supported if the word sustained is changed to analytical. As noted, after about 250 ms, attention and the operation of the analytical mode have been directed elsewhere. The operations of the analytical mode are sustained, but they no longer relate to the task of the experiment.

Bedford presented and discussed a response to the 1995 paper159. She described two alternate explanations to the situation of A. H. regarding grasping objects without benefit of even a sketch. The paper describes two interpretations of how A. H. could misinterpret the angular direction to an object introduced suddenly into the visual field, the so-called Helmholtz model of the middle 1800's and the Gibson model of the 1950's. The Gibson model relies on the angular position of an object within a larger object space (described as a field). The overall field is taken as a frame of reference. Bedford introduces an additional irrelevant question concerning the orientation of the frame of reference relative to a second individual. She also introduces comments about the relative illumination of the object and its surrounding field. The Helmholtz model relies upon the angular position of an object relative to the sagittal plane of the subjects body. If the head is positions so its sagittal plane is the same as that of the body (the azimuth angle between the head and body is zero), the overall task is taken as simpler. Neither model treats the subject of interpretation and information extraction concerning the character of the object, only the position of the object in the perceived field. Bedford does note the character of A. H.’s problem in the grasping problem, A. H. estimates the visual angles relative to the sagittal plane but frequently gets the sign of the angle wrong, at least initially. After two pages of discussion, Bedford offers no conclusion concerning the character of A. H.’s malfunction. The following discussion will concur that the approach of Helmholtz (who was a trained optometrist of his day) is the correct one.

Based on this work, the position of an object in visuotopic space is totally independent of any other objects, including the edges of a scene within the visual field. The surrounding scene can be dark or in fact empty as in the typical white screen of a perimetry measuring machine. The animal first determines the position of an object within its visuotopic field of view through operation of the awareness mode on a routine basis, and through operation of the awareness (or potentially the alarm) mode upon the sudden appearance of the object. The angular position is always made with reference to the body under the assumption of the single Cyclopedian eye model. The situation discussed in these papers only involves the version, or angle of the eyes, from the straight ahead position (Section 7.4.2). No discussion of vergence or the convergence of the eyes is included (Section 7.7.4.3).
The alarm and awareness modes are both based on the relative horizontal angular position of the object relative to the sagittal plane of the animal. This angle is the sum of the angle between the sagittal plane of the head and the sagittal plane of the body, and the angle between the horizontal angle between the line of fixation of the eyes and the sagittal plane of the head. The experiments of McCloskey & Rapp primarily involved objects displaced in the horizontal plane.

The following material will specifically recognize the difference between the alarm, awareness and analytical modes developed in this work.

19.12.1.2 Signaling errors in dyslexia of A. H.

The documentation accumulated relative to A. H. provides excellent material for further defining the signaling between the PGN/pulvinar couple and the orthodromic engines of the visual modality. The Top Level Block Diagram of interest was presented in Section 1.5.1. The Top level schematic was presented in Section 11.6.1. The functional diagram of major interest was presented in Section 15.2.5. The more detailed top level flow diagram, including feedback via the environment of Section 19.11.6 is particularly relevant. Figure 19.12.1-4 presents an expansion of the top level flow diagram focused on the signals emanating from both the LGN/occipital pathway and the PGN/pulvinar pathway. It was originally prepared to discuss the derivation of gross motion signals, convergence of the two eyes, depth perception and the focusing of the eyes (accommodation. See Sections 7.4.2 & 7.4.3. By replacing the words vergence (located at the extreme right and left of the figure) with version, it can be used as a starting point for discussing the long term dyslexia of A. H.

Figure 19.12.1-4 Expanded signal flow diagram between the diencephalon and orthodromic engines ADD. The words vergence should be replaced with version when discussing the dyslexia of A. H.

The key points developed in this figure are the following:

1. The peripheral fields of view are processed separately from the central 1.2 degree diameter fields of view.
2. The peripheral fields of view are processed by the LGN/occipital pathway, specifically the striated portion of the occipital lobe described as V1 (or area 17).

3. The central fields of view surrounding the line of fixation are processed by the PGN/pulvinar pathway.

4. The delay associated with the LGN/occipital pathway is significantly greater than that associated with the PGN/pulvinar pathway.

5. The two LGN’s and two PGN’s can be considered portions of the TRN and the pulvinar is largely surrounded by the TRN.

6. The TRN acts as a switching center controlling the flow of signals to the different engines of the above pathways and integrating the results of their information extraction activities.

7. The initial, and coarse, version signals are developed by the LGN as part of its responsibility to merge the two hemifields received from the two retinas before passing the merged image to one of the two occipital lobes.

8. It is not yet clear how the right and left images within the occipital lobe are assembled into one overall perception of the visuotopic field.

9. The precision version signals are developed by the PGN/pulvinar couple as part of the process of combining the left and right halves of the imagery from the foveola following initial (coarse) version of the two eyes. Precision version is performed in conjunction with information extraction from the signals obtained via the foveola. The precision version signals lag the coarse version signals in time.

10. Both the coarse and precision version signals are delivered to the Precision Optical Servo System for establishing the desired pointing of the eyes.

11. There appear to be several locations along the signaling paths where erroneous computations might be performed.

Additional important points include:

12. Information related to the visuotopic scene are also extracted from the LGN/occipital and PGN/pulvinar and forwarded to the association areas of the cortex after processing.

13. Significant error of position are forwarded to the association areas by long term dyslexics for individual elements of the visuotopic scene (Figure 1 of McClosky et al., 1995)

Current open questions concerning this figure are:
A. How the version signals are combined to represent a complete visuotopic field?
B. What signaling format is employed before the above merging and after the merging?
C. Is the failure causing dyslexia in A. H. due to signaling errors or computational errors?
D. What type and where does the failure occur in A. H. that leads to her long term dyslexia?

19.12.1.2.1 Static perceptions of A. H. vs the visuotopic scene

Figure 1 of McCloskey et al. is reproduced here as Figure 19.12.1-5. The direct copy of the stimulus pattern can be interpreted as what A. H. actually perceived. Although data was not provided on the geometry involved when A. H. viewed the stimulus, the discussion provided by McCloskey et al. indicate the copying involved several steps in the PEEP procedure (Section 19.10). McCloskey et al. did note, “A.H. made both localization and orientation errors in direct copying, . . .” However, they did not comment on whether the errors were global or local. The errors are clearly not global. As a general rule, A. H. correctly reproduced the central portion of the stimulus (within the 1.2 degree diameter area about the point of fixation associated with her foveola-PGN-pulvinar). However, when she then performed the PEEP procedure to bring other areas of the stimulus to the center of her foveola, she routinely represented them on the opposite side of the figure (left-to-right) from where they actually were. Both the diamond on the right and the cross on the left were perceived as both displaced and inverted. However, objects near the centerline of the stimulus were generally not perceived erroneously (such as the box with diagonal and cross in the lower field of view as well as the triangle above the central field of the stimulus). These errors in perception would suggest the version signals associated with the glimpses to the left and right of center are forwarded to the association areas of the cortex with both the azimuth and elevation showing errors similar to those encountered in the open loop experiments following information extraction by the PGN/pulvinar couple. However, when the PEEP procedure causes a glimpse to be acquired from above or below the central image, minimal errors are passed to the association areas. No rotation of these glimpses is apparent in the direct copy. It appears likely that the PEEP procedure was extended to analyze the horizontal bars at upper left and the hash marks at lower right. As a result, these features were perceived as inverted with respect to the bars, and both inverted and displaced left-to-right with respect to the hash marks relative to the original center of the image.
142 Processes in Biological Vision

The interpretation suggested above could be confirmed or falsified if eye-tracking instrumentation were employed to record the steps in the PEEP procedure used by A. H.

