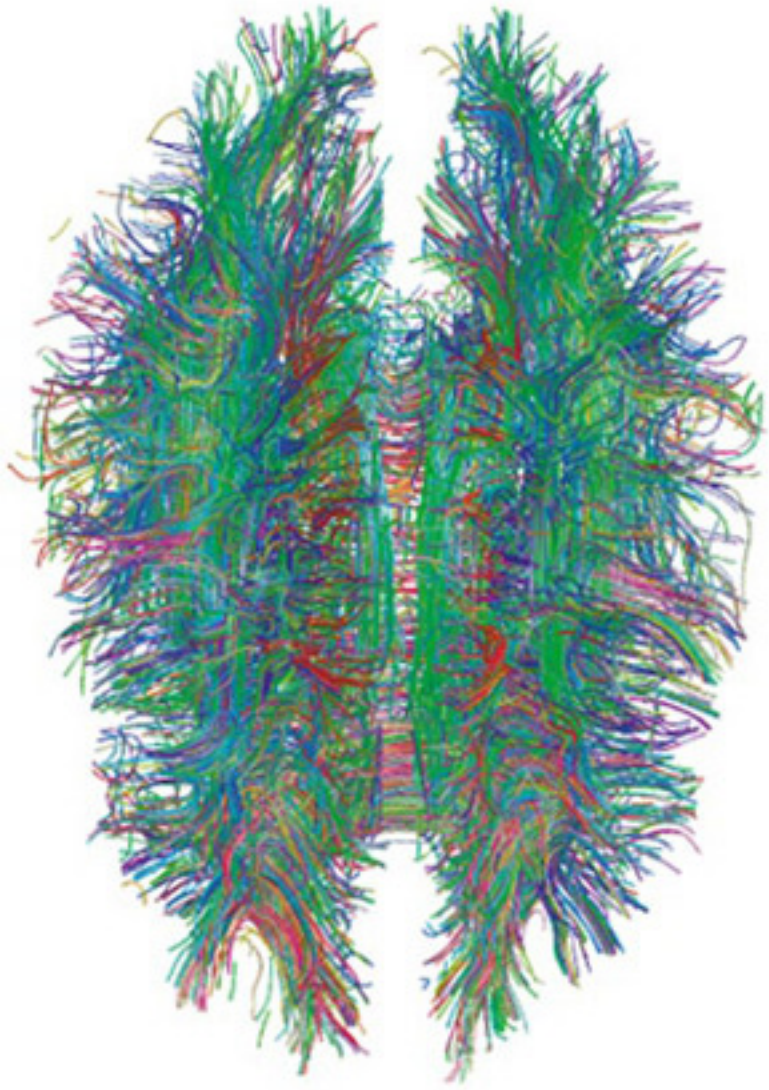
Neural Systems

NEUR 1030 | Dr. Linden



Brown University

Originally written by:

James McIlwain, *Sidney A. and Dorothea Doctors Fox Professor Emeritus of Ophthalmology and Visual Science, Professor Emeritus of Surgery, Professor Emeritus of Neuroscience*

Edited by:

David Sheinberg, *Professor of Neuroscience*
Monica Linden, *Sr. Lecturer in Neuroscience*

Layout and Graphic Design by:

Emma Margulies ’18
Morgan Cheatham ’17

Acknowledgements:

Nora Orton, TJ Kalaitzidis, Ph.D., Morgan Cheatham ’17, all NEUR 1030 students who helped error-check the reader

NEUR 1030: Neural Systems
Table of Contents

Models and Methods in Systems Neuroscience 1

Table of Contents

How are neural systems studied?

Neuroscientists use many different techniques to analyze neural systems. These methods help ascertain which neurons are interconnected, what the circuits do, and how the circuitry translates into behavior. The connectivity of a system can be established by anatomical methods, such as pathway mapping with chemical tracers that move retrogradely or anterogradely along axons by normal axonal transport mechanisms. Pharmacological agents, immunological probes or artificial stimulation can be used to determine the functionality of these connections. Electrical recordings and activity-sensitive imaging methods measure the dynamic, real-time behavior of neural circuits. Combining such measurements with behavioral observations sheds light on perceptual processes otherwise inaccessible to external observers.

Because these methods can be quite complex, when studying neural systems we use models. A model is a compact, economical way to represent the essential features of a phenomenon, process, mechanism or aspect of reality. A model must strip away detail to get to the heart of a concept, so it generally presents an impoverished view of the real thing. Many think that models always involve mathematical expressions, but in reality phenomena can be represented at many levels of abstraction. Even as we speak and write, we are using language to model various things. For example, the statement ‘the sun always rises in the east’ is a model that uses words and syntax to represent a phenomenon. Because the phenomenon is observable, the model can be tested by getting up in the morning and seeing if the sun really does rise in the east. If it rises in the east, the model is supported by the evidence, ‘confirmed’, or ‘validated’, at least for the time being. If the sun happens to rise in the west that day, the model is invalidated. All theories, hypotheses and so-called scientific laws are models of one sort or another.

Models in Systems Neuroscience

While language serves as a model for communication, neuroscientists employ types of representations to capture the essential properties of a system. A list of the kinds of models that are used might include the following:

1. Idealized Neurons and Their Connections

   A diagram of the myotatic reflex that connects a single Ia muscle spindle afferent to a single alpha motor neuron would be such a model. It would capture the idea of the myotatic reflex circuitry very well, but would ignore a number of anatomical details, such as the fact that a single Ia afferent from a muscle contacts all the alpha motor neurons to that muscle.

2. Mechanical Devices

   Mechanical devices of known properties can be used to model phenomena such as muscle contraction. We will encounter this later when we represent a muscle and its neural connections by a spring of variable stiffness. The behavior of springs is well understood in quantitative, physical terms, and this understanding can be applied directly to a living system that behaves like a spring.

3. Electric Circuits

   Neural processes often lend themselves to representation in terms of the properties of electrical circuits made up of resistors, capacitors and batteries. In the next chapter, we will model the technique of extracellular recording in this way. Such a model is often called the equivalent circuit of the phenomenon being represented.

4. Black Boxes

   Systems can be represented as a set of interconnected 'black boxes' symbolizing various kinds of operations such as amplification, integration, summation and so forth. This method identifies a set of operations that appear to be going on in a system without worrying about exactly how the operations are actually performed. This type of model is very important in understanding motor systems, and we will use it a lot.

5. Mathematical

   Perhaps the most abstract model is the mathematical one which represents a system in terms of equations. For example, an efficient, compact model of a voltage amplifier is simply the ratio of its output to its input:

$$V_{out} = 10 * V_{in}$$

   without reference to whether the amplifier is built of integrated circuits, discrete

transistors or old fashioned vacuum tubes. If all that one is concerned with is the amplification or *gain* (G) of the amplifier, this model is a sufficient description.

$$G = \frac{V_{out}}{V_{in}}$$

and in this case is 10.

The match between a model and what it represents is captured by the term *isomorphism* (Gr., *iso*-same, *morphe*-form). The model may be isomorphic with the function of a system, as in the equation

$$V_{out} = 10 * V_{in}$$

or there may be a detailed correspondence between the elements of the model and the components of the neural system.

Models are most useful if they are testable. If a model cannot be falsified, it is too general. When a model is proved wrong, parts of the model are often amended to account for the new empirical results and thus the model evolves.

Methods in Systems Neuroscience

Electrical recording from single neurons: intracellular recording

Measurements of the electrical activity of the nervous system are an important way to assess the dynamic behavior of the system. To gain knowledge at the single-cell level, there are two basic recording techniques: *intracellular* and *extracellular*. It is important to distinguish between intracellular and extracellular recording because the two methods yield different kinds of information about events occurring at the level of single neurons. Figure 1.1 schematizes some of the key features of intracellular recording and a typical intracellular record is shown in Figure 1.2. A fine-tipped micropipette filled with a conducting salt solution is introduced into the cytoplasm through the cell's membrane. At the moment of penetration, the electrode and amplifying equipment record the resting membrane potential, E_m. Remember that the value of E_m results largely from the differential permeability of the cell membrane to potassium. The intracellular electrode is in a position to register the effects of a number of physiological events that modify the membrane potential.

There are multiple cellular events that are read in an intracellular recording. Activation of excitatory synaptic inputs can result in depolarization of the cell resulting from the opening of cation channels at the synaptic region, which is known as an excitatory postsynaptic potential (EPSP). Inhibitory synaptic inputs may result in a rapid and relatively short lasting

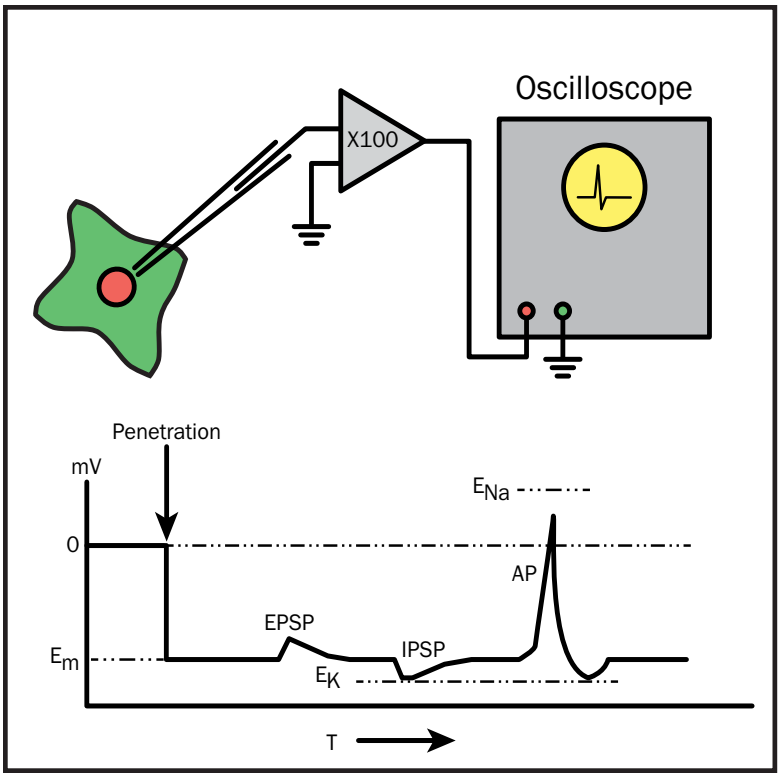


Figure 1.1

Schematic illustration of intracellular recording of synaptic potentials and action potentials. Intracellular recordings provide information about changes in membrane potential by measuring difference potentials between the inside and outside of single neurons.

hyperpolarization of the cell known as an inhibitory postsynaptic potential (IPSP). IPSPs result from the opening of potassium and chloride channels, which usually drives the membrane toward the potassium equilibrium potential, E_K. In some cases, when only chloride channels are opened and E_m is close to E_{Cl}, there is no hyperpolarization, but the increased membrane conductance has an inhibitory effect by short circuiting or shunting the depolarizing currents. Certain neurotransmitters and neuromodulators can also modify the membrane potential through intracellular messenger systems. These processes can result in excitation or inhibition of the cell.

If the membrane becomes sufficiently depolarized, voltage-gated sodium channels open, current flows into the cell carried by sodium ions, and an all-or-none action potential is initiated. The E_m is driven toward E_{Na} (the sodium equilibrium potential), although it does not reach it before repolarization begins. Repolarization occurs as voltage-gated potassium channels open, potassium flows out of the cell, restoring E_m to near E_K, or the Na/K pump restores the resting concentration gradients of the cell.

Because intracellular electrodes enter the cell, they can sense small-scale changes in the ionic environment. Thus, intracellular recordings can yield information about the membrane voltage during both action potentials and synaptic potentials (See Figure 1.1)

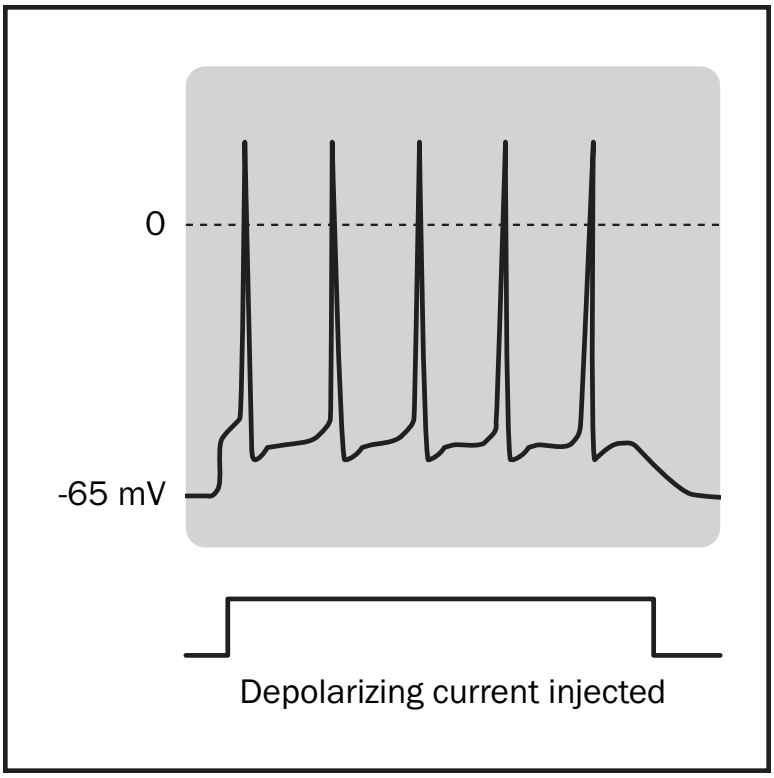


Figure 1.2

Intracellular recordings from a neuron of the cerebral cortex. Depolarizing current injected through the intracellular electrode causes the cell to generate action potentials accompanied by slower potentials.

One can use the intracellular technique to manipulate the membrane. One method is called *voltage clamp*, in which the experimenter sets or "clamps" the voltage at a certain value and observes how current flows at a given voltage. The *current clamp* method, on the other hand, entails injecting current into the cell to and observing how the membrane potential changes (see Figure 1.2).

In recent years it has become possible to study a small patch of the neural membrane in isolation and to measure its responses to changes in the transmembrane voltage and electric gradients. This method, known as *patch recording*, even permits the study of single ion channels. In essence, a micropipette with a specially prepared tip is used to detach a small piece of membrane and leave it sealed to the electrode tip for study, making it easy to vary the ionic environment on either side of the patch. A variation on this method is to patch on to the cell with an electrode containing an antibiotic that will punch holes in the patch and leave the whole cell attached. Such *whole cell* or *attached cell* patch recording gives the pipette electrical access to the interior of the cell without actually having to penetrate the cell.

Electrical recording from single neurons: Extracellular recording

Action potentials and synaptic potentials result in a distinctive flow of current in the

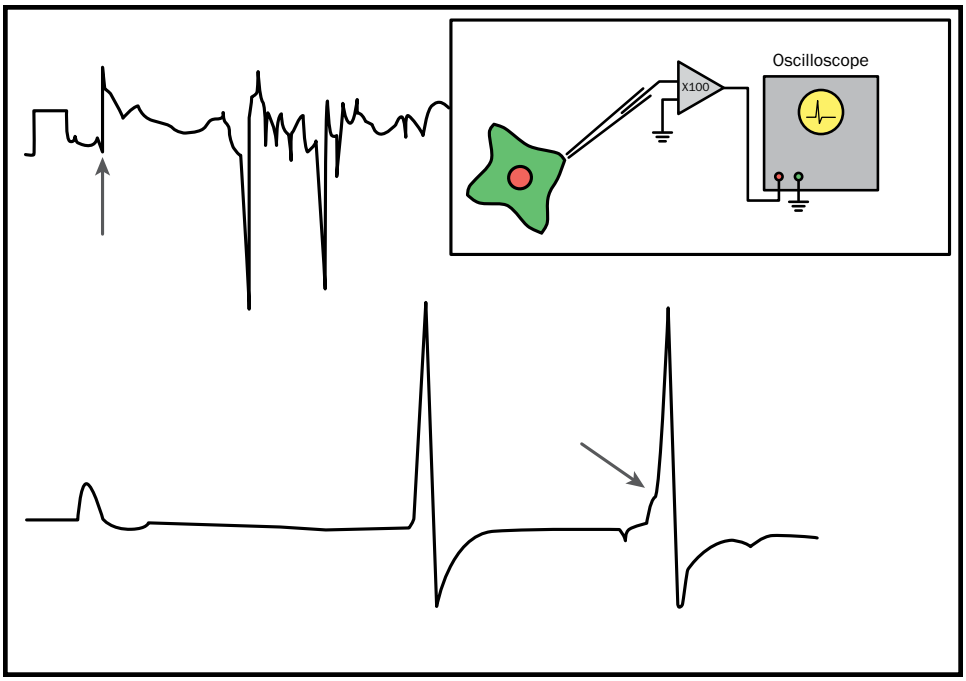


Figure 1.3

Extracellular recordings. Top: neuron in the superior colliculus responding to electrical stimulation of the optic tract (at arrow). The action potentials are largely negative in polarity. Bottom: Neuron in the lateral geniculate nucleus of the cat. Electrode is near the source of current for the action potential and also records smaller amplitude currents due to EPSPs that sometimes trigger action potentials (arrow). Inset illustrates that in extracellular recordings both the microelectrode and reference are outside the cell lines.

extracellular space, and a suitably placed microelectrode can detect the voltage changes produced by these currents. Electrodes for extracellular recording may be made of glass micropipettes but are more commonly fabricated from various metals in the form of sharpened rods or microwires, insulated except at the very tip.

It is conventional in neuroscience to arrange the recording apparatus so that the recording device produces an upward deflection of the oscilloscope beam (or paper trace in the old days, this may all be done digitally with computers now) when the tip of the electrode registers a voltage that is positive with respect to a reference electrode placed remotely, say on the scalp (bottom trace in Figure 1.3). Such a deflection is said to be *positive-going*. The extracellular spikes in the top trace of are *negative-going*. The reasons for these differences in recorded spike *polarity* will be discussed in the next chapter.

Rasters and Histograms

The neuroscience literature has adopted several conventional ways of illustrating the electrical behavior of cells measured by extracellular recordings. In a typical experiment a

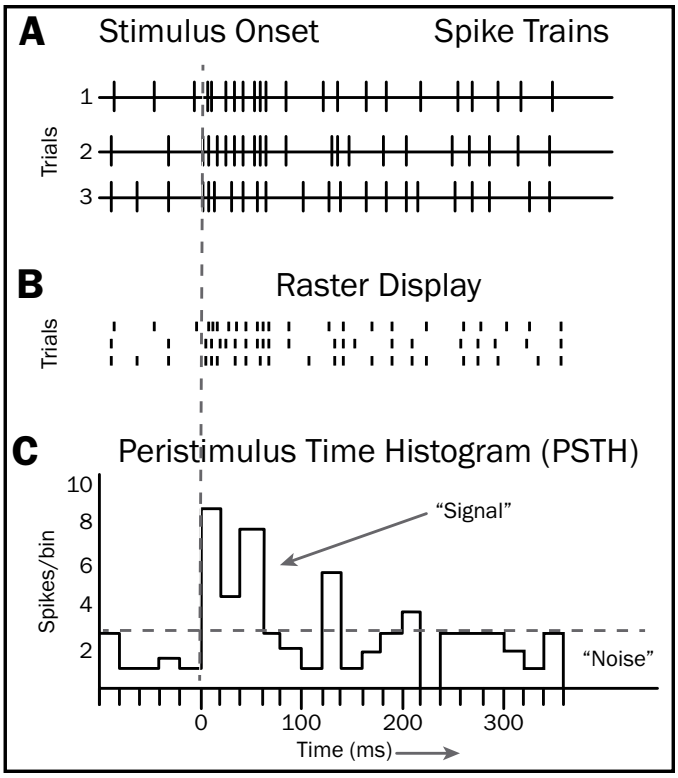


Figure 1.4

Conventional displays of extracellular spike trains. A. Schematic drawing of three responses of a single neuron to three presentations of the stimulus. The vertical lines represent spikes when they would appear if recorded by an oscilloscope as the trace moves slowly from left to right in time (although a real trace would be noisier). The stimulus occurs 100 milliseconds after the oscilloscope sweep begins. B. Raster display of the responses to three stimulus presentations. Tic marks are shown at each time of occurrence of a spike. All other details of the signal are discarded. C. PSTH that sums the number of spikes in each 20 ms epoch or bin across trials. 'Noise' is defined here as the level of discharge prior to the occurrence of the stimulus.

stimulus may be presented several times, either in succession or interspersed with other stimuli. The responses can be displayed as 'spike trains' from different trials, as in Figure 1.4, which schematically shows a single cell's response to three stimulus presentations or trials. Each vertical line represents an extracellular recording of an action potential as it might appear if recorded on an oscilloscope.

To compare larger numbers of repeated presentations, the spike trains can be transformed (usually by a computer) into a *raster display*, a series of dots or tic marks (short vertical lines), where each dot represents the occurrence of an action potential, and each row represents a single trial. Figure 1.4 is a raster display of the spike trains of panel A, with each spike replaced by a tic mark and each trial appearing as a row of tic marks.

One way to summarize the responses to repeated presentations of a stimulus or some other experimental manipulation is a *peristimulus* or *poststiumulus-time* histogram (PSTH). The recording period is divided into time segments, called "bins." The experimenter creates a histogram by counting the number of spikes occurring in each bin as in Figure 1.4. These histograms are "*peristimulus*" because they display spike events *around* the time of stimulus presentation.

Signal averaging and the signal-to-noise ratio

In a typical situation some spikes in a trace are a direct result of the stimulus and others are unrelated to the stimulus and usually of unknown cause. Spikes that arise because of the stimulus are called the *signal* and the unrelated spikes are the *noise*.

Determining which spikes are "signal" and "noise" is important for understanding how a cell responds to a stimulus. Because signal spikes regularly occur after the stimulus, there are ways to reduce the effect of the randomly occurring 'noise' spikes. The 'noise' spikes are as likely to occur at one time as another, so they will be distributed evenly across the time-bins in the histogram after a large number of repeated trials. In contrast, the 'signal' spikes will add up more quickly in certain bins than in others because they tend to be *time-locked* to the stimulus. As repeated trials are accumulated, the peaks caused by the synchronized 'signal' spikes will rise above the evenly distributed 'noise' spikes, and the presence of the signal will be enhanced. This technique can be used to detect very weak effects of a stimulus in the presence of a lot of noise.

The average response of a cell is one useful way to summarize the behavior of the cell. This estimate is found by dividing each bin in the histogram by the number of trials. Yet, the summation operation is the key to detecting the effect of the stimulus. In Figure 1.4, there is a background activity of 1-3 spikes per bin before the presentation of the stimulus, which would provide an estimate of the 'noise' of the recorded activity. Three spikes is the maximum 'noise' expected in each bin, so number of spikes above that is likely related to the stimulus the stimulus and reveals the presence of the 'signal'. The horizontal dashed line represents the demarcation between signal and noise.

$$\text{Average response} = \frac{\frac{\text{spikes}}{\text{bin}}}{\frac{\text{time}}{\text{bin}}} = \frac{\text{bin}}{\text{bin size}} = \frac{\text{spike}}{\text{time}}$$

The relative magnitude of the observed 'signal' compared to the 'noise' is called the signal-to-noise ratio. There are various ways to estimate this ratio. In general, the noise is measured by spiking activity records over periods when no stimulus is present (or in the presence of a fixed background stimulus). The amount of activity detected in response to a signal can then be compared to this background level of activity. The ratio of the mean activity during the signal period to the standard deviation of the noise level is one way to quantify the signal to noise ratio.

The Equivalent Circuit 2

Table of Contents

Electric Circuits and Ohm's Law

Action potentials and synaptic potentials are associated with current flow into and out of the neuron from the extracellular space (Figure 2.1). At the initiation of the action potential, positive current (carried by Na^+ ions) flows into the soma, charging the inside of the active patch of membrane strongly positive. As the action potential propagates, current flows through the cytoplasm and exits the cell through adjacent, inactive areas of the membrane. The exiting current is carried mostly by K^+ ions. Because of the ionic flow that occurs during action potentials and synaptic potentials, we can model a cell as an electric circuit.

Figure 2.2 demonstrates a model of an electric circuit. The battery causes a separation of charge (think of the + and – terminal of household batteries). The separation of charge creates a potential difference (like a potential energy). This energy is the voltage, measured in units of "volts" (V_b, for "battery voltage" in the diagram). Technically defined, one volt is the energy required to move a unit of positive charge a distance of one meter. Sometimes, this energy is called the "electromotive force" and is often represented in equations by the variable $\mathcal{E}$ (see Equation 1, below).

If we created a path for the charge to flow (for example, by adding a wire to our circuit), the positive charge would be attracted to the negative charge (and vice versa). This would induce a current in the circuit. Current is the net flow of electrical charge per unit time. Note: by a physics convention, when we talk about current we are referring to the movement of positive ions.

The resistor, which is represented by a zig-zag, disrupts the current flow. It is somewhat similar to the idea of friction, as it opposes the passage of the electrical current. Sometimes

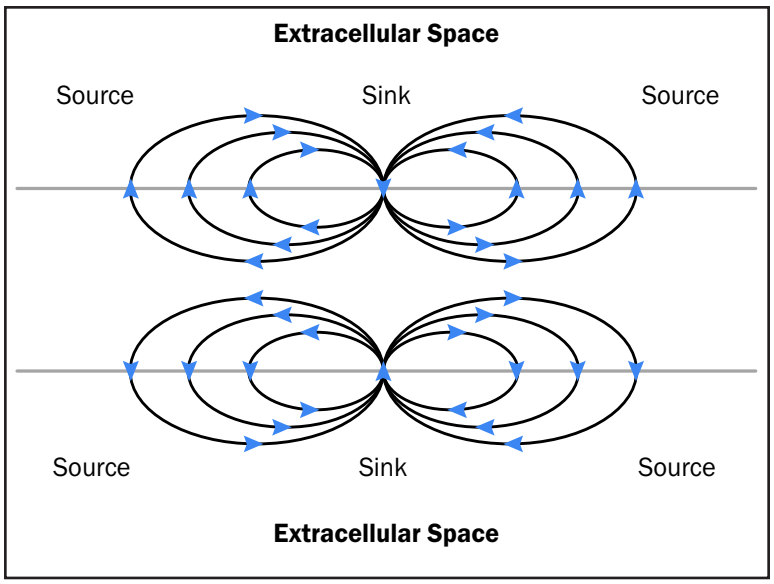


Figure 2.1

Current flow across the cell membrane at the site of an action potential. Arrows indicate the direction of positive current flow. Current density is highest at the sink labeled here, which represents the location of the ion channel.

it is useful to consider the reciprocal of resistance, called the conductance (represented by the variable G). Conductance is essentially how easily the current can pass.

These attributes serve as a good analogy for the cell. The lipid bilayer of the neuron acts as a selectively permeable barrier separating different ion concentrations. Across this barrier we establish the equilibrium potentials for our various ions, just as a battery establishes a potential by separating charge. If we can create a path for the ions to flow, for example, by opening ion channels (raising their conductance), we will have a current. Anything that resists the flow of the current, be it cytoplasm, extracellular fluid, or a closed channel, can be represented by a resistor because it is preventing the current from moving freely.

Once the analogy is established between the circuit and the cell, it is important to understand which factors affect current. We can quantify the relationships between current, voltage and resistance using Ohm's Law. Because a voltage difference drives current to flow, we can intuit that the two are directly proportional. Resistors, however, restrict the flow, so it seems that the two would be inversely proportional. Indeed, we find that current (I) varies directly with voltage (E) and inversely with resistance (R):

$$\text{Equation 1: } I = \frac{E}{R}$$

As current moves through a resistor, we say that voltage is "dropped." In other words, the energy supplied by the voltage dissipates as the current travels through the resistor. This concept seems abstract when thinking about electric circuits, but considering the movement

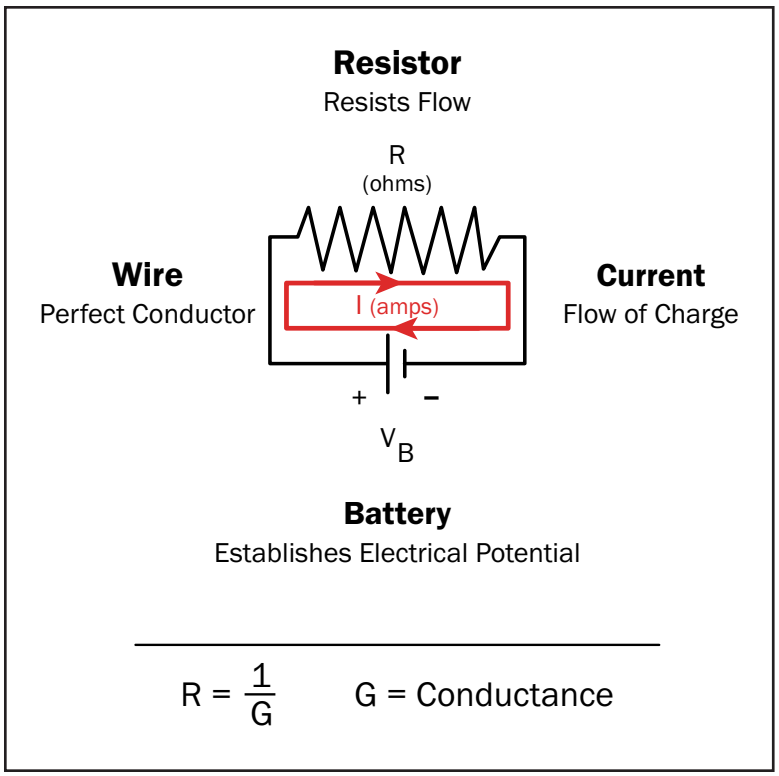


Figure 2.2

Important elements of a circuit. Current flows through the wire after the battery establishes electric potential. The resistor resists the flow of charge.

of something familiar like balls rolling down hills often feels more intuitive.

A useful analogy for thinking about voltage is to consider the potential energy resulting from gravity. While this analogy works for the circuits we will look at in this class, please note that this model should not be considered a perfect comparison for all electrical circuits and components.

Consider a ball sitting on top of a hill of height h (Figure 2.3). That ball has a gravitational potential energy equal to the work done against gravity to lift it to that height, and is indeed, directly proportional to the height, h. We can create paths for our ball to roll down the hill. If the hill magically disappeared, and the ball fell towards the Earth, we could say it encountered a path of no resistance. We could also allow the ball to travel down paths of varied steepness. For the purposes of our analogy, we will consider gradually sloping paths to be like high resistance, with very steep paths analogous to low resistance. Regardless of the path the ball takes, as the ball drops, its potential energy drops, as a function of height. So when the ball is at a height of h/2 for example, the potential energy will have dropped by half. Different paths may mean the ball takes a different amount of time to get to that height, but the potential energy, once at that height, is always the same.

As you can see, the gravitational potential energy depends on the height of the object.

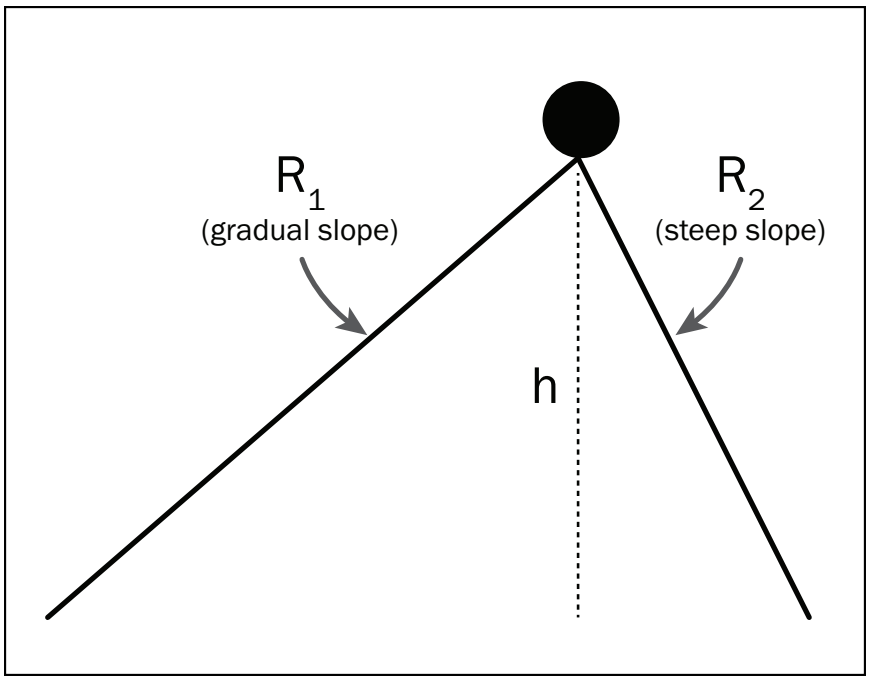


Figure 2.3

A ball sitting on top of a hill of height h. The potential energy of the ball is directly proportional to its height. Paths of different steepness model slopes of varied resistance, with steeper slopes indicating less resistance ($R_2 < R_1$).

As the potential energy is dropping, we could determine the change in height by measuring the relative distance between heights. Our potential energy is always calculated relative to some point. The initial height, h, was measured relative to the ground level. Similarly, in our electrical circuit, we will always calculate the voltage drop between two points. Sometimes, we calculate the voltage relative to "ground" which is the technical name given to our reference point. This can be a physical connection to the Earth, or some other reference, depending upon our recording setup.

Equivalent Circuits, Intracellular Recordings and Synaptic Potentials

We can apply these models to help us understand how current behaves during synaptic potentials. During an excitatory post-synaptic potential (EPSP), current flows into the cell (Figure 2.4). The current encounters resistance from extracellular fluid, cytoplasm and the cell membrane, so resistors are drawn in to represent each of these. A battery is drawn to represent the equilibrium potential created by the different ionic concentrations inside and outside of the cell, and we can think of the location of the battery as the site of our open ion channel, where current will move freely into the cell.

Note that we have the highest concentration of current flow across the cell membrane

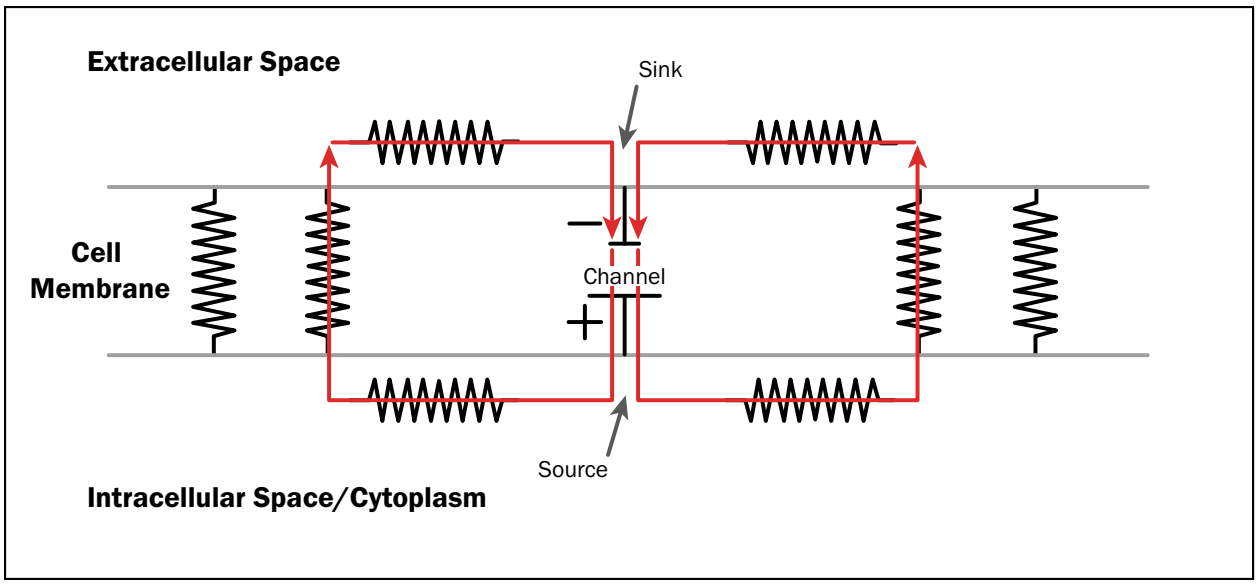


Figure 2.4

Flow of current during synaptic potentials. Across the ion channel, the concentrated current flows away from the sink and towards the source. To complete the circuit, the current can flow diffusely away from sources and towards sinks. Current flow during EPSP's and IPSP's can be recorded using the intracellular technique but not the extracellular technique because the voltage changes are too subtle for an extracellular electrode to sense.

from extracellular sink to intracellular source. In both extracellular and intracellular space, the ions flow in a more diffuse manner to complete the circuit. If the current is coming toward the electrode in a concentrated fashion, you are at the source. If the current is moving away from the electrode in a concentrated fashion, you are at the sink. Importantly, the electrode will record a positive-going deflection (as viewed on an oscilloscope) at the source, and when the electrode is placed at a sink, a negative-going deflection will be recorded.

Synaptic potentials from individual neurons can only be recorded via intracellular recordings because the change in the ionic environment is too subtle for an extracellular recording to pick up. When recording, researchers place the intracellular electrode near a channel. If current flows into the cell (an EPSP), the electrode location is a source, and the recording will be positive-going. If current flows out of the cell, the electrode resides at a sink, and the deflection will be negative.

Current flow during an inhibitory post-synaptic potential (IPSP) looks the same as in an EPSP, but the arrows go in the opposite direction. This is easy to understand in the case of IPSPs mediated by the outward flow of K^+, but also holds true for IPSPs mediated by the inward flow of Cl^-. Recall that negative current, carried by anions, flows in the opposite direction from the positive current carried by cations.

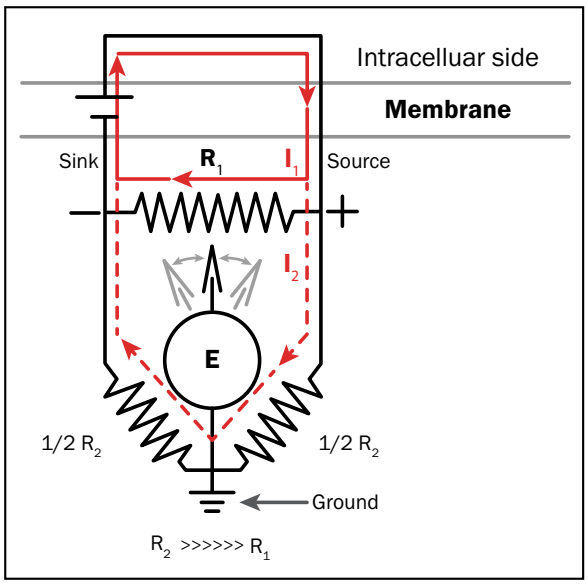


Figure 2.5a

Setup for extracellular recordings. Extracellular recordings require an electrode, which can sample the voltage at different locations across the membrane. This figure represents one electrode in three locations, not three electrodes.