19.12.1.2.2 Open Loop Experiments

The low error rate for the initial pointing situation of figure 4 as modified (first 100 ms or less) suggests the dyslexia of A. H. does not arise in the initial (coarse) pointing obtained from the LGN/occipital pathway. The error rate for A. H. is not accompanied by statistical range bars so her error rate maybe similar to that of the “controls” for an equal number of trials. What is clear from the text is that when exposed to a short term stimulus (250 ms), She proceeds to accomplish a ballistic (open loop) approach to the location of the stimulus that is in error 60% of the time regardless of whether the stimulus is on the left or right. The error is one of direction relative to the sagittal plane. This error rate occurs after information extraction by the PGN/pulvinar has begun and a first precision signal has been provided to the POS.

McCloskey et al. note, “These results indicate that A.H.’s errors in localizing visual stimuli reflect a disorder of visual perception, and not a general spatial deficit affecting localization from all sensory modalities, or an apraxia or ataxia affecting the ability to carry out intended movements.” Based on the observations and the block diagrams, it is not clear that the error is one of perception at the stage 5 cognition level. It may be an error at the nonconscious, stage 4, level that is initially passed to the stage 6 command implementation level and only passed to the stage 5 engines subsequently. The fact that the error occurs at about the 50% level whether to the left or right suggests the problem is not in the signaling circuits of stage 3 but in the actual information extraction and/or information assembly engines of stage 4. The approximate error rate of 50% occurring after 250 ms stimulation in the ballistic experiments (regardless of direction of stimulus) suggests the error occurs in the computations performed by the PGN/pulvinar engines and not in the signaling pathways associated with those engines.

The first (coarse) saccade that attempts to bring the target to the point of fixation in visuotopic space is typically accurate to about ±2 minutes of arc in both the horizontal and vertical directions (Section 7.4.2 & Appendix L). The overall coarse saccade signal path appears to involve a type 0 servomechanism (position). The analytical mode of vision employs the much higher spatial resolution of the combined foveola and associative correlator of the PGN. The accuracy of the second (precision) saccade that is provided by the PGN/pulvinar is on the order of ±2–6 arc seconds (about 1/4 the diameter of a single photoreceptor). The PGN/pulvinar attempts to provide both position and velocity of the stimulus object using multiple glimpses in the PEEP procedure. The specific accuracy of these signals is subject to stimulus intensity, contrast and uniformity of reflection coefficient of the object. The PGN/pulvinar appears to use a type 1 servomechanism (velocity).

In the closed loop experiments testing the ability of A. H. to grasp the stimulus while the stimulus is continually applied presents as different operational scenario. xxx add here

19.12.1.2.3 Closed Loop Experiments
When A. H. was shown an object at one of the off axis positions and instructed to reach for the object, she initially began to reach toward the opposite position about 50% of the time until both the object and her hand were within her visuotopic field of view. At that point, she immediately transitioned into a type 1 servomechanism operating mode with the goal of reducing the distance between the object and her hand to zero. This operating mode did not rely upon the location signals (or the analytical information extraction process) from the PGN/pulvinar couple and the task proceeded to a normal successful result. Typically, no new information extraction process using the analytical mode would be instituted until the hand closed on the object (and possibly reoriented the object relative to the point of fixation).

19.12.1.2.4 Discussion

The major finding based on the above material is that the errors in position and orientation observed in long term dyslexics occur within the PGN/pulvinar signaling path.

1. To the extent the PGN operates similarly to the LGN, it is likely that the errors arise in the PGN.

2. The first conclusion is supported by at least two facts.
   A. The errors are passed to both the precision optical servomechanism controlling version and to the association areas processing interpretation of the complete visuotopic field of vision.
   B. The information extracted by the pulvinar relating to the character of the scene elements are not disturbed in the dyslexic.

3. The errors associated with the long term dyslexic appear to be nominally random since the reported angular errors between two possible positions occur roughly 50% of the time.

4. The simplest assumption that the errors are the result of computational failures rather than failures in the stage 3 signal propagation errors.
   A. The computational errors may involve errors of the same type as observed in taking the square root of a number where the answer where both a plus and minus value are obtained and the selection of the correct answer requires other information.

5. The information provided to both the precision optical servomechanism and to the association engines of stage 4 by the LGN/occipital pathway appears to be correct and dependable in all of the cases investigated by McCloskey et al.

6. No reason has appeared to suggest the errors occur in the association engines located at the junction of the temporal and occipital lobes of the cortex.
   A. The association engines appear to be accepting the erroneous location and orientation information provided to them via the PGN/pulvinar pathway and provide a mis-mapping to the stage 5 engines for further cognition arising from the PEEP procedure and

Discovering the character of the error in information extraction within the PGN and/or pulvinar is difficult to pursue at this time due to a lack of information concerning the processing architecture within these engines.

The tremor of the eye plays an important role in converting the photoreceptors of the fovea and foveola from change detectors into imaging detectors. The tremor involves a programmed motion of the eyes with an amplitude of about 36 arc seconds amplitude at a mean frequency on the order of 30 Hertz. It is due to the twitch portion of each pair of conventional eye muscles (Section 7.3.4 & overall servo figure). By employing synchronous detection techniques in stage 2 using the output of the stage 1 photoreceptors of the foveola, the system is able to achieve spatial acuities smaller than the diameter of a single receptor (the commonly defined pixel). Because of the summation of signals from individual photoreceptors in stage 2 signal processing of the peripheral field of the retina, tremor does not play a significant role in the operation of the LGN/occipital pathway. It does play a major role in image extraction within the pulvinar.

The source of the tremor signal used to drive the twitch portion of the eye muscles and to demodulate the stage 1 photoreceptor output signals is not currently known. It may be within the superior colliculus or within the precision optical nucleus. If the reference signal from this source to the demodulating circuits should be interrupted, it is possible an error in polarity of the information provided by the PGN could be passed to the POS and the pulvinar.

As noted by McCloskey et al., “Individuals categorized as dyslexic probably vary considerably at the level of underlying cognitive dysfunctions, and efforts to identify a single cause of dyslexia seem unlikely to prove fruitful.” They also note, “Although impaired performance bearing some resemblance to that of A.H. has been observed in a few brain-damaged patients (e.g., Feinberg & Jones, 1985; Halligan,
Marshall, & Wade, 1992; Jacobs, 1980), we have found no studies in which the patient's performance suggested an underlying cognitive deficit similar to A.H.'s. This does not necessarily mean, however, that A.H.'s impairment is unique or even extraordinarily rare.” The analysis above supports these statements. The variety of failures that can be lumped under the current medical definition of long term (as opposed to developmental) dyslexia is quite large. As also noted by McCloskey et al., “Results from several tasks demonstrate that A.H.’s localization errors reflect a deficit specific to vision. In contrast to her severely impaired performance with visual stimuli, A.H. showed intact localization from auditory, kinesthetic, and tactile information.”

19.13 Time delays among sensory modality tasks

A number of investigators have studied the difference in end-to-end time delays between “similar” tasks. The term similar is used with extreme caution as most of the experiments assumed a similarity between the tasks without the benefit of any top level schematics or block diagrams. The fact that significantly different stage 4 information extraction engines were employed was not considered or evaluated in most of the studies.

19.13.1 The classic Stroop Test and Fraisse

Fraisse studied several pairs of situations in 1969 involving what he considered similar tasks. The fact the paper was translated from native French may have caused some awkwardness with the terminology of the time period. Introducing Figure 19.13.1-1 may make it easier to introduce the experiments Fraisse carried out. In the first experiment, he prepared 15 sets of words and corresponding drawings. There were 5 animal names, 5 plant names and 5 inanimate object names of the type shown. He asserted that “The recognition threshold is a good index of the relative difficulty in perceiving a stimulus, citing his earlier work. He detailed the characteristics of each pair of stimuli. The India ink drawings were circumscribed within a hypothetical 5 cm square, with particular attention being paid to the level of detail as shown. The type was described as Letraset, 4.2 mm high. Both were observed from a distance of one meter. Twenty psychology students were shown the words and 25 others were shown the drawings.

---

The Stroop Test plays a fundamental role in Fraisse’s experiments\textsuperscript{161}. Stroop originally carried out three experiments\textsuperscript{162};

- **Experiment 1**: The Effect of Interfering Color Stimuli Upon Reading Names of Colors Serially,
- **Experiment II**: The Effect of Interfering Word Stimuli upon Naming Colors Serially,
- **Experiment III**: The Effects of Practice (Learning) upon Interference

Stroop provided excellent statistics. However, his terminology surfaces a key problem, the maturity of terminology over time. xxx definition of terms.

More recently, the Stroop Test has taken on several alternate forms. Fraisse notes,,

“A classical example is provided by the two first parts of the Stroop test in which the subject has to read color names and to name a series of colored areas corresponding to the color names. In such a situation the stimuli are different (words and colors) but the verbal response is the same: the color name.

\[ S_1 \text{ (name)} \rightarrow \text{reading} \rightarrow R \]
\[ S_2 \text{ (color)} \rightarrow \text{naming} \rightarrow R \]

Therefore it may be thought that the difference results either from the difficulty in associating the

\[ \begin{array}{l|cc}
\text{Manual reaction time} & \text{Verbal reaction time} \\
\hline
\text{Drawings} & 249 \text{ msec} & 563 \text{ msec} \\
\text{Words} & 259 \text{ msec} & 473 \text{ msec} \\
\hline
\end{array} \]

*Figure 19.13.1-1 Reaction times to identify matched drawings and labels ADD. From Fraisse, 1969.*


verbal response to the stimulus or from the difficulty in perceiving $S_1$ and $S_2$.” There is a complication in using words for naming colors if non-standard colors are used. Japanese has around ten words for the word red! This may introduce an additional delay in the color selection process.

Summarizing Dalrymple-Alford & Budayr: “In the Stroop Test, color naming is found to be slower when the colors are those in which non-corresponding color names are written. A similar result was found when the incongruent word-color combinations were presented individually. When presented in lists, the structure of the lists was found to contribute to impairment of color naming.” The number of variables related to the training of the subjects and the language skills of the subjects makes these experiments very difficult to repeat.

Fraisse indicated “The stimuli with which the 20 students had first been familiarized were presented in a tachistoscope (Scientific Prototype model) with the method of ascending limits. The complexity of the protocol developed by Fraisse and the amount of variability of the previous learning relative to the figures and selected words make confirmation of the experiments problematical.”

The discussion at the end of Fraisse shows the state-of-the-art at that time. Many more recent investigators have suffered the same problem, not having an adequate neural network schematic for reference. Fraisse’s experiment 3 did establish the high significance of training (learning) in the responses to his tests.

The manual reaction times in the above figure refer to pressing a button upon completion of the task by the subject. These are seen to be very similar for the case of the drawings and words. However, the development of a verbal response to the stimulus involves entirely different signaling paths through the neural system as outlined in Figure 19.10.6-1 and is very much longer in both cases. The differences within the verbal reaction time probably reflects the difference between object recognition and word/syllable recognition as well.

Ater trying to prove several negatives, the conclusion of Fraisse was, “The results, some of them old, others recent, show that on the whole neither training nor discriminability of the stimulus can explain the longer duration of the naming process as compared with the reading process.”
19 Reading, listening and Other Complex Operations