The Equivalent Circuit, Action Potentials and Extracellular Recordings

Understanding the equivalent circuit informs the way we interpret extracellular recordings. As mentioned in Chapter 1, sometimes an extracellular recording will yield an initial upward deflection, while other times it yields an initial downward deflection when viewed on an oscilloscope. The equivalent circuit, as well as our discussion of source and sink, helps to understand this phenomenon.

Note that in this section we are only dealing with the recording of action potentials. Single-unit extracellular recordings are not sensitive to synaptic potentials because EPSPs and IPSPs do not create big enough changes in the ionic environment for the electrode to pick up. However, this type of recording is an important way to study action potentials.

Figure 2.1 demonstrates the flow of current into an axon.The initiation of the action potential is at the sink. This is where the highest concentration of ions flows into the cell. In both extracellular and intracellular space, the flow of the ions is more diffuse.

When recording extracellularly, the shape of the action potential depends on the electrode's position relative to the source and the sink. Figure 2.5a demonstrates the experimental set up for such recordings.

To understand this setup, it's helpful to understand how current behaves in various circuits. The circle with an "E" represents a voltmeter. This machine is used to measure the voltage difference, E. The voltage difference is literally a difference – so experimentally we find our value by recording the electrical potential relative to another point. In our setup above, we have indicated that point as "ground." (Ground is symbolized by three lines of decreasing lengths). Ground is a reference point, and in the case of neural recordings, we often get our

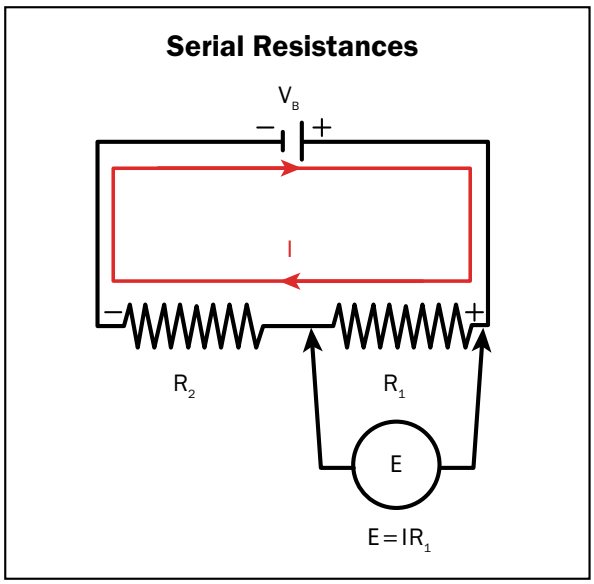


Figure 2.6

Resistors in series. When multiple resistors reside on the same current path, they are said to be "in series." When calculating overall resistance, their resistances (in Ohms) is added arithmetically because the current will encounter both resistive elements.

reference value by placing an electrode in a brain area far away from the neuron of interest. To find the voltage difference we will apply Ohm's Law. Because the setup is somewhat complex, we will break it down into components.

Resistors in Series

The experimental setup has resistors arranged in series and in parallel. Figure 2.6 shows a circuit with two resistors in series. To be in series means the resistors are in same current path, and thus current must flow through both resistors to return to the battery.

We can use our gravitational potential model to think about resistors in series (Figure 2.7a). This would just be two different slopes in our path down the hill. Similarly to how our voltmeter is measuring just the voltage drop across resistor R_1 in figure 2.6, we could measure just the drop in potential energy once the ball reached the end of the uppermost slope.

When resistors are in series, the resistance of the system is simply the arithmetic sum. In the case of the figure,

$$\text{Equation 2: Total Resistance} = R_1 + R_2$$

Given Ohm's Law, for the entire circuit,

$$\text{Equation 3: } I = \frac{E_{tot}}{R} = \frac{V_b}{R_1 + R_2}$$

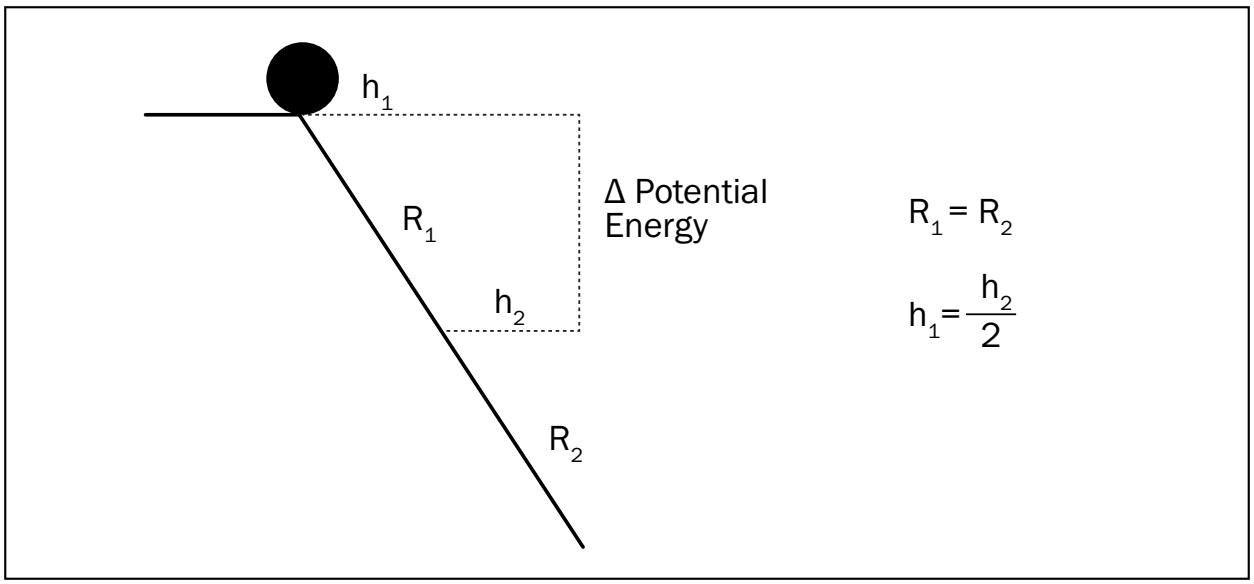


Figure 2.7

We can measure the change in potential energy from the top of the hill, h_1, to the middle of the hill, h_2. When the slope is constant, we are measuring the potential energy change from the top of the hill to halfway up the hill. This decrease of half the height would drop our potential energy by half.

We can use Ohm's Law again to create a system of equations and solve for E (measuring just across R_1). We start again with

$$\text{Equation 4: } I = \frac{E}{R_1}$$

This, however, is not in terms of our battery voltage. To get this into those terms, we can just set our two equations for current (3 and 4) equal to each other.

$$\text{Equation 5: } \frac{E}{R_1} = \frac{V_b}{R_1 + R_2}$$

Now, let's consider the case where R_1 and R_2 are equal. If $R_1 = R_2$, then we can substitute in R_1 for R_2:

$$E = \frac{V_b R_1}{2R_1}$$

$$\text{Equation 6: } E = \frac{V_b}{2}$$

Similarly, we can see this in our gravitational potential model (Figure 2.7). If our two

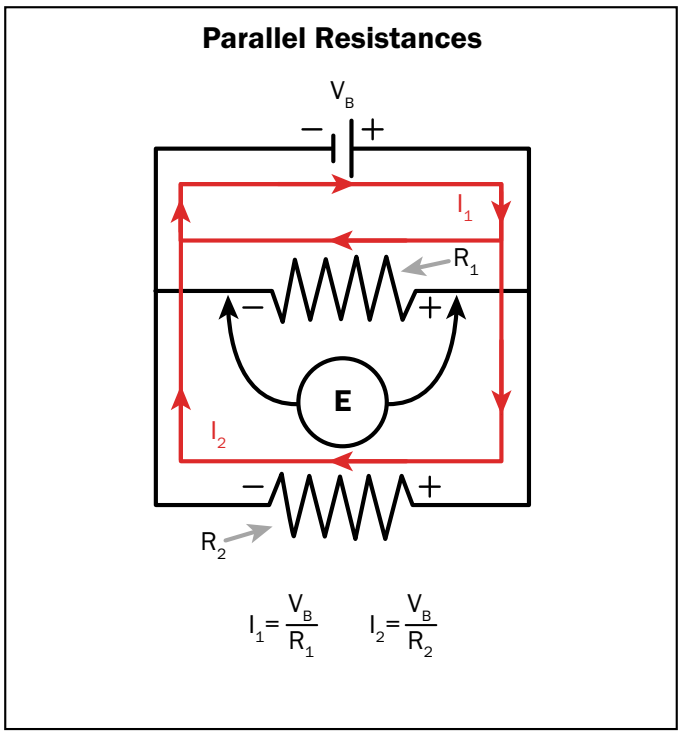


Figure 2.8

Resistors in parallel. When resistors reside on different current paths of the same circuit, they are said to be "in parallel." When calculating overall resistance, their resistances (in Ohms) are not added arithmetically. Because the current can go through either path, you cannot simply add the resistance values of the two resistors. Instead, The total resistance is found by summing the reciprocals of each resistor and then taking the reciprocal of this value.

slopes were equal, we'd be measuring the potential energy change from the top of the hill to halfway up the hill. This decrease of half the height would drop our potential energy by half.

Now let's see how this information relates to the experimental set up for extracellular recordings. (Figure 2.5a)

In this setup, we can see that there is a resistor, divided in half into $1/2R_2$ and $1/2R_2$. Because the electrode divides R_2 into two "separate" resistors, we can think of them as being in series. Based on the reasoning above, the voltage drop across each resistor is $V_b/2$.

We can also see that there is another resistor labeled R_1. To understand how current runs through this resistor, and how that movement relates to the R_2 resistors, we need to investigate how current behaves when there are resistors in parallel.

Resistors in Parallel

Current always flows from the positive terminal of the battery to the negative terminal. When resistors are set up in parallel, there are two possible paths for the current to take. In

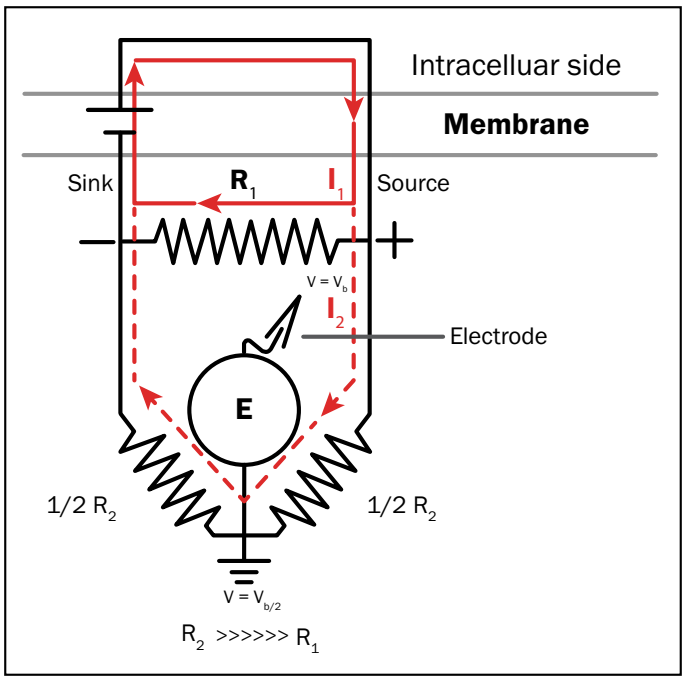


Figure 2.5b

Extracellular recordings with the electrode by the source. Here, the electrode is placed by the source. At this location, the voltage is equal to the voltage of the battery. When the electrode takes the difference between this value and the ground, whose voltage is half that of the battery, it will read +V/2.

Figure 2.8 we see that some of the current (I_1) goes through R_1, and some of the current (I_2) goes through R_2. Most of the current will go through the path of least resistance, so if R_1 is less than R_2, then I_1 will be greater than I_2. Resistance is also greater in paths that require current to travel farther, simply due to the fact that there is more wire to travel through. So long paths will have less current, as well. (Unless your wires are perfect conductors. But we're actually talking about biology here, so nothing is perfect!)

Importantly, even though the current through each path is different, the voltage drop across each path is the same. This is easiest to understand by considering our gravitational potential model (Figure 2.3). The ball starts at height h, and can choose to roll down multiple paths. Regardless of the path the ball takes to get to the bottom of the hill, the drop in potential energy from the top to the bottom is always the same.

Now let's apply this concept to the experimental setup.

Calculating Voltmeter Readings

When an action potential occurs, current will initially flow into the cell at the sink and out of the cell at the source. The voltmeter (E) reads the voltage difference between the electrode's

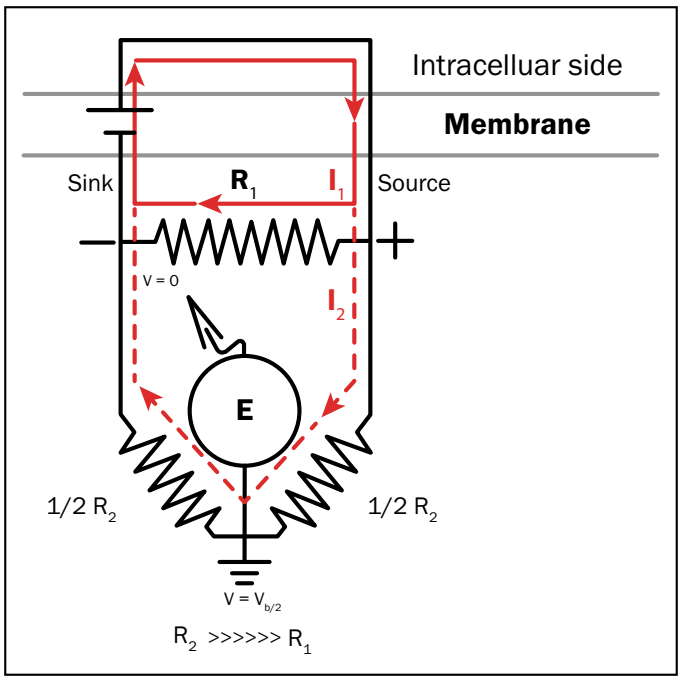


Figure 2.5c

Extracellular recordings with the electrode by the sink. Here, the electrode is place by the sink. At this location, the voltage is zero because all of it has dropped. When the electrode takes the difference between this value and the ground, whose voltage is half that of the batter, it will read -V/2.

location and the ground. In Figure 2.5b, the electrode is near the source.

To calculate what the voltmeter will read, we must find the voltage at the electrode and subtract the voltage at the voltmeter's ground. Assuming the total voltage drop is V_b, we know that the voltage at the source is V_b because the voltage has not yet dropped over the resistor. Also, based on our above calculations of resistors in series, we know that half of the voltage is dropped over each R_2. Because half of the voltage is dropped over the first R_2 resistor, the voltage at the ground will be $V = V_b/2$. Thus, the voltmeter will calculate:

$$\text{Equation 7: } E = V_{electrode} - V_{ground} = V_b - \frac{V_b}{2} = \frac{V_b}{2}$$

This equation yields the result that **when the electrode is near the source, the voltmeter will display a positive reading.**

Now we will examine what happens when the electrode is at the sink. When the electrode is near the sink, all of the voltage had dropped. This means that the voltage at this location will be $V = 0$. Meanwhile, the voltage at the ground is still the same, $V_b/2$. Thus, the voltmeter will calculate

$$\text{Equation 8: } E = V_{electrode} - V_{ground} = 0 - \frac{V_b}{2} = -\frac{V_b}{2}$$

This equation yields the result that **when the electrode is near the sink, the voltmeter will display a negative reading**.

What happens when the electrode is exactly in the middle of R_1? We can apply the same ideas we've seen before to this situation. We know that over R_1 the voltage will go from V_b to 0. Halfway between these values is $V_b/2$. When we calculate the voltage difference the voltmeter would read, we find:

$$\text{Equation 9: } E = V_{electrode} - V_{ground} = \frac{V_b}{2} - \frac{V_b}{2} = 0$$

Thus, **when the electrode is exactly between the source and sink, the voltmeter will show no deflection at all**.

Other Aspects of Reading Extracellular Recordings

The amplitudes of the potentials recorded extracellularly are related to the amount of current flowing through the extracellular recording resistance, R_1, and are never as large as those recorded intracellularly. Also, the amplitudes tend to be larger near the cell than far from it, because the closer the electrode is to the cell, the shorter the distance the current must travel through extracellular space. Shorter distances mean less resistance and larger potentials.

Generally, currents produced by more distant cells will have little effect on the recording, but the action potentials (actually the action currents) produced by several nearby cells may be detected with a single extracellular electrode. When a spike can clearly be attributed to a single cell, it is said to be well isolated. A good criterion for this is that the recorded action potential has a constant size and shape.

It is very unlikely that the extracellularly recorded action potentials of two adjacent cells would have identical profiles because of the geometric factors illustrated in Figure 2.5. Sometimes it is useful to use relatively blunt electrodes to record the extracellular currents generated by a population of neighboring cells. Such multiunit recordings often provide sufficient resolution to do mapping experiments like those we will see throughout the course.

Intensity Coding in Sensory Systems

3

Table of Contents

As stimulus intensity increases, so does the magnitude of the perception, so a major goal of sensory research is to relate stimulus intensity to neural activity, and the latter to perception. The general strategies used by sensory systems to encode increased stimulus intensity are ***increased firing rate*** in the responding neurons and ***recruitment*** of additional neurons to the responding population. Historically, the study of sensory processes began by correlating the perception of stimuli with their physical characteristics. This approach is the domain of *psychophysics*.

Signaling Sensory Magnitude: What has to be accounted for?

In 1852 E.H. Weber handed his subjects a series of objects of different weights and asked them to report when two objects differed in weight. He noted that the ***just-noticeable-difference*** (JND) in weight between a test and a comparison object was proportional to the absolute weight of the comparison object. On this basis he proposed what has come to be known as ***Weber's law***: $\Delta I / I = \text{constant}$, where I = baseline intensity and ΔI = JND (Figure 3.1).

Elaborating on this in 1860, the physicist G. Fechner assumed that a JND in stimulus magnitude corresponded to a constant increment or decrement in some perceptual dimension, which he called Psi. Thus, even though the JND was increasing, the nervous system only took notice of the change after it had exceeded a certain constant amount in the perceptual system (Figure 3.1).

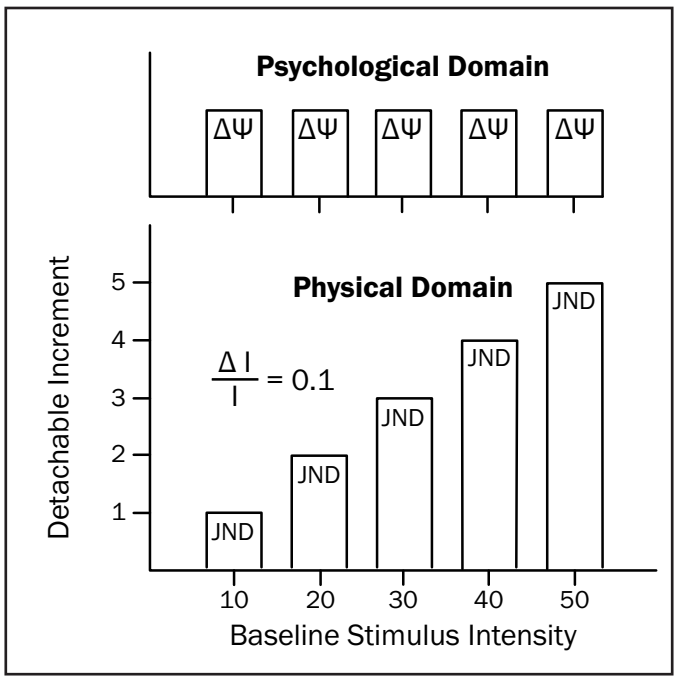


Figure 3.1

Schematic relationship of proportionality between JND and baseline stimulus intensity in physical terms (Weber's law) and Fechner's conception of the constancy of just noticeable differences in the perceptual or psychological dimension. The proportionality constant of 0.1 between ΔI and I is by way of example.

Observe that the process is accessible only to the subject, who must report its magnitude through some kind of behavior, verbal or otherwise. In other words, Fechner hypothesized that the sensation of a JND always corresponded to an equivalent psychological increment (even though, as we have already stated, the absolute physical intensity of that increment depends on the baseline).

Based on Weber's assertion that $\Delta I/I$ is constant and assuming that $\Delta\psi$ for any JND is also a constant, Fechner proposed what has come to be called the ***Weber-Fechner law***:

$$\Delta\psi = \frac{K\Delta I}{I} = \text{constant}$$

Let's Integrate: At threshold (I_0), $\psi = 0$, so $Q = -K \ln (I_0)$

$$\text{So, } \psi = K \ln(I) - K \ln(I_0) = K \ln(\frac{I}{I_0})$$

While these equations may seem complicated, you must keep in mind that logarithms are not difficult! Logarithmic scales can seem unintuitive, but they actually help us understand behavior over extremely large ranges. For example, going from a power of 0 to a power of

80 spans the range between one single item to the all of the molecules in the entire universe. This type of is scale useful when thinking about processes like hearing, because the ear can detect sound anywhere from $1 \times 10^{-12} \, W/m^2$ (the threshold intensity of hearing) to $1 \times 10^{4} \, W/m^2$ (the intensity at which the eardrum spontaneously perforates). To better grasp this change by a power of 16, we use the logarithmic decibel scale and characterize the range as 0 to 160 dB.

To understand a logarithmic graph it helps to change how we think about the axes. For instance, say we are measuring change in sound intensity over time. If the Y-axis is logarithmic, we can think that it is counting not the change in units of W/m^2 but rather the change in ***multiplications*** or ***digits*** in W/m^2. Going from 100 W/m^2 to 1000 W/m^2 is a change from 120 to 130 dB. This change connotes one multiplication by 10. If the data forms a line on a logarithmic graph, then the data are not linear, but the ***number of digits*** of the data is linear.

The Weber-Fechner law was a helpful tool for psychophysics because sensory experiences vary over very large ranges. His derivation led to the remarkable conclusion that the subjective, or psychological, experience of any stimulus can be precisely quantified in terms of its physical intensity. Despite the seeming universality of this claim, the Weber-Fechner law has been shown to be true only under highly restricted conditions, if at all.

In a lifelong pursuit of these questions, S.S. Stevens more directly measured the psychological magnitude of stimulation in a variety of modalities by asking subjects to assign numbers to the strength of various stimuli. Data from these experiments in ***magnitude estimation*** usually give straight lines when plotted on log-log axes, meaning that they are best described by power functions. Thus, Stevens proposed that an appropriate way to characterize the relationship between the physical and psychological sense of magnitude was his ***power law of sensory perception***:

$$\log \psi = K \log (I - I_0), \text{ so } \psi = (I - I_0)^K$$

This law states that the perceived magnitude of a stimulus is a function of the intensity of the stimulus, I (less a threshold, I_0), raised to some power, K. Different kinds of stimuli are characterized by different exponents. Some typical exponent values are given in Table 3.1.

The exponents give the slope of the function if plotted in log-log coordinates (Figure 3.2). If the exponent is greater than one, the function is said to be ***expansive*** or ***accelerating***, as revealed when such a function is plotted in linear coordinates (Figure 3.2). In systems with accelerating functions, constant increments in stimulus intensity yield increasingly greater increments in the magnitude of the perception. This makes sense for the representation of, for example, painful stimuli because the capacity of such stimuli to cause tissue damage increases with the intensity of the stimulus so a more vigorous behavioral response is required.

If the exponent is less than one, the function is ***compressive*** or ***decelerating*** (e.g. brightness). Sensory mechanisms with compressive functions occur in modalities in which we make discriminations over a large range of stimulus intensities, such as vision and audition. In

Stimulus	Exponent
Loudness (binaural)	0.6
Brightness (dark-adapted eye)	0.33
Temperature (warm on arm)	1.6
Taste (salt)	1.3
Smell (coffee odor)	0.55
Electric shock (to fingers)	3.5

Table 3.1

Some representative exponents for Stevens' power law model of several sensory perceptions. The exponents are values of K in the equation above and are the slopes of the functions plotted in log-log coordinates.

these domains, the systems are sensitive to small changes in weak stimuli but detect only larger changes in intense stimuli. For instance, when we are in a dark room, we can sense a tiny pinprick of light, but when we are out in the sun that bit of light would be imperceptible.

When the slope or exponent is one, the function plots as a straight line in linear coordinates (not shown in Figure 3.2B).

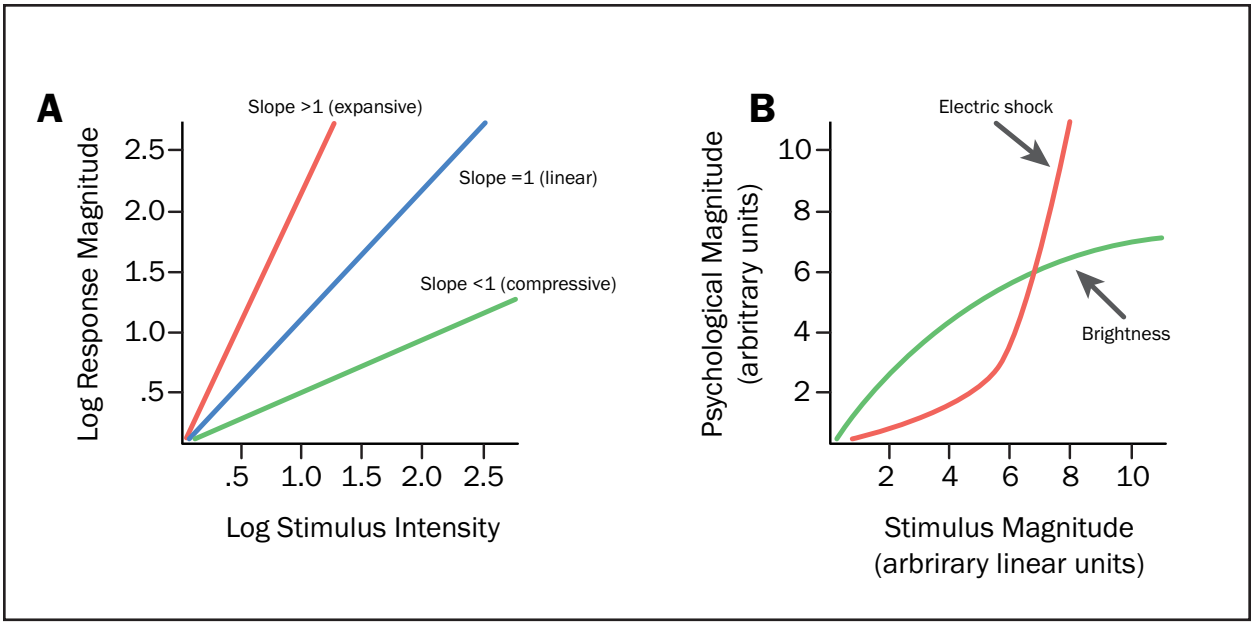


Figure 3.2

Relationship between slope of log-log plot and compressive/expansive psychological magnitude functions. B. Psychological magnitude functions plotted on linear axes for electric shock (expansive) and brightness (compressive). (Data from Stevens in *Sensory Communication*, ed. A. Rosenblith, MIT/Wiley, New York, 1961.)

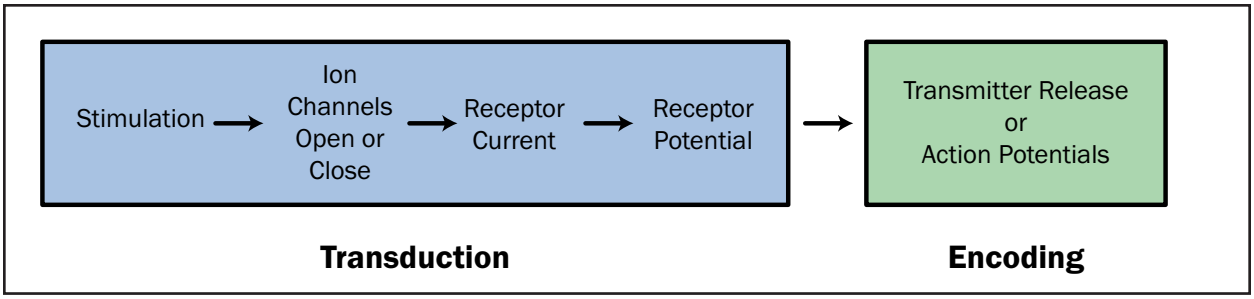


Figure 3.3

Schematic representation of the processes of transduction and encoding.

Transduction and encoding by sensory receptors

Neurophysiological studies of sensory systems reveal that sensory receptors make a substantial contribution to the quantitative properties of a particular modality revealed by psychophysics. The receptor stage of a sensory process may usefully be divided into the steps of ***transduction*** and ***encoding*** (Figure 3.3). Transduction is the conversion of stimulus

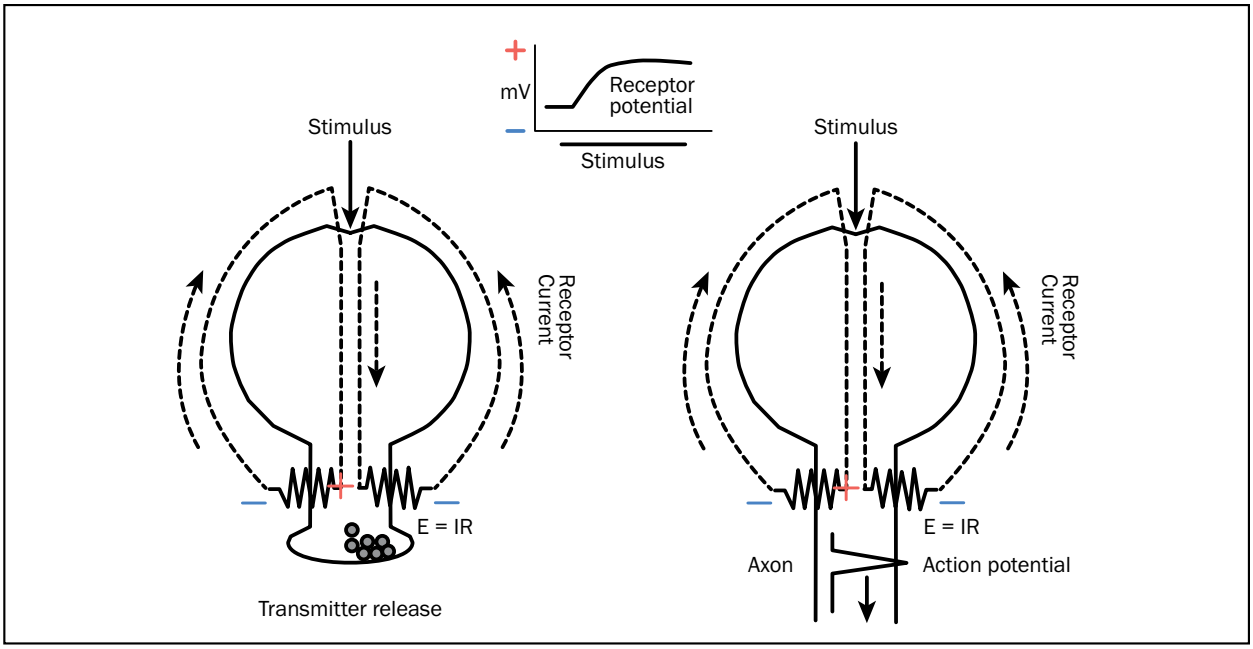


Figure 3.4

Transduction and encoding by sensory receptors. Generation of the receptor potential and encoding of transmitter release (left) or action potential generation (right). Inset: depolarizing receptor potential.

energy into a change in membrane potential. For example, when skin is touched, the message is transduced by a depolarization of mechanoreceptors in the skin. All transduction involves some type of modification of a protein in the membrane: an ion passes through a channel, or some second messenger system operates to open or close ion channels. The current flowing through these channels produces a change in membrane potential called the ***generator potential*** or ***receptor potential*** (Figure 3.4). In most systems stimuli depolarize the receptor, but exceptions exist: recall that vertebrate photoreceptors are hyperpolarized by light. It is evident that the quantitative properties of the transduction process will have a major impact on the way information about the stimulus is represented in the neural signal.

Depending on the receptor involved, the receptor potential may directly modulate the release of a neurotransmitter, as it does in non-spiking photoreceptors and auditory receptors, or it may modulate the rate of action potential generation, as it does in cutaneous mechanoreceptors. This encoding step provides another stage at which the quantitative characteristics of the neural representation may be determined. The properties of the synaptic release mechanism or the voltage-dependent ion channels responsible for the generation of action potentials affect the perception of a stimulus.

Operating characteristic of a receptor

If one plots the discharge frequency of a spike-generating receptor against stimulus intensity, one obtains a curve that describes the conversion of stimulus energy to a neural

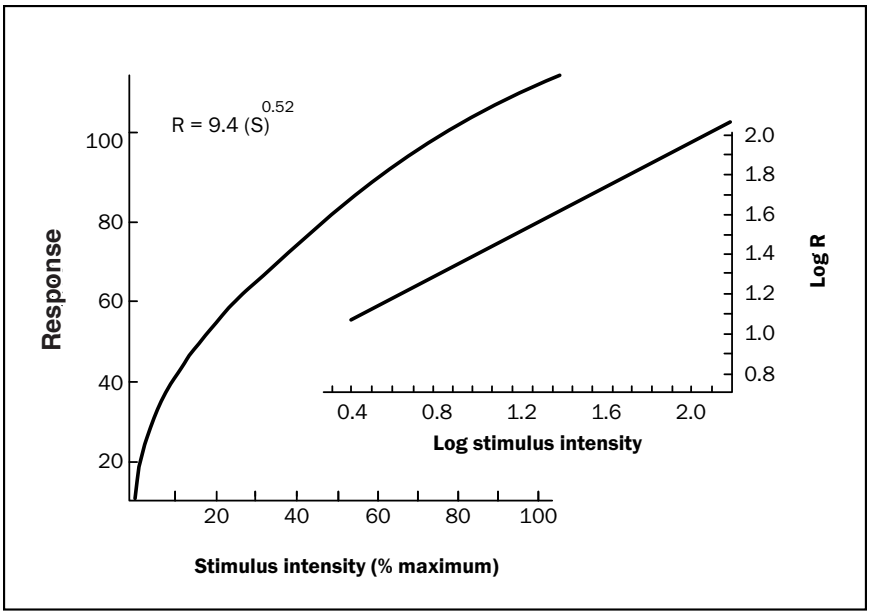


Figure 3.5

Stimulus-response relationship of a peripheral mechanoreceptor plotted in linear and logarithmic coordinates. (Data from Werner and Mountcastle, ***J. Neurophysiol.*** 28:359-397, 1965).

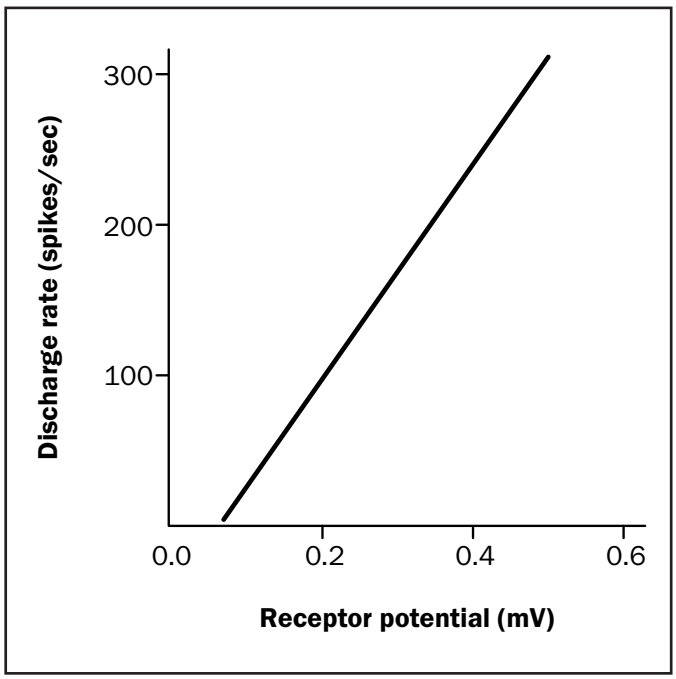


Figure 3.6

Hypothetical relationship between receptor potential and impulse frequency in a cutaneous mechanoreceptor. If this relationship were shown to be linear then the compressive relationship shown in Figure 3.5 must be due to a compressive transduction stage (i.e. in the relation between stimulus intensity and receptor potential).

representation of that energy at the level of transduction and encoding. This is the ***operating characteristic*** of the receptor. Figure 3.5 shows an example from the somatic sensory system. Note that this mechanoreceptor has a compressive operating characteristic (slope < 1) and that the transformation can be fit by a power function, as indicated by the straight-line plot in log-log coordinates.

The curves of Figure 3.5 incorporate both transduction and encoding stages of the receptor process so it is not clear at which stage the compressive characteristic occurs. This question could be answered by comparing the receptor potential to the rate of action potential discharge. For example, if the peripheral mechanoreceptor of Figure 3.5 were found to exhibit a linear relationship between receptor potential and impulse frequency, as in Figure 3.6, then we could conclude that the compression must occur at the transduction stage, not at the encoding stage.

Figure 3.7 identifies certain critical features of the operating characteristic of a receptor. The ***absolute threshold*** is the stimulus intensity that reliably evokes a response from the receptor. Because a weak stimulus may not evoke responses on every presentation, 'threshold' is actually a probabilistic parameter. Investigators therefore usually set some criterion for a 'reliable' response, such as an observable response on 50% of trials. There are more sophisticated approaches, which describe the receptor's behavior at these low stimulus levels while accounting for the statistical fluctuations of the response.

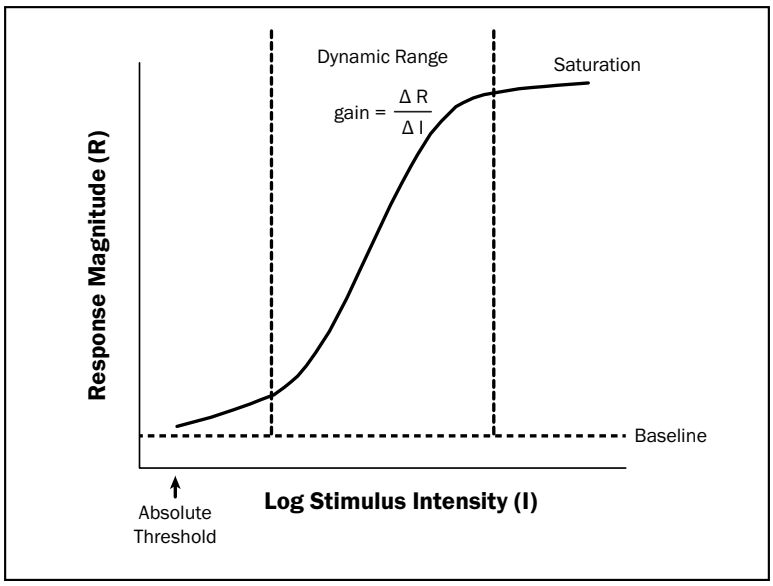


Figure 3.7

Schematic illustration of the operating characteristic of a receptor. Δ = 'change in'.