related to object recognition ....................................................... 1
19.1 Overview of the general process of perception and cognition ........................................... 2
  19.1.1 General perception of a scene–role of the awareness channel ........................................ 4
  19.1.2 Detailed perception of a scene–role of the analytical channel ......................................... 4
  19.1.3 Detailed perception of a scene–role of the Precision Optical Servomechanism ................. 4
  19.1.4 Important and specific terminology in perception ......................................................... 4
  19.1.5 A The routine versus learning states of operation (old 15.4.1) ........................................ 5
    19.1.5.1 The generic aspects of learning and routine operation EDIT ................................. 5
    19.1.5.1.1 “Experienced” operation of the pulvinar in vision ............................................ 6
    19.1.5.1.2 “Learning mode” of operation of the pulvinar ................................................. 7
  19.2 Perception and cognition of a general scene ........................................................................ 7
    19.2.1 Frequency of ocular motions during perception ......................................................... 7
    19.2.2 Planimetry of eye motion supporting interpretation of detail ........................................ 7
  19.3 Perception and cognition in reading .................................................................................... 8
    19.3.1 The development of writing ........................................................................................... 9
    19.3.2 Higher Levels of Perception–Introduction to reading .................................................. 9
      19.3.2.1 Modes of vision using nomenclature of psychology ................................................. 10
      19.3.2.2 Operating modes extended to reading ................................................................. 11
      19.3.2.3 Operating modes extended to verbalization ......................................................... 11
  19.4 Smooth and Salutatory Pursuit EMPTY .............................................................................. 11
  19.5 Interp & Percept formation within the diencephalon EXPAND ............................................ 11
  19.6 Reserved ............................................................................................................................ 12
  19.7 Reserved ............................................................................................................................ 12
  19.8 Reading and the interpretation of fine detail ........................................................................ 12
    19.8.1 Background .................................................................................................................. 12
      19.8.1.1 Glossary ................................................................................................................ 12
    19.8.2 Overview scenario of the reading process REWRITE .................................................... 13
      19.8.2.1 State of knowledge concerning the circuits of reading .......................................... 14
      19.8.2.2 Setup procedures prior to reading ......................................................................... 14
      19.8.2.3 Typical sequence of minisaccades during reading .................................................. 15
        19.8.2.3.1 The familiarity default procedure in reading .................................................... 18
        19.8.2.3.2 Typical time sequence in reading a paragraph ............................................... 21
        19.8.2.3.3 Difference between good and poor readers ..................................................... 21
      19.8.2.4 The fundamental phenomenon and mechanisms of reading .................................... 21
        19.8.2.4.1 The mechanism of detailed character group analysis ........................................ 21
        19.8.2.4.2 Typical sequences of microsaccades ................................................................. 24
        19.8.2.4.3 Suggestions for future research in reading ......................................................... 24
      19.8.2.5 The Strategy Employed in Reading ......................................................................... 25
        19.8.2.5.1 Examples of the reading process ..................................................................... 25
      19.8.3 Extension of the model of the visual servosystem ........................................................ 26
      19.8.4 Extension of the concepts leading to cognition ............................................................ 27
        19.8.4.1 Role of memory in reading ..................................................................................... 27
        19.8.4.2 Reading alone is achromatic .................................................................................. 28
      19.8.5 The mechanics and mechanisms of reading .................................................................... 28
        19.8.5.1 Functional definitions needed to understand the reading process ......................... 28
        19.8.5.2 Major functional steps in the reading process ....................................................... 29
        19.8.5.3 Memory is crucial to the reading process ............................................................... 30
          19.8.5.3.1 Types of short term memory ............................................................................ 30
          19.8.5.3.2 Types of long term memory .......................................................................... 32
        19.8.5.4 The input to the cognitive portions of the reading mechanism .................................. 32
          19.8.5.4.1 The Precision Optical System (aka Auxiliary Optical System) ......................... 32
          19.8.5.5 The role of the control point in reading EMPTY ................................................. 32
          19.8.5.6 The role of the thalamus in reading EMPTY ....................................................... 32
          19.8.5.7 The role of the cerebellum (and/or pulvinar) in reading EMPTY ......................... 32
          19.8.5.8 The ultimate result of the cognitive portion of the reading mechanism .................. 32
      19.8.6 Performance in reading and analysis of fine detail EMPTY ........................................... 32
        19.8.6.1 Reserved ................................................................................................................ 32
19.8.6.2 Exploratory research related to crowding ........................................ 32
19.8.6.3 Reserved ...................................................................................... 33
19.8.7 Progress in machine vision related to reading ..................................... 33
19.9 Listening to speech .............................................................................. 33
19.9.1 Historical description of Broca’s and Wernicke’s areas EMPTY............ 33
19.9.2 More recent descriptions of brain areas EXPAND ............................... 34
19.10 Generic information extraction leading to object and face recognition .... 34
19.10.1 Historical background focused on object (face) recognition .............. 35
19.10.1.1 Mathematical concepts developed in connection with Neural Networks ............................................... 38
19.10.1.2 Recent conceptual approaches to understand face recognition ....... 42
19.10.1.3 Features in presumed order of importance BRIEF ...................... 46
19.10.1.4 Efforts to show upper or lower field preference in recognition ....... 47
19.10.2 Capability of eye-trackers in face recognition experiments ................ 48
19.10.2.1 Capability of non-invasive eye-trackers ................................. 48
19.10.2.2 The capability of invasive eye-trackers .................................. 52
19.10.2.3 New classes of eye-trackers ............................................... 52
19.10.3 The PEEP procedure used in object recognition .............................. 53
19.10.3.1 The signal manipulations used in object (face) recognition .......... 54
19.10.3.2 The coarse information extraction by the LGN/occipital couple ....... 60
19.10.3.2.1 The unordered saccades among primates ........................ 60
19.10.3.3 The fine information extraction by the PGN/pulvinar couple ...... 63
19.10.3.4 The dimensionality of visual information stored in the saliency map ...................................................... 64
19.10.3.5 Potential means of encoding visuotopic targets of interest EARLY 64
19.10.3.6 The role of savants as related to the plasticity of the brain EMPTY .... 65
19.10.3.7 Current state of knowledge about face recognition EARLY .......... 65
19.10.4 Establishing a multi-discipline object recognition framework ............ 66
19.10.4.1 Overview ................................................................................. 68
19.10.4.1.1 Recent progress in reading of script by machine vision ........ 99
19.10.4.1.2 The papers of Livingstone & Hubel .................................. 102