The *dynamic range* of the receptor is that region of stimulus intensities over which the receptor will change its output when the stimulus intensity changes. Even though the cell may respond at very high stimulus intensities, if it cannot modulate its response it cannot serve a discriminatory function for the rest of the nervous system. When the receptor no longer gives a response for a more intense stimulus, it is considered *saturated*.

A concept of great importance in discussing the behavior of neural systems is that of *gain*. *The gain of a system is the ratio of its output to its input*. Gain can be expressed in various ways depending on what is most useful for a particular situation. In Figure 3.7, the gain of the receptor is defined as the change in output for a given change in input. If this were a mechanoreceptor, the gain might be expressed as (change in firing rate)/(mm distortion). Note that the gain of the receptor of Figure 3.7 differs in different parts of its operating range. Also, the abscissa has a logarithmic scale, which has the effect of graphically shrinking the scale of stimulus intensity as intensity increases, so the gain actually changes even when the slope does not.

Receptor adaptation

Receptors are normally subject to background stimulation, which can modify their responsiveness. A change in the gain of a receptor produced by background stimulation is called *adaptation* and is dependent on both the intensity and length of exposure to the adapting

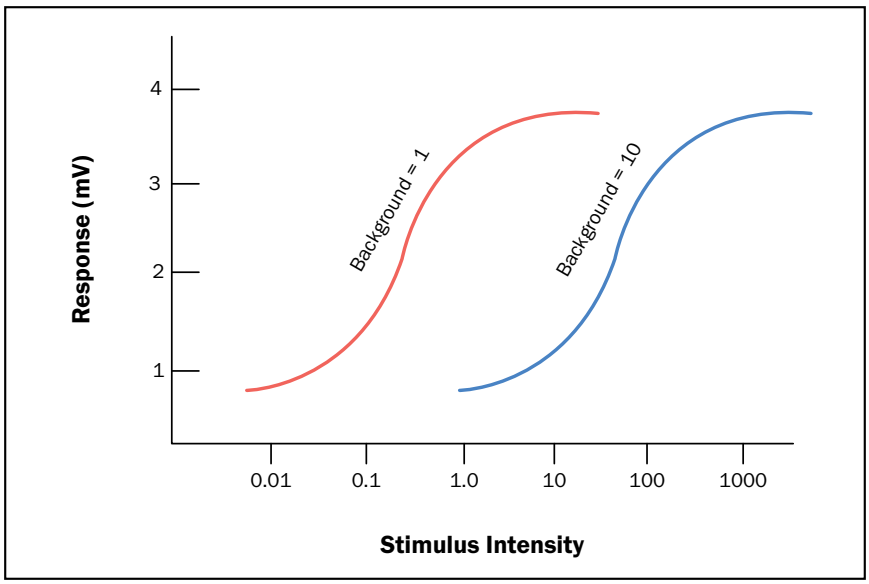


Figure 3.8

Effect of adaptation on the operating characteristic of a receptor. The receptor is first exposed to a constant background and the responses to stimuli of increasing intensity are then measured. The background intensity is increased and the measurements repeated. The rightward shift of the operating characteristic at increasing background levels indicates adaptation.

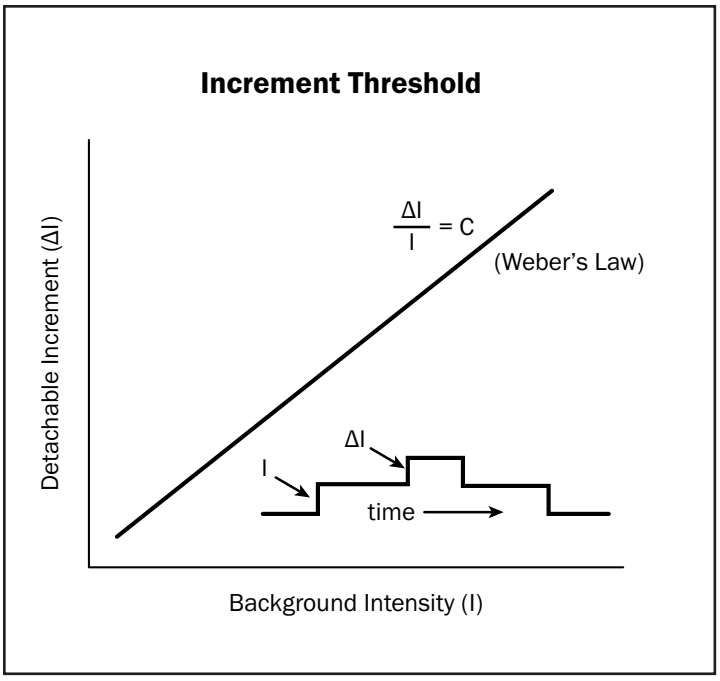


Figure 3.9

Schematic illustration of an increment threshold curve. The increment threshold is clearly related to the notion of Just Noticeable Difference. The function illustrated here obeys Weber's Law, but this need not always be the case.

stimulus. Adaptation can be studied by determining the operating characteristic of a receptor or system at different levels of background stimulation. Figure 3.8 illustrates schematically what one might see if the operating characteristic (see Figure 3.7) of a photoreceptor were measured at two different levels of background light. At the higher background level, the

operating characteristic is shifted to the right. Thus, to evoke a response of 2 mV, for example, a stronger stimulus is needed at the higher background level. The receptor has adapted.

Note that these curves retain their shape as they shift rightwards. Because the axis is logarithmic, the shift means that the dynamic range of the receptor increases at higher levels of adaptation, while its absolute sensitivity decreases. In other words, the receptor will be able to respond to a greater range of luminance, but it will be less able to detect a given change in luminance.

Another way in which the effect of adaptation is studied is by determining the ***increment threshold*** at set levels of background intensity (Figure 3.9). In this method, a constant background is turned on, and at some point a small probe stimulus representing an increment (or decrement) of intensity, I, is applied. The value of this increment that evokes a reliable response from the receptor is the increment threshold. When there is zero background, the increment threshold has the same value as the absolute threshold.

The increment threshold is clearly related to the JND, as described in the various 'laws' of psychophysics. For systems with compressive characteristics, one would expect the increment threshold to rise as the background intensity rises (Figure 3.9). Note that the difference in the experiments of Figure 3.8 and Figure 3.9 is that the former is measuring the increase in response to increases in stimulus intensity, whereas the latter is measuring the changes in threshold at different levels of adaptation.

Mechanisms of receptor adaptation

Receptor adaptation can usually be traced to a mechanical and/or biochemical mechanism. A Pacinian corpuscle provides a good example of mechanical adaptation. Pacininian corpuscles contain onion-like wrappings around their endings. When these fluid-filled wrappings are compressed by a probe, a deformation of the stretch-sensitive channels of the underlying nerve creates a receptor potential. If the compression persists, fluid eases back in between the lamellae, and the deformation of the nerve is relieved.

An excellent example of molecular or biochemical adaptation is provided by photoreceptors. The transduction process in these cells is based on an enzymatic cascade, which begins when a photopigment called rhodopsin absorbs a quantum of light (Figure 3.10). In the dark, high levels of intracellular cyclic GMP (cGMP) keep cation channels in the open state, allowing a large current (the ***dark current***) to flow into the cell. This current is carried by sodium and calcium ions. When there is light, however, the enzymatic cascade decreases the level of intracellular cGMP. When there is less cGMP, cation channels that are normally open close, and the membrane hyperpolarizes. This hyperpolarization results in a decrease in synaptic transmitter release onto the bipolar and horizontal cells, which are the next neurons in the

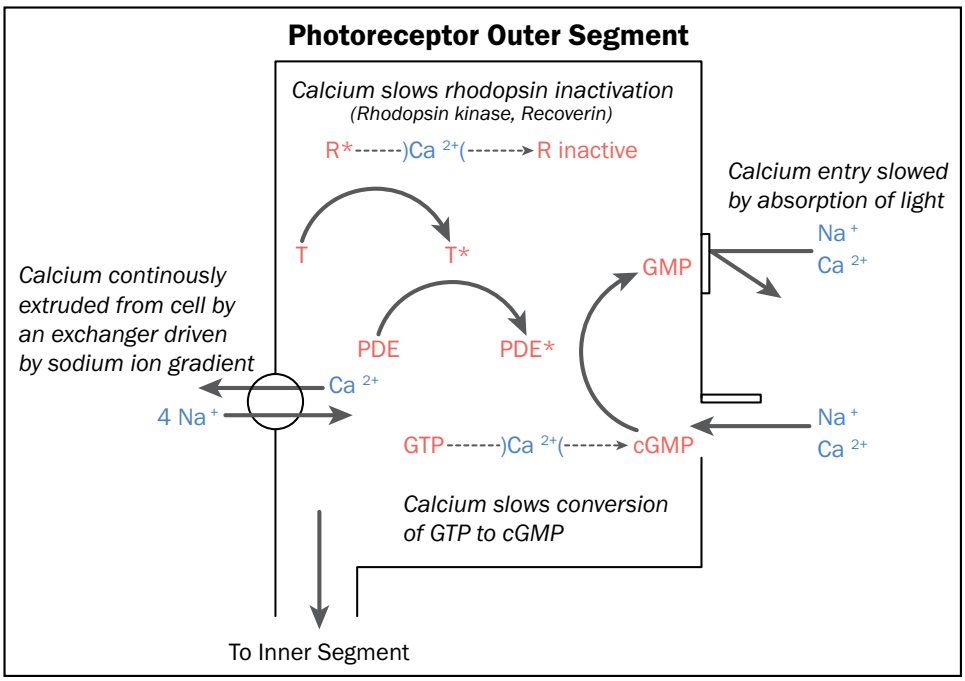


Figure 3.10

The enzymatic cascade of phototransduction. A quantum of light activates a molecule of rhodopsin (R*) which converts many molecules of the G-protein transducin (T) to the active form (T*). Each T* activates a phosphodiesterase (PDE -->PDE*), and each PDE* converts many cGMP molecules to GMP. The reduction of [cGMP] closes the cation channels and reduces the inward flow of positive ions, hyperpolarizing the photoreceptor and slowing transmitter release. Ca^{2+} slows the rate at which R* is inactivated and GTP is converted to cGMP. Reduction of intracellular Ca^{2+} has the effect of decreasing the gain or sensitivity of the photoreceptor. See text.

retinal circuit. Thus, the vertebrate photoreceptor is actually inhibited by light. This mechanism is important in photoreceptor gain control.

Because humans can see over vastly different levels of light, there must be a way for the eye to control the gain of photoreceptors. When levels of light are low, the eye must have a large response to small increases of light. When light levels are high, it must adjust the gain by not responding so strongly to a small stimulus. One definition of the gain of the photoreceptor is the change in cGMP concentration caused by the absorption of a single quantum, i.e. $\Delta[cGMP]/quantum$. The cGMP concentration at any time is a function of the rate at which it is produced from GTP by the enzyme guanylyl cyclase and the rate at which it is removed by conversion to GMP by the activated form of the enzyme phosphodiesterase (PDE*) (Figure 3.10). Both of these processes are sensitive to the intracellular concentration of calcium ions, which is reduced by light. When cGMP concentration decreases and the cation channels close, the rate at which Ca^{2+} enters the cell decreases. At the same time Ca^{2+} continues to

be extruded from the cell by a cation exchanger that is unaffected by light. This reduction in intracellular calcium concentration has two main effects that decrease the ability of a quantum of light to lower the cGMP concentration.

The process of gain control can be summarized as follows:

- Quantum absorbed reduces intracellular cGMP via the enzymatic cascade
- Calcium (and sodium) entry slows
- Calcium extrusion continues though the Na^{+}/Ca^{2+} exchanger
- Intracellular calcium concentration drops
- Effect 1
    - Rhodopsin inactivation accelerates; active life of rhodopsin shortens
    - Thus, fewer Transducins and fewer PDEs are activated per quantum
- Effect 2
    - Calcium inhibition of guanylyl cyclase is reduced; cGMP synthesis accelerates
    - Thus, absorbed quantum now convert fewer cGMP to GMP and these are quickly replaced
- Subsequent quantal absorptions have less effect on cGMP concentration; *[cGMP]/quantum is decreased as are the change in membrane potential and the change in transmitter release per quantum absorbed.*

Terms and Techniques

absolute threshold	increment threshold	JND
transduction	encoding	adaptation
Weber's law	Weber-Fechner law	Steven's power law
expansive function	compressive function	operating characteristic
dynamic range	sensitivity	receptor gain
receptor potential	saturation	cGMP
G-protein	second messenger	enzymatic cascade
transducin	PDE	opsin
11-cis retinal	rhodopsin	dark current

Spatial Coding: Receptive Fields and Tactile Discrimination

4

Table of Contents

The nervous system maintains distinctions between different kinds of information by putting that information in different places. The most obvious example is the initial compartmentalization of the sensory modalities, each being assigned its own lobe in the brain. The same basic strategy applies to those sensory modalities that represent some continuous dimension of a stimulus as a spatial map in neural tissue. Such representations can be called ***topographic*** and are found in the somatic sensory, visual and auditory systems, where the geometric relations of the sensory surfaces are preserved in the central pathways. Neighboring points on the sensory surface (skin, retina, organ of Corti) are represented at neighboring points in the central maps. Even the motor system exhibits this topographic orderliness to some degree in cortical area 4 and elsewhere.

The discriminative capacity of these systems in the spatial domain is intimately related to the characteristics of their topographic maps. Olfaction appears to be ***non-topographic***, as there is no evidence in this system that a receptor surface or some chemical dimension is mapped continuously onto nervous tissue. The gustatory system has now become an area of controversy, as researchers are now beginning to see the possibility for some topography in the cortex! We will begin our discussion of spatial coding in topographic modalities by examining cutaneous mechanoreception.

Figure 4.1 summarizes the two major somatic sensory pathways arising from the spinal cord. Equivalent systems exist for the head and enter the brainstem through the trigeminal nerve. The dorsal column-medial lemniscal system (DC/ML, left) mediates conscious perception of discriminative touch and position sense (left). It is the neural substrate of discriminative touch that is critical for the perception and analysis of surface texture. Its major inputs are from

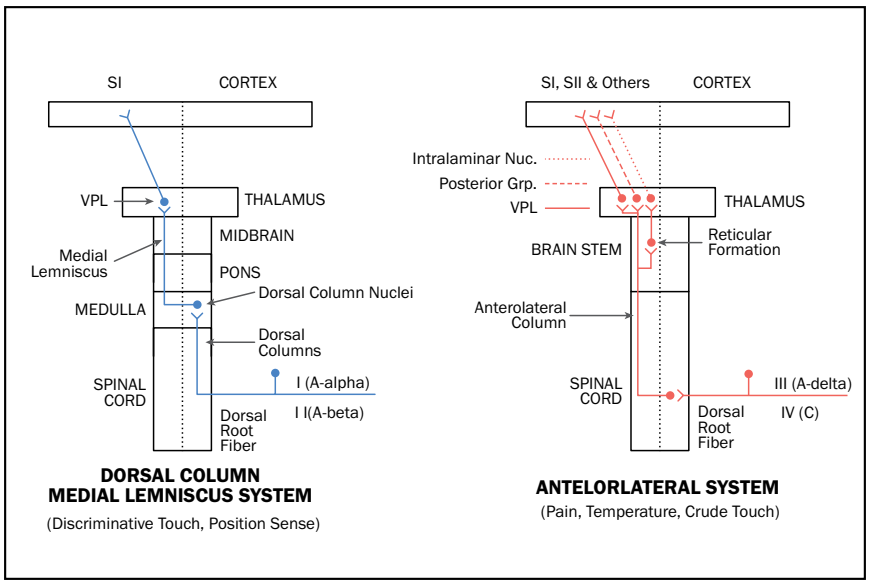


Figure 4.1

The two major pathways of the somatic sensory system. Left. Schema of the dorsal column-medial lemniscus system. Right. The anterolateral (spinothalamic) system. Dashed line is the midline of the brain and spinal cord. The equivalent components of the trigeminal innervation serving the head are not illustrated. Afferent axon categories are shown on the dorsal root fibers.

primary afferents of class I (A-alpha) and II (A-beta).

The anterolateral system (or spinothalamic system, right) mediates the sensations of pain, temperature and crude, poorly localized touch and receives its major input from primary afferents of class III (A-delta) and IV (C). Pain and position sense will be treated in later chapters.

Cutaneous receptors are, with one possible exception, specialized endings of the sensory axons whose cell bodies lie in the dorsal root ganglia. If one records from such a neuron and stimulates the skin appropriately, the region where stimulation modifies the discharge of the cell is by definition the ***receptive field*** of that cell. This is the most general definition of the term ”receptive field.” Given this definition, the modification could be an increase or a decrease in discharge rate; but in the case of cutaneous receptors, activation of the receptive fields is excitatory.

For a neuron terminating as a single cutaneous mechanoreceptor of, say, the Meissner’s corpuscle type, the receptive field would correspond generally to the area just around the location of the receptor. If the cell’s axon branches, the receptive field would be the entire region to which it supplies receptors. The size of a cutaneous receptive field depends to some degree on stimulus intensity, because a strong stimulus may distort the receptors at a distance while a weaker stimulus would not.

The most familiar test of spatial discrimination asks how close together two stimuli can be and not be mistaken for a single stimulus. In the somatic sensory system this is called the ***two-point discrimination threshold*** and is usually expressed in units of millimeters. Two factors are critically important to this kind of cutaneous spatial resolution: ***receptive-field size***

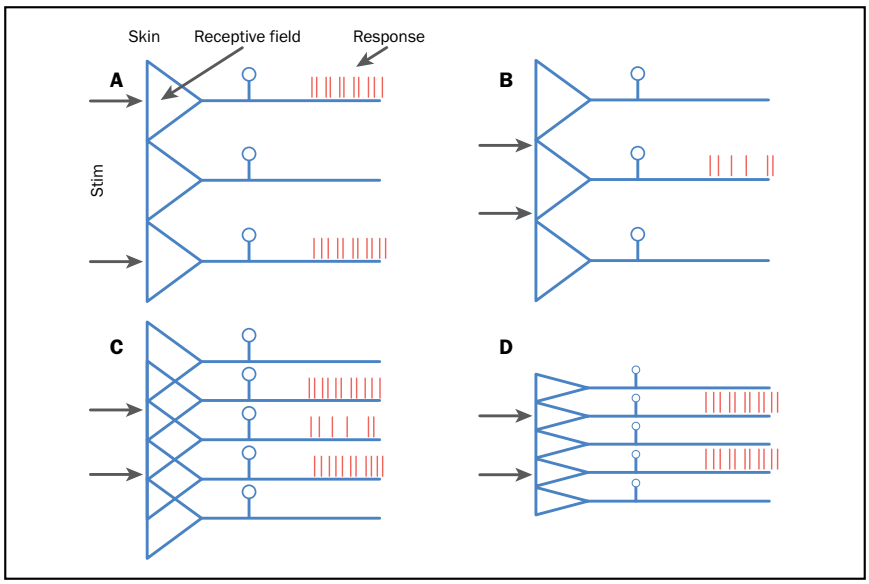


Figure 4.2

Effect of receptive field size and innervation density on two-point discrimination. The skin is to the left and is innervated by three (A and B) or five (C and D) dorsal root afferents. The bases of the triangles represent the area of the cells' receptive fields. The height of the triangles indicate the sensitivity of the cell to a stimulus at the corresponding location on the skin. Horizontal arrows indicate the location of punctate stimuli.

and the *innervation density* of the area tested, i.e. the number of receptive fields per unit area. Figure 4.2 illustrates the effect of varying receptive-field size and innervation density on two-point discrimination in an array of cutaneous mechanoreceptors. In panel A the two stimuli lie in the receptive fields of the two lateral cells, but not in that of the middle one. Thus, in the array of receptor cells, there are two peaks of activity separated by a trough. In panel B, the two stimuli are so close together that they lie only in the receptive field of the middle cell. The resulting pattern of activity would look to the brain like a single stimulus. In panel C, more receptive fields of identical size have been added to the same patch of skin, increasing the innervation density. Now the two stimuli, spaced the same distance apart as in panel B, activate two flanking cells more than the center cell, producing once again two peaks of activity separated by a trough, so the brain receives a signal that two separate stimuli are present. Panel D shows that the same effect can be achieved by decreasing the size of the receptive fields, so that there is at least one cell in the middle that is little, if at all, affected by the stimuli.

Here, in diagrammatic form, we see how spatial resolution can be increased by reducing the receptive field size and by increasing innervation density. In general, the nervous system applies both strategies simultaneously, with receptive-field size usually being inversely related to innervation density. In the real situation the skin is supplied with a continuous array of overlapping receptive fields, partially shifted with respect to one another. One consequence of the high innervation density of certain parts of the skin is that their representation in the central pathways is magnified relative to other areas up to the level of the cerebral cortex.

Another device employed by the nervous system to increase the resolving power of a sensory surface is lateral inhibition, which occurs throughout the CNS including at the first

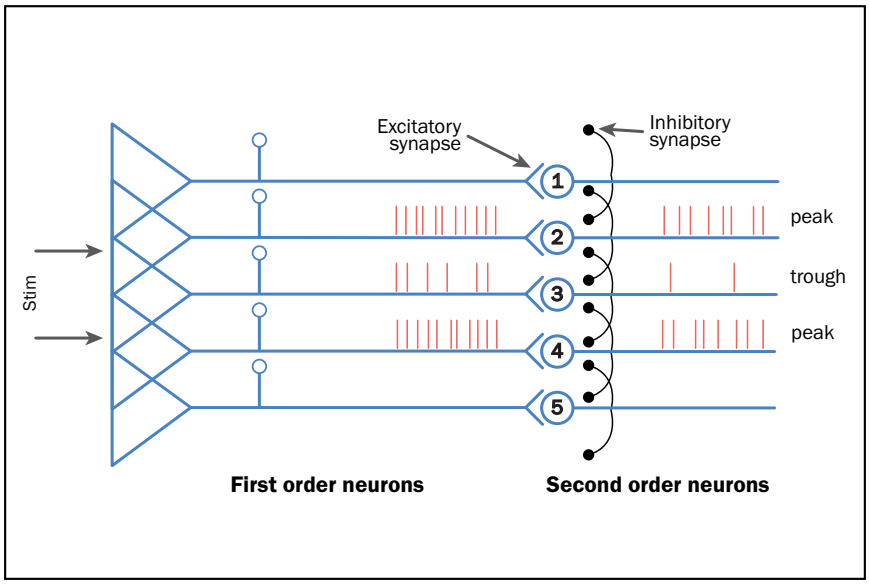


Figure 4.3

Contribution of lateral inhibition to spatial resolution. The inhibitory synapse would, in actuality, be supplied by an interneuron activated by a collateral of the postsynaptic cell. In this configuration the inhibition 'travels' back or retrogradely to the affected cell and is called ***feedback inhibition***. If the inhibition of the second order neurons had been initiated in the orthograde direction via axon collaterals of the first order neurons, the pattern would be called *feedforward*.

and subsequent synaptic relays. In Figure 4.3 each second order, postsynaptic neuron has recurrent collaterals that inhibit the neighboring cells. The two punctate stimuli to the skin produce two peaks of activity in first order neurons synapsing on neurons 2 and 4 and smaller response (or "trough") in the neuron presynaptic to neuron 3. Observe that ***second order neuron 3*** is inhibited from both sides by neurons ***2*** and ***4***, whereas these two second order 'peak' neurons are inhibited only by 3 but not by ***1*** and ***5***. Because second order neuron 3 is laterally inhibited by two neurons, while second order neurons 2 and 4 are only laterally inhibited by one each, the trough between second order neurons 2 and 4 is accentuated. This contrast enhancement increases the probability that the CNS will know there are two separate stimuli.

The Perception of Touch

Two-point discrimination on the skin is best in those parts of the body having the highest innervation density. This is illustrated in Figure 4.4 for the Meissner's corpuscle (a type of receptor) in the hand. For this type of receptor, about 20 receptive fields overlap at a given point on the fingertip as compared with nine on the main part of the finger and palm. There is also a corresponding increase in the sizes of the receptive fields as one proceeds from fingertip to palm.

The receptive fields and innervation densities of other types of receptors demonstrate

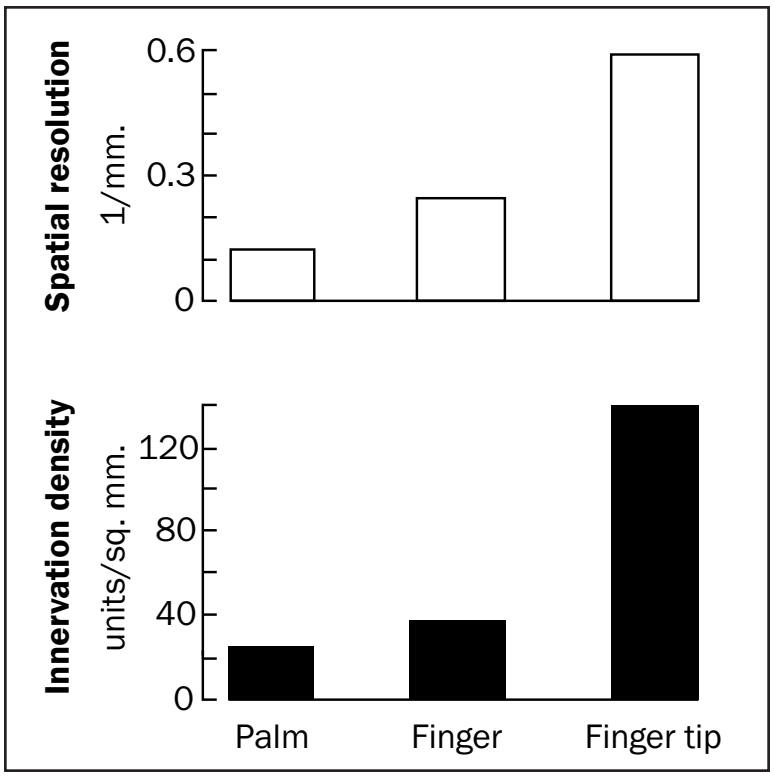


Figure 4.4

Comparison of innervation density of the Meissner's corpuscle receptor with spatial resolution on the hand and finger. (data from Johansson and Vallbo, *Trends in Neurosciences*, 1983).

similar patterns, though the absolute values differ. The Meissner's corpuscles are not the only receptors present in the skin, and each class of receptor may be thought of as segregating a particular aspect of the stimulus for transmission to the central nervous system. Recordings from single axons of human cutaneous nerves indicate that there are at least four physiological types of low-threshold mechanoreceptors distinguishable on the basis of their receptive-field size and the rate at which they adapt to a sustained deformation of the skin. About 17,000 such units innervate the human hand, and of these about 44% are slowly adapting and 56% are rapidly adapting. Table 4.1 summarizes the characteristics of these receptors.

One may ask whether these four types of receptors are associated with four distinct and familiar kinds of sensation. Even though all function as mechanoreceptors, the answer to this appears to be "no." In 1987, Torebjörk, Vallbo, and Ochoa reported the effects of intra-axonal electrical stimulation of human cutaneous afferents. Stimulation of a single axon always gave the same sensation regardless of frequency. Increasing the frequency of stimulation increased perceived intensity but not the quality of the perception. Interestingly, single fiber stimulation never yielded a pure, familiar sense of touch. This suggests that the perceptions triggered by natural stimulation are synthesized from the inputs of different classes of afferents.

Physiological Class	Adaption	Receptive Field	Probable Morphology
PC (RA I)	Fast	Large	Pacinian corpuscle
RA (RA II)	Fast	Small	Meissner's corpuscle
SA I	Slow	Small	Merkel's disc
SA II	Slow	Large	Ruffini's ending

Table 4.1

Classes of low-threshold cutaneous mechanoreceptor afferents. Axons are of the Aβ(Group II) class. Note that thermo- and nociceptors of axon classes Aδ(Group III) and C (Group IV) are not included here. PC, Pacinian corpuscle; RA, rapidly adapting receptor, SA, slowly adapting receptor. Parentheses in column 1: alternative nomenclature applied to these axons. (Adapted from Johnson and Hsiao, *Ann. Rev. Neurosci*. 15:227-250, 1992.)

Parallel/Distributed Processing of Cutaneous Input

The term *parallel processing* was introduced in the early 1970s to capture the idea that different aspects of the retinal image are encoded by specialized classes of ganglion cells and transmitted to the brain over essentially independent channels. It is clear from Table 4.1, and the fact that stimulation at a particular point on the skin will activate a number of receptor types, that a similar strategy is used in the somatic sensory system as well. The natural stimulus is 'dissected' by the receptor types into four attributes that are processed in parallel as the encoded representation of the stimulus. The ***presence*** of texture is signaled by the exquisitely sensitive but poorly localizing Pacinian corpuscles, while the SAI and RA receptors encode the ***structure*** of the stimuli with relatively high fidelity. It should also be remembered that active touch also incorporates position sense and motor signals into the information matrix available to the brain, and this may augment the capacity of the brain to 'read' the patterns.

The notions of parallel and distributed processing are similar in meaning and are often lumped together in the term *parallel/distributed processing*. The distinction can be made if you consider a sheet of receptors all of the same type. If a point stimulus activates a population of such cells, information about the point is ***distributed*** across the active population. If there are several different classes of receptor, different aspects of the stimulus are being signaled by *parallel* systems. As a rule, most situations involve both processes and are aptly described by the phrase parallel/distributed processing.

Microelectrode recordings from the primary somatic sensory cortex (S1) indicate that the subdivisions of this area continue to segregate the information originally encoded by the four types of low-threshold cutaneous mechanoreceptors. Area 3b of cortex exhibits a columnar organization with rapidly and slowly adapting sensory responses represented in alternating columns. Cells in area 1 respond preferentially to input from rapidly adapting cutaneous and deep receptors. Cells in area 2 are activated by rotation of joints and mechanical stimulation of deep tissues, while those in area 3a receive input from muscle receptors. Information at this level does not stem only from primary afferents because there is also evidence of integrative activity. Areas 1 and 2 receive projections from area 3 as well as from the thalamic relay nuclei VPM and VPL. Neurons in these areas exhibit properties such as selectivity to motion, direction and orientation that are not seen in the dorsal column nuclei or thalamus. Therefore, information flowing in the parallel primary afferent channels is combined and transformed at the cortical level in complex ways that are not yet understood.

Punctate stimulation of the skin is also known to activate different areas of S1 simultaneously, as might be expected from the existence of multiple maps. Using an optical recording method

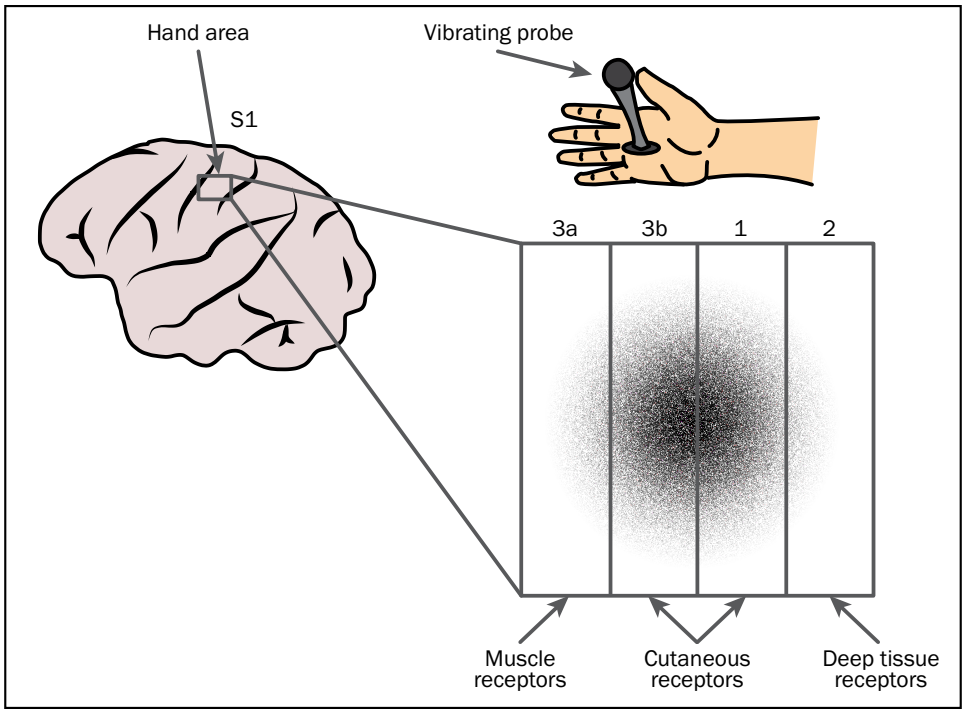


Figure 4.5

Activation of primary somatic sensory cortex (including areas 1, 2, 3a, and 3b) by a 2 mm vibrating probe applied to a spot on the palm of the hand. The activity (dark areas) was measured optically using a special infrared imaging technique. (After M. Tommerdahl, et al., J. Neurophysiology, 1996.)

that images the activity of a region of cortex (see technical appendix), Tommerdahl and colleagues estimated the extent of somatic sensory cortex that responded to application of a 2 mm vibrating probe to the hand. As schematized in Figure 4.5, the stimulus activated a large patch that was centered in areas 1 and 3b.

Technical Appendix: Imaging Neural Activity

In recent years it has become possible to map neural activity by exploiting changes in the optical properties of active tissue. In one version of this technique, the cortex is soaked with a dye which absorbs a certain amount of light of a particular wavelength as the membrane potentials of the cortical cells vary. The absorption patterns, presumably reflecting patterns of neural activity, are recorded by special optical equipment while natural stimuli are presented. A variation on this approach dispenses with dyes and measures intrinsic light scatter in the tissue during stimulation, perhaps due to local changes in blood flow. The spatial resolution of these techniques is high (see Figure 4.5), but their temporal resolution is not as good as that of electrical recordings. Also, the signals can be quite weak, so it is often necessary to use repeated presentations of the stimulus and compare large numbers of stimulus and non-stimulus periods statistically.

Because neurons also respond to increasing activity by increasing levels of expression of immediate early genes (e.g. ***c-fos*** and ***zif268***), another method for identifying neurons activated by sensory inputs is to track levels of gene expression. While this has typically been done in fixed tissue, recent transgenic developments have coupled activity-dependent gene expression with expression of a labeling protein (green fluorescent protein, or GFP), allowing active cells to be imaged in–vivo. Modern techniques also allow for the imaging of calcium flow, where increased flow of calcium indicates an increase in neural activity.

Terms and Techniques

topographic modality	non-topographic modality	lateral inhibition
innervation density	cortical magnification	receptive field
two-point discrimination	feedback inhibition	feedforward inhibition
DC/ML system	dorsal column nuclei	anterolateral system
central sulcus	nociceptor	proprioception
Brodmann's areas 1, 2, and 3	Merkel's disk	Meissner's corpuscle
Ruffini organ	Pacinian corpuscle	free nerve ending
rapidly adapting	slowing adapting	mechanoreceptor
axon groups I, II, III, IV	axon groups Aα, Aβ, Aδ, C	recruitment
frequency coding	parallel/distributed processing	optical imaging of neural activity

Spatial Coding: Sampling, Magnification, and Visual Activity 5

Table of Contents

In the previous chapter we discussed how the spatial resolution of a sheet of cutaneous receptor cells depends on the size and spacing of their receptive fields. In vision, the acuity measure that is equivalent to two-point discrimination is called the ***minimal angle of resolution*** (MAR), which is the smallest visual angle between two points that can reliably be distinguished as separated in visual space (Figure 5.1). We will use the example of the MAR to illustrate how the topography of the central sensory pathways can affect the spatial resolution of a sensory surface. We will then introduce the notion of sampling to provide a quantitative treatment of spatial resolution by an array of detectors. Distances and positions in visual space and on the retina are most conveniently expressed in terms of angle, as illustrated in Figure 5.1.

Central Topographic Maps and Spatial Resolution

As neurons from the receptor surfaces of the skin, retina and organ of Corti project centrally, they preserve their topographic order. Said another way, the terminals maintain the same near-neighbor relationships as the cells of origin. This arrangement persists through the central projections and is responsible for the topographic maps of the body found in the somatic sensory cortices, of visual space found in visual cortices, and the maps of frequency found in the primary auditory cortex.

These maps, however, are distorted in that not every part of the sensory surface is represented equally. This is evident in the ***somatotopic map*** of S1, which magnifies the face

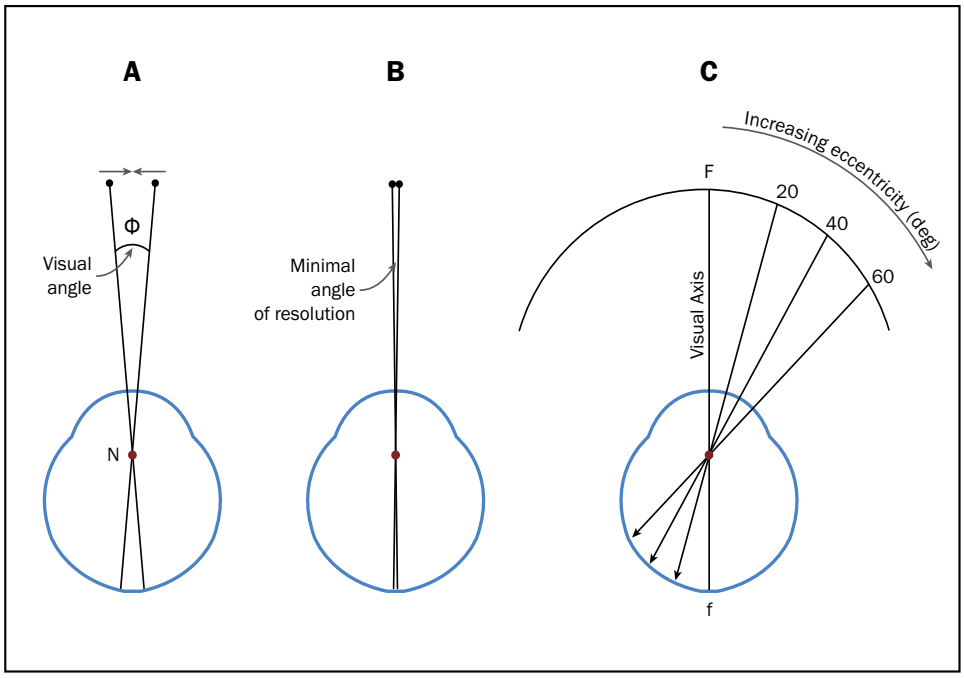


Figure 5.1

Angular positions and distances in visual space and on the retina. A. Visual angle. The separation of two points in visual space can be expressed as the angle formed by lines (rays) from the points in visual space to the retina that pass through the nodal point, N, of the eye. Rays passing through the nodal point of an optical system continue in the same direction. Observe that the separation of the images of the two points on the retina can also be expressed as an angle. B. MAR. The minimal angle of resolution is the angular separation of two features, dots in this case, such that the two dots are resolved as separate, i.e. where the two dots would not be seen as one. C. Eccentricity. Units of angle can also be used to specify the position of an object in visual space or its image on the retina. The reference is the "line of sight" or ***visual axis***, which projects from the fovea (f) to the fixation point (F).

and hand relative to other areas. Also, the ***retinotopic map*** of V1 magnifies the foveal area corresponding to the central part of the visual field (Figure 5.2).

In S1 and V1 this differential magnification arises in part from the variation in innervation densities across the sensory surfaces. For example, more ganglion cells represent a square millimeter in the foveal region of the retina than a comparable area in the peripheral retina, reflecting the large concentration of cone photoreceptors in the fovea. If the terminals of each retinal ganglion cell were allotted the same amount of space in the lateral geniculate nucleus, the central part of the field would claim the most territory simply because it is represented by the largest number of cells. The projection from lateral geniculate to cortex would preserve this pattern.

This variable expansion of the representation of visual space can be expressed quantitatively in terms of a ***magnification factor*** defined as the number of linear millimeters in the cortex (or other visual projection area) allotted to a given linear degree of visual space. Thus, one degree

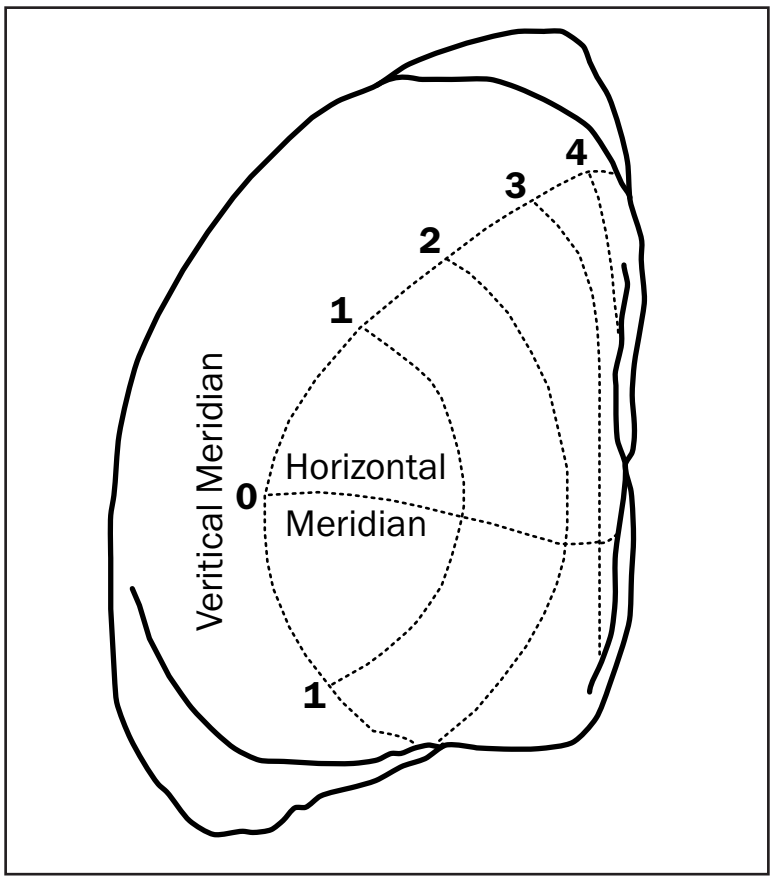


Figure 5.2

Expansion of the area of central vision in the primary visual cortex. The view is of the back of the left hemisphere of the squirrel monkey. The polar coordinates of the central part of the visual field are projected onto the cortex as lines. Note that the central 1 degree has the largest representation. (Redrawn from Cowey, ***Journal of Neurophysiology*** 27:66-93, 1964.)

of visual angle might occupy 6 mm of cortex near the foveal representation (magnification factor M = 6 $^{mm}/_{deg}$) and only 1 mm in peripheral parts of the map (M = 1$^{mm}/_{deg}$). One can also think of magnification as the area of cortex representing a degree of solid angle in visual space.

Cortical organization is closely linked to spatial acuity because of this unequal magnification of the sensory surfaces in the topographic maps. While the magnification factor in visual cortex falls off continuously with distance from the fovea, the minimal angle of resolution or MAR increases at the same rate as the cortical magnification factor decreases. This relationship makes sense because as spatial acuity decreases (as demonstrated by the increase in MAR), the amount of cortex devoted to that area also decreases.

Because the two measures vary reciprocally with eccentricity (i.e. with distance from the fovea), their product is more or less constant. When we multiply magnification factor (M) by MAR, the product of $^{mm}/_{deg}$ by deg has units of distance. Table 5.1 shows what happens when MAR measurements at several eccentricities are multiplied by the corresponding magnification factors. The effect is impressive. While the MAR increases up to 11.7 times larger between

Eccentricity	2.5°	5°	10°	20°	30°	40°	50°
M ($^{mm}/_{deg}$)	3.87	2.35	1.31	0.70	0.48	0.36	0.29
MAR (deg)	.018	.028	.044	.080	.116	.163	.210
MAR x M (mm)	.070	.066	.058	.056	.056	.059	.061

Table 5.1

MAR scaled by cortical magnification factor M. Values of M are from Cowey & Rolls (1974) as plotted by Foster et al. (1981). Values of MAR are from Wertheim (1894).

2.5° and 50°, the product of MAR and M is relatively constant at ~.06-.07 mm. Thus, while the value in degrees of the MAR varies as the test is applied at different locations in visual space, the measurement in cortical space is ***translationally invariant***. As a general rule, when a particular transformation converts a variable phenomenon into an invariant one, it is telling you something important about the mechanism of the phenomenon. (It should be noted that in primates this invariance may not hold for eccentricities less than 2.5 deg. where magnification factor rises more rapidly than MAR declines).

Multiplying the MAR for a given eccentricity by the cortical magnification factor M for that eccentricity is essentially transferring the MAR measurements to the retinotopic cortical map

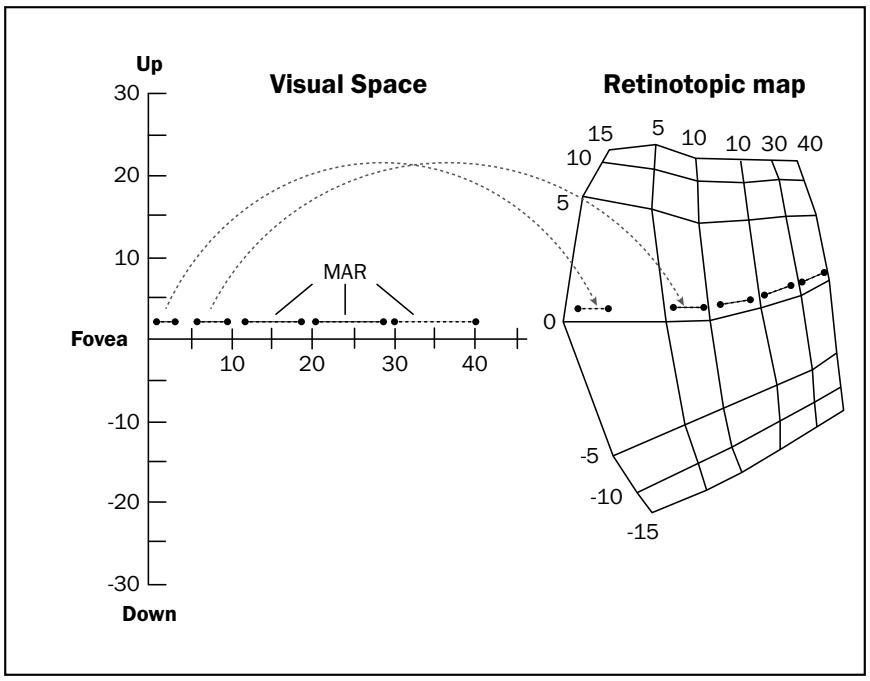


Figure 5.3

Graphical illustration of the concept of equivalent cortical distance. Values of the minimal angle of resolution at different eccentricities in the visual field are transferred to the corresponding coordinates in the cortical retinotopic map. The MAR (and its equivalent cortical distance) is the distance between the dots joined by horizontal dotted lines.

coordinates (Figure 5.3). For example, if the MAR at an eccentricity of 20° is 0.08° and cortical M is 0.7 mm/deg for that eccentricity, the product of the two numbers is .056 mm. This is the distance across the cortical retinotopic map corresponding to an 0.08° line located 20° from the fovea. The product of MAR x M can be said to estimate the ***equivalent cortical distance*** for the MAR. Thus, the results of Table 5.1 indicate that, if the space between two point stimuli is to be resolved, the patches of cortical activity associated with these stimuli must be centered about .06-.07 mm apart, regardless of where the two stimuli are positioned in visual space. This finding suggests that the neural machinery involved in the determination of the MAR is the same size everywhere in the cortex, and the MAR accordingly varies in the outside world because the mapping from retina to cortex is distorted.

Similar translation-invariant equivalent cortical distances are found for other visual discrimination tasks. While the cortical distances might be different in these tasks than in the task for determining MAR, the idea of scaling from one coordinate system to another is quite similar. This finding suggests that different recognition or detection tasks engage cortical regions of different size during their execution. Comparable quantitative correlations have not been established for the topographic cortical maps in the primary auditory and somatic sensory cortices, but these maps do not have the same degree of uniformity as the visual map

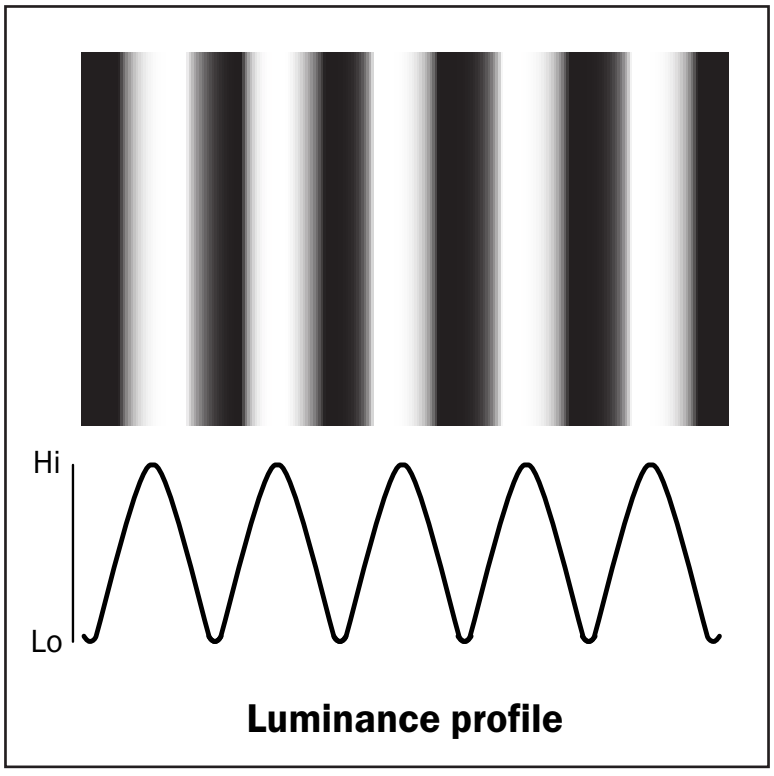


Figure 5.4

A one-dimensional sinusoidal grating and its luminance profile.

of V1. Nonetheless, the underlying principle probably applies. The price of higher resolving power is more neurons.

Because the MAR and two-point discrimination are spatial concepts with spatial units, they are said to describe resolution in the ***spatial domain***. A more versatile and sophisticated approach to spatial resolution is to consider stimuli and the responses to them in the ***frequency domain***. This is usually done for vision by presenting the eye with gratings, arrays of dark and light bars in which the brightness of the bars varies across the array. If this variation in brightness is continuous and sinusoidal, such a stimulus is called a ***sinusoidal grating*** (Figure 5.4).

The ***period*** of such a grating is one complete cycle of dark and light areas and its ***spatial frequency*** is the number of cycles occurring within a certain distance. The frequency is the reciprocal of the period:

$$f = \frac{1}{P}$$

These parameters are usually expressed in terms of visual angle, so the units for period are degrees or fractions of a degree. Because spatial frequency is the reciprocal, its units are in cycles per degree. For example, a grating of 60 cycles per degree would have a period of $(1/60)^{o}$ or 1 minute (1') of visual arc. (There are 60 minutes (60') in a degree and 60 seconds (60") in a minute).

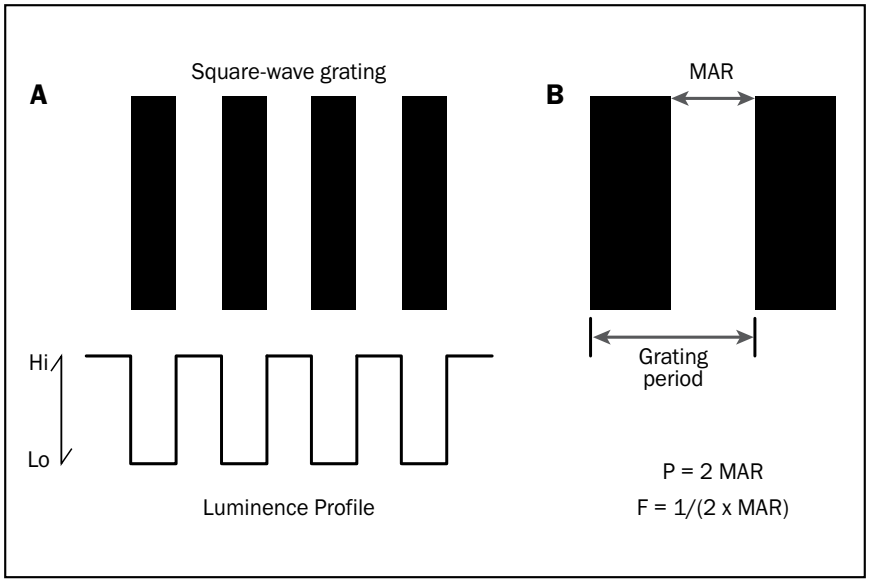


Figure 5.5

Square-wave gratings and the MAR. A. Example of a square wave grating and its luminance profile. B. Relationship between the MAR and the frequency of the square-wave grating. The period of the grating is one complete cycle of dark and light bars. The MAR would be measured as the separation of the black bars, i.e. half the period of the grating.

The third parameter used to describe the grating is its ***contrast***, which is a measure of the difference in brightness of the dark and light bars. This is often expressed as a percentage or fraction based on a formula such as

$$\frac{\text{Hi - Lo}}{\text{Hi + Lo}}$$

where Hi and Lo refer to the luminance of the brightest and darkest areas, respectively. As noted above, the contrast varies continuously in a sinusoidal grating, whereas a grating with sharp boundaries between dark and light areas would be a ***square-wave grating*** (Figure 5.5).

Two-point discrimination or MAR thresholds can be related to a just-resolvable, high-frequency square-wave grating if one considers that the separation between the test points of the MAR corresponds to the space between two dark bars of the grating. If the dark and bright bars are the same width (Figure 5.5), then the period of the grating is twice the MAR because this distance would represent one dark and one light bar. The frequency would again be the reciprocal:

$$\frac{1}{2 * \text{MAR}}$$

For example, if the MAR is 0.5 minutes (0.5'), then the period of the corresponding grating would be 1'. The frequency of the grating would then be one cycle per minute or 60 cycles per degree. (Note that it is important to keep track of the units used in such calculations).

While square wave gratings seem much simpler sine wave gratings, the mathematician Fourier demonstrated that a square-wave grating can be decomposed into a series of sine wave gratings of different frequency, amplitude and phase. Fourier's finding demonstrates that they are actually a complex sum of multiple sinusoidal gratings, and therefore when researchers want to understand frequency, they will usually employ sinusoidal gratings, not square wave gratings. Nonetheless, the calculation relating the MAR to the highest spatial frequency resolvable by the eye gives a good approximation, because the eye's optics tend to filter out the very highest frequencies present in a square-wave grating. This filtering phenomenon plays an important role in spatial processing by the retina as we will discuss below.

The Contrast Sensitivity Function

Two main parameters that determine whether someone can detect a pattern are the spatial frequency and the contrast between the bars. One way to establish the threshold contrast

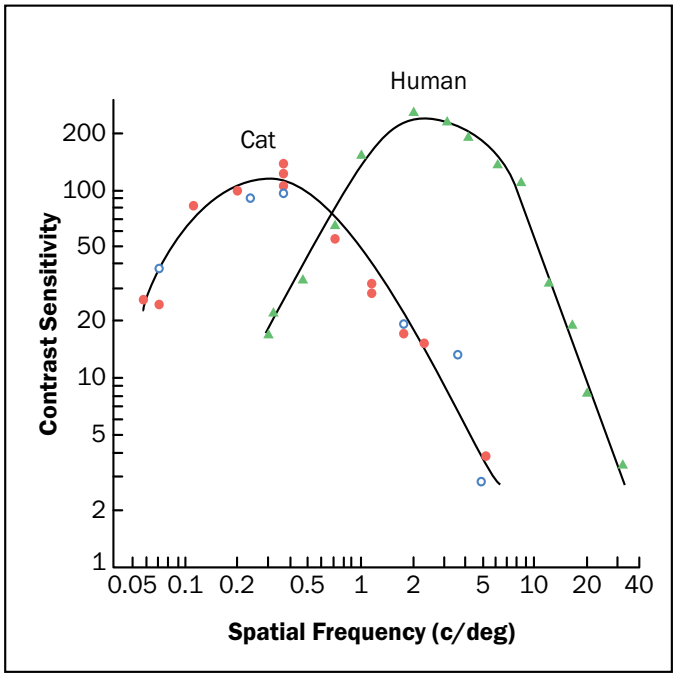


Figure 5.6

Contrast sensitivity functions of humans and cats. (Data from S. Bisti and L.Maffei, Behavior contrast sensitivity of the cat in various visual meridians. *J.Physiol.* 241:201-10, 1974.)

for a spatial frequency is to allow a subject to vary the contrast of a sinusoidal grating until it is indistinguishable from a uniform gray background. Taking the reciprocal of that contrast, one obtains a measure of the ability (sensitivity) of the subject to sense that particular spatial frequency. A higher sensitivity connotes that the amount of contrast required to sense a frequency pattern is very low—the eye can easily sense this frequency, so high contrast is not needed. Accordingly, low sensitivity means the contrast between adjacent bars must be very high in order for the subject to distinguish the bars.

When this measurement is taken over a range of spatial frequencies, one can plot the values as the Contrast Sensitivity Function or CSF (Figure 5.6). Note that, with appropriate changes in the experiment, the CSF can be obtained from a reporting human subject, a behaving laboratory animal or a responsive single neuron in the visual system. The range of spatial frequencies to which a system will respond is called the ***bandpass*** of the system.

As shown in Figure 5.5, the bandpass of the human visual system contains higher frequencies than does that of the cat, which means that humans can resolve finer spatial detail than cats. The graph plots sensitivity over frequency, so a rightward shift means the human eye has a greater capacity to see many light/dark bars per degree. This capability means the human has a higher ***grating acuity***. The highest frequency that a human can detect is about 60 cycles/degree, which constitutes the ***cutoff frequency*** of the human CSF. At the cutoff frequency, the contrast between bars must be 100% for subjects to distinguish the pattern.

Sampling Theory and Spatial Frequency Analysis

The ability of a system to reliably detect and characterize the spatial frequency components of a real stimulus is determined by two key parameters: how often it samples the visual scene and the how well the retina responds to the frequencies present in the scene. The general theory that applies to such situation is called ***sampling theory***.

We may think of photoreceptors as sampling the retinal image at points separated by the distances between the receptors. (For the moment we will ignore that the receptors actually sample a small area of the image, though this will become an important consideration later). Important limitations arise purely on the basis of the ***sampling interval***, which is the distance between the sample points. Consider the array of tiny receptors spaced X degrees apart in Figure 5.7 and their capacity to encode sinusoidal gratings of different frequencies. The spacing, X, between the cells is the sampling interval, and 1/X is the ***sampling frequency***.

Let's assume that the three sine waves represent the luminance profiles of three sinusoidal gratings of equal amplitude presented one at a time to a row of photoreceptors, and that the dotted lines in Figure 5.7 indicate the point at which each sinusoid is sampled by a given receptor. Grating B is sampled twice in every cycle, grating A many times and grating C only once in every few cycles. Observe that the outputs of the receptors will be different for gratings A and B but identical for gratings B and C. In other words, any observer knowing only these outputs would not be able to tell grating C from grating B. In contrast, there is no

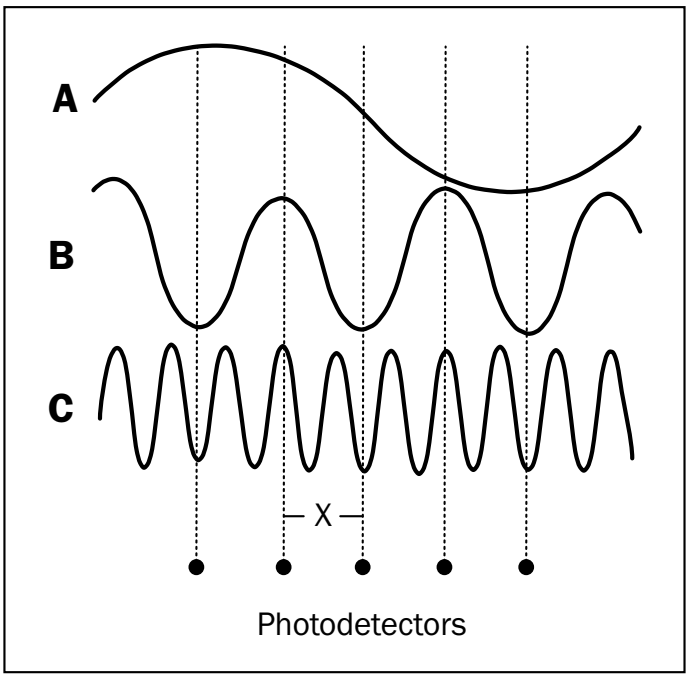


Figure 5.7

Schematic illustration of aliasing. Sinusoids are the luminance profiles of three gratings presented one at a time to the array of photocells below. The outputs of the photocells are identical for gratings B and C, so C is aliased as B. No frequency lower than that of B can mimic B in the outputs of the detector array. X is the sampling interval of the array. (Redrawn from McIlwain, An Introduction to the Biology of Vision, Cambridge UP, 1996.)

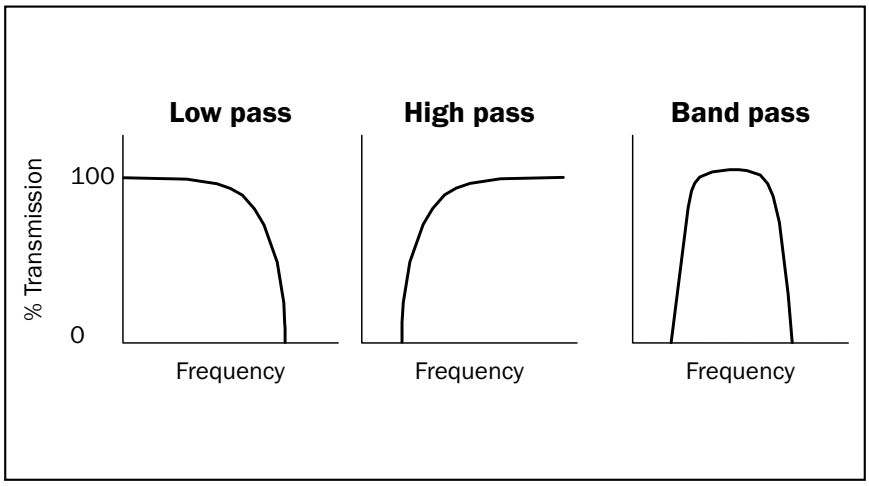


Figure 5.8

Graphical representation of low-pass, high-pass and band-pass filters.

way to draw a grating of lower frequency than B that will give exactly the same output values as that grating (try it!). Thus, at this sampling interval, grating B can be distinguished from all gratings of lower frequency, but not from some gratings of higher frequency, e.g. grating C.

The phenomenon in which a sampling array confuses higher frequencies with lower frequencies is called ***aliasing*** because the higher frequencies look like the lower frequencies, i.e. they produce identical outputs from the sampling array. Biological sensory systems are generally constructed to avoid aliasing because it can give rise to misleading signals about the a stimulus's orientation and direction of motion. A convenient example of aliasing is provided by the phenomena of Moiré patterns, which appear, for instance, when a set of lines is placed behind and viewed through a screen of the right coarseness. Aliasing in temporal sampling is what makes the wheels of wagons appear to roll backwards in old cowboy movies, because the camera is not capturing frames as fast as the spokes are changing position.

To avoid aliasing of biologically relevant stimuli, sensory systems must first possess the right sampling interval. The ***sampling theorem*** states that the highest frequency that can be identified ***unambiguously*** is one that is sampled twice in each cycle of the waveform, i.e. there must be at least one sample for each peak and one for each trough, although not necessarily located at those positions on the waveform. Thus, if X in Figure 5.7 equals $1/120$ deg. of arc, then gratings of 60 cycles/degree or lower will be sampled at least twice and none will yield exactly the same sample values and be confused with each other. The frequency $1/2x$, 60 cycles per degree in this case, is called the ***Nyquist*** frequency of the sampling array, after the man who introduced the term.

If frequencies higher than the Nyquist frequency are present in the signal, the output of the system could be ambiguous because of aliasing. To prevent the ambiguity, higher frequencies

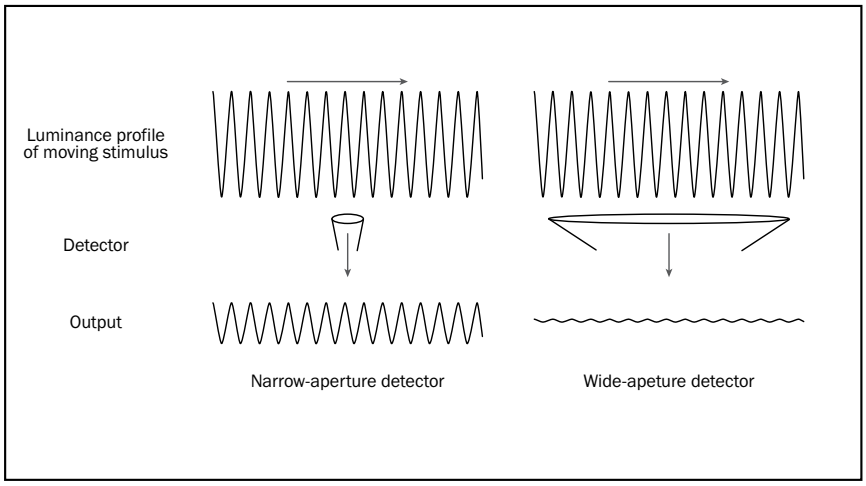


Figure 5.9

Effect of receptive field size on high spatial frequency transmission. The receptor to the left samples a smaller area of the moving grating than the receptor to the right. If there are equal numbers of light and dark stripes in the receptive field of the large receptor, there will be little or no modulation of the output.

subject to aliasing must not be allowed to affect the receptors. The solution is some kind of filtering mechanism between the physical stimulus and the receptors and/or in the receptors themselves. The eye's optical elements are said to act as a "low pass filter," because the retinal image tends to blur fine textures. The blurring occurs because these receptors pass the low but exclude the high spatial frequencies (Figure 5.8). This appears to be an important factor for preventing aliasing of foveal images where the photoreceptor and effective ganglion-cell spacing (and consequently the Nyquist frequency) is matched to the filtering properties of the eye's optics.

In the peripheral parts of the retina the ganglion cells, which function as sampling elements, are spaced farther apart and the Nyquist frequency is lower. Here, aliasing can be avoided if these detectors are made relatively insensitive to high frequencies in the retinal image. There are multiple ways to accomplish this, and in this case the detectors themselves function as low pass filters. One way in which the detectors omit high frequencies is sampling an area instead of a point. If a high frequency grating is moved over a sampling detector with a very small, point-like receptive field, the receptive field will be filled alternately by dark and light and the output of the cell will be modulated (Figure 5.9, left). However, if many cycles of the stimulus lie within the receptive field (Figure 5.9, right panel), the cell will integrate across the dark and light areas, and its output will reflect the average luminance across the grating. Any modulation of the output will be due to slight discrepancies in illumination at the edges of the receptive field. The larger the aperture of the receptor, the more incapable it will be of encoding the luminance structure of the stimulus and the more pronounced will be its low-pass filtering characteristics.

Sampling theory thus provides an important insight into why receptive field size increases as innervation density decreases. If each receptive field is considered a sampling point, the

farther apart they are, the lower the Nyquist frequency will be. To reduce the possibility of aliasing, as the receptive fields become sparser they increase in size and become less sensitive to higher spatial frequencies.

Note that these concepts also apply at any level of the nervous system where an array of sampling elements must transmit spatial information about an extended stimulus. These principles also apply when a given structure has more than one class of sampling elements. For example, one could analyze the somatic sensory system in terms of the sampling array of Meissner's corpuscles, or free nerve endings or Merkel's discs. Similarly, retinal ganglion cells in the monkey come in M and P varieties that form independent arrays, each of which can be considered as a separate sampling system with its own Nyquist frequency.

What is the relationship between the Nyquist frequency of the retina and the cut-off frequency of the CSF?

The short answer to this question is "nothing." The Nyquist frequency is determined purely by the spacing of the photoreceptors. Thus, one could measure the spacing between cones or rods and calculate Nyquist frequencies for each array. One could also calculate a Nyquist frequency for a particular class of ganglion cells. This measure is a theoretical frequency based on a calculation and is therefore conceptually analogous to the equilibrium potential (E_{ion}) calculated from the Nernst equation. The number may be relevant to a particular situation, or it may not. The value of the Nyquist frequency simply reflects the highest frequency that the array ***could*** unambiguously characterize or identify, without the possibility of aliasing. It does not say that the array actually does this.

The CSF is a set of measurements describing the actual behavior of a subject. The subject may be an intact organism or a single neuron. The cut-off frequency is the frequency at which a grating of 100% contrast can just barely be distinguished from a uniform gray background. Any higher frequency looks gray, even at 100% contrast. As noted earlier, this frequency can be estimated from the MAR. The calculations relating the Nyquist frequency to the sampling interval of an array, and that relating the cut-off frequency of the CSF to the MAR are very similar in form. It is easy to confuse the concepts because of this similarity, but it is important not to do so.

What is striking is that the separation of our foveal cones predicts a Nyquist frequency that has the ***same value in cycles per degree*** as the cut-off frequency of our CSF. Having the same numerical value does not mean that they are the same thing, but rather that evolution has optimized the visual system so that it only responds to frequencies that cannot be aliased by the photoreceptor array. As we have discussed, it accomplishes this by using low pass filtering to prevent aliasable frequencies from reaching the array and/or by rendering the

sampling elements insensitive to the aliasable frequencies. Virtually all known visual systems do this, be they vertebrate or invertebrate. Aliasing is maladaptive, and nature has contrived to avoid it when building visual systems.

Terms and Techniques

spatial domain	frequency domain	spatial period
spatial frequency	square-wave grating	sinusoidal grating
equivalent cortical distance	eccentricity	degrees of arc
visual angle	minutes of arc	seconds of arc
MAR	receptive field image	retinal coverage
grating acuity	sampling interval	aliasing
sampling theorem	Nyquist frequency	spatial contrast
sampling frequency	low pass filter	high pass filter
band pass filter	translational invariance	luminance
contrast	Contrast Sensitivity Function	cutoff frequency

Spatial Coding: Distributed Processing in Vision

6

Table of Contents

Any stimulus, no matter how localized or punctate, will affect many cells at a sensory surface. Thus, information about a point on the sensory surface is encoded in activity spread over a distribution of neurons rather than being confined to the activity of only one cell. The task of two-point discrimination is to discern the presence of two possibly overlapping patterns of activity. This process is enhanced through the action of lateral inhibition.

As with a sheet of cutaneous receptors, information about a point of light on the retina is encoded across multiple cells in the retina. The pattern of retinal cells that respond to one point in space is called a ***point image***. While receptive fields refer to a cell and all the points in visual space it can "see," a point image refers to a point in space and all the cells that "see" the point.

Let's consider a class of retinal ganglion cells that have perfectly circular receptive fields centered on their cell bodies (Figure 6.1). We will ignore the presence of on- and off-areas in the receptive fields. Cells in the point image can only be so far from the point of light. One receptive field radius is the farthest away the cell can be while still having its circular receptive field overlap with the point. Under these geometric assumptions, the point image will also be circular and centered on the stimulus point and the same size as the receptive field. There is therefore ***geometric reciprocity*** between the point image and the receptive field.

Observe that we have made one more critical assumption here namely that the receptive fields are the same size and shape on all sides of the stimulus point. In other words, we have assumed that the receptive fields have the same size and shape no matter where they are measured. This is the important assumption of ***translational invariance*** of the receptive fields. When finding point images, the receptive field profiles must only be locally translationally

invariant—that is, they are invariant in the immediate vicinity of the stimulus point. As you move increasingly further from the fovea, the change in receptive field size may not even be detectable over short distances, so assuming local translational invariance will not introduce large errors.

We may now ask how many cells lie inside the point image and convey information about that point. This number is the product of the area of the point image and the spatial density of the cells, i.e. the number per unit area. For example, if the point image has an area of 2 mm^2 and there are 50 cells per square mm near the point, the number of cells in the point image is 2 mm^2 x 50 $^{cells}/_{mm^2}$ or 100 cells. The number of cells in the point image is sometimes referred to as the ***coverage*** provided by that class of cells, as it refers to how many cells' receptive fields cover that point. (In the scientific literature on the retina, 'coverage' can also refer to the number of cells whose dendritic fields overlap at a given point. This anatomical concept is analogous to the functional concept based on the notion of receptive field).

Suppose now that, as in a real retina, the receptive fields of the cells get larger and their spatial density decreases with increasing eccentricity—that is, as distance from the fovea increases. Because of the increased receptive field size with increasing eccentricity, the point image must also get larger; yet, the spatial density of the ganglion cells decreases at the same rate. Therefore the product of point-image area and spatial density can remain constant (Figure 6.2). As one moves from the fovea to the periphery, it is possible for the point image among ganglion cells to increase in size while the coverage remains constant. Coverage indeed appears relatively constant across the retina for certain types of ganglion cells, even though their spatial density decreases (and sampling interval increases).

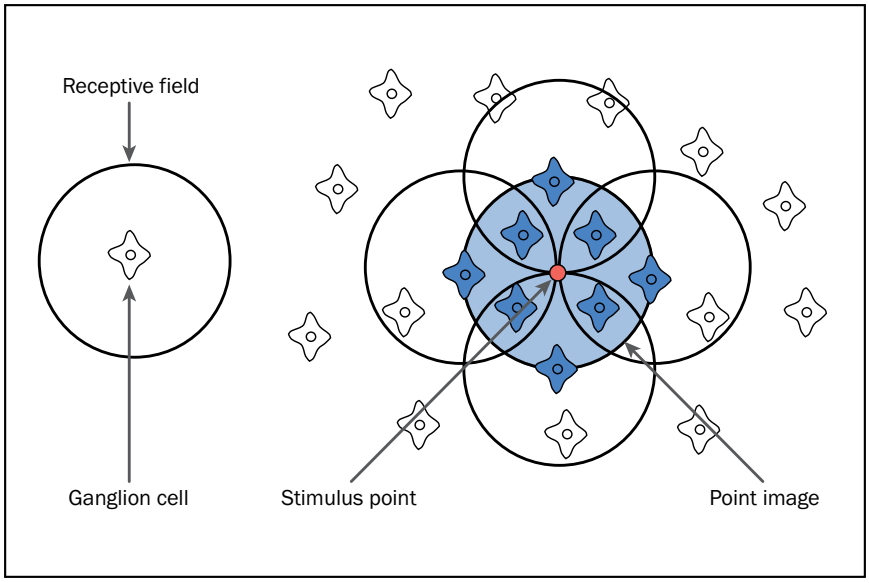


Figure 6.1

Point image in retinal ganglion cells. Blue cells have receptive fields containing the stimulus point. All such cells lie within the shaded area. Note that the point image has the same size and shape as the receptive field.

Receptive field size varies among the different classes of retinal ganglion cells, so the overall point image size at any locus on the retina will be largest in the class of cells with the largest receptive fields.

The retina is also carpeted with overlapping receptive fields of every type. For example, a given point near the primate fovea will lie in the receptive fields of ON and OFF M-type ganglion cells, as well as color-opponent P-cells of every variety. Because of this overlap, when studying point images, it is often helpful to think of each class of cells separately. Information about luminance and chromaticity at a given point enters each of these channels simultaneously and is sent to the LGN, and then to the cerebral cortex. So, in addition to the ***distributed coding*** implied by the finite size of the point image, there is ***parallel processing*** of the information through several specialized channels. This is perfectly analogous to the situation in the somatic sensory system.

Representation of Points in the Central Visual System

Remember that after light hits the retina, the signal must be conveyed to cortex. Just as the retina codes information from visual space, the striate cortex must receive a retinotopic projection of the visual world. Here arises the major problem: the cortical cells are in one spatial

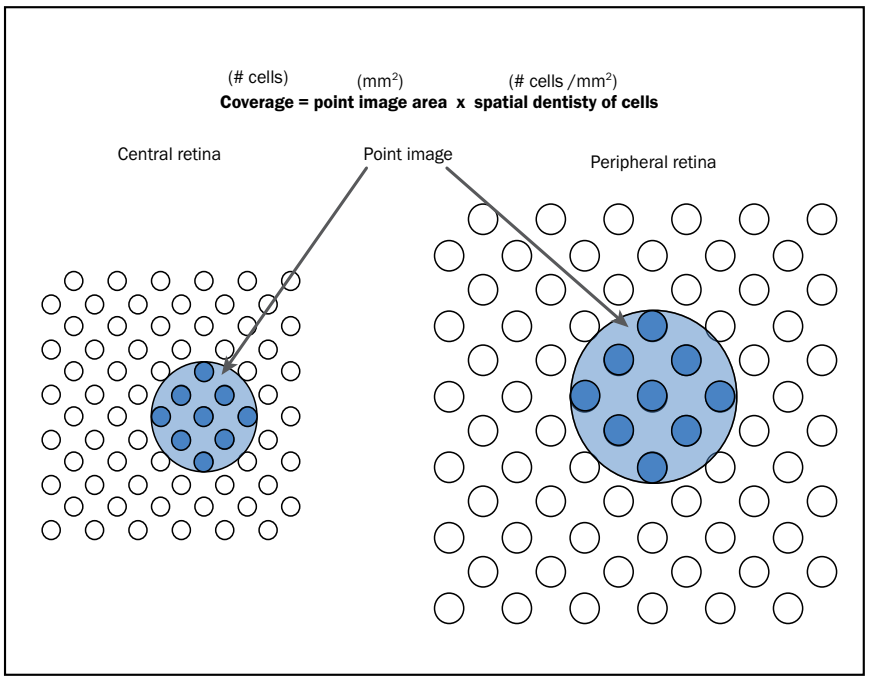


Figure 6.2

The relationship between retinal coverage, point image size and spatial density of ganglion cells (i.e. number per unit area).

coordinate system, but their receptive fields have been plotted in retinal coordinates related to visual space. We cannot directly relate cell location to the boundaries of the receptive fields to determine coverage and point-image size like we did with the retina. We require a way to relate points on the retina (or visual space) where the receptive fields are plotted, to points in the cortex, where the cells actually reside. This relationship is provided by a retinotopic map that accurately reflects the retinal origins of signals arriving at a given cortical point.