19.10.4.1.3 Potential implementation of an engine with hidden layers in V1 .................................................. 103
19.10.4.1.4 Potential introduction of engines with hidden layers in V2 through V5/MT .................................. 104
19.10.4.2 A major problem associated with V2 through V5/MT ................. 106
19.10.4.2.1 Potential implementation of an engine with hidden layers in V1 .................................................. 104
19.10.4.2.2 Potential introduction of engines with hidden layers in V2 through V5/MT .................................. 106
19.10.4.2.3 A major problem associated with V2 through V5/MT ............. 108
19.10.4.2.4 The “inclusive” evaluation of Felleman & Van Essen re: V1 to V5/MT ................................................. 108
19.10.4.2.5 The RBM as a potential engine for an engine(s) of the coarse awareness path ........................................ 103
19.10.4.3 Performance of 256 x 256 Boltzmann machine arrays ................. 112
19.10.4.3.1 Performance of the 352 x 264array of Muller et al. .............. 112
19.10.4.3.3 Thoughts on resolving a historical paradox–occlusions of the target .................................................. 116
19.10.4.4 An alternate school using alternate terminology–SVM as a BM .... 119
19.10.6 Plausible model of biological object recognition .............................. 122
19.10.6.1 Neurological background specific to recognition task ............. 122
19.10.6.2 Schematic of human learning—supervision or no supervision .... 123
19.10.6.3 Adaptation of the MV framework to biological vision ............ 127
19.10.6.4 An initial baseline using a mean field version of a BM or HM ... 128
19.10.6.5 Modifications to the baseline using a mean field version of a BM or HM ................................................................. 129
19.10.7 The minimum dimensionality required to index recognized objects .......... 130
19.10.7.1 Rhetorical questions concerning object recognition ............. 130
19.10.7.2 Current estimate of the minimal-dimensionality record of a scene ............................................................ 131
19.10.8 Protocol for evaluating gaze content within PEEP procedure PRELIMINARY .......................................................... 131
19.11 Utilization of extracted information by the Intellectual & Limbic Systems ... ADD 132
19.12 A compendium of abnormal conditions associated with high level vision .......... 132
  19.12.1 Behavioral data on reading problems—summarizing Dehaene ........ 133
    19.12.1.1 Dyslexia occurs in three forms during development ........ 133
    19.12.1.2 Signaling errors in dyslexia of A. H .......................... 140
    19.12.1.2.1 Static perceptions of A. H. vs the visuotopic scene ...... 141
    19.12.1.2.2 Open Loop Experiments ................................. 142
    19.12.1.2.3 Closed Loop Experiments ................................. 142
    19.12.1.2.4 Discussion ............................................. 143
19.13 Time delays among sensory modality tasks ............................ 144
  19.13.1 The classic Stroop Test and Fraisse ............................. 144
**LIST OF FIGURES 4/30/17**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 19.1.4-1</td>
<td>Definition of specific term used in the perception process of this work EMPTY</td>
<td>5</td>
</tr>
<tr>
<td>Figure 19.1.5-1</td>
<td>Candidate state diagram describing the operation of the pulvinar and thalamic reticular nucleus</td>
<td>6</td>
</tr>
<tr>
<td>Figure 19.8.2-1</td>
<td>Hypothetical eye movement record showing the time in milliseconds</td>
<td>16</td>
</tr>
<tr>
<td>Figure 19.8.2-2</td>
<td>The procedure of perceiving and interpreting a sentence</td>
<td>18</td>
</tr>
<tr>
<td>Figure 19.8.2-3</td>
<td>Distribution of fixation duration for two subjects</td>
<td>20</td>
</tr>
<tr>
<td>Figure 19.8.2-4</td>
<td>The horizontal component of eye movements when reading</td>
<td>21</td>
</tr>
<tr>
<td>Figure 19.8.2-5</td>
<td>The Stellen E placed on the retina at 20/20.</td>
<td>22</td>
</tr>
<tr>
<td>Figure 19.8.4-1</td>
<td>The functional organization of the reading mechanisms</td>
<td>31</td>
</tr>
<tr>
<td>Figure 19.10.1-1</td>
<td>Misleading caricature of prosopagnosia</td>
<td>35</td>
</tr>
<tr>
<td>Figure 19.10.1-2</td>
<td>A graphical comparison of ANN compared to the biological neuron</td>
<td>38</td>
</tr>
<tr>
<td>Figure 19.10.1-3</td>
<td>A tabular comparison of artificial and biological neural networks</td>
<td>40</td>
</tr>
<tr>
<td>Figure 19.10.1-4</td>
<td>Basic networks of ANN theory</td>
<td>41</td>
</tr>
<tr>
<td>Figure 19.10.1-5</td>
<td>Figure 2 of Gosselin &amp; Schyns overlaid with gaze circles</td>
<td>44</td>
</tr>
<tr>
<td>Figure 19.10.1-6</td>
<td>Areas associated with face recognition based on fMRI</td>
<td>45</td>
</tr>
<tr>
<td>Figure 19.10.1-7</td>
<td>The significance of facial hair in real face recognition investigations</td>
<td>46</td>
</tr>
<tr>
<td>Figure 19.10.1-8</td>
<td>Brute force approach to computer-vision face recognition REPLACE</td>
<td>47</td>
</tr>
<tr>
<td>Figure 19.10.2-1</td>
<td>Illustration of classification flow for five eye-movement algorithms</td>
<td>50</td>
</tr>
<tr>
<td>Figure 19.10.2-2</td>
<td>Raw eye movement signal with classified fixations and saccades</td>
<td>51</td>
</tr>
<tr>
<td>Figure 19.10.2-3</td>
<td>A prototype active contact lens for precision eye tracking</td>
<td>53</td>
</tr>
<tr>
<td>Figure 19.10.3-1</td>
<td>The general flow of visual information extraction in mammals MODIFY</td>
<td>55</td>
</tr>
<tr>
<td>Figure 19.10.3-2</td>
<td>The PEEP procedure sequence</td>
<td>56</td>
</tr>
<tr>
<td>Figure 19.10.3-3</td>
<td>Initial steps in object (face) recognition using a Yarbus template, “the girl”</td>
<td>57</td>
</tr>
<tr>
<td>Figure 19.10.3-4</td>
<td>Initial steps in object recognition prior to subsequent object identification</td>
<td>58</td>
</tr>
<tr>
<td>Figure 19.10.3-5</td>
<td>Figure 2 of Gosselin &amp; Schyns overlaid with gaze circles</td>
<td>59</td>
</tr>
<tr>
<td>Figure 19.10.3-6</td>
<td>Comparison of saccades sets between primate species</td>
<td>62</td>
</tr>
<tr>
<td>Figure 19.10.3-7</td>
<td>Typical scanning paths related to attention for facial pictures</td>
<td>63</td>
</tr>
<tr>
<td>Figure 19.10.4-1</td>
<td>Which of the objects shown are NOT tiles of a legitimate domino set?</td>
<td>67</td>
</tr>
<tr>
<td>Figure 19.10.4-2</td>
<td>A framework for discussing object recognition</td>
<td>69</td>
</tr>
<tr>
<td>Figure 19.10.4-3</td>
<td>Traffic flow to and from the prefrontal cortex related to the motor function</td>
<td>71</td>
</tr>
<tr>
<td>Figure 19.10.4-4</td>
<td>A prosopagnosiac and normal scanning the same object</td>
<td>78</td>
</tr>
<tr>
<td>Figure 19.10.4-5</td>
<td>Comparison of human (a) and macaque (b) eccentricity maps</td>
<td>83</td>
</tr>
<tr>
<td>Figure 19.10.4-6</td>
<td>A revised functional diagram of human vision ca 2002</td>
<td>85</td>
</tr>
<tr>
<td>Figure 19.10.4-7</td>
<td>Passive membrane simulation in a mixed technology model of the adaptation amplifier</td>
<td>87</td>
</tr>
<tr>
<td>Figure 19.10.5-1</td>
<td>Tree structured hidden Markov models</td>
<td>91</td>
</tr>
<tr>
<td>Figure 19.10.5-2</td>
<td>Three networks of interest in machine vision</td>
<td>93</td>
</tr>
<tr>
<td>Figure 19.10.5-3</td>
<td>The architecture of a Helmholtz Machine</td>
<td>98</td>
</tr>
<tr>
<td>Figure 19.10.5-4</td>
<td>A flow chart for supervised script recognition</td>
<td>100</td>
</tr>
<tr>
<td>Figure 19.10.5-5</td>
<td>Concatenation of analysis cycles using HMM</td>
<td>101</td>
</tr>
<tr>
<td>Figure 19.10.5-6</td>
<td>Designated areas of the cerebrum serving awareness vision</td>
<td>103</td>
</tr>
<tr>
<td>Figure 19.10.5-7</td>
<td>Optical flow extraction on a column by column basis simulating V1</td>
<td>105</td>
</tr>
<tr>
<td>Figure 19.10.5-8</td>
<td>Expansion of the coarse (scene) recognition pathway of the visual modality</td>
<td>107</td>
</tr>
<tr>
<td>Figure 19.10.5-9</td>
<td>Percentage of cells responsive to various stimuli in six visual areas</td>
<td>109</td>
</tr>
<tr>
<td>Figure 19.10.5-10</td>
<td>Preprocessing and feature extraction steps in recognition path</td>
<td>113</td>
</tr>
<tr>
<td>Figure 19.10.5-11</td>
<td>The subsampling raster of Muller et al.</td>
<td>113</td>
</tr>
<tr>
<td>Figure 19.10.5-12</td>
<td>Query sketches and retrieved images</td>
<td>115</td>
</tr>
<tr>
<td>Figure 19.10.5-13</td>
<td>Two common illusions</td>
<td>117</td>
</tr>
<tr>
<td>Figure 19.10.6-1</td>
<td>Schematic of the recognition function in analytic pathway of biological vision</td>
<td>125</td>
</tr>
<tr>
<td>Figure 19.12.1-1</td>
<td>Behavioral aspects of mirror symmetry problems in A. H. ADD</td>
<td>135</td>
</tr>
<tr>
<td>Figure 19.12.1-2</td>
<td>Top Level Modes overlaying the Top Level Stages of vision ADD</td>
<td>137</td>
</tr>
<tr>
<td>Figure 19.12.1-3</td>
<td>Localization error rate as a function of exposure duration ADD</td>
<td>138</td>
</tr>
<tr>
<td>Figure 19.12.1-4</td>
<td>Expanded signal flow diagram between the diencephalon and orthodromic engines ADD</td>
<td>140</td>
</tr>
<tr>
<td>Figure 19.12.1-5</td>
<td>The Rey-Osterrrieth complex figure and A. H.’s direct copy</td>
<td>141</td>
</tr>
<tr>
<td>Figure 19.13.1-1</td>
<td>Reaction times to identify matched drawings and labels ADD</td>
<td>145</td>
</tr>
</tbody>
</table>
### Active Subject Index (using advanced indexing option)