The first step in locating the boundaries of the cortical point image is to transfer the receptive field boundaries from the coordinate system of the retina to that of the central map. Note that the central map is a theoretical map used to represent cortex or superior colliculus; however, this model lacks many of the specific features of each brain area. That said, the central map is a useful way to model coordinate system transformations.

This procedure "images" the receptive field into the new coordinate system, so the resulting profile has been called the ***receptive-field image*** or RFI (Figure 6.3). Observe that each point inside the RFI corresponds to a point inside the receptive field. A point on the boundary of the receptive field will have a corresponding map point at the edge of the RFI. Points outside the receptive field in visual space will project outside the RFI in the central map.

Consider now the theoretical situation of Figure 6.4 in which electrode penetrations have been made in the central map, and the largest receptive fields recorded at four of these locations are those mapped on the left of the figure. These fields were selected because their boundaries just touch the stimulus point. Replotting these receptive fields in the central map yields the RFIs at the right. While the receptive fields are only ***locally*** translationally invariant,

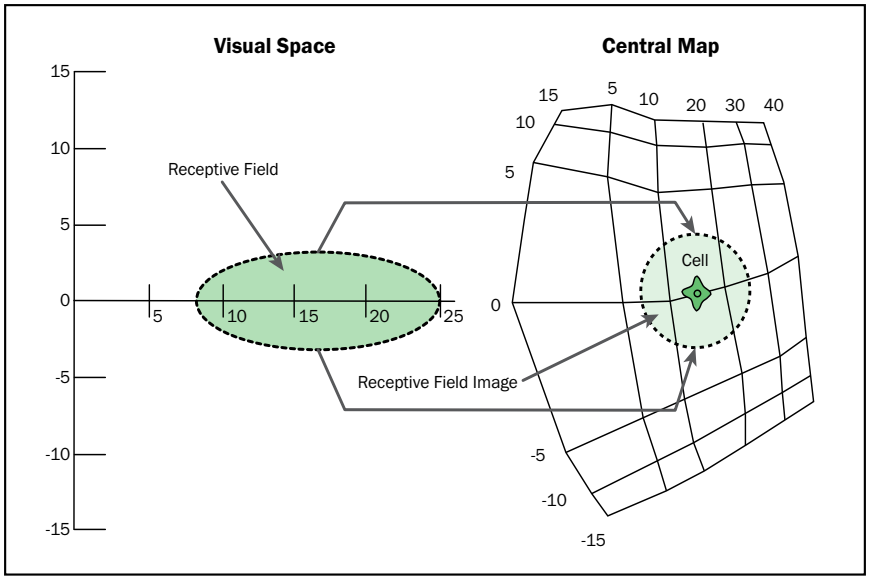


Figure 6.3

Receptive fields and their central images. The receptive field image is obtained by replotting the boundary of the field in the retinotopic coordinate system of the central structure. The operation is analogous to that of transferring the MAR points to the cortical map as in Figure 4.2.

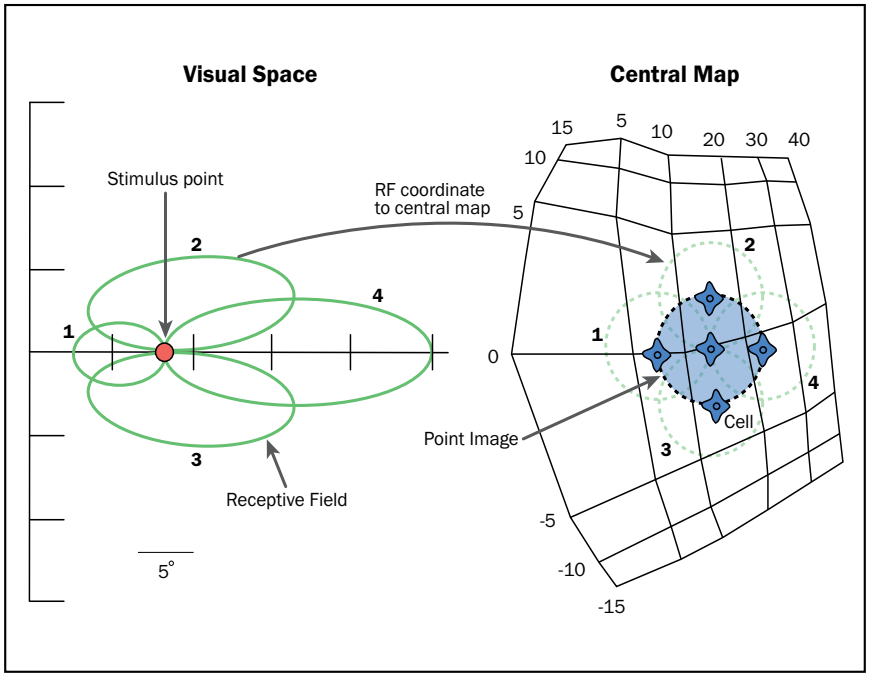


Figure 6.4

Receptive field images (dashed circles) and point image (shaded circle) in a central map. The receptive field images are obtained by transferring the boundary coordinates of the largest receptive fields recorded in four electrode penetrations to the coordinate system of the retinotopic map. (See text) Redrawn from McIlwain, *An Introduction to the Biology of Vision,* Cambridge, 1996.

the diagram shows that these RFIs turn out to be the same size and shape over the central part of the map. Even though the receptive fields vary across visual space, when transferred to the central map they become translationally invariant.

In this idealized situation we have the geometric relationship of the idealized retina of Figure 6.1, except that now a cortical cell can lie no farther from a map point than the radius of its receptive field ***image*** instead of receptive field (Figure 6.4). Any point in the central map outside its RFI will correspond to a retinal point outside its receptive field.

Once again, the cells with the largest RFIs can lie farthest from a map point while maintaining the retinal point inside their receptive fields. Thus, the cells with the largest RFIs lie at the edge of the central point image. By the same logic that holds in the retina, the point image will be identical in size and shape to the largest RFIs of the central cells.

This theoretical case demonstrates that, if the largest RFIs of cells in the central map are translationally invariant (at least locally), then the point image among these cells has the same size and shape of the RFIs—i.e., there is geometric reciprocity. In practice, this situation has been experimentally shown to hold in certain central visual structures, such as the superior colliculus. In the visual cortex, the RFIs change in size very slowly as one records farther from the foveal projection, but locally they may be assumed to be translationally invariant in size.

To transfer the retinal receptive fields to the coordinate system of the central map, one simply scales them by the local magnification factors (Chapter 5). One can estimate the size

of the point image at a given location on the visual cortex by simply multiplying the diameter of the largest recorded receptive field by the local magnification factor. This estimates the size of the largest RFI, which is geometrically identical to the point image.

Estimates of the size of the point image in the monkey's striate cortex indicate that it is larger in the foveal area than in the representation of the periphery. A point image's diameter may reach 10 mm in the foveal representation but is about 1 mm in peripheral parts of the map. Because neighboring visual points have overlapping point images, nearby points in space could both activate the same neuron. It is also clear that the foveal point image contains many complete cycles of ocular dominance columns and orientation columns, since each cycle occupies only about 1-2 mm of cortex. Thus, the analytic superiority of foveal over peripheral vision is reflected by not only a larger magnification factor but also the greater number of cells recruited to "see" a foveal image than in the periphery. The reason that the cortical point image increases in size toward the foveal representation is that the local magnification factor increases faster than the size of the receptive fields decrease, so the product increases.

Thus far we assume there was only one point image in the striate cortex. We must now assess the point image in each of the functional subsystems of the cortex, as these differ in their afferent and efferent connections and in the geometry of their receptive fields. For instance, the cells in layer 4 have small receptive fields, and information from the two eyes remains relatively separate in the map, which implies there are two small point images for a given point in the visual field. This fact may be important for stereoscopic vision or spatial acuity. In layer 5, the large size of the single point image may ultimately be related to the control of gaze through the corticotectal projection arising from this layer.

You may be wondering why the receptive fields of the four cells in Figure 6.4 have such odd shapes. Imagine that the RFIs of these cells (on the right) represent the spread of their dendrites across the central map. While the cells and their dendrites are identical in size and shape (as might well be the case), they occupy areas of the map that magnify the visual field to different degrees. Thus, the dendrites of cell #4 sweep out a larger part of the visual field, as represented in the map, than do the dendrites of cell #1, so in visual space the receptive field of cell #4 is larger than that of cell #1. The shape of the receptive field also depends on the interaction between the dendrites and the geometry of the afferent map. The dendritic arbor is not the only determinant of RFI size. Just as the receptive field of a retinal ganglion cell reflects whatever connections exist between that cell and the photoreceptors, the RFIs of the central neurons reflect the connections of those cells to the photoreceptors, complex and remote as these might be.

Point Images and Sensory-Motor Transformations in the Superior Colliculus

The superior colliculus, a midbrain structure involved in the control of gaze direction, provides an example using a point image to understand a sensory transformation. The superior colliculus receives a direct retinal input and its inactivation results in severe deficits in visual orienting responses. These orienting responses partially depend on high-velocity eye movements called ***saccades***. Most superior colliculus neurons, particularly those in the superficial layers, respond to visual stimulation and have large receptive fields. Neurons in the deeper collicular layers discharge prior to voluntary saccadic eye movements toward a restricted region of visual space called the ***movement field*** of the cell.

Movement fields can be found for any cell involved in controlling action. If a neuron fires and causes an animal to move, the movement field is mapped by noting an arrow from the default position to the final position. The movement field is analogous to the sensory receptive field, except that it is mapped by noting the endpoint of a movement and is preceded by activity from the recorded cell. The collection of these endpoints constitutes the movement

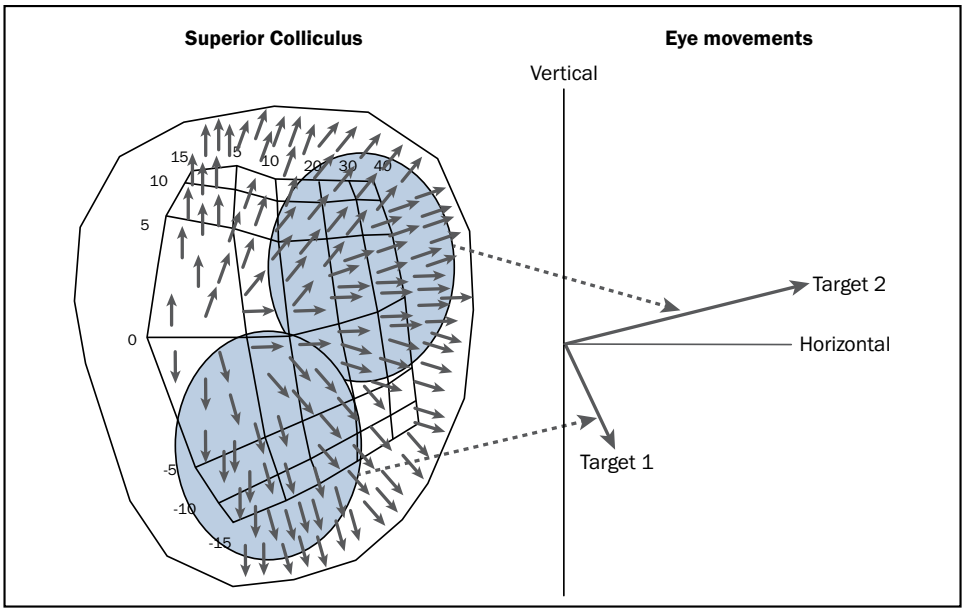


Figure 6.5

Sensory-motor transformation in the superior colliculus. Small arrows represent gradients of functional connectivity of the superior colliculus to brainstem generators of the upward, downward and horizontal components of electrically evoked saccades. An arrow symbolizes the contribution of the local site to the vector of the evoked saccade. The shaded oval profiles represent the distribution of sensory and motor activity prior to two visually evoked saccades, whose directions and amplitudes are indicated by the large arrows in the diagram at right. V-vertical, H-horizontal.

field. Many cells have both sensory receptive fields and movement fields that overlap in visual space. Some of these cells project to the brainstem regions that provide premotor input to the oculomotor neurons. Focal electrical stimulation of the colliculus produces saccades whose directions and amplitudes are related to the retinotopic sensory coordinates at the stimulus site. For example, if the receptive fields of the recorded cells are located 10 degrees to the right of the fixation point and 10 degrees above the horizontal meridian, electrical stimulation at that site produces a saccade to that location. Thus, the sensory and motor maps of the colliculus are said to be ***in register***, i.e. spatially aligned with one another.

Slight displacements of the stimulating electrode result in small changes in the vector of the evoked saccade, corresponding to the shift in sensory coordinates. These results suggest that the upward, downward and horizontal components of saccades are coded by spatial gradients in the connectivity of the colliculus to the brainstem. Each point in the colliculus contributes a component of this vector, or a sort of "***mini-vector***." The distribution of the arrows in Figure 6.5 symbolizes the pattern of these mini-vectors that would account qualitatively for the effects of electrical stimulation.

The point image in the sensory layers of the colliculus has been estimated to be quite large, perhaps occupying as much as 25-30% of the tissue. Similarly, prior to each saccade, movement-related cells in an equally larger patch of the deep layers discharge a strong burst of activity. Thus, when a point stimulus appears in the visual field, the large point image ***assembles*** just the right combination of mini-vectors to direct the eye in the right direction and for the right distance (Figure 6.5). As the stimulus or saccade target moves around in visual space, the zone of activity moves around in the colliculus, changing the composition of the output signal appropriately.

Other ways have been suggested for implementing the transformation from a large zone of activity to a precise motor command and the scheme of Figure 6.5 is shown here simply as a way of illustrating that such a transformation is possible. No one disputes, however, that the direction and amplitude of saccades are encoded by the activity of a large population of cells spread throughout a significant portion of collicular tissue.

Terms and Techniques

receptive field image	point image	coverage
sensory-motor transformation	geometric reciprocity	movement field
saccade	mini-vector	superior colliculus
translational invariance	electrical stimulation	registration of maps

Visual Object Analysis 7

Table of Contents

How objects in the external world are represented in the nervous system is a central question in neuroscience. An object (car, book, tree, person, dog, etc.) may have distinguishing qualities such as shape, size, color, weight, odor or taste, and may make characteristic sounds or movements. All of these properties, together with an individual's memories, emotions and concepts associated with the object, constitute the object's identity for that individual. Students of visual perception often divide the process of visual object analysis into two levels called low-level (or early) vision and high-level vision. The general idea is schematized in Figure 7.1.

The study of low-level vision is concerned with how the retinal image is processed by the neural machinery of the retina and central visual system. This is believed to occur through a number of channels operating in parallel to dissect the image along specific dimensions (Figure 7.1). High-level vision is carried out by machinery that attempts to make sense of the several aspects of the retinal image presented to it by the low-level processes. This may involve not only computations on the image information itself, but also associations with information the brain already has about the world.

Low-Level Visual Processing

We will begin with a discussion of how low-level visual processes can account for certain phenomena in color and pattern vision. The three normal photopigments of human cones sample wavelengths in three overlapping regions of the visual spectrum (Figure 7.2) and light of any given wavelength will affect the three cone types in a unique ratio. By encoding the

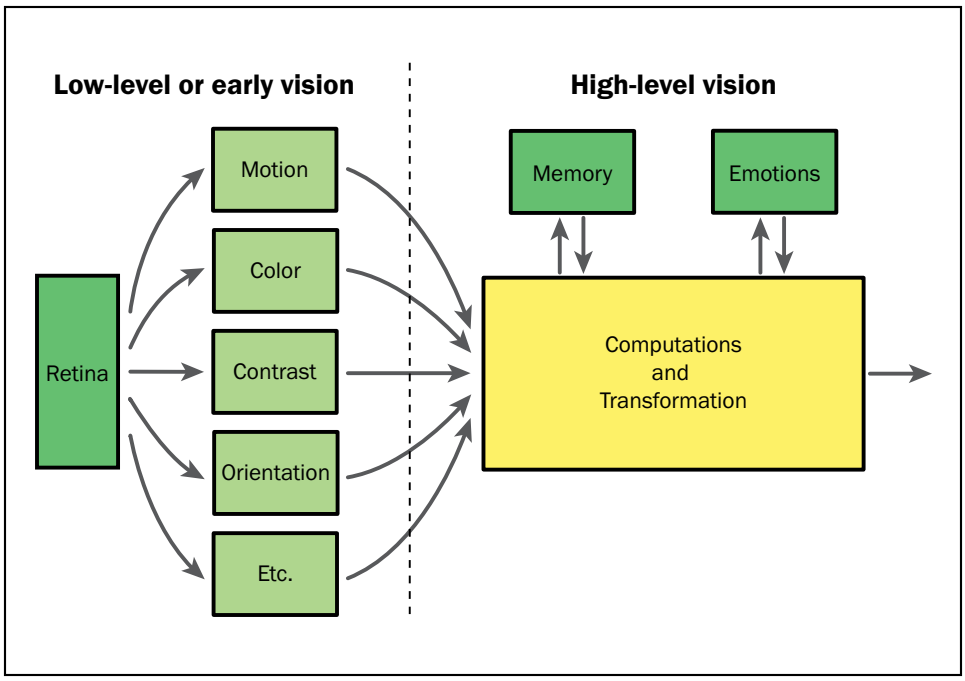


Figure 7.1

The visual process as it might be viewed by a cognitive scientist.

relative activity in each cone class, the retina can specify the wavelength present and enable the brain to distinguish it from other wavelengths.

This is a prime example of an ***ensemble*** or ***distributed code***. The output of any one cone is ambiguous with respect to the wavelength of the stimulus, but the distribution of activity among the cones is unique for each wavelength. Such a system allows the eye to cover the spectrum with only three cone types, rather than having to place a large number of narrowly tuned receptors at each retinal location. The resolving power of this system is extraordinary. Near the middle of the visible spectrum we can distinguish two monochromatic lights differing by only half a wavelength. The three classes of cones make unique connections with the CNS so that the brain knows how much a given light is affecting members of each class.

A yellow light of 550 nm is absorbed by both the long-wavelength and medium-wavelength cones in about equal proportions, and this unique ratio of L to M cone involvement is associated with the perception of yellow. If a deep red light of 600 nm is combined with a very green light of 500 nm, the combination can be adjusted in intensity to produce equal 'activation' of the L and M cones. When this is done, the ratio is the same as that produced by a light of 550 nm and the mixture appears yellow. Adding blue light to this mixture creates a situation in which all three cone types are activated and the combination appears white. These effects

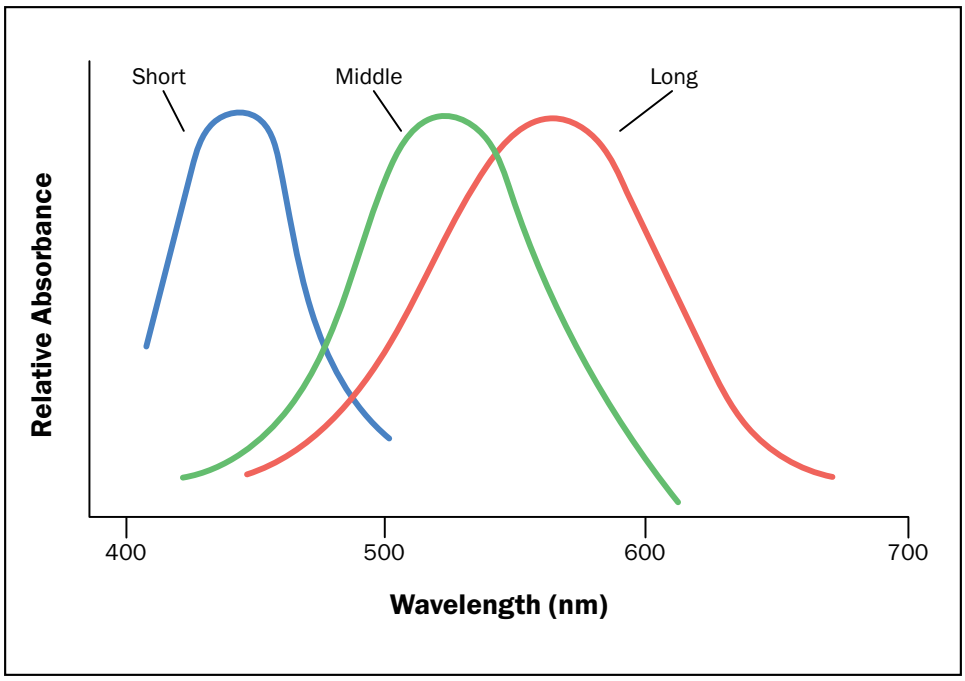


Figure 7.2

Relative spectral absorbance of the three human cone photopigments.

are completely accounted for by processes occurring at the level of photoreceptors and therefore fall in the category of low-level vision.

In the primate retina there are three major classes of ganglion cells. M-cells constitute about 10% of retinal ganglion cells, have larger receptive fields than P-cells and are very sensitive to motion and contrast. They take their name from the fact that they project to the magnocellular layers of the lateral geniculate nucleus (LGN). P-cells compose 80% of retinal ganglion cells, project to the parvocellular layers of the LGN and have relatively smaller receptive fields, which are often color-opponent in nature. The remaining 10% of ganglion cells have small cell bodies and fine axons and appear to be heterogeneous in their response properties. Some are thought to constitute a third channel, the koniocellular pathway, which synapses on cells scattered between and below the laminae of the LGN. This channel probably plays an important role in color vision, but, because little is known about it, it will not concern us here.

Both the M- and P-pathways incorporate object encoding. These both arise at the very first synapse in the visual pathway, i.e. that between the photoreceptor and the bipolar cell. A cone photoreceptor contacts more than one bipolar cell, depolarizing one class when light is present (ON bipolars) and inhibiting the other, which is depolarized when light is removed (OFF bipolars). Thus emerge the ON and OFF systems whose influence propagates throughout the

visual pathway. At the level of the ganglion cells these two systems are so deployed that they encode not only *'darker here'* vs *'lighter here'* but also, through lateral inhibitory processes, provide sensitivity to spatial contrast, i.e. *"darker here than there," "lighter here than there."* This process gives rise to the characteristic center-surround receptive field, which may be ON-center:OFF-surround, or OFF-center:ON-surround (Figure 7.3). The mutual antagonism between the center and surround of these *concentrically organized* receptive fields gives them the ability to enhance the contrast at a border between dark and light areas of the retina. This arrangement represents a form of *lateral inhibition* analogous to that described earlier for the somatic sensory system (see Figure 4.3).

On- and Off-Responses, Lateral Inhibitions and Contrast Enhancement

The functional significance of lateral inhibition together with the carpeting of the retina by ganglion cell receptive fields may be appreciated by observing what happens at a black-white

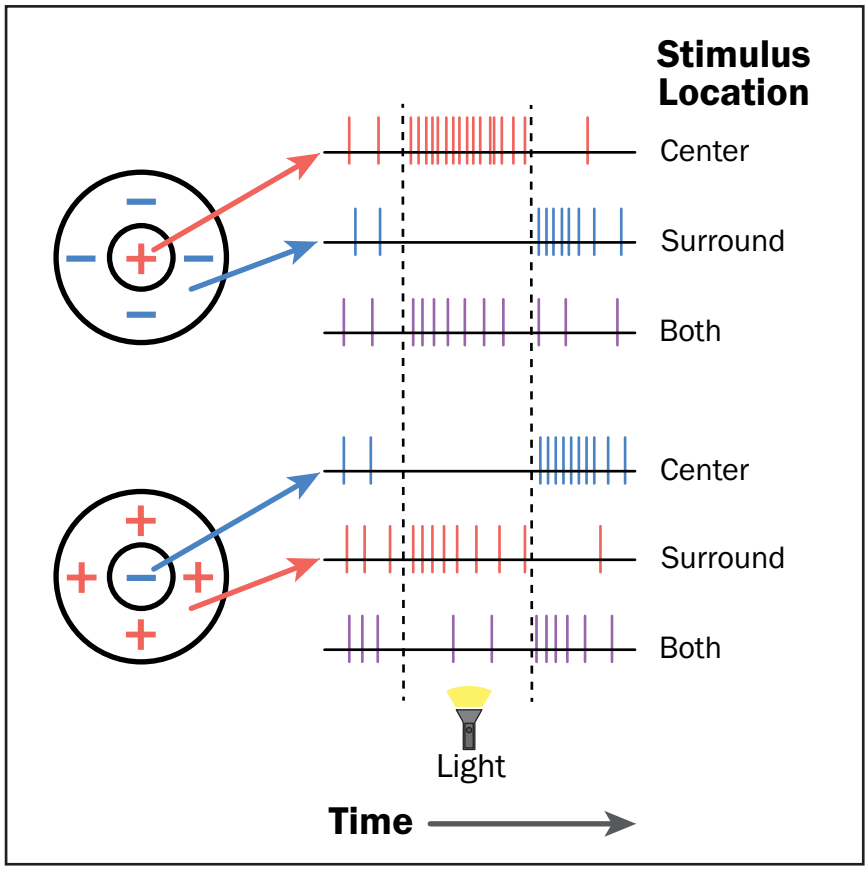


Figure 7.3

Two types of ganglion-cell receptive fields. Top: ON-center, OFF-surround field. Bottom: OFF-center, ON-surround field. Traces in each panel illustrate the effects of shining a small light spot on the center or surround and a diffuse light on both simultaneously. Plus signs, ON responses; minus signs, OFF responses. Note that diffuse light evokes weaker responses that are dominated by the center mechanism.

border. Consider a population of ON-center ganglion cells of which four are depicted in Figure 7.4. Two of these are near a black/white border and two are some distance from it. Observe that Cell A will fire at a higher rate than Cell D because the receptive field center dominates a cell's response to diffuse illumination (Figure 7.3). Furthermore, Cell B will have a higher rate of discharge than Cell A because only part of its inhibitory surround is illuminated. Conversely, Cell C will have a lower discharge rate than Cell D because one part of its inhibitory surround is illuminated.

The net effect is to create a pattern of activity that accentuates the difference between the firing rates of the ON-center cells just at the border. If an increase in the activity of these ON-center cells is interpreted by the brain as "there is more light here", then one should see a very bright zone to the left of the border and a very dark one to the right. Such zones, called ***Mach bands***, are readily observed when viewing a sharp black/white border. If one considers the carpet of OFF-center cells, and assumes that an increase in their discharge means "there is less light here", the same bands of contrast enhancement are predicted. Mach bands (Figure 7.5) can be attributed to low-level visual processes.

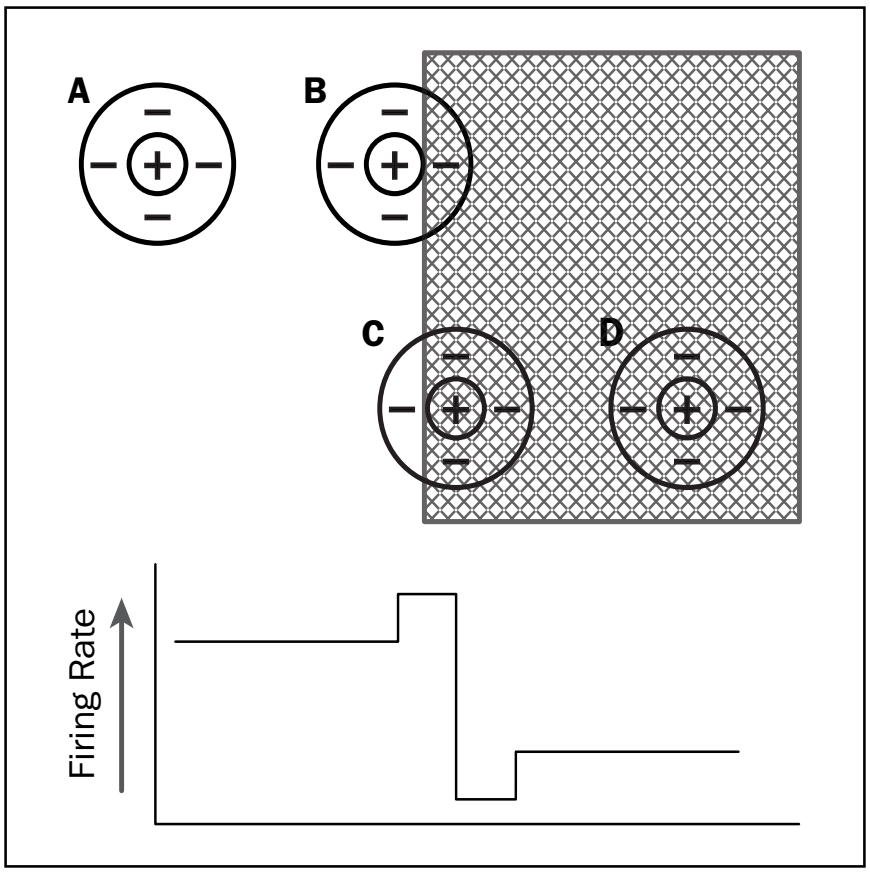


Figure 7.4

Contrast enhancement by center-surround receptive fields. See text for details. Plus signs, ON responses; minus signs, OFF responses.

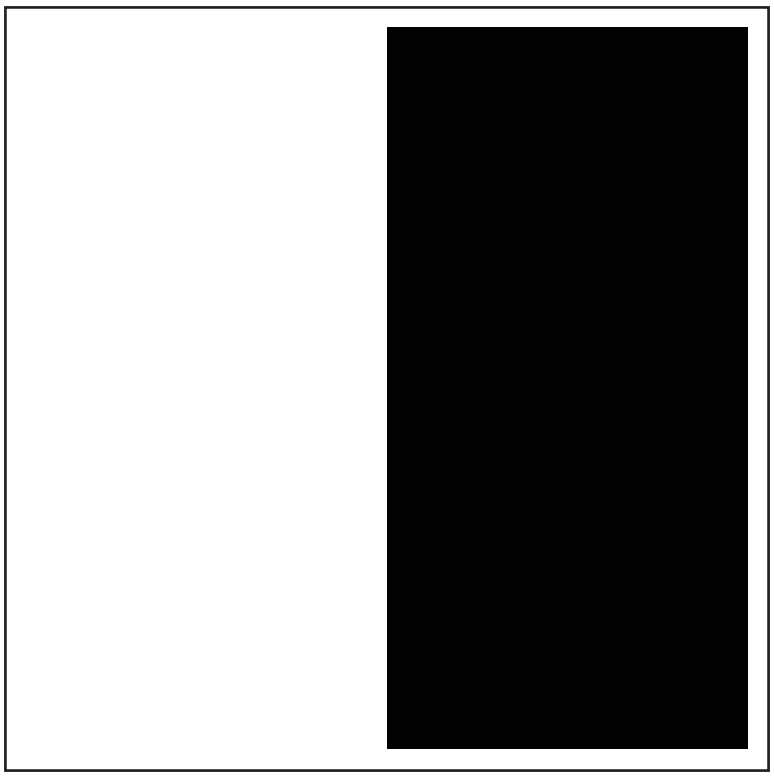


Figure 7.5

Mach band. Look carefully at the black-white border and you will see a thin, intense light line just to the left and a blacker-than-black line just to the right.

Figure 7.6 illustrates another perceptual phenomenon that can possibly be traced to the center-surround organization of retinal receptive fields. In this Hermann grid, dark spots appear at the intersection of the white strips separating the black squares. Consider the two ON-center cells with receptive fields located as illustrated. The one centered in the intersection will have a lower discharge rate than the one outside the intersection because more of its OFF-surround is illuminated. If the discharge of this cell means "light here", the brain would interpret the different firing rates of these two cells to mean "more light between the squares than at the intersections," which is what one sees. The OFF-center cells have a complementary distribution of activity but would send the equivalent message to the brain. When the Hermann grid is constructed of white squares on a black background, the illusion has the opposite contrast; there are light spots in the intersections.

Mach bands and the Hermann grid and their presumed neural substrates show how the brain enhances or makes explicit certain kinds of information that may only be implicit in the distribution of light on the retina. The receptive fields of the ganglion cells are designed to combine both spatial and luminance information in order to highlight a parameter that is critical to object segregation, namely the presence of edges. Indeed, lines that mark edges can carry the essential information in the retinal image, as is illustrated by the effectiveness

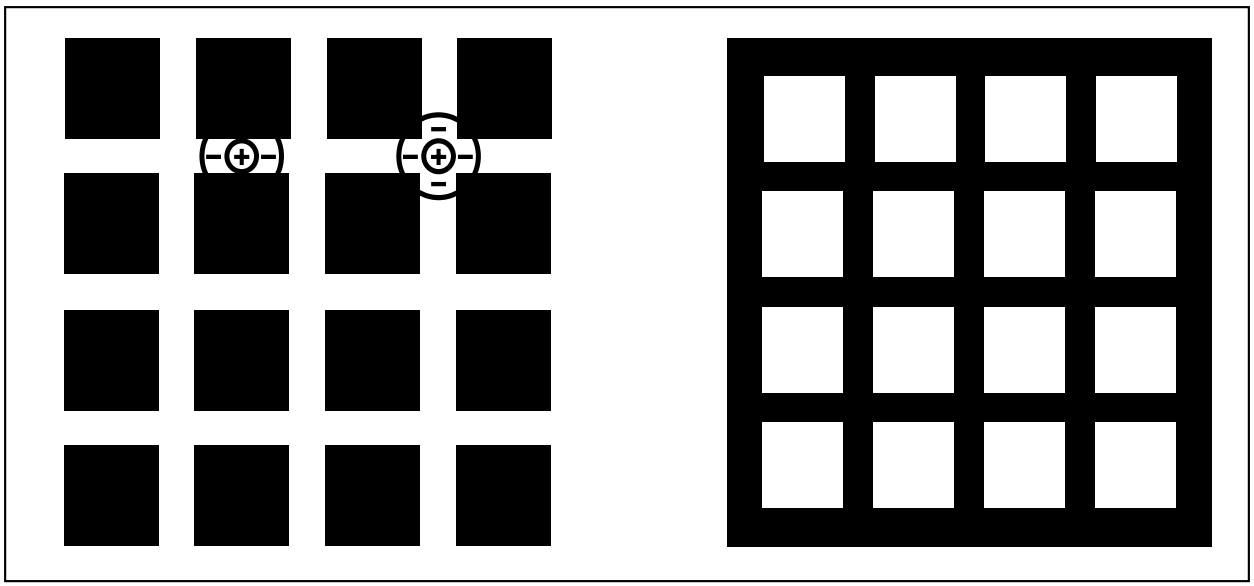


Figure 7.6

The Hermann grid. ***Left***: Note the darker zones at the intersections. The receptive fields of two ON-center ganglion cells are illustrated, one in 'the middle of the block', the other at an 'intersection' of the white lines. ***Right***: A similar illusion with light areas at the intersections.

of line drawings in cartoons and textbook figures. This aspect of an image is not encoded in the discharge of ganglion cells taken one at the time, but in the instantaneous pattern across a sheet of such cells. In other words, it is a distributed code.

The center-surround receptive field organization observed in retinal ganglion cells is preserved at the level of the lateral geniculate nucleus, the main thalamic target of the retina. In this nucleus the inputs from the two eyes remain separate and the segregation of M- and P-pathways is evident. Because of this segregation it is possible to selectively eliminate one pathway or the other for a particular part of the visual field and assess the behavioral deficits that result. This was done by Schiller who observed that the major effects of parvocellular lesions were on color and fine spatial vision, while magnocellular lesions affected motion and flicker perception predominantly.

On reaching the primary visual cortex, information in the M- and P-pathways undergoes a number of transformations, including some that combine the two inputs to varying degrees. Information from the two eyes is also combined at this level, in that the majority of cells in striate cortex can be activated by stimulation of either eye. Hubel and Wiesel identified two major classes of cortical neurons, ***simple*** and ***complex***, that are generally accepted by researchers in the field. The receptive fields of these cells are not of the center-surround type, as in the retina and LGN, but are elongated. The cells respond best to oriented lines, bars and edges (Figure 7.7). ON areas and OFF areas are spatially separate in the receptive fields of

simple cells and superimposed in complex cells. The position of the stimulus in the receptive field is critical for the simple cells, but not for the complex cell. Changing the orientation of the elongated stimulus from its optimum reduces the responses of both kinds of cell. The "complexity" of a complex cell refers, among other things, to the inability to predict its preferred orientation from the distribution of ON and OFF areas within the receptive field, while that of the simple cell can. The response of a cortical neuron to the orientation of a line or bar can be quantified by systematically varying the stimulus orientation and recording the response. When response magnitude is plotted against orientation angle, an ***orientation tuning curve*** results (Figure 7.8).

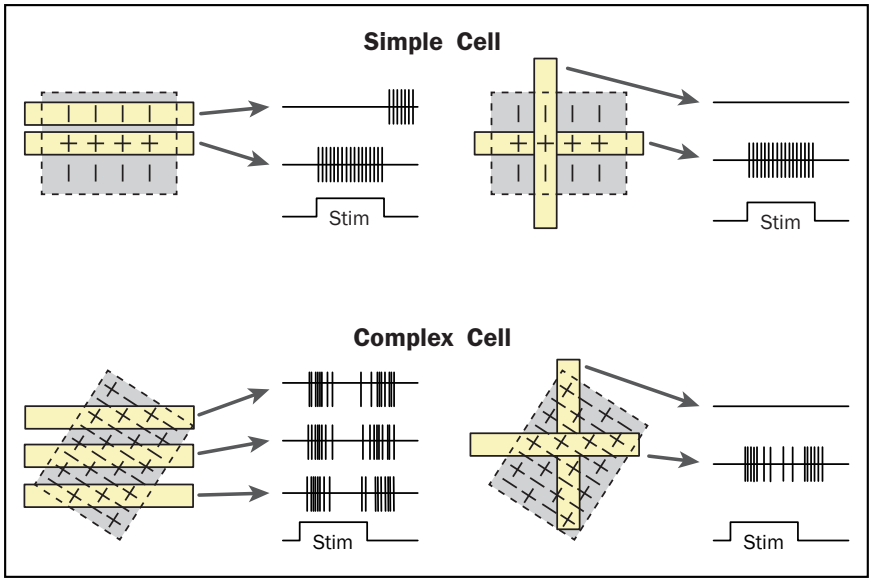


Figure 7.7

Schematic example of simple and complex cortical receptive fields. The bars represent a moving bar of light. Plus signs, ON responses; minus signs, OFF responses. ± means that stimulation at this position yields both an ON and an OFF response.