<table>
<thead>
<tr>
<th>Term</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1010</td>
<td>87</td>
</tr>
<tr>
<td>3D</td>
<td>3, 64, 65, 90</td>
</tr>
<tr>
<td>95%</td>
<td>38</td>
</tr>
<tr>
<td>action potential</td>
<td>10, 38, 91, 119</td>
</tr>
<tr>
<td>Activa</td>
<td>10, 38, 39, 86-89</td>
</tr>
<tr>
<td>active diode</td>
<td>89</td>
</tr>
<tr>
<td>adaptation</td>
<td>80, 86, 87, 127</td>
</tr>
<tr>
<td>adaptation amplifier</td>
<td>86, 87</td>
</tr>
<tr>
<td>alarm mode</td>
<td>4, 10, 15</td>
</tr>
<tr>
<td>amplification</td>
<td>86</td>
</tr>
<tr>
<td>analytical mode</td>
<td>6, 10-12, 28, 51, 129, 133, 138, 139, 142, 143</td>
</tr>
<tr>
<td>anatomical computation</td>
<td>34, 105, 129</td>
</tr>
<tr>
<td>area 7</td>
<td>6, 19, 26, 27, 54</td>
</tr>
<tr>
<td>association areas</td>
<td>6, 46, 81, 106, 107, 126, 133, 141, 143</td>
</tr>
<tr>
<td>associative correlator</td>
<td>142</td>
</tr>
<tr>
<td>associative memory</td>
<td>11, 94</td>
</tr>
<tr>
<td>attention</td>
<td>3, 4, 10, 11, 15, 17, 32, 37, 55, 58, 60-63, 74, 80, 117, 126, 127, 132, 134, 139, 144</td>
</tr>
<tr>
<td>awareness mode</td>
<td>10, 12, 28, 104, 138, 139</td>
</tr>
<tr>
<td>basal ganglia</td>
<td>71</td>
</tr>
<tr>
<td>baseball</td>
<td>90, 126</td>
</tr>
<tr>
<td>bat</td>
<td>39</td>
</tr>
<tr>
<td>Bayesian</td>
<td>91, 92, 95, 128</td>
</tr>
<tr>
<td>bifurcation</td>
<td>102</td>
</tr>
<tr>
<td>bi-stratified</td>
<td>90</td>
</tr>
<tr>
<td>blob</td>
<td>103</td>
</tr>
<tr>
<td>BOLD</td>
<td>22, 133</td>
</tr>
<tr>
<td>Boltzmann machine</td>
<td>83-85, 92-96, 110, 112, 116, 120, 128</td>
</tr>
<tr>
<td>broadband</td>
<td>24, 123</td>
</tr>
<tr>
<td>Brodmann</td>
<td>6, 54, 81, 82, 104</td>
</tr>
<tr>
<td>calibration</td>
<td>43</td>
</tr>
<tr>
<td>cerebellum</td>
<td>2, 4-6, 9, 11, 28, 32, 68, 70, 71, 90, 122-129, 138, 139</td>
</tr>
<tr>
<td>cerebrum</td>
<td>4-7, 13, 19, 28, 55, 79, 81, 103, 104, 106, 133</td>
</tr>
<tr>
<td>CIE</td>
<td>116</td>
</tr>
<tr>
<td>coaxial cable</td>
<td>86</td>
</tr>
<tr>
<td>colliculus</td>
<td>3, 9, 70, 102, 132, 143</td>
</tr>
<tr>
<td>coma</td>
<td>7</td>
</tr>
<tr>
<td>commissure</td>
<td>14, 111, 112</td>
</tr>
<tr>
<td>computation</td>
<td>34, 35, 42, 44, 60, 73, 81, 105, 129</td>
</tr>
<tr>
<td>computational</td>
<td>5, 9, 11, 13, 23, 27, 34, 37, 42, 73, 78, 80, 82, 89, 92, 95, 97, 122, 128, 141, 143</td>
</tr>
<tr>
<td>computational anatomy</td>
<td>42</td>
</tr>
<tr>
<td>conduction velocity</td>
<td>119</td>
</tr>
<tr>
<td>confirmation</td>
<td>2, 78, 92, 122, 146</td>
</tr>
<tr>
<td>conscious executive</td>
<td>58</td>
</tr>
<tr>
<td>consciousness</td>
<td>2, 74</td>
</tr>
<tr>
<td>conspecifics</td>
<td>62</td>
</tr>
<tr>
<td>continuum</td>
<td>105</td>
</tr>
<tr>
<td>cross section</td>
<td>90, 102</td>
</tr>
<tr>
<td>data base</td>
<td>100</td>
</tr>
<tr>
<td>database</td>
<td>13, 25, 47, 53, 65, 113, 122, 126, 130, 131</td>
</tr>
<tr>
<td>declarative memory</td>
<td>4, 27, 57, 66, 77, 122, 127, 131</td>
</tr>
<tr>
<td>declaratory memory</td>
<td>3, 6, 28, 130, 131</td>
</tr>
<tr>
<td>depth perception</td>
<td>140</td>
</tr>
<tr>
<td>determinants</td>
<td>36</td>
</tr>
<tr>
<td>diencephalon</td>
<td>11, 34, 71, 121, 140</td>
</tr>
<tr>
<td>diode</td>
<td>79, 89</td>
</tr>
<tr>
<td>disparity</td>
<td>48, 52</td>
</tr>
<tr>
<td>dyslexia</td>
<td>133, 134, 137, 140-144</td>
</tr>
<tr>
<td>equilibrium</td>
<td>97</td>
</tr>
<tr>
<td>evoked potential</td>
<td>3</td>
</tr>
<tr>
<td>evolution</td>
<td>94</td>
</tr>
<tr>
<td>executive</td>
<td>58, 139</td>
</tr>
</tbody>
</table>
152 Processes in Biological Vision