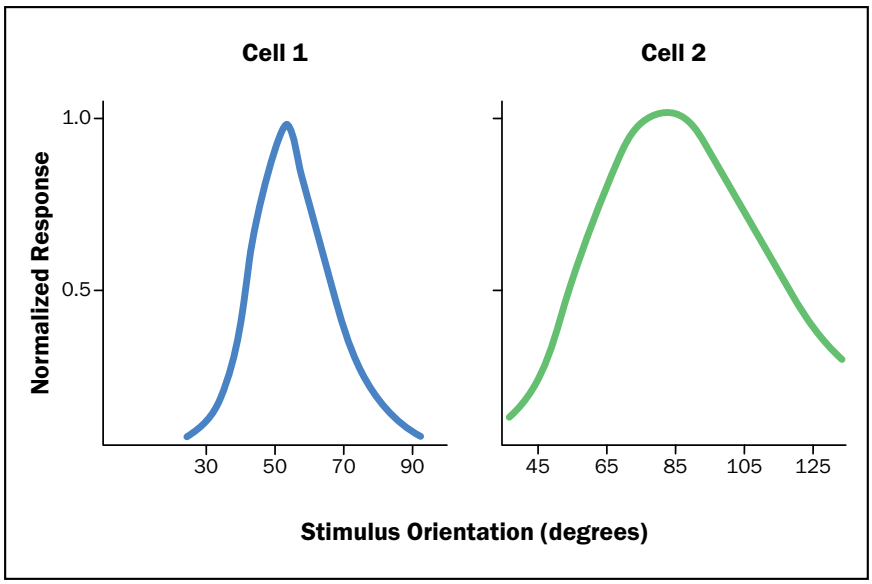


Figure 7.8

Orientation tuning curves for two simple cells. The magnitudes of the cells' responses were measured as a moving bar traversed the receptive field at different orientations. (Data from Heggelund and Albus, Experimental Brain Research. 32:197-212, 1978).

Neurophysiological studies have shown that repeated exposure of the eye to a high contrast grating of one particular orientation will reduce the responses of cells tuned to that orientation. This is said to represent "fatigue" of the cells, although the mechanism for such a decreased responsiveness is not known. If you consider a vertical grating after repeated presentation of a slightly oblique grating to which vertically orientation cells also respond, then one might expect some change in the perception of the vertical grating. Such a test is illustrated in Figure 7.9. If, indeed, the 'fatigue' of orientation-tuned cells accounts for such adaptation effects, then this phenomenon is a low-level visual process. The brain adds nothing to the information present in the retinal image and its direct effect on the cortical neurons.

Though orientation tuning in cortical neurons plays an unknown role in such adaptation effects, these findings support to what is called the "feature detector" theory of neuronal coding. According to this theory, the discharge of a tuned cell signifies the presence of its preferred stimulus. A problem with this concept is that the striate neurons respond to lines of more than one orientation (Figure 7.8), and their responses can depend on parameters such as speed/direction of motion, contrast, color and distance from the observer. No compelling evidence demonstrates that cells tuned for orientation are essential for the detection of oriented lines or that this is the only role they might play in vision.

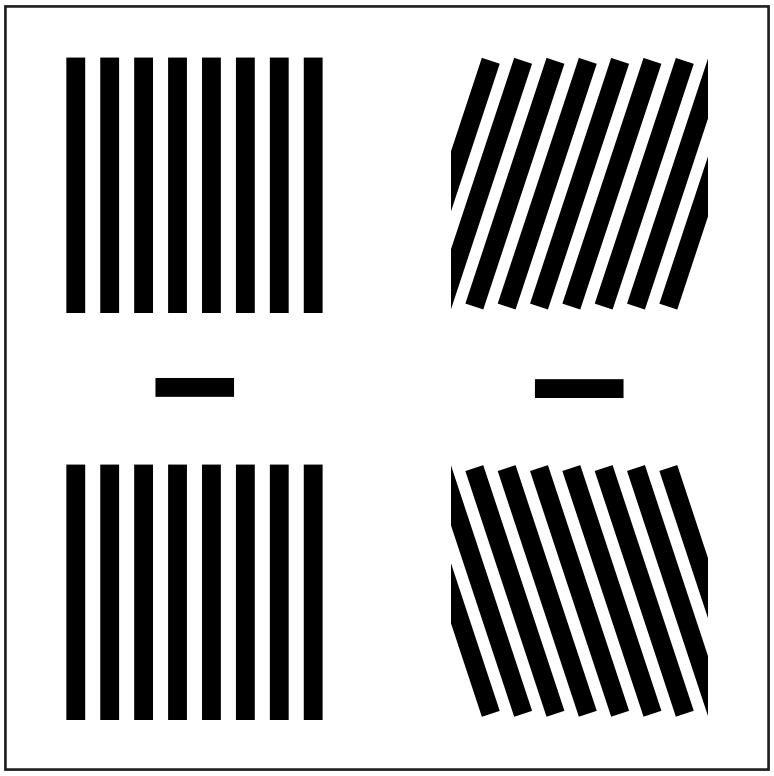


Figure 7.9

Orientation adaptation. Stare at the line between the two patterns at the right for a minute and then shift your gaze to the line between the two patterns at the left.

It should not be forgotten that an oriented line evokes activity in a population of cells distributed across a retinotopic map of visual space. The brain "knows" the form of this map and can "read" the orientation of the line in visual space from the macroscopic distribution of the active cells.

High Level Vision and the Ventral Stream

Beyond V1 visual information is channeled into two ***extrastriate*** cortical systems, which, though not completely separate, nonetheless exhibit distinct properties (Figure 7.10). Spatial information flows toward the posterior parietal cortex in what is called the ***dorsal*** or ***parietal*** stream. A ***ventral*** or ***temporal*** stream, directed at the inferior part of the temporal lobe, appears to be specialized for the recognition of objects. Note that only feed-forward connections are illustrated in Figure 7.10; for each of these there is a reciprocal projection. The two streams are further linked by reciprocal connections between many of the specific areas (see the two headed arrow between MT and V4, for example). Thus, while there is clear evidence of an anatomical divergence into two major processing streams, there is not complete isolation of either pathway. By the same token, different mixtures of M-cell and P-cell information are found in the responses of cells in the two streams. Although the dorsal stream is more or less dominated by M-cell input, some P-cell 'signatures' are observed in cellular responses. Neurons in the ventral stream reflect both significant M- and P-cell input in their responses.

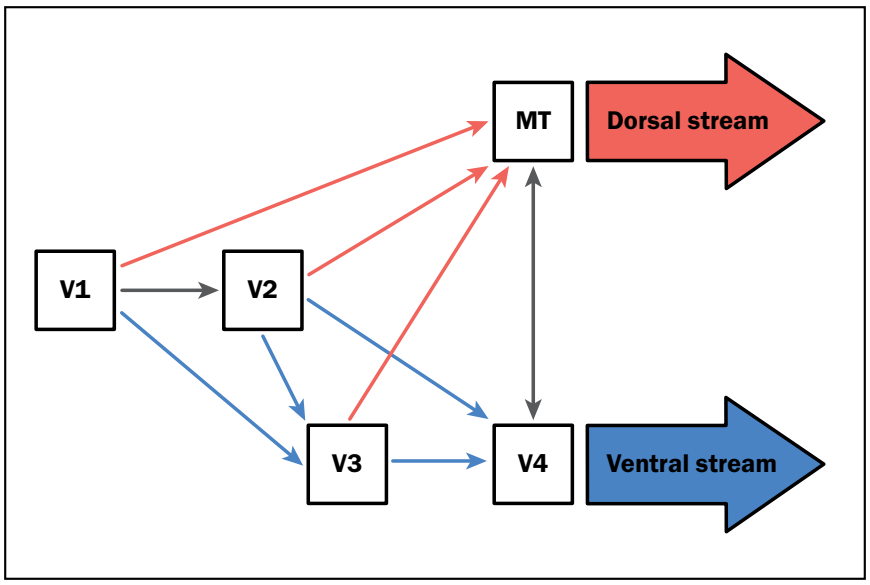


Figure 7.10

Schematic of information flow from V1 to the ventral (temporal) and dorsal (parietal) streams.

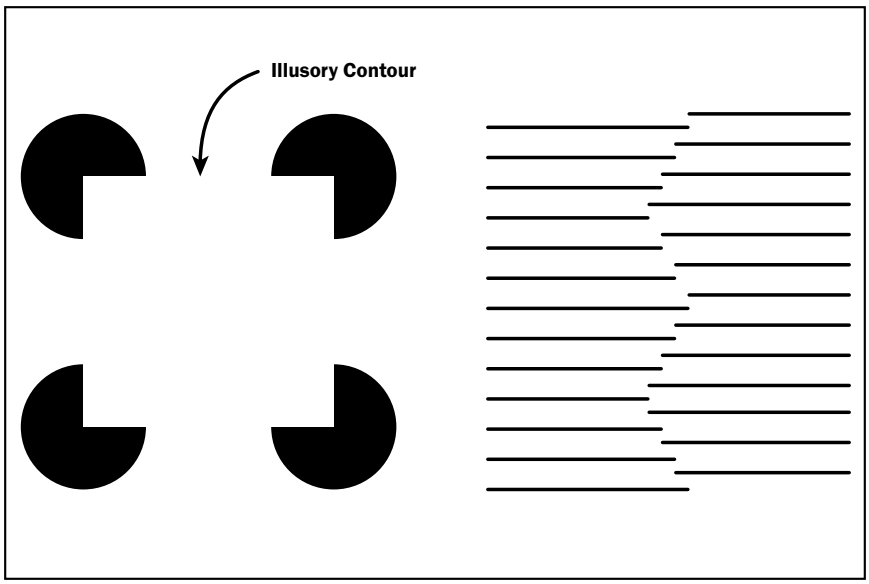


Figure 7.11

Illusory contours. Left: a Kanizsa figure. Right: An illusory contour formed by offsets of line termini appears to arise from an irregular cutting and displacement of two pieces of lined paper.

The existence of the two streams was first came to light after experimenters observed behavioral deficits following lesions in monkeys. Monkeys with lesions of the inferotemporal cortex exhibited impairment in visual pattern discrimination and recognition, but were able to perform behavioral tasks that depended mainly on accurate localization of cues. Posterior parietal lesions had little effect on pattern discrimination but severely impaired the animals' visuospatial performance. Physiological recordings in experimental animals were consistent with this kind of compartmentalization. Neurons in areas belonging to the dorsal stream have large receptive fields and are very sensitive to motion and contrast but relatively indifferent to stimulus shape and color. In contrast, color and shape are usually critical to the responses of neurons in the ventral stream, where phenomena readily associated with form vision are seen.

As early in extrastriate cortex as area V2 cells begin to exhibit properties associated with high-level vision. One such property is their ability to respond to contours that are not physically present but that are implied by the retinal image, so-called ***illusory contours*** such as those in Figure 7.11.

Figure 7.12 illustrates schematically an experiment by von der Heydt and Peterhans in which they recorded the responses of cells in area V2 of the macaque to illusory figures of the type shown in Figure 7.11. The illusory contour activated the cell when it was at the cell's preferred orientation, even when the thin lines composing the figure were ineffective in exciting the cell. One interpretation of this result is that the representation of the visual stimulus in area V2 is 'constructed,' meaning it integrates the details of the retinal image over a larger area and imposes a meaning on the encoded information. In the case of the illusory contour

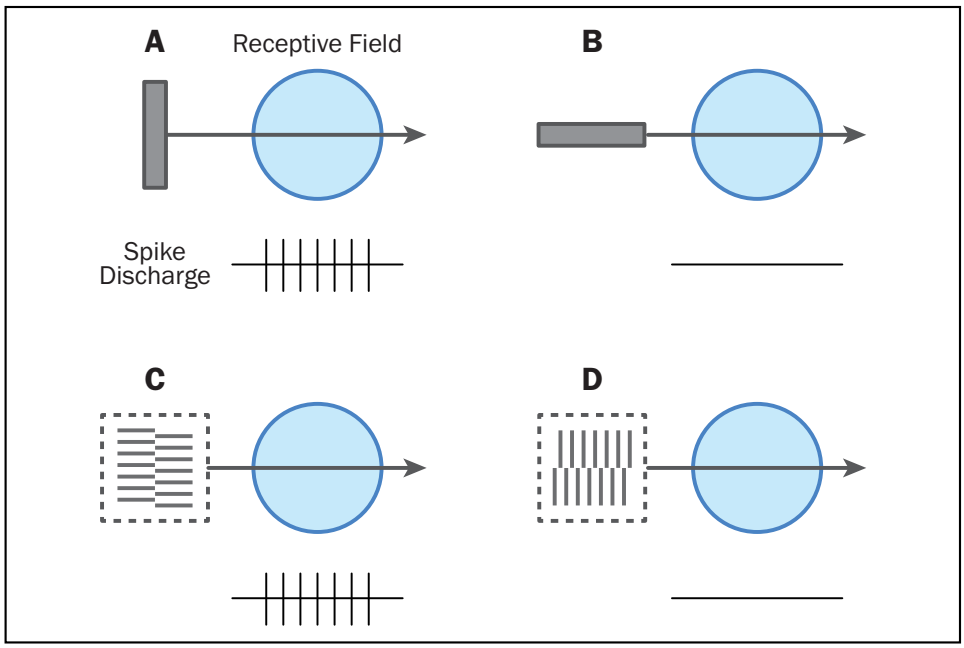


Figure 7.12

Response of a cell in V2 to an illusory contour. Schematic illustration of the experiment of von der Heydt and Peterhans, ***J. Neuroscience***. 9:1731-1748, 1989. A. Movement of a vertically oriented bar through the receptive field excites the cell. B. A horizontally oriented bar has no effect. C. A vertically oriented illusory contour activates the cell. D. The same contour oriented horizontally has no effect. Observe that the cell also does not respond to the very thin lines at its preferred orientation.

of Figure 7.12, the pattern made by the thin lines implies that they terminate at a boundary, which is essentially added to the visual perception, perhaps in part because of the responses of the cells in V2. This is an example of the phenomenon of 'filling-in' or 'completion'. Such high level integrative processes may also be responsible for the transformation of luminance information that make a gray central square appear darker against a light background than against a dark background.

Representation of Complex Visual Forms

In simian primates the ventral stream composes a number of cortical areas in the superior temporal sulcus and the inferior temporal gyrus. Certain cells in these areas respond preferentially to complex shapes, most notably faces. Because facial expressions drive primate interactions with conspecifics, it is thought that a significant amount of tissue in the ventral stream may be devoted to recognizing faces and interpreting facial expressions. It is also

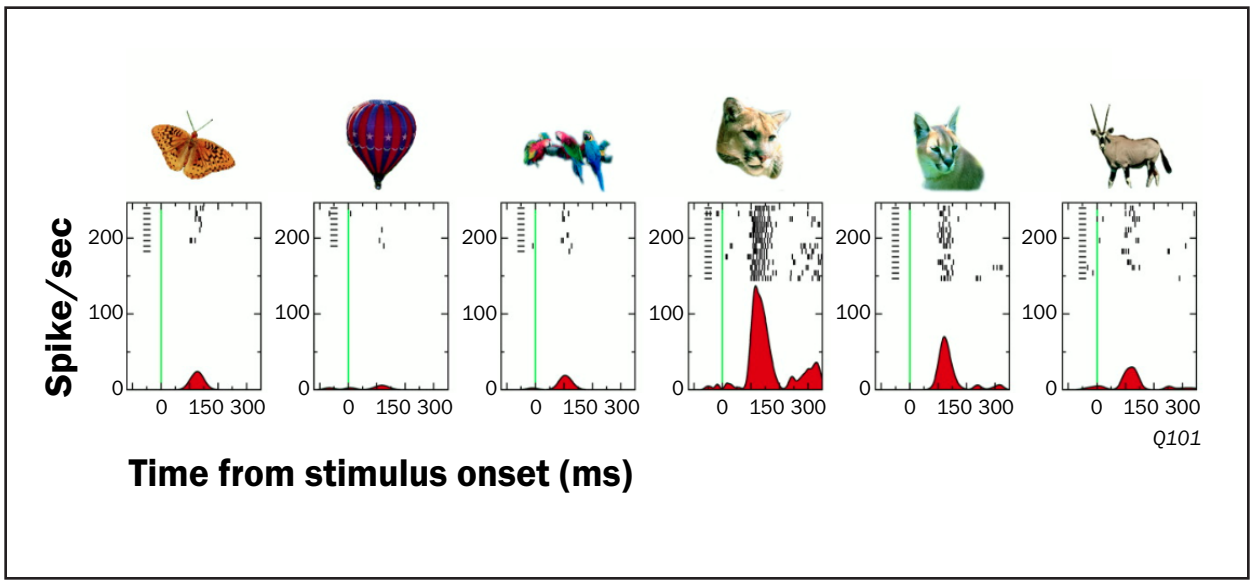


Figure 7.13

Activity of a cell in monkey inferior temporal cortex selective for complex visual forms. (Data from Sheinberg and Logothetis, *J Neurosci*, 21:1340-1350).

clear, however, that object selectivity is not limited to faces of conspecifics (as illustrated in Figure 7.13), and there is mounting evidence that the receptive fields of cells in these regions may be tuned by experience well into adulthood. This idea makes sense given that visual expertise for new objects classes can be acquired throughout one's life. Related single cell studies have also established tight links between the activity of these pattern sensitive cells and visual awareness.

The functionally homologous region of the human brain appears to lie on the inferior surface of the occipital and temporal lobes, occupying the fusiform gyrus and neighboring regions (Figure 7.14). Obviously it is not possible to record routinely from single neurons in this area or to make experimental lesions in humans. Nonetheless, studies of naturally occurring lesions (such as those caused by strokes), results from magnetic resonance imaging, and, rarely, recordings of field potentials and single unit activity have supported the correspondence between this area of the human brain and the ventral stream structures identified in monkeys.

It has long been known that bilateral lesions in the inferior occipito-temporal region of the human cortex are associated with a phenomenon known as ***prosopagnosia***, from Gr. ***prosopon*** (face) and ***agnosia*** (no knowledge). Because of its etymology, prosopagnosia is often described as a failure to recognize faces, but this is a misleading description. Faces are recognized as faces, and the individual can tell that two faces are different. What is lost is the ability to discern the identity of the individual to whom the face belongs. Thus, a patient with

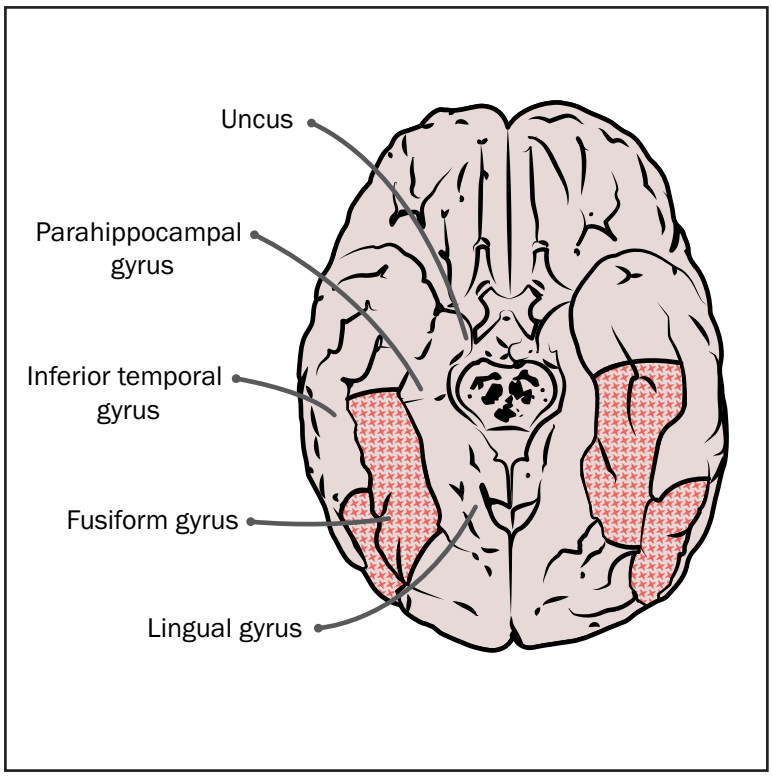


Figure 7.14

Base of the human brain. The shaded regions are the general areas where increased activity as measured by fMRI and large amplitude N200 waves are evoked by faces. Damage to these areas, particularly bilaterally, is associated with prosopagnosia. (Data from Allison et al. ***J. Neurophysiol.*** 71:821-825, 1994, and ***J. Neurophysiol.*** 74:1192-1199,1995)

prosopagnosia often cannot identify family members or famous persons from their pictures. The images are no longer familiar. The patient can, however, identify the family members by the sounds of their voices, so the complex of other characteristics (voice, history, relationship, etc.) by which known individuals are identified is still intact in memory.

Neuroscientist Antonio Damasio has argued that the identification problem extends beyond the loss of familiarity of faces because patients fail to identify objects within other classes of images as well. For example, the patient can tell a dog from a horse, but not a shepherd from a poodle or a dachshund. Also, the category 'car' remains intact, but the patient cannot distinguish one brand from another (assuming that they once could). Thus, the deficit relates to disconnection — the percept of face has been disconnected from other material that collectively constitutes the identity of a particular member of a class of objects. For anatomical reasons, a condition called ***cerebral achromatopsia*** is often associated with prosopagnosia. This is a disturbance in color vision which ranges in magnitude from a reduction in saturation (i.e. increased paleness) to complete color blindness.

Functional imaging studies in patients presented with pictures of faces also indicate activation in a part of the occipitotemporal lobe called the fusiform gyrus (Figure 7.14). In selected patients scheduled for operative treatment of intractable epilepsy, strips of electrodes

have been placed across the inferotemporal and inferior occipital cortex in humans by a team headed by Truett Allison at Yale. When the subjects viewed a series of pictures that included human faces, electronically scrambled faces, cars and scrambled cars, several electrodes recorded a large negative potential at a latency of 200 ms when the faces were shown. From its polarity and latency, the potential was named N200.

Much evidence points to the fusiform gyrus as the area for facial recognition. In the N200 experiment, the sites corresponding to the electrodes were centered in the area of the fusiform gyrus and surrounding tissue. Electrical stimulation through electrodes recording large N200s interfered with face identification. Recent functional MRI experiments have also highlighted regions of the fusiform gyrus particularly sensitive to faces (fusiform face area or FFA).

In summary, several lines of evidence point to the importance of the inferior temporal areas of the brain in the processing of visual form. Neighboring structures such as the hippocampus, parahippocampal gyrus, and perirhinal cortex are also implicated in memory storage and retrieval. These elements all contribute to the high-level visual processing postulated in Figure 7.1, whereby information provided by low-level processes is matched to the stored experience of the organism. Nearby structures such as the amygdala, known to be involved in emotion may also be a factor in this high-level processing. This structure can attach emotional valence to visual stimuli, such as particular facial expressions.

Terms and Techniques

fovea	optic disk	blind spot
optic nerve	optic chiasm	optic tract
optic radiation	lateral geniculate nucleus	primary visual cortex
Area 17	VI	extrastriate visual cortex
S-cone	M-cone	L-cone
center-surround organization	M ganglion cells	P ganglion cells
magnocellular laminae	parvocellular laminae	inferior temporal cortex
simple cell	complex cell	fusiform gyrus
ocular dominance column	orientation column	orientation tuning curve
Mach bands	Hermann grid	Kanizsa figure
illusory contour	prosopagnosia	

Auditory Processing: Central Control of Input 8

Table of Contents

Central control of sensory input is a general strategy in the nervous system, and the auditory system provides a particularly rich example of this. The initial stages of hearing can be divided into three major processes: (1) mechanically coupling sound pressure waves from air to the endolymph of the inner ear; (2) transforming this energy into vibrations of the basilar membrane; and (3) transducing these mechanical vibrations into electro-chemical neural signals. Each of these processes is subject to feedback control from the central nervous system, illustrating how the brain uses its input to modify the machinery responsible for that very input.

Coupling from Air to Inner Ear

The principal function of the tympanic membrane and the ossicular chain of the middle ear is to couple the acoustic energy from the air to the endolymph in the cochlear duct. Because of their low spatial density, molecules of air are much easier to displace than are the molecules of a liquid such as endolymph. Thus, air impedes (or resists) the propagation of sound less than a liquid and is said to have a lower acoustic ***impedance***. This property means that transfering the energy from the rarefactions and condensations of the air to a liquid is quite difficult. (A comparable situation exists when you swim underwater and the sounds coming from the air above the surface are muffled.)

The outer ear fixes this mismatch by taking advantage a principle of physics: energy transfer from one medium to another is most efficient when the impedances of the media are the

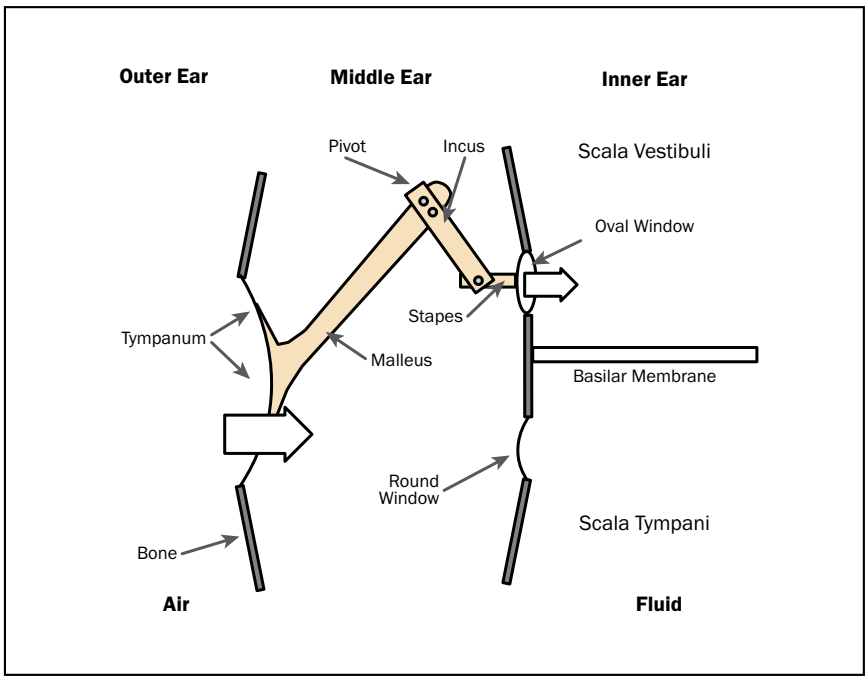


Figure 8.1

Arrangement of the ossicular chain (malleus, incus and stapes) for coupling acoustic energy from air to the inner ear. (See text for details).

same, or nearly so. The tympanic membrane and ossicular chain of the middle ear perform an ***impedance matching*** function, which facilitates energy transfer from the air to the fluid of the inner ear.

There are two mechanisms of impedance matching (Figure 8.1).The cross-sectional area of the human tympanum is about 90 mm^2, whereas that of the footplate of the ***stapes*** is only 3.2 mm^2. Thus, a force acting over the large area of the tympanum is transferred to a small piston. There is also a lever arm action between the ***malleus*** and the ***incus***, which pivot around their point of articulation. The large, low-energy displacement of the malleus results in a small, high-energy movement of the stapes. This lever action is most important at frequencies below 3000 Hz, because above this frequency the tympanic membrane does not move as a unit.

After sound travels through the tympanic membrane and ossicles, it must be transduced in the inner ear. Two small muscles in the middle ear influence the transmission of acoustic energy to the inner ear. The ***stapedius*** is innervated by cranial nerve VII (the facial nerve) and pulls the stapes footplate out of the ***oval window***, tightening the fibrous band attaching it to the temporal bone (Figure 8.2). The ***tensor tympani*** is innervated by cranial nerve V (the trigeminal nerve), inserts on the malleus and increases the tautness of the eardrum when it contracts. Some individuals are able to contract their middle ear muscles voluntarily and this has been shown to reduce the sensitivity of the ear to low frequency sounds.

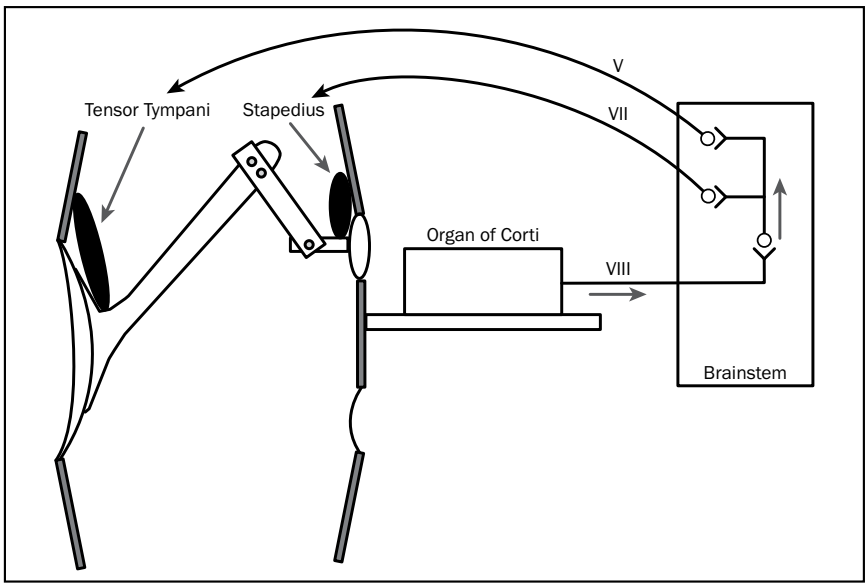


Figure 8.2

Mechanism of the acoustic reflex. The afferent limb of the reflex is cranial nerve VIII and the efferent limbs are nerve V to the tensor tympani muscle and nerve VII to the stapedius muscle. (See text).

Activation of the auditory component of the VIIIth cranial nerve causes the tensor tympani and stapedius muscles to contract, restricting the motion of the tympanum and stapes. This ***acoustic reflex*** (sometimes called the ***attenuation reflex***) has a protective function against mechanical injury due to large, low frequency displacements of the ossicular chain, which can be very damaging. However, its latency is relatively long since central synaptic connections are involved, and the reflex will not protect the ear against single loud sounds, such as when a hunter fires a rifle near the ear. The muscles are normally active during vocalization, presumably modulating the motion of the ossicular chain in response to sounds produced inside the head as well as those emanating from the environment. Thus, the circuitry of this reflex represents a ***negative feedback*** pathway that automatically adjusts the 'gain' of the tympanum-ossicular chain system to low frequency sounds in response to sound input. This is an example of ***automatic gain control***.

Following facial nerve lesions, which denervate the stapedius muscle, patients sometimes complain of ***hyperacusis***, i.e. ordinary sounds seem much louder than usual. In principle, damage to the trigeminal nerve could also cause hyperacusis by paralyzing the tensor tympani, but such lesions are exceedingly rare. Also patients recovering from facial nerve damage sometimes experience loud clicking sounds when they move their facial muscles. This may occur because branches normally destined for the muscles of facial expression instead reinnervate the stapedius.

Because the acoustic reflex is mediated by the Vth, VIIth and VIIIth nerves and the brainstem pathways interconnecting these, it can be used clinically to test the integrity of each of these components, even when the patient cannot hear or is unable to cooperate. Changing the tension on the tympanic membrane alters the degree to which it absorbs or reflects different

acoustic frequencies introduced into the ear canal. Various devices exploit this fact to measure the ability of sounds to modify the tension on the membrane caused by activation of the acoustic reflex and contraction of the tensor tympani muscle.

Motion of the Basilar Membrane and Transduction in the Organ of Corti

The basic anatomy of the cochlea and the early mechanical events in auditory transduction are only briefly described here. When the tympanum is displaced into the middle ear, the footplate of the stapes bulges into the inner ear near the scala vestibuli. This causes the round window, at the end of the scala tympani, to bulge out into the middle ear. The pressure pulse gives a flick to the basilar membrane, which sets up a traveling wave on the membrane, just as a flick of the wrist creates a traveling wave in a jump rope or garden hose. The basilar membrane is much stiffer near the stapes, at the basal turn of the cochlea, than it is near the apex at the helicotrema (where the scala tympani and scala vestibuli communicate).

The relative floppiness of the apical turns of the membrane makes them less able to transmit traveling waves set up by high frequency vibration of the stapes. The high frequency traveling waves move quite well along the stiffer membrane of the basal turn, but as they move into the more flexible apical turns they are damped out as the flexible membrane attempts to move through the surrounding fluid. Because of this gradual change in the stiffness of the basilar membrane, traveling waves due to different acoustic frequencies produce their maximal displacements at different distances from the basal (stapedial) end. Thus, stimulus frequency is coded along the length of the basilar membrane in the sense that one place on the membrane is displaced maximally by a particular frequency and not by others. This is referred to as ***place coding*** of frequency on the basilar membrane, with frequency being laid out tonotopically from base to apex. It should be remembered, though, that although only one frequency will have a maximal effect at given locus on the membrane, that locus is displaced by a large number of other frequencies as well.

Motion of the basilar membrane leads to bending of the cilia or 'hairs' on the hair cells of the organ of Corti, which modulates the membrane potentials of the cells and the rate at which they release neurotransmitter onto the terminals of the VIIIth cranial nerve. The endolymph of the scala media is about 80 mV positive to the perilymph in the other scalae. This ***endocochlear potential*** depends on a vascular area on the lateral wall of the cochlea called the ***stria vascularis***. The inside of the hair cell is about 60 mV negative to the perilymph or about 140 mV negative to the endolymph bathing the cilia. Thus, a very large potential difference exists across the apical ends of the hair cells where the cilia are located. Bending the cilia opens channels at their tips, allowing a current to flow into the cell. This current is carried by potassium ions, even though there is essentially no concentration gradient for potassium across the membrane. The potassium is driven into the cell by the 140 mV electrical gradient

established by the endocochlear potential and the transmembrane potential of the receptor cell. This is why damage to the stria vascularis by disease, which abolishes the endocochlear potential, can disable cochlear transduction and cause deafness.

The magnitude of the receptor current depends on the displacement of the basilar membrane and the resulting shear on the cilia of the hair cells. The greater the bending in one direction, the greater the depolarization. Bending in the opposite direction reduces the secretion of transmitter. The varying depolarization of the hair cells regulates the release of a neurotransmitter that excites the afferent nerve fibers and modulates their firing rates. Note that the hair cell itself does not generate spikes. It is not even a neuron but rather a specialized epithelial cell.

The Cochlear Amplifier

It has been recently discovered that motion of the basilar membrane is not simply a passive response to sound pressure waves communicated to the inner ear, but rather functions by electromechanical properties of the ***outer hair cells***. These specialized epithelial cells contain a contractile protein called ***prestin***, and as the cells depolarize and repolarize with the rise and fall of the basilar membrane, these motor proteins shorten and lengthen. This oscillation

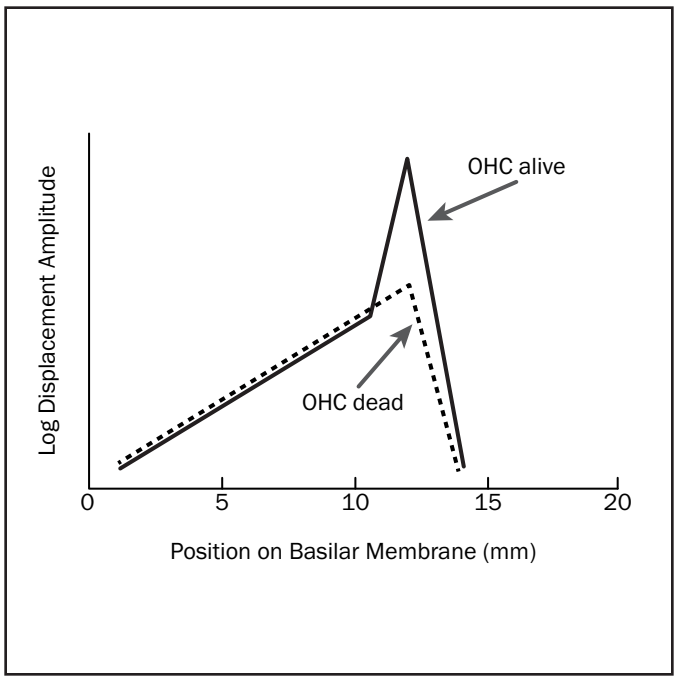


Figure 8.3

Diagram of the effect of the cochlear amplifier on motion of the basilar membrane to a particular sound frequency. When the outer hair cells (OHC) are alive, the motion of the basilar membrane is amplified at the site in the tonotopic map of the cochlea representing the cell's characteristic frequency.

in the length of the outer hair cells is coupled mechanically to the basilar membrane, resulting in an amplification of the displacement initiated by the traveling wave (Figure 8.3).

The electromechanical interaction of the outer hair cells and the basilar membrane represents a form of ***positive feedback***, because the two processes of motion and contraction mutually reinforce each other. This coupling constitutes an electro-mechanical ***cochlear amplifier.*** Obviously, this process in turn enhances the signals originating in the inner hair cells, which provide the main input to the auditory component of the VIIIth cranial nerve. (Of the 30,000 afferent auditory nerve fibers, about 27,000 innervate the inner hair cells and 3,000 contact the outer hair cells.)

The presence of the cochlear amplifier can be demonstrated by selectively destroying the outer hair cells with furosemide (Figure 8.3). This drug markedly reduces the amplitude of traveling waves on the basilar membrane and profoundly depresses the sensitivity of the ear to sound, whether measured psychophysically or by recording electrical responses from the auditory nerve and central pathways. Other drugs have been found to also destroy the outer hair cells, including several antibiotics, some of which are still employed in clinical medicine.