expanded ................................................................. 6, 27, 68, 70, 80, 81, 92, 124, 140
external feedback .......................................................... 10, 86, 87
feedback .......................................................... 10, 13, 41, 86, 87, 91, 92, 98, 128, 136, 139, 140
feedforward ........................................................... 3, 91, 92, 98, 120, 128
fill-in ........................................................................ 117, 118
fMRI ........................................................................ 1, 14, 34, 44-46, 60, 76, 81, 82, 106
Fourier transform ....................................................... 129
frequency of occurrence ...................................................... 7, 18
Gaussian ........................................................................ 42-44, 58, 59
Gestalt ........................................................................... 77
Grandmother .................................................................. 7
Helmholtz machine ............................................................. 91, 97, 98, 110, 128
HMM ........................................................................... 50, 92, 95, 96, 99-101, 116
homogeneous ................................................................ 75, 76, 113
horopter ........................................................................... 2
illusion ........................................................................... 66, 77, 117
inflated ........................................................................... 111
intelligence ..................................................................... 12, 34, 90, 102, 122-124, 127
internal feedback ........................................................... 86
interp ........................................................................... 1, 5-7, 9, 11, 14, 15, 23, 27-31, 34, 35
inverting ........................................................................... 37-39, 41
latency ........................................................................... 16, 17, 19, 52, 54
lateral geniculate ................................................................. 4, 121
lgn/occipital ................................................................. 4, 34, 36, 55, 56, 60, 64, 66, 70, 79, 92, 104, 131, 133, 136, 138, 140-143
limbic system ................................................................... 132
lips ................................................................................ 72
listening ........................................................................... 1, 2, 9, 33, 34, 73
long term memory ........................................................... 30, 32, 133, 134
lookup table ..................................................................... 5, 16, 30, 68, 72, 123
LOT ........................................................................... 45, 50, 103
machine vision ................................................................. 33, 39, 52, 80, 81, 84, 89, 90, 93, 94, 99, 102, 104, 110, 112, 119-122
marker ........................................................................... 64
Markov ........................................................................... 33, 50, 90, 91, 95, 96, 99, 101, 116, 128
medial geniculate ................................................................. 112
MEG ........................................................................... 81
metadata ........................................................................... 34, 60, 64-68, 70, 72, 76, 78, 81, 90, 104, 105, 122, 123, 126, 130, 131
midbrain ........................................................................... 4, 14, 26, 28
modulation ...................................................................... 11
monopulse ........................................................................ 40, 88, 127
monopulse oscillator .............................................................. 40, 88
morphogenesis ................................................................ 23
MRI ........................................................................... 1, 11, 14, 46, 82, 133
multi-dimensional ............................................................. 28, 30, 31, 36, 64, 67, 80, 92
myelinated ...................................................................... 86, 119
narrow band ................................................................... 108
neurite ........................................................................... 89
neurites .......................................................................... 111, 127
Node of Ranvier ................................................................ 39, 40, 88
noise ........................................................................... 2, 16, 28, 32, 37, 48, 51, 65, 110, 139
nonconscious executive ......................................................... 139
non-declarative memory ....................................................... 27
non-declaratory memory ......................................................... 3, 28
non-inverting ................................................................. 38, 39
OCT ........................................................................... 52
P/D equation ................................................................... 131
parametric ...................................................................... 95
parietal lobe ..................................................................... 2, 6, 54, 55, 106
parvocellular .................................................................... 102, 139
PEEP ........................................................................... 34, 43, 48, 53-56, 58-61, 64-66, 68, 70, 72-74, 78, 80, 83-86, 89, 90, 92, 97, 104, 105, 110, 131, 136, 141-143
percept ................................................................. 1, 5, 7, 11, 14, 15, 27-31, 66
perigeminate ................................................................. 4, 10, 112, 132
perigeminate nucleus ........................................................... 10, 132
<table>
<thead>
<tr>
<th>Term</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>perimetry</td>
<td>139</td>
</tr>
<tr>
<td>PET</td>
<td>1, 11</td>
</tr>
<tr>
<td>pgn/pulvinar</td>
<td>4, 34, 36, 46, 55-57, 60, 63, 64, 66, 70, 79, 86, 92, 123, 131, 133, 136, 138-143</td>
</tr>
<tr>
<td>plasticity</td>
<td>65, 119</td>
</tr>
<tr>
<td>pnp</td>
<td>86, 87</td>
</tr>
<tr>
<td>poda</td>
<td>10</td>
</tr>
<tr>
<td>podites</td>
<td>39</td>
</tr>
<tr>
<td>POS</td>
<td>4, 8, 12-15, 18, 19, 23, 26, 27, 30, 84, 86, 105, 110, 112, 123, 124, 126, 128, 131, 142, 143</td>
</tr>
<tr>
<td>precision optical servomechanism</td>
<td>4, 86, 104, 105, 143</td>
</tr>
<tr>
<td>Pretectum</td>
<td>5, 6, 13, 14, 23, 26-29, 31</td>
</tr>
<tr>
<td>probabilistic</td>
<td>92, 94, 95</td>
</tr>
<tr>
<td>procedural memory</td>
<td>28</td>
</tr>
<tr>
<td>protocol</td>
<td>33, 42, 48, 67, 73, 75, 102, 110, 120, 123, 131, 132, 146</td>
</tr>
<tr>
<td>pulvinar</td>
<td>3-7, 9-11, 13, 19, 23, 28, 30, 32, 34, 36, 46, 53-58, 60, 63, 64, 66, 70, 79, 81, 82, 86, 90, 92, 95, 102, 104, 110-112, 120-129, 131-133, 136, 138-143</td>
</tr>
<tr>
<td>Pulvinar pathway</td>
<td>4, 6, 53, 58, 70, 79, 82, 104, 120-122, 126, 131, 140, 141, 143</td>
</tr>
<tr>
<td>raster scan</td>
<td>80</td>
</tr>
<tr>
<td>RBM</td>
<td>84-86, 93, 94, 96, 97, 102-106, 110-112, 120, 128</td>
</tr>
<tr>
<td>reading</td>
<td>1-3, 5, 7-9, 11-18, 20, 21, 24-33, 35, 48, 50, 52, 53, 67, 73, 76, 85, 88, 90, 92, 99, 101, 110, 111, 129-134, 144-147</td>
</tr>
<tr>
<td>residue</td>
<td>123</td>
</tr>
<tr>
<td>resonance</td>
<td>34, 82</td>
</tr>
<tr>
<td>Restricted Boltzmann Machine</td>
<td>84, 85, 93, 94, 120</td>
</tr>
<tr>
<td>reticulated</td>
<td>60, 63, 79, 80</td>
</tr>
<tr>
<td>Riemann transform</td>
<td>42</td>
</tr>
<tr>
<td>saliency map</td>
<td>4, 10-12, 18, 25, 27, 28, 34, 53-55, 57, 60, 64, 72, 81, 91, 122, 126-128, 138, 139</td>
</tr>
<tr>
<td>scotoma</td>
<td>118</td>
</tr>
<tr>
<td>segregation</td>
<td>102</td>
</tr>
<tr>
<td>servo loop</td>
<td>27, 52</td>
</tr>
<tr>
<td>servomechanism</td>
<td>4, 8, 26, 27, 72, 86, 104, 105, 131, 142, 143</td>
</tr>
<tr>
<td>signal-to-noise ratio</td>
<td>28, 139</td>
</tr>
<tr>
<td>sleep</td>
<td>97</td>
</tr>
<tr>
<td>stage 0</td>
<td>42, 60</td>
</tr>
<tr>
<td>stage 1</td>
<td>4, 70-72, 80, 84, 86, 103, 104, 117, 118, 122, 128, 143</td>
</tr>
<tr>
<td>stage 2</td>
<td>4, 38, 39, 60, 63, 64, 72, 104, 118, 143</td>
</tr>
<tr>
<td>stage 3</td>
<td>38, 39, 81, 88, 91, 110, 127, 142, 143</td>
</tr>
<tr>
<td>stage 4</td>
<td>1, 4, 11, 34, 35, 42, 53, 54, 58, 60, 64, 66-69, 71-73, 75, 78, 79, 81, 84, 86, 90, 92, 94, 103-106, 108, 110, 112, 118, 122, 124, 127, 128, 130, 132, 133, 142-144</td>
</tr>
<tr>
<td>stage 4b</td>
<td>68, 70, 138</td>
</tr>
<tr>
<td>stage 5</td>
<td>4, 9, 11, 57, 66, 69, 70, 72, 73, 75, 78, 90, 122-124, 126-129, 139, 142, 143</td>
</tr>
<tr>
<td>stage 6</td>
<td>9, 68-70, 72, 73, 75, 90, 123, 126, 127, 129, 139, 142</td>
</tr>
<tr>
<td>stage 7</td>
<td>68, 70</td>
</tr>
<tr>
<td>stratified</td>
<td>90</td>
</tr>
<tr>
<td>stress</td>
<td>35, 43, 54</td>
</tr>
<tr>
<td>Stroop Test</td>
<td>76, 144-146</td>
</tr>
<tr>
<td>superior colliculus</td>
<td>3, 9, 70, 102, 132, 143</td>
</tr>
<tr>
<td>SVM</td>
<td>104, 119</td>
</tr>
<tr>
<td>synapse</td>
<td>38-40, 80, 89, 98</td>
</tr>
<tr>
<td>Talairach</td>
<td>9, 34, 81</td>
</tr>
<tr>
<td>temporal lobe</td>
<td>34, 54, 55, 79-81, 103, 106</td>
</tr>
<tr>
<td>thalamic reticular nucleus</td>
<td>6, 11, 85, 86, 105, 111, 121</td>
</tr>
<tr>
<td>thalamus</td>
<td>2, 3, 14, 28, 32, 60, 111, 121</td>
</tr>
<tr>
<td>threshold</td>
<td>29, 48-50, 94, 108, 144</td>
</tr>
<tr>
<td>TMS</td>
<td>46</td>
</tr>
<tr>
<td>tomography</td>
<td>34, 52</td>
</tr>
<tr>
<td>top hat</td>
<td>69</td>
</tr>
<tr>
<td>topography</td>
<td>79, 81, 108</td>
</tr>
<tr>
<td>topology</td>
<td>88, 89, 92, 110, 123, 128</td>
</tr>
<tr>
<td>Transcranial Magnetic Stimulation</td>
<td>46</td>
</tr>
<tr>
<td>translation</td>
<td>5, 7, 9, 17, 68, 73, 80, 120</td>
</tr>
<tr>
<td>transposition</td>
<td>133</td>
</tr>
<tr>
<td>tremor</td>
<td>1, 4, 7, 8, 13, 15, 17, 22, 23, 27, 52, 53, 65, 75, 80, 90, 91, 110, 112, 122, 123, 127, 132, 143</td>
</tr>
<tr>
<td>Term</td>
<td>Page Numbers</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Turing</td>
<td>93, 94</td>
</tr>
<tr>
<td>type 0 servomechanism</td>
<td>142</td>
</tr>
<tr>
<td>type 1</td>
<td>142, 143</td>
</tr>
<tr>
<td>type 1 servomechanism</td>
<td>142, 143</td>
</tr>
<tr>
<td>V2</td>
<td>3, 79, 103, 105-108, 117</td>
</tr>
<tr>
<td>V5/MT</td>
<td>106, 108</td>
</tr>
<tr>
<td>VEP</td>
<td>3, 106</td>
</tr>
<tr>
<td>visual acuity</td>
<td>2, 108</td>
</tr>
<tr>
<td>visual cortex</td>
<td>40, 60, 92, 108, 120, 121, 131</td>
</tr>
<tr>
<td>Viterbi algorithm</td>
<td>96, 101</td>
</tr>
<tr>
<td>volition mode</td>
<td>10</td>
</tr>
<tr>
<td>von Neumann</td>
<td>90, 92-94, 105</td>
</tr>
<tr>
<td>Wernicke’s area</td>
<td>33</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>46, 94, 95</td>
</tr>
<tr>
<td>word serial</td>
<td>67, 130</td>
</tr>
<tr>
<td>word serial/bit parallel</td>
<td>67, 130</td>
</tr>
<tr>
<td>working memory</td>
<td>48</td>
</tr>
<tr>
<td>X-ray</td>
<td>34</td>
</tr>
</tbody>
</table>