The outer hair cells also have other specializations that increase the local gain and augment capacity for frequency coding. Proceeding from the basal (high frequency) turn of the cochlea to the apical (low frequency turn), the outer hair cells and their cilia increase in length. The length of these elements mechanically tunes the cells in a way that predisposes them to respond best to those frequencies represented tonotopically at their location on the cochlea. In addition, interactions of several ion channels in the membrane permit these cells to generate

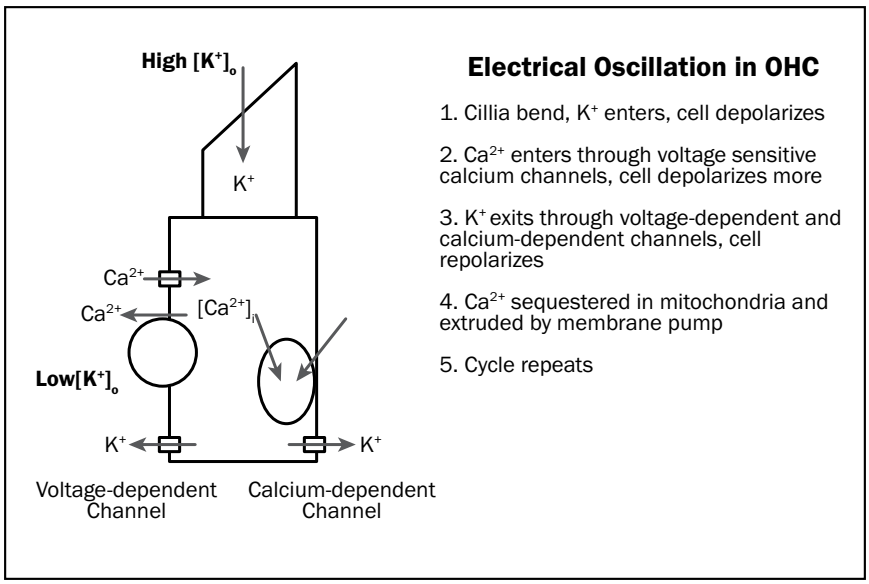


Figure 8.4

Ionic basis of membrane potential oscillations in outer hair cells. (See text)

oscillatory membrane potentials. The natural frequency of these oscillations for a given cell is also correlated with its position along the basilar membrane. This oscillatory tendency further amplifies or reinforces the electrical response of the outer hair cell to the rise and fall of the basilar membrane, thereby 'tweaking up' its sensitivity to particular sound frequencies, but not to others.

The cellular mechanism of the oscillatory membrane potential is summarized in Figure 8.4. When the stereocilia bend in their 'on' direction due to upward motion of the basilar membrane, K^+ flows down its voltage gradient into the cell and depolarizes the cell. This leads to the opening of voltage-sensitive Ca^{2+} channels and a rise in the intracellular calcium concentration, which activates the contractile proteins in the outer hair cell. The cell is therefore depolarized by potassium entry followed by voltage-dependent calcium entry.

As this phase of the cycle progresses, the depolarization and rise in calcium concentration activate voltage-gated potassium channels at the base of the hair cell. K^+ exits the cell into fluid bathing the surface of the cell that has a lower K^+ concentration. This fluid is probably perilymph that has seeped through the basilar membrane into the spaces around the basal ends of the hair cells. Meanwhile, the Ca^{2+} is actively extruded from the cell by a metabolically driven pump and sequestered by mitochondria. As a result, the membrane potential is driven in the direction of repolarization. There is a time delay between the influx of K^+ at the apical end and exit of K^+ and Ca^{2+} from the basal end, which leads to a cyclical process of depolarization followed by repolarization and so on.

Otoacoustic Emissions

Systems that employ positive feedback face a potentially serious problem: the system can run away into a state of continuous activation. This runaway feedback occurs when one speaks into a microphone that is very close to the loudspeaker; the amplified signal returns through the air to the microphone, is amplified further and eventually sets off a loud, screeching, self-sustaining oscillation.

This phenomenon can occur in the inner ear either spontaneously or in response to a properly configured auditory stimulus. The spontaneous occurrences are presumably caused by some kind of damage or abnormality that causes the membrane potential of certain outer hair cells to undergo large amplitude oscillations. Responding to this, the contractile proteins cause the basilar membrane to vibrate at a particular frequency. Because the process is self-reinforcing, it grows in amplitude and sets the ossicular chain and tympanic membrane in motion. The result is that sound comes out of the ear, a *spontaneous otoacoustic emission* or SOAE. These sounds can be sufficiently loud to be heard by someone standing nearby.

Otoacoustic emissions can also be evoked. If a sharp click is presented to the normal ear,

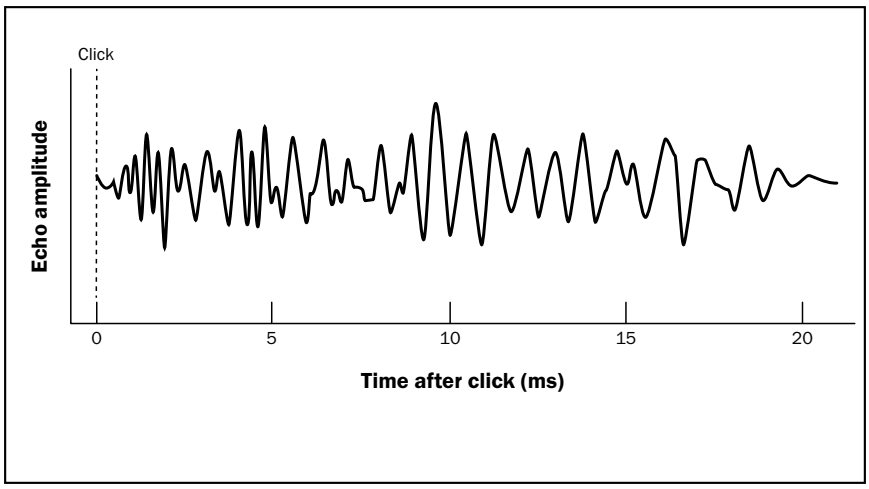


Figure 8.5

Transient evoked otoacoustic emissions produced by a click and recorded with a microphone in the ear canal. Traces are average responses to three presentations of the stimuli. (Adapted from Hood et al., *Hearing Res.* 101:113-118, 1996).

it can trigger a short period of oscillation like that just described, and the sound emitted by the ear can be recorded by sensitive microphones and analyzed (Figure 8.5). The click 'rings' the organ of Corti as a hammer would ring a bell. These oscillations die down after a while because of intrinsic damping properties of the healthy inner ear, like resistance of the fluids in the basilar membrane.

Such ***transient evoked otoacoustic emissions*** or TEOAEs have a number of experimental and clinical uses. For example, TEOAEs normally have a particular frequency distribution that reflects the widespread activation of the outer hair cells and the complex mechanical response of the basilar membrane that results. If a particular part of the organ of Corti has been damaged by disease or acoustic trauma, its characteristic contribution to the recorded TEOAE will be absent or diminished. Thus, the TEOAE can be used to confirm that a frequency-specific hearing loss has been caused by localized cochlear damage.

This technique is particularly useful in assessing the hearing of infants who cannot be tested with standard adult audiographic methods. It is very important that congenital hearing deficits be detected in children before they reach the age at which they would normally develop language. Abnormal hearing at this stage can have a devastating effect on language acquisition. So, when an infant is thought to be at risk from, say, a case of measles suffered by the mother during pregnancy, the infant can have its cochlea evaluated by the TEOAE technique. If damage is detected, appropriate intervention (e.g. fitting hearing aids) can occur before the child needs to hear in order to acquire language.

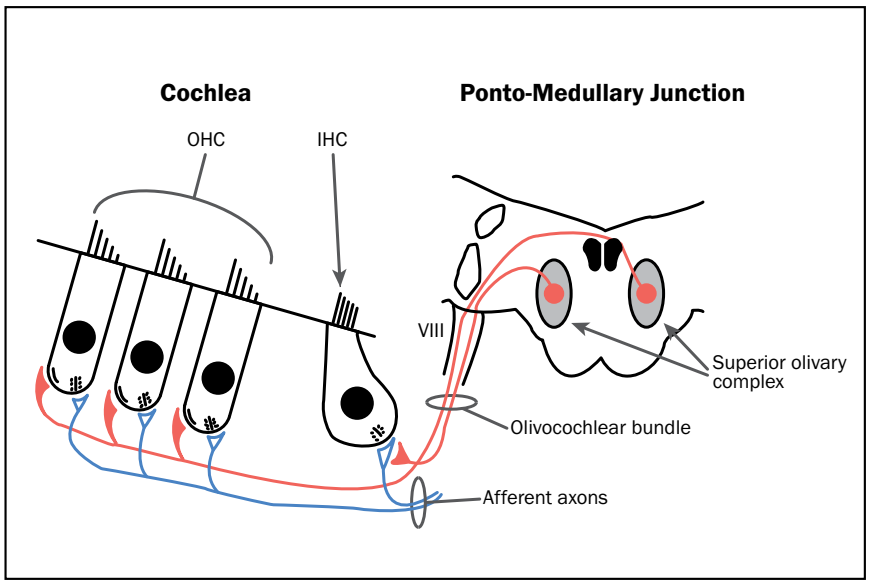


Figure 8.6

Olivocochlear bundle. This projection arises from nuclei at the ponto-medullary junction and reaches the cochlea over the VIIIth nerve where it makes synaptic connection with hair cells and some axo-axonic synapses on afferent axons. OHC, outer hair cells; IHC, inner hair cells.

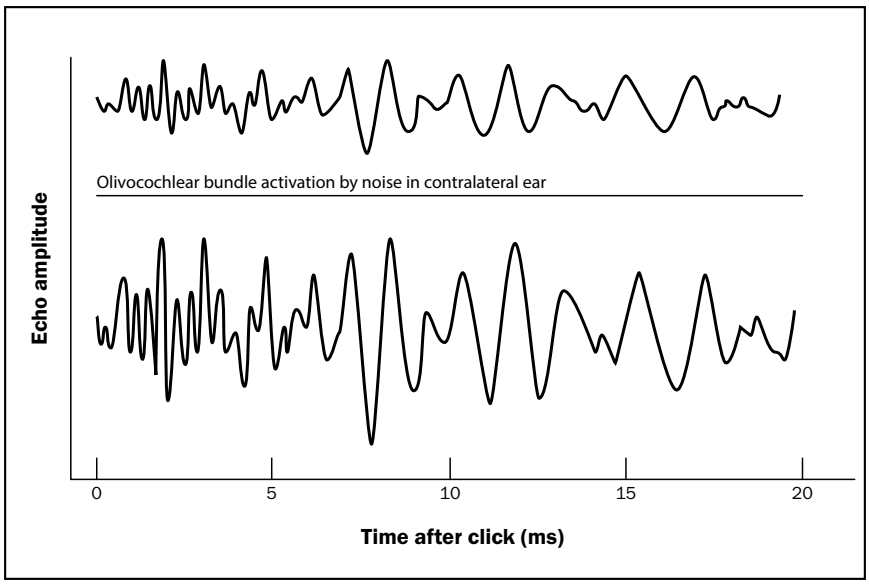


Figure 8.7

Activation of the OCB by stimulation of the one ear with noise attenuates the amplitude of the TEOAE recorded from the other ear.

Central Gain Control of the Cochlear Amplifier

In addition to carrying information from the cochlea to the cochlear nuclei, the VIII cranial also carries information from the superior olive back to the cochlea. Each auditory nerve contains some 600 efferent fibers arising from the ipsi- and contralateral superior olivary nuclei (Figure 8.6). These fibers proceed to the cochlea and constitute the olivocochlear bundle. The fibers of the ***olivocochlear bundle*** are excited by acoustic stimuli to either ear and their

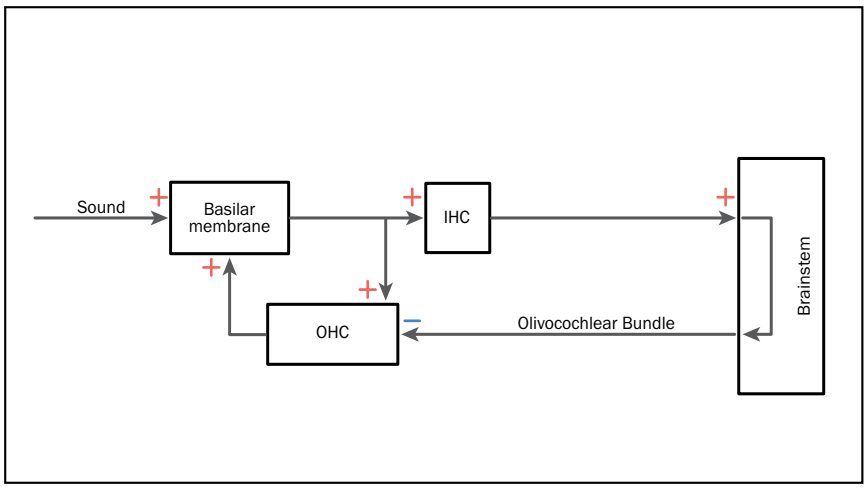


Figure 8.8

System diagram of the feedback pathways of the inner ear.

projection is tonotopic; fibers excited by high frequencies project to the basal turns while those responding to low frequencies end in the apical turns. Olivocochlear efferent fibers form medial and lateral bundles, the former targeting inner hair cells and the latter the outer hair cells. The greater part of the projection is to the outer hair cells.

Activation of the olivocochlear fibers decreases the responsiveness of the inner ear to sound. In experimental animals electrical stimulation of the bundle as its crosses the floor of the 4th ventricle reduces the motion of the basilar membrane to sound. In humans, the function of the olivocochlear system can be studied by presenting noise to one ear (activating the crossed olivocochlear bundle) while recording the TEOAE from the other ear. The contralateral noise reduces the amplitude of the ipsilateral TEOAE, which reflects the response of the ipsilateral outer hair cells (Figure 8.7).

These effects appear to be mediated largely by cholinergic synaptic inhibition of the outer hair cells and a reduction in the gain of the cochlear amplifier. In experimental animals the effects can be blocked by antagonists to acetylcholine. It is important to note, however, that the terminals of the olivocochlear fibers contain a variety of neurotransmitters and neuromodulators that may contribute to the observed effects. Also, some of the fibers terminate on inner hair cells and on the endings of the afferent VIIIth nerve fibers. Thus, the functions of this projection are complex and many remain to be identified.

Figure 8.8 represents the olivocochlear system as a negative feedback device. Sound moves the basilar membrane, which excites both inner and outer hair cells (as the membrane moves upward). The outer hair cells cause the basilar membrane to move even more (the cochlear amplifier effect). The inner hair cells also project to the brainstem, which sends back feedback signal through the olivocochlear bundle that inhibits the outer hair cells. The olivocochlear

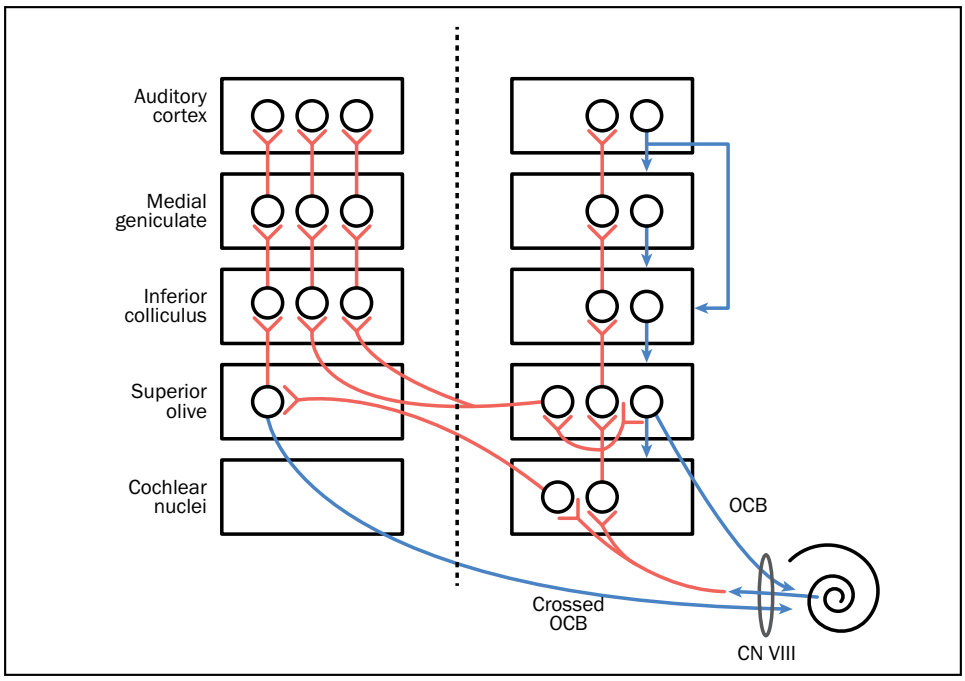


Figure 8.9

Ascending and descending (arrows) pathways of the auditory system. Several smaller nuclei and many details of the projections have been omitted. Note the origin of the OCB from the superior olivary nucleus.

system, like the acoustic reflex, represents a negative feedback pathway that allows sound levels to automatically adjust the sensitivity of the peripheral auditory apparatus.

An interesting question is whether this system can be used for consciously mediated selective auditory attention. Behavioral studies that manipulate a subject's attention to particular sounds have not succeeded in modifying either the TEOAE evoked by that sound or the ability of contralaterally presented noise to affect that TEOAE. These results suggest that the olivocochlear bundle is not involved in selective auditory attention.

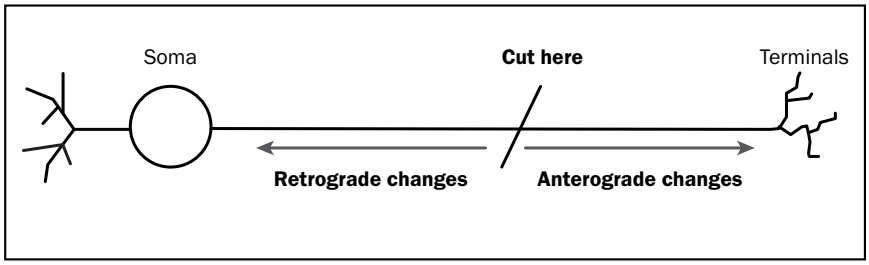
Descending Control of the Central Pathways

In addition to the olivocochlear system, the auditory system has many synaptic stations inside the CNS (Figure 8.9) and each of these is subject to control from one or more of the stations rostral to it. The functions of these central and peripheral feedback projections are not fully understood, but there is little doubt about their importance. The auditory system must deal with noisy acoustic environments, including the sounds produced inside the head

by speaking, chewing and so forth. Also, the projections may have something to do with the system's strategies for recognizing important acoustic objects, such as those contained in speech sounds, a task discussed in Chapter 9.

Technical Appendix A: Effects of Cutting a Peripheral Nerve or Central Pathway

One of the standard techniques used in neuroscience is to interrupt pathways to, from and within the central nervous system. This method may be useful in answering a number of questions, so it is important to understand what happens to the nerves and their cell bodies.



Appendix Figure 8.1

Effects of cutting a nerve.

If the nerve is cut at the position shown in the figure above, certain changes occur. First, the axon usually seals at the cut, so that the membrane potentials are maintained in the segments on either side of the cut. Natural or electrical stimulation of the distal segment (farthest from the soma) will continue to elicit action potentials, at least for a while. If this were a motor nerve, electrical stimulation of the distal segment would continue to cause the muscle to contract. Similarly, electrical or synaptic activation of the proximal segment will continue to activate the cell body. (Remember that an axon can conduct action potential in both directions). Eventually (often over a matter of many hours) the distal segment will degenerate (*anterograde* or *Wallerian degeneration*) because of the interruption in anterograde axonal transport will prevent the distal segment from receiving necessary proteins.

The proximal segment and cell body will remain functional as long as the cut is not too close to the cell body. The metabolic and genetic machinery of the cell body is activated by the interruption of axoplasmic flow, presumably in an effort to restore the missing distal segment through regeneration of the axon, a process requiring weeks or months. However, if the cut is made too close to the cell body (on the order of mm), the cell body may die (***retrograde degeneration***). This process also may take hours.

Consider now a case in which the VIIIth cranial nerve is cut somewhere between the cochlea and the brainstem. This nerve contains both afferent and efferent axons. The axons going to the CNS (afferent axons) will die on the CNS side of the cut, and the olivocochlear bundle will die on the inner-ear (distal) side of the cut. As long as the cut is far enough away, the cells of the spiral ganglion of the cochlea will continue to respond to appropriate stimulation, but the signals will not reach the CNS. The distal axons of the olivocochlear bundle will die, eliminating the negative feedback to the organ of Corti, but the cells of origin of the OCB in the superior olive will continue to respond to sound input to the contralateral ear.

Technical Appendix B: The Decibel (dB) Scale

As discussed in Chapter 3, the relationship between stimulus intensity and subjective sensory experience is generally nonlinear. Like other sensory systems, the auditory system has an extremely wide dynamic range. Humans perceive approximately equal increments in loudness for every 10-fold increment in sound intensity. The ***decibel*** (dB) is used to express sound levels in a way that scales more closely with perceived loudness. It is important to note that the decibel scale is a relative (ratio) scale. The dB relation for intensity, I, compared to a reference, I_0, is:

$$dB = 10 * \log_{10}\left(\frac{I}{I_0}\right)$$

For sound levels, it is conventional to use as the reference the level corresponding to the lowest intensity detectable by the human ear. Note that because sound is most often measured in units of pressure and because intensity is proportional to pressure squared ($I \propto p^2$), dB sound pressure level (SPL) is commonly expressed in terms of pressure:

$$dB\ SPL = 10 * \log_{10}\left(\frac{p^2}{p_0^2}\right) = 10 * \log_{10}\left(\frac{p}{p_0}\right)^2 = 20 * \log_{10}\left(\frac{p}{p_0}\right)$$

Terms and Techniques

malleus	incus	stapes
basilar membrane	organ of Corti	traveling wave
tympanic membrane	middle ear	inner ear
scala media	scala vestibuli	scala tympani
endolymph	perilymph	helicotrema
spiral ganglion	acoustic reflex	round window
oval window	inner hair cells	outer hair cells
endocochlear potential	stria vascularis	hyperacusis
positive feedback	negative feedback	olivocochlear bundle
otoacoustic emissions	cochlear amplifier	automatic gain control
SOAE	TEOAE	decibel (dB SPL)

Acoustic Object Analysis

9

Table of Contents

Verbal or spoken communication is based on a stream of ***acoustic objects*** (words or syllables) formed of three elements: frequency, intensity and time. Figure 9.1 illustrates such a stream in the form of a ***spectrogram***, i.e. the spectral profiles of the objects over time. In this chapter we will discuss how the nervous system analyzes an acoustic stream.

As in all other sensory systems, two basic mechanisms are used in the auditory system to encode an increase in stimulus intensity: recruitment of fibers to activity and increased discharge frequency in fibers already recruited. Frequency is represented in at least two ways

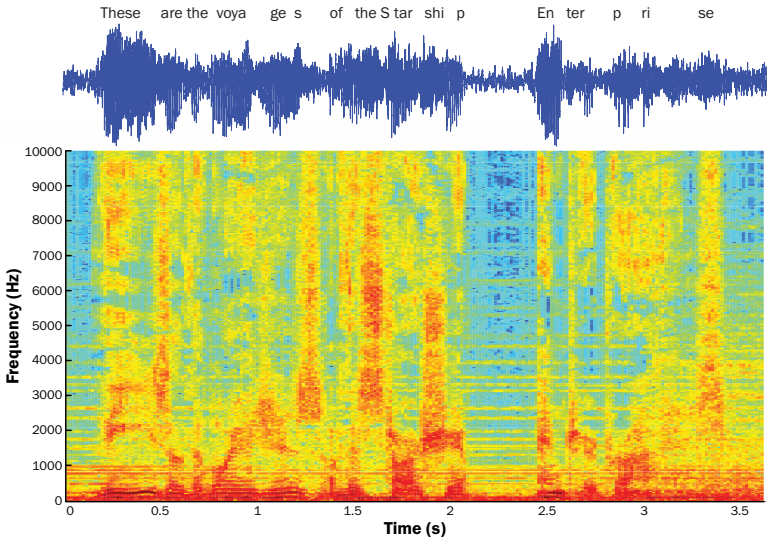


Figure 9.1

A sound spectrogram of human speech. The top trace plots sound intensity as a function of time. The graph plots time on the abscissa and acoustic frequency on the ordinate. The hotter the color in the spectrogram, the more intense the sound.

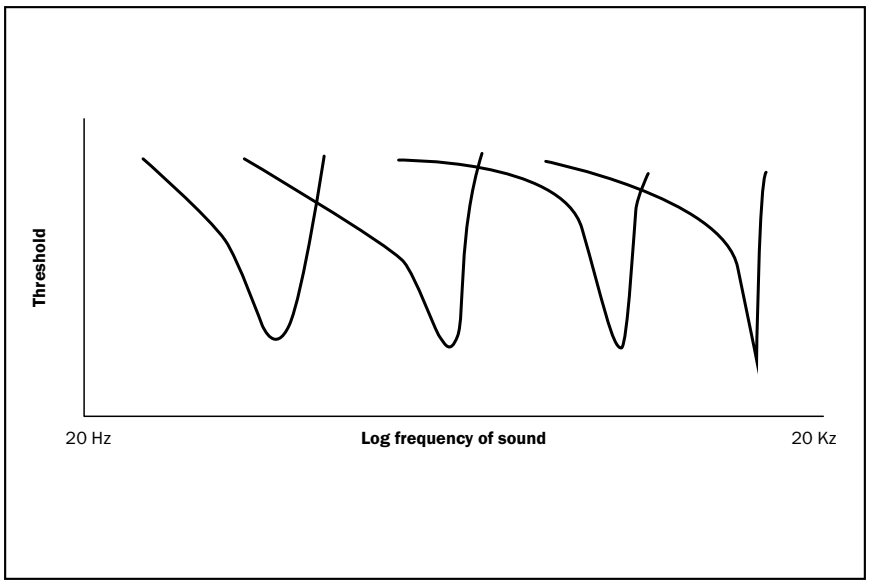


Figure 9.2

Schematic ***spectral tuning curves*** for four VIIIth nerve axons. Sounds of different frequency are presented and the intensity identified which just produces a reliable response. This threshold intensity is a measure of the sensitivity of the cell. The lowest point of the tuning curve locates the ***characteristic frequency*** or ***best frequency*** of the cell.

in the lower reaches of the auditory pathways: ***place coding*** discussed briefly in Chapter 7 and ***volley coding*** described below. Place coding appears to be the dominant mechanism for frequencies above about 5000 Hz, while volley coding plays an increasingly important role at frequencies below about 400 Hz. Between these two regions both mechanisms appear to contribute.

Place Coding

Place coding refers to transduction based on a stimulus's spatial location over some intrinsic map. Transduction is topographic—neighboring points in the map are coded at neighboring points in a structure. The retinotopic map in V1 is an example of place coding.

As described in Chapter 7, the properties of the basilar membrane and its fluid environment result in the formation of a tonotopic map of frequency along the basilar membrane. The projections from the organ of Corti to the brain preserve this spatial order, so frequency continues to be encoded as place in the various relays of the auditory system from the medulla to the cortex. The major synaptic stations in this pathway are the cochlear nuclei, the superior olive, the inferior colliculus, the medial geniculate nucleus and Heschl's gyrus on the superior aspect of the temporal lobe.

While the place coding mechanism is topographic, it is far from 'point-to-point.' As early as the VIIIth nerve itself, frequency information is distributed among a number of fibers, i.e. as a

population or ensemble code. This phenomenon is apparent if one records from several axons in the nerve and plots their thresholds, varying the frequency of the stimulus and determining the minimal intensity of sound required to produce a response (Figure 9.2).

These tuning curves may be thought of as the 'receptive fields' of the cells in the space of acoustic frequency. Like the receptive fields of cutaneous afferents, they increase in size up to a point as the stimulus gets stronger. A suprathreshold stimulus will clearly activate many cells because of the overlap of their tuning curves. It is estimated that a 1 KHz stimulus at 40dB above threshold excites about 25% of the fibers in the auditory nerve.

As the stimulus increases in intensity, more and more axons are recruited into the active zone, but the pitch does not change. This means that frequency is signaled by the location and the composition of the patch of excited fibers in the auditory nerve. When the frequency shifts, the patch changes position in the population of auditory nerve fibers (Figure 9.3) Note that the axes in Figure 9.3 have changed from Figure 9.2.

Tuning curves like those schematized in Figure 9.2 are generated when acoustic stimuli at single frequencies are presented one at a time. Thus, they do not reveal possible interactions between two stimuli presented simultaneously. It is now certain that such interactions occur even at the cochlear level. Complex stimuli such as speech or music are encoded as a dynamic pattern of peaks and valleys of activity along the length of the organ of Corti. Damage to a local area may affect single frequency testing differently than when multiple frequencies are present simultaneously. For example, a person with cochlear dysfunction may complain that voices are hard to understand or that familiar music sounds dissonant and unpleasant. The damage has essentially established a new neural code that represents familiar sounds in new

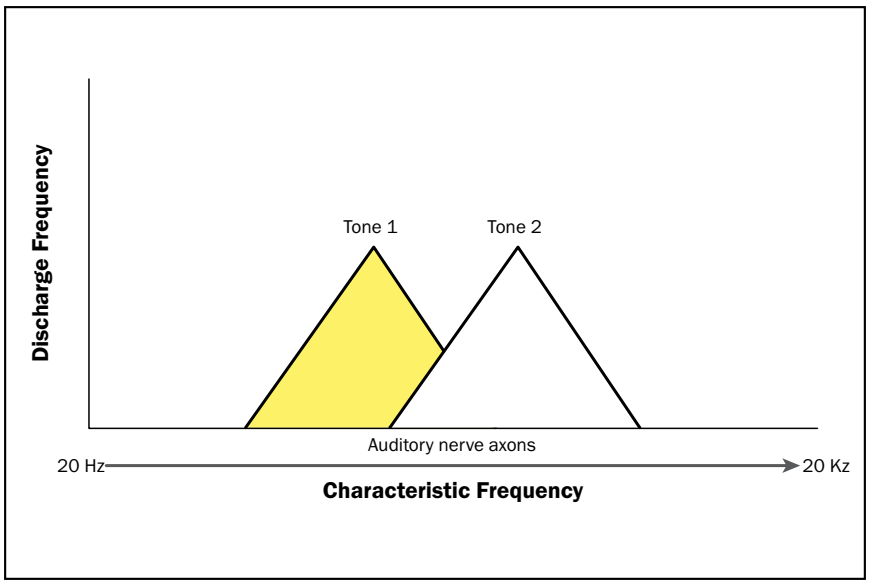


Figure 9.3

Ensemble coding of acoustic frequency in the VIIIth nerve. The nerve fibers are laid out in this diagram in order of their characteristic frequencies. The height of the envelope of active fibers represents the intensity of discharge. As the input frequency shifts, the active population migrates to a new position. Neighboring tones are encoded by overlapping populations of nerves.

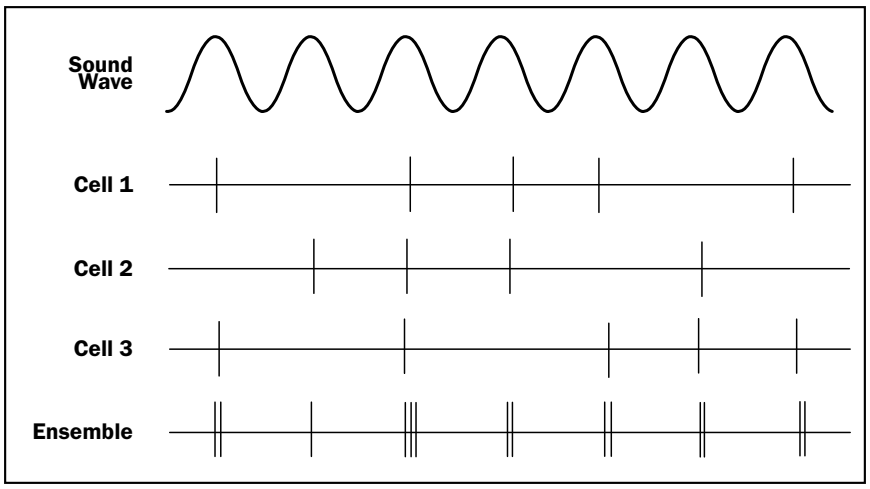


Figure 9.4

Volley coding in the auditory nerve. The sound wave intermittently activates the three nerves but the action potentials are phase locked to the sound. Together, the discharge of the three nerves faithfully reproduces the frequency of the sound wave.

and unfamiliar patterns of neural activity.

Volley Coding

Evidence suggests that at stimulus frequencies below a few hundred Hz, the CNS uses a second parameter of the neural discharge to encode frequency. In this frequency range, action potentials in a large number of auditory nerve axons become phase-locked to the signal, occurring at identical times in each cycle (Figure 9.4). This produces volleys of impulses that collectively have the same period as the acoustic signal. The nervous system appears to be able to time the volleys and extract frequency information.

The auditory system can locate sources of sounds with great precision, suggesting it is exquisitely sensitive to temporal differences. Similar sensitivity is required for making fine distinctions among complex acoustic objects, such as the difference between voiced and unvoiced consonants of words, e.g. between the initial sounds of the words ***bat*** and ***pat*** or ***van*** and ***fan***, which differ only in the delay between the release of the consonant and the onset of the vowel. Echolocation in bats also relies on this capacity for resolving temporal cues in the microsecond range and below.

The extent to which the mechanisms used in sound localization are also employed in acoustic object encoding is unknown, but some evidence suggests its role is significant. For example, the primary medullary structures involved in localization, the superior olivary nuclei, are also part of the main ascending pathway for conscious auditory perception. This finding suggests that sound localization machinery might also be important for resolving speech.

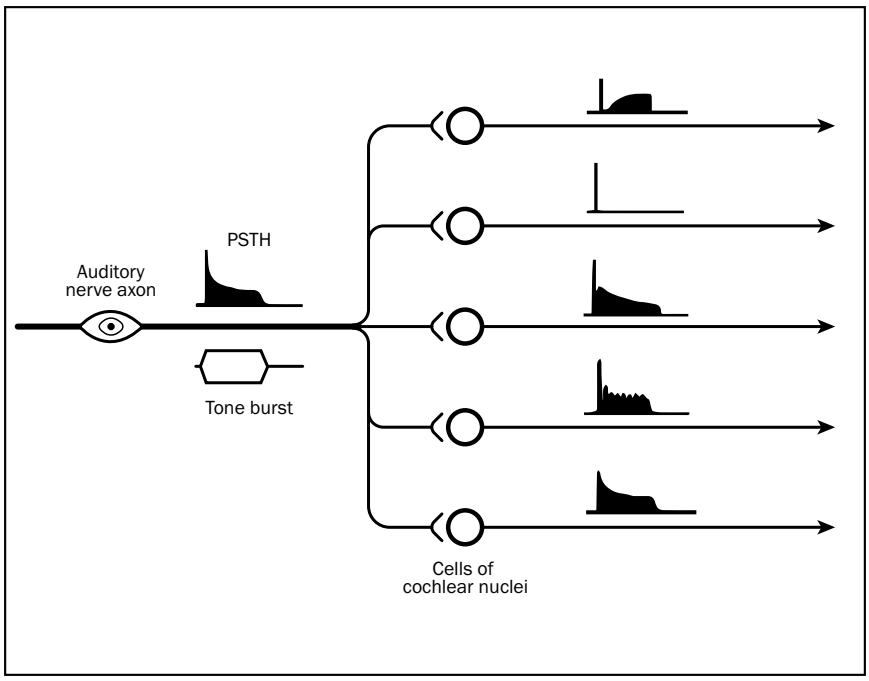


Figure 9.5

Schematic of responses of cells in the cochlear nuclei. (Redrawn from N. Kiang, Stimulus representation in the discharge patterns of auditory neurons. In *The Nervous System: Human Communication and Its Disorders*, Vol. 3, E.L. Eagles, ed. Raven Press, New York, 1975).

Transformations at Higher Stages in the Auditory Pathway

Acoustic information encoded in the discharge of auditory nerve fibers undergoes detectable transformation at the very first relay in the brainstem, the cochlear nuclei. For example, as schematized in Figure 9.5, a tone burst produces a phasic-tonic pattern of discharge in the VIIIth nerve fiber that is transformed into a variety of patterns by the cells of the cochlear nuclei. These cells appear to be seizing upon important features of the incoming signal and isolating them for presentation to higher levels of the auditory system.

It is as though the acoustic object, i.e. the tone burst, is being 'viewed' through a range of ***filters*** that discern features that are importance for identifying the object. (Filter, used in this sense, refers to the selective transmission by the cell of a particular aspect of the acoustic signal). This selective processing is not unlike what happens in the retina as the ON-center and OFF-center M and P cells modulate signals transmitted to them from the photoreceptors via the bipolar cells. This process of transformation probably continues at each step in the ascending auditory pathway, modulated perhaps by the descending projections discussed previously (Figure 8.9).

Cortical Processing and Categorical Perception

There is good evidence that the human brain scans the acoustic stream of speech and extracts linguistically significant objects it knows ***a priori.*** For example, one can test a person's ***categorical perception***, or the ability to identify stimuli and place them into discrete categories. This ability can be demonstrated in the laboratory by manipulating the spectral and temporal patterns of artificially produced sounds, such as 'ba' and 'pa', to make them vary by small acoustic steps. Under these conditions, the psychological transition from perceived 'ba' to 'pa' takes place only within a very narrow window of the physical continuum, although the subjects can perceive the differences in different versions of 'ba' and 'pa.' In other words, the acoustic structure of the sounds can vary over a wide range, but the brain seems to put the sound in either the 'ba' box or the 'pa' box, as though these are the only two objects it is familiar with and inclined to recognize.

This categorization allows the brain to compensate for wide variations in the frequency and volume range of a speaker. Thus, the decoding apparatus does for the acoustic objects of speech what the visual system does for physical objects seen at various distances and positions: it recognizes and categorizes the patterns despite a wide variation in the physical acoustic signals themselves.

Locating the Sources of Sounds

The brain uses interaural differences in timing and intensity to identify the direction from which sounds are coming. At frequencies above 2 KHz, the head produces a "shadow" which results in intensity differences at the two ears. These provide a major clue to direction of the sound source. The nervous system also detects time delays in the microsecond range because sound from a source off the midline reaches one ear slightly before the other. This second phenomenon--known as interaural time delay--can also occur with continuous sounds, even though we may not hear the onset of these sounds. This is because a continuous tone played from a source off the midline will be in different phases when arriving at different ears. As long as one cycle of the tone is greater than the distance between your ears (which is true for frequencies below ~2 KHz), the brain can use the phase difference to determine which ear is closer to the tone's source.

When stimuli are fed independently to the two ears by earphones, clicks of variable intensity and interaural delay can be presented. It is found that intensity and time delay cues may be played-off against each other. Thus, a small time delay, which causes the sound to appear to come from the right, can be counteracted by making the click to the left ear louder.

The above-mentioned cues are useful in determining the horizontal location, or azimuth,

of a stimulus. Determining the vertical location of a sound source is more difficult because a source can produce the same temporal delays and intensity differences from several elevations at a given azimuth. Although some information is provided by the filtering properties of the pinna, good localization in elevation requires that the head be able to move. By tilting the head one can zero in on the source using the very accurate azimuth information determined in several planes.

Neurons at many stations in the CNS are exquisitely sensitive to small delays and intensity differences between the two ears. The superior olive and inferior colliculus are key brainstem nuclei in the reflex orientation of organisms to sound stimuli. The primary auditory cortex is also important in this process. Because of the bilateral projections of the two ears (see Figure 8.9), damage to the auditory cortex does not produce deafness, but it does interfere with localization of the source of a sound. In fact, any damage to one side of the auditory pathway will have this effect.

Brain Mechanism of Language

The original medium for linguistic communication was undoubtedly sound. Thus, it is not surprising that language is intimately connected with mechanisms of audition and vocalization, and the brain regions involved include the perisylvian cerebral cortex where one finds the primary receiving area of the auditory system as well as the cortical motor areas controlling muscles of the lips, tongue and larynx. These regions are necessary for the production and analysis of the acoustic signals serving spoken language.

While the comprehension of spoken language has received the bulk of scientific study, it is likely that the neuromuscular mechanisms for producing speech evolved in parallel with cognitive capacity. Some evidence demonstrates that earlier human forms such as ***Homo neanderthalensis*** may not have been able to produce the range of vowels because the shapes of their vocal tracts prevented movement of which modern humans are capable. Great apes and human infants share this limitation, but the latter eventually achieve a full phonetic capacity as their vocal tracts mature to the adult configuration.

The production of interpretable speech requires neural and articulatory systems capable of varying frequency and intensity in the correct temporal patterns. This involves not only controlling the vocal cords (a task of the vagus nerves), but also precisely regulating flow of air through the larynx by actions of the respiratory muscles of the diaphragm and thoracic cage (a task of several spinal nerves). Speech production ultimately depends on the proper functioning of the somatic motor system. Recognition of the acoustic objects forming speech requires that the brain 'bind' together neural signals of frequency, intensity and timing transmitted by the auditory pathways. To make sense of these signals, the brain must map the entire array

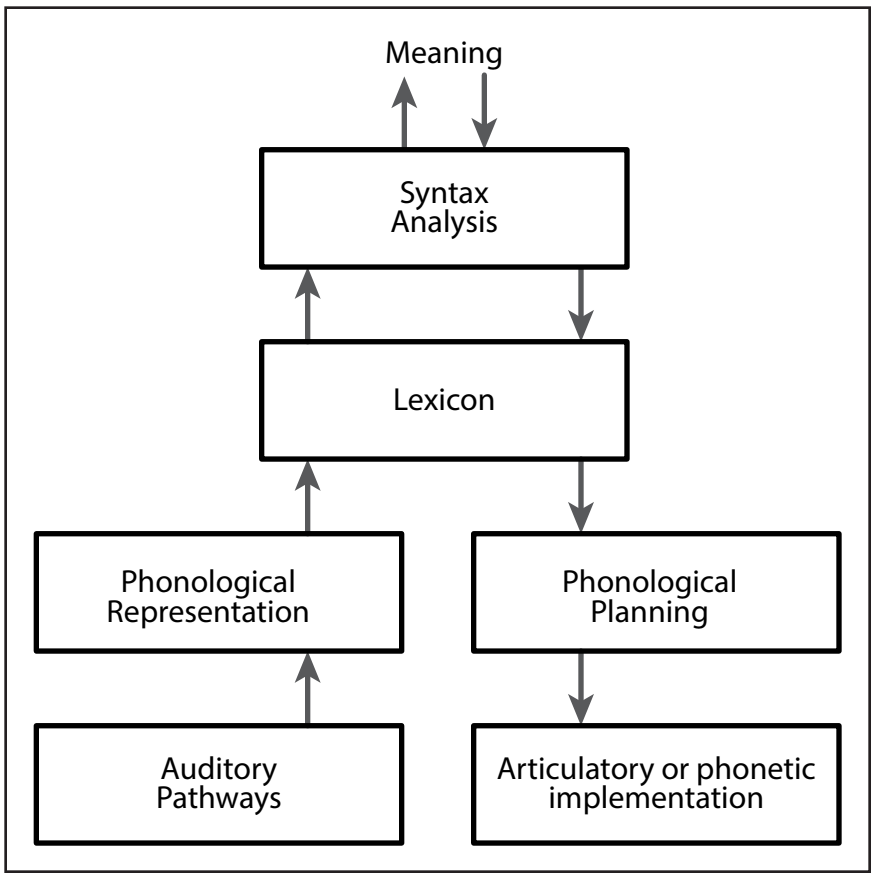


Figure 9.6

A simplified linguistic model of the functional elements of speech production and comprehension.

into a language system that identifies objects in the acoustic stream and assigns meanings to them, both individually and collectively.

Language is more than speech, even though the word itself is derived from ***lingua***, the Latin word for tongue. It arises from the capacity of the human brain to attach meaning to symbols that bear no explicit relation to what is meant. The association of symbol and meaning is learned and to a considerable degree arbitrary. Meaning can be communicated from one brain to another by means of these symbols, whether they are instantiated as acoustic signals, marks on paper, hand gestures, etc. The production and comprehension of speech are manifestations of language, but so are writing and reading, and perhaps even mathematics and music and art in general, all relatively recent additions to the human repertoire. When language is disturbed by damage to the brain, deficits may appear in all of these areas.

It is obvious that microelectrode recordings cannot be made routinely from the human brain as speech sounds or other linguistic symbols are processed or produced, so the greater part of our understanding of the neural substrate of human language comes from 'experiments of nature', i.e. lesions produced by disease states. These cases present all the interpretive

difficulties of any lesion experiment, mainly that one can only observe the capacities of the remaining structures, not those of the part that has been removed or rendered inoperable. Nonetheless, these cases have demonstrated that the encoding of speech sounds, either for production or analysis, is only one step in the brain's implementation of language. Current research using functional imaging techniques, such as PET and MRI, is adding new insights into the brain areas underlying language.

Two distinct scientific traditions have concerned themselves with how the human brain subserves language. Because the professional care of brain-damaged patients falls principally to physicians, the earliest tradition is associated with clinical medicine. The other tradition arises from the scholarly field of linguistics, which is the study of natural languages, including their historical development.

Students of ***neurolinguistics*** attempt to identify isomorphisms between models developed from the study of natural languages and neural mechanisms of the brain. Like clinical neurologists, neurolinguists attempt to learn about language from those experiments of nature that disturb language function. Such disturbances are called ***aphasias*** (Gr. phasis - speech) and those who study them are ***aphasiologists***.

Linguistic models of language production and comprehension distinguish several operations or stages (Figure 9.6). First, we can consider the process of speaking. To encode some complex "meaning" or idea in a stream of acoustic signals, the brain must isolate the functional elements of the idea (the syntax: agents, actions, objects, etc.), select the appropriate word forms for these elements (from the lexicon), look up the sounds that stand for the words (phonological planning) and then produce these sounds correctly via the vocal tract (articulatory or phonetic implementation).

Similarly, to comprehend speech, the brain must encode the acoustic signal (via the auditory pathways), identify those acoustic objects that are words (phonological representation), associate those words with their possible referents (lexical mapping), identify the linguistic functions the words are performing (syntax) and associate a meaning with the result.

Aphasias

To understand aphasic speech it can be helpful to break the process down into the elements shown in this model. For example, aphasic errors commonly occur at the phonetic implementation level, causing the patient to produce sounds that do not occur in his or her natural language. In rare cases, such errors can give the patient's speech what sounds like

a foreign accent. If a native speaker of Italian began to substitute a French or German 'r' for his normal rolled 'r', he would be making a ***phonetic*** error. Phonetic deficits result in problems with sound production and may be traced to impairments in oral-facial movement control. Aphasic patients with ***phonemic*** deficits may be capable of generating isolated sounds but make errors in assembling the constituent sounds (***phonemes***) to form meaningful words. Examples of such phonemic errors would be saying 'keams' instead of 'teams', or 'geen' instead of 'green', or 'gedree' instead of 'degree'. Here the sounds occur normally in the patient's language, but are being used incorrectly to form a word. Lexical problems can also occur, such as ***anomia*** (difficulty finding words), or word substitution (e.g. 'father' for 'mother'; 'sofa' for 'chair') or overuse of non-specific, empty words (e.g. thing, be, like). Impairment at the syntactic level may be reflected in the infrequent use of function words (and, the, a, very, have etc.), giving the speech a telegraphic quality. There may be loss of inflections and other markers of case, person, number, time and mood. Thus, such speech may contain agrammatical constructions such as "yesterday he go" or "Joe book" instead of "Joe's book." While the term "aphasia" is used to describe this entire family of deficits, it should be clear that the underlying mechanisms and associated symptoms are extremely complex.

From the tradition of clinical neurology has arisen a classification of aphasias that attempts to associate the different types with lesions in particular areas of the brain. These associations are usually less secure than standard textbooks imply because patients present with complex combinations of deficits and their brain lesions are often large and poorly defined. Furthermore, the anatomical studies of aphasia have usually focused on the cortical location of lesions and have ignored damage to subcortical structures, which are usually involved as well. In general, though, a useful distinction can be made between two major classes of deficit: ***non-fluent aphasia*** and ***fluent aphasia.*** These terms refer to the rate and smoothness with which speech is produced by aphasic patients.

Non-fluent aphasia is also called expressive aphasia, motor aphasia, or anterior aphasia. The classic example of non-fluent aphasia was described by Broca, and this subtype is also called Broca's aphasia. Patients with a non-fluent aphasia manifest some or all of the following signs:

- Speech is labored and slow. Small grammatical words and endings are absent and high content words are strung together, giving the speech a telegraphic quality.
- Musical tasks are usually carried out well.
- Comprehension is intact or relatively spared.
- The patient is acutely aware of the deficit and distressed by it. Affect is appropriate.
- Reading is intact but laborious; writing is abnormal in much the same way as speech.
- This form of aphasia is commonly associated with a lesion near the posterior end of the inferior frontal gyrus in ***Broca's area***.

Fluent aphasia is also called receptive aphasia, sensory aphasia or posterior aphasia. The classic description is from Wernicke, and this subtype is also called Wernicke's aphasia. A fluent aphasia may have some or all of the following characteristics:

- The amount of speech is excessive (logorrhea) with normal melody or prosody, but it is empty of semantic content. Content words are replaced with words empty of meaning (thing, this, etc). There may be much jargon and gibberish.
- Comprehension is impaired. The patient has difficulty interpreting spoken and written words, or words and letters traced on the skin.
- Writing is disturbed and parallels deficits in spoken language.
- Patients are euphoric and do not appear disturbed by their deficits. Affect is inappropriate and some patients exhibit paranoia.
- Fluent aphasia is commonly associated with a lesion of the posterior superior temporal gyrus or *Wernicke's area*.

Functional imaging studies using PET or fMRI generally confirm the anatomical localizations based on long clinical experience with aphasic patients, but there is considerable overlap in the areas affected. The contributions of subcortical structures like the striatum and thalamus are still poorly documented. Memory is also a key element in language function, and knowledge of its neural mechanisms will be critical to an understanding of how the brain implements language. Both fluent and non-fluent aphasias appear to affect all categories of linguistic function (phonetic, phonemic, lexical and syntactic) but not in the same proportions. Thus, neither the clinical model nor the linguistic model currently provides a "Rosetta stone" for a complete understanding of aphasia or of the neural basis of language.

Terms and Techniques

place coding	volley coding	categorical perception
characteristic frequency	best frequency	aphasia
fluent aphasia	non-fluent aphasia	anomia
anterior aphasia	posterior aphasia	aphasiologist
Broca's area	Wernicke's area	filter bank
syntax	lexicon	phoneme
acoustic stream	acoustic object	neurolinguist

Structure and Function of the Vestibular System

10

Table of Contents

The vestibular system informs the brain about the position and movements of the head. It is a major contributor to conscious perception of body position and motion, and it also plays a critical role in motor control. The receptor organs are located in the ***vestibular labyrinth***, a complex of membranous bags and tubes that are closely associated with the cochlea. The sensory organs in each of these components are presynaptic to the axons of neurons located in Scarpa's ganglion. The centrally directed axons of these cells proceed over the VIIIth cranial nerve to the brainstem where they synapse in the vestibular nuclei and cerebellum.

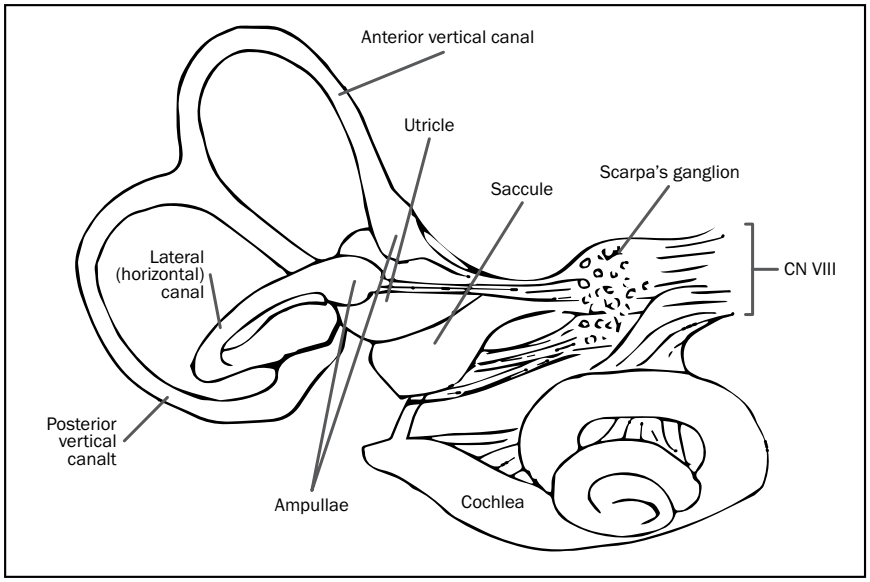


Figure 10.1

The membranous vestibular labyrinth occupies spaces within the temporal bone of the skull and comprises two functional divisions. Three semicircular canals are specialized to signal angular acceleration of the head. Receptor cells of the utricle and saccule transduce linear accelerations of the head, including that due to gravity The labyrinth is filled with endolymph, a special extracellular fluid rich in K^+.

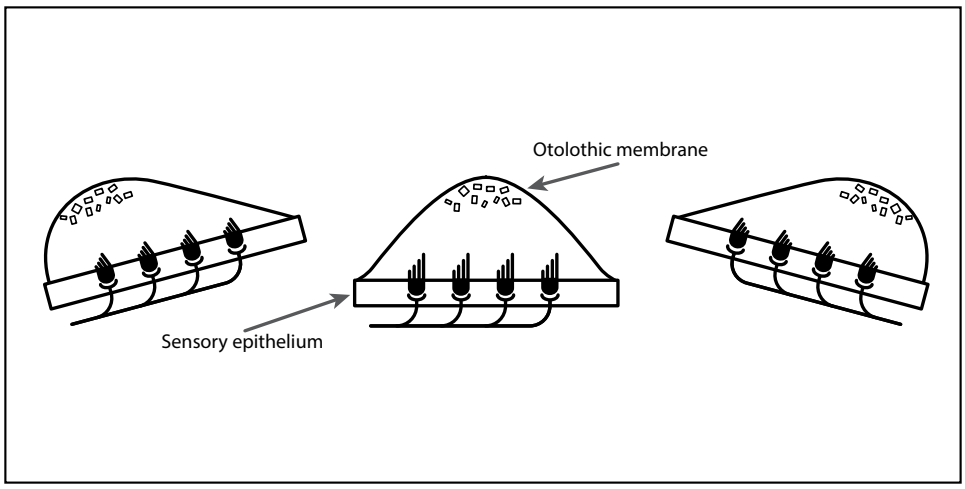


Figure 10.2

Functional anatomy of the utricular macula and its response to tilt of the head. (Redrawn from Schmidt. Fundamentals of Sensory Physiology. 3rd edition. Springer. 11086.)

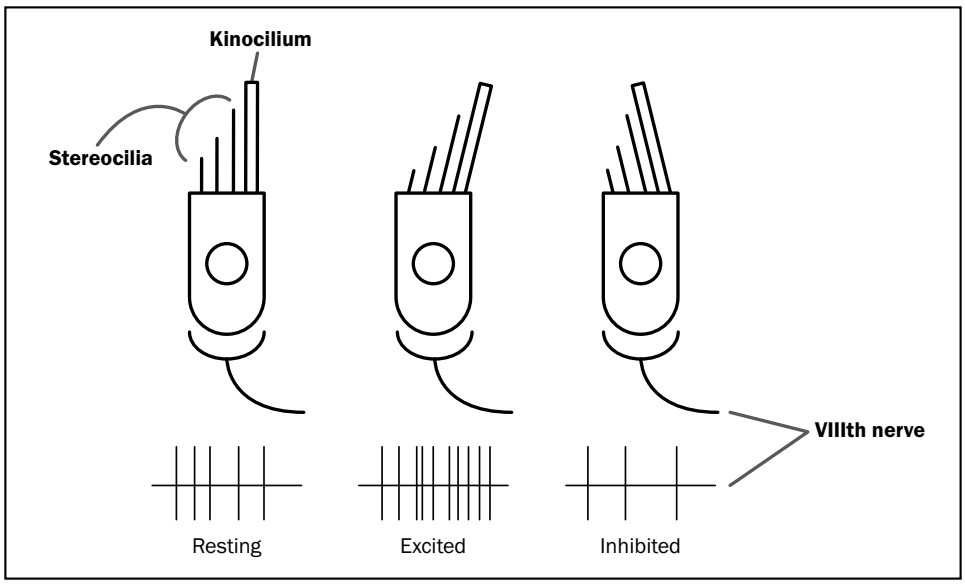


Figure 10.3

Bending of the hair cell's stereocilia toward the kinocilium allows K^+ to enter and depolarize the cell's membrane potential, causing it to secrete a transmitter on the terminals of the VIIIth nerve afferents.

Gross Structure of the Labyrinth

Structure and Function of the Utricle and Saccule

The receptor organs of the utricle and saccule are located on regions of their membranes called ***maculae*** (singular is ***macula***, which means spot). These organs are specialized for the

detection of linear acceleration, including that due to gravity. Hair cells of the macula extend their cilia into the gelatinous ***otolithic membrane*** above them (Figure 10.2). This membrane contains granules of calcium carbonate, the ***otoliths*** or ***otokonia***, which give utricle and saccule one of their common names, the ***otolith organs.*** The presence of the granules makes the otolithic membrane denser than the surrounding endolymph, so when the head is tilted or undergoes linear acceleration, the otolithic membrane shifts relative to the sensory epithelium and this bends the cilia of the hair cells. The utricle and saccule differ in the orientation of their maculae, but their functions are essentially the same.

The hair cells are modified epithelial cells, not neurons, and they do not generate action potentials. Each bears a single ***kinocilium*** (the tallest cilium) and several rows of ***stereocilia*** (Figure 10.3). The hair cells are depolarized when the stereocilia are bent toward the kinocilium and hyperpolarized when the stereocilia are bent away from the kinocilium, in a manner analogous to what happens in the organ of Corti ( ).

The orientation of the stereocilium and the kinocilium is said to ***polarize*** the cell for detection of linear acceleration in a particular direction. The term 'polarize' is used here to indicate the directional sensitivity of the cell, not the effect of ciliary bending on its membrane potential. Their membrane potentials and the rate at which they secrete their transmitter are modulated continuously as a function of the amount and direction of bending of the stereocilia.

As with neurons, depolarization causes the transmitter to be released. The axons of cells of Scarpa's ganglion are post-synaptic to the hair cells and generate action potentials at a frequency determined by the rate at which the hair cells release their transmitter. The axons of Scarpa's ganglion neurons normally have a resting discharge that is modulated up and down by the action of the hair cells.

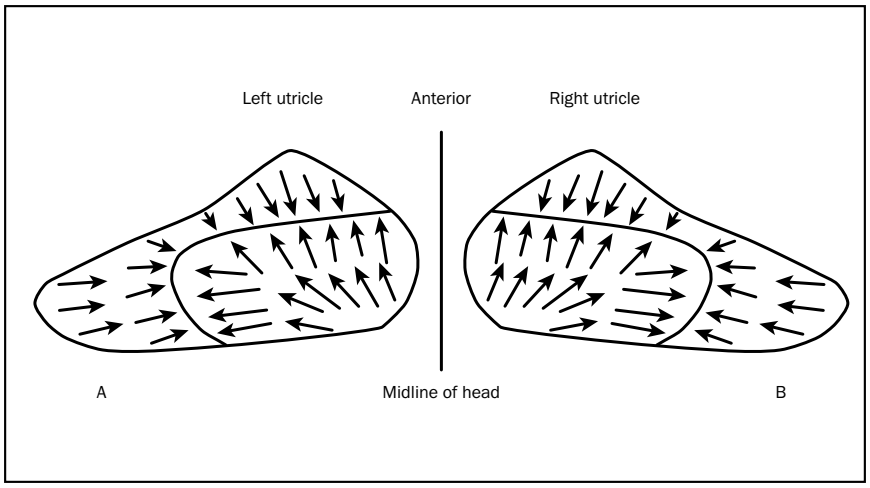


Figure 10.4

Schematic of the polarization directions of macular hair cells in the utricles. The diagram depicts the maculae of the two utricles as they might be viewed, greatly magnified, through the top of a transparent skull. The arrows indicate the on-directions of bending.

The diagram of Figure 10.4 represents the maculae of the two utricles as they might be viewed through the temporal bone from the top of the skull. The arrows point in the direction of head tilt that would depolarize local hair cells (the ***on-direction***). It can be seen here that the macular hair cells of the right and left utriculi are polarized in such a way that all directions of head tilt can be signaled by either side. For this reason, destruction of one utricle does not deprive the person of vestibular information about head tilt. Note also that the distributions of 'preferred directions' are mirror symmetric for the two sides of the head. Head tilt to the right, for instance, excites cells near A and inhibits cells near B. The macula of the saccule is located on its lateral wall, which would make its hair cells prominently influenced by gravity. The geometry of the 'on-directions' of the hair cells in these two organs ensures that linear acceleration of the head in any direction will be detected.

Structure and Function of the Semicircular Canals

The semicircular canals are capable of detecting rotation of the head about any axis. Figure 10.5 illustrates the mechanism. Imagine a circular, tube-shaped vessel made of transparent material and filled with water. On one outer wall is attached a rubber flap, which is bent if the water swirls around in the tube. If the tube is moved along a linear axis (Figure 10.5a), water will not flow around inside the tube. Because there will be no relative motion of the fluid with respect to the walls of the tube, the flap will not bend.

If the tube is rotated clockwise about an axis through its center, the water will lag behind

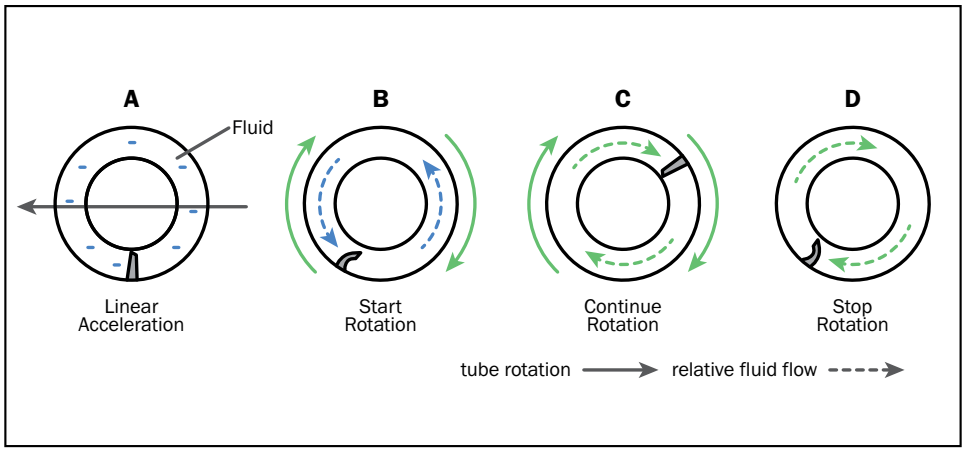


Figure 10.5

Mechanics of stimulation of the semicircular canals. Solid arrows - directions of acceleration of the head. Dotted arrows - direction of endolymph flow. See text for details.

at first because of its inertia. Thus, the outer walls will initially move relative to the water and the rubber flap will bend (Figure 10.5B).

Eventually, frictional forces between the tube and the water will cause the water to 'catch up', eliminating the relative motion, and elastic forces in the rubber flap will return it to the vertical position (Figure 10.5C). If the rotation is now stopped, the water will tend to keep moving clockwise for a while, and the rubber flap will be bent in the direction opposite to that at the beginning of rotation (Figure 10.5D).

The tubes, water and flap serve as a good model for the semicircular canals. The rubber flap is synonymous with the sense organs that lie inside the ***ampulla***, which include the ***cupula*** and ***cilia***. Movement of the endolymph relative to the wall of the canal causes deflection of the gelatinous cupula, into which the cilia of the hair cells are inserted (Figure 10.6). Unlike in the utricle and saccule, all the hair cells on a ***crista ampullaris*** are polarized in the same direction so they are all depolarized by a movement of the endolymph in one direction and hyperpolarized by movement in the opposite direction. Thus, the output of the ampullary nerve signals both the direction and the magnitude of angular acceleration in the plane of the canal.

The canals are arrayed in pairs so as to detect all directions of angular acceleration and to

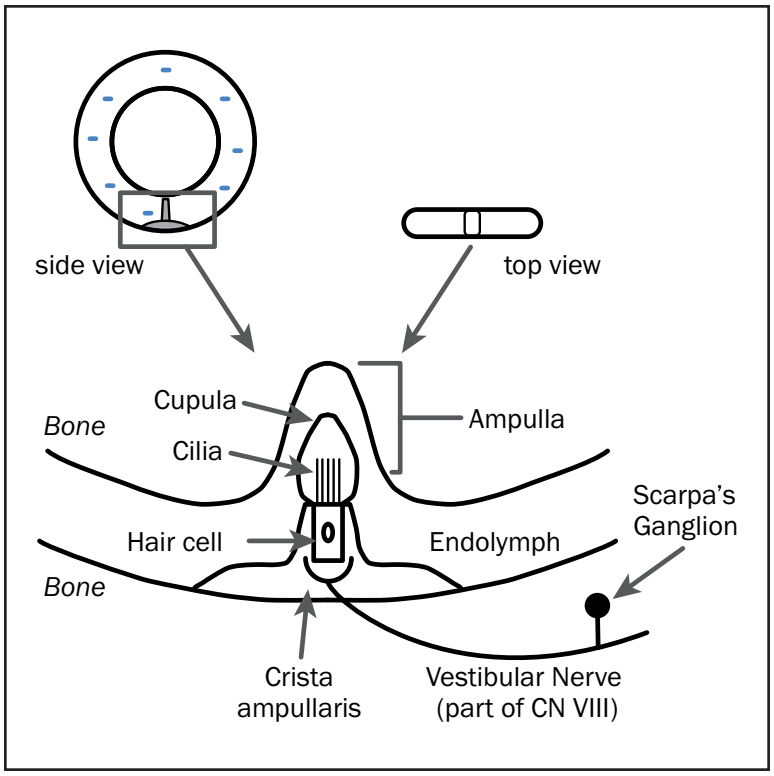


Figure 10.6

Anatomy of the receptor in the ampulla of a semicircular canal. The cupula is a gelatinous substance analogous to the otolithic membrane but without the calcium carbonate granules. Flow of the endolymph within the canal bends the cupula and cilia. This leads to excitation or inhibition, depending on the direction of flow.

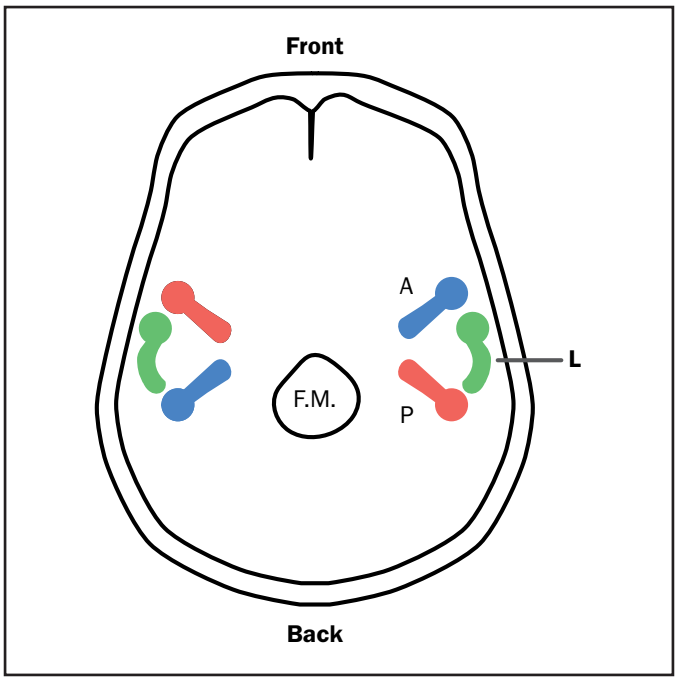


Figure 10.7

Top view of the base of the skull showing the geometry of the paired semicircular canals. The canals are in color, indicating paired canals. Note that the anterior and posterior pairs are crossed and complementary, such that the anterior canal on one side is in line with the posterior on the other and vice versa. A-anterior, P-posterior, L-lateral. F.M.-foramen magnum.

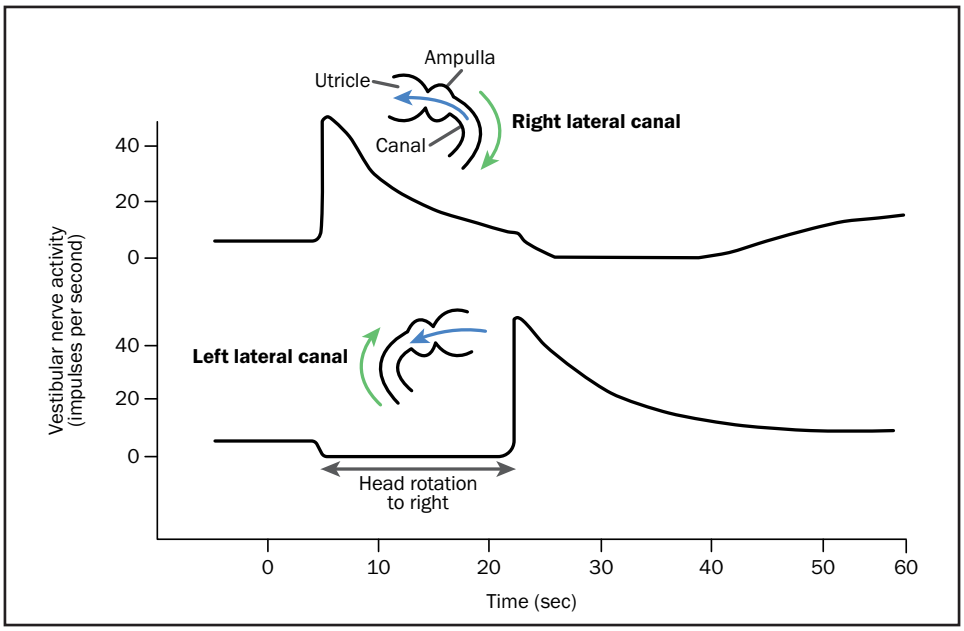


Figure 10.8

Effects of acceleration on the right and left lateral semicircular canals and discharge rates of fibers in the VIIIth cranial nerve. Note the resting discharge from the two canals. Green arrows: direction of head rotation. Blue arrows: relative flow of endolymph. Note that for head rotation to the right, the endolymph moves toward the utricle on the right (utriculopetal flow) and away from the utricle on the left (utriculofugal flow). Utriculopetal flow is excitatory in the lateral canals.

resolve the direction of the acceleration in terms of three vectors (Figure 10.7). Note that the two horizontal canals form one pair while each anterior canal is paired with the contralateral posterior canal. Figure 10.7 shows how these pairs lie in or parallel to the same plane.

The pairs of canals operate in 'push-pull' fashion like a differential amplifier. An acceleration (change in velocity) that excites one member of the pair inhibits the other. Figure 10.8 demonstrates how rotation of the head to the right excites hair cells of the right lateral canal and inhibits those of the left.

Note that the discharge of the right ampullary nerve declines during the plateau of rotation when velocity is constant (acceleration is zero). When the fluid moves at the same rate as the walls of the canal, the cupula does not deflect, so the cilia neither depolarize nor hyperpolarize. This situation is modeled in Figure 10.5C.

When the rotation stops, note that the left canal is excited as though the head were turned to the left. When the canal's wall suddenly stops moving, the endolymph's inertia will keep it flowing, so that the relative motion between the endolymph and the walls is opposite of how it started. This relative movement deflects the cupula in the opposite direction, as modeled in Figure 10.5D.

Normally the head does not experience continued rotation such as is illustrated in Figure 10.8. This type of movement only occurs when constantly turning, as in a spinning office chair, and then suddenly stopping. Natural movements of the head produce short-lasting motions with a limited range of frequencies. Because of the physical properties of the system, the crista moves slower than the head. This drag creates a time lag that allows hair cells to add up responses over time. The hair cells act as integrators so that the information conveyed to the VIII cranial nerve is in velocity, not acceleration units. (Remember that the integral of acceleration is velocity.) Thus, even though the canals detect acceleration, their signal is 'integrated' by the mechanics of the cupula to yield a signal proportional to head velocity.

Central Connections of the Vestibular Apparatus

Scarpa's ganglion is a collection of cell bodies that collects information from the vestibular organs and sends it to the central nervous system. Cell bodies in *Scarpa's ganglion* send an axon ***distally*** to the saccule, utricule and semicircular canals, where they innervate the hair cells. These function as the 'dendrites' of the approximately 20,000 primary afferents, which proceed ***proximally*** (i.e. toward the CNS) together with the cochlear afferents of the VIIIth cranial nerve. They enter the brainstem at the pontomedullary junction, where the majority synapse in the complex of vestibular nuclei (Figure 10.9). Some project directly to the cerebellum, where a specific segment called the flocculonodular lobe processes vestibular information.

The central projections of the vestibular nuclei reflect three major functions of the vestibular system:

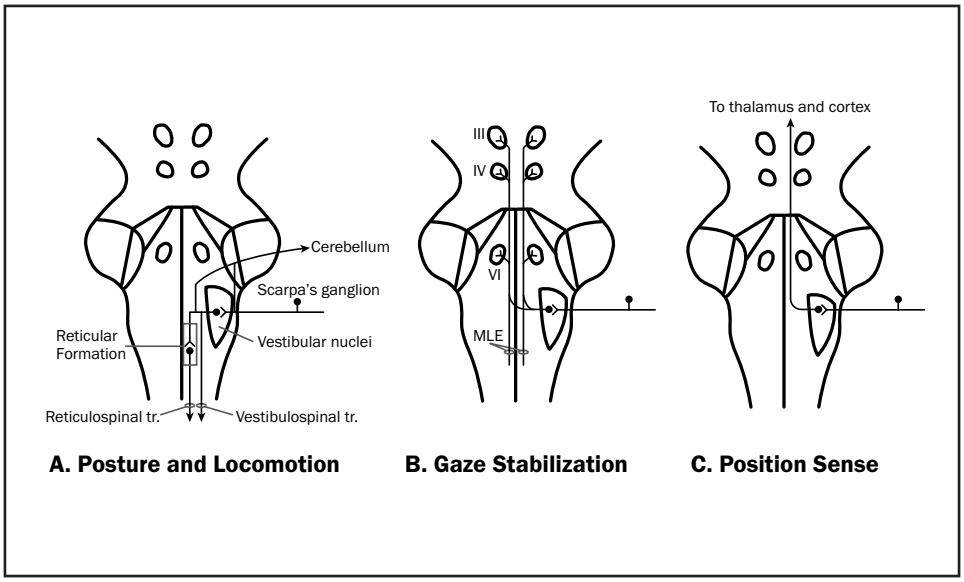


Figure 10.9

Schematic of projections subserving the three major functions of the vestibular system. A. Projection to cerebellum, vestibular nuclei and reticular formation for control of posture and locomotion. B. Projection to oculomotor nuclei via the MLF for gaze stabilization (VOR). C. Projection to cortex (parietal and temporal lobes) via the thalamus for sense of head position and motion in space.

- control of posture and locomotion
- stabilization of gaze direction during movements of the head
- provision of a sense of orientation in space

Adjustment of somatic motor mechanisms to linear and angular accelerations of the body mass and head depends on input from the vestibular system. Scarpa's ganglion projects largely to the vestibular nuclei, which in turn send second order fibers to the ***cerebellum, reticular formation*** and ***spinal cord***. The vestibular nuclei and reticular formation have major descending motor connections, the ***vestibulospinal tracts*** and the ***reticulospinal tracts*** (Figure 10.9A). Dysfunction of the vestibular system leads to an inability of the subject to maintain normal posture and to move efficiently. This state is called ***disequilibrium*** and may be associated with falling (see Chapter 23).

The vestibular nuclei also participate in circuits that automatically adjust eye position to maintain a steady gaze direction in the face of head movements. This circuitry subserves the ***vestibulo-ocular reflex*** or VOR. You can demonstrate the VOR by looking at a piece of paper that has writing or a picture on it. If you shake the piece of paper right and left, the words/ picture will be difficult to discern; however, if you shake your head left and right and keep the

paper stable, you should be still be able to focus your gaze. This exercise demonstrates that, as you turn your head, your eyes automatically shift to maintain your gaze.

The main central pathway for the VOR is the ***medial longitudinal fasciculus***, which links the vestibular nuclei to the nuclei of cranial nerves III, IV and VI (Figure 10.9B). The MLF must activate cranial nerves connected to muscles in both eyes so that, for instance, if you turn left, the right eye can abduct (go away from the midline) and the left line can adduct (go towards the midline) simultaneously. This will be discussed further in Chapter 15.

Dysfunction in the vestibular system is translated into a signal indicating, erroneously, that the head is turning. This signal triggers the VOR, and the eyes move slowly in one direction (slow phase) and then jerk back in the other direction (quick phase) (Figure 10.10). This slow phase occurs when trying to stabilize an image on the retina during head movement. The quick phase represents when eyes snap back because they are at extremes of rotation (the eyes would otherwise roll into the back of one's head). This is ***vestibular nystagmus***. We will study this concept in greater detail later in the course (see Chapter 16).

Conscious sense of orientation in space differs from other sensations in that we are usually unaware of it. We become very aware of its disturbance, however, when there is vestibular dysfunction. The vestibular system projects to several areas of the cortex, including the ***temporal lobe*** near the primary auditory cortex, and the ***parietal lobe***, which is concerned with the relation of the body to extracorporeal space (Figure 10.9C). Presumably these cortical projections contribute to our conscious sense of orientation in space. Dysfunction of the vestibular system leads to a loss of this orientation sense, and we feel that we are spinning or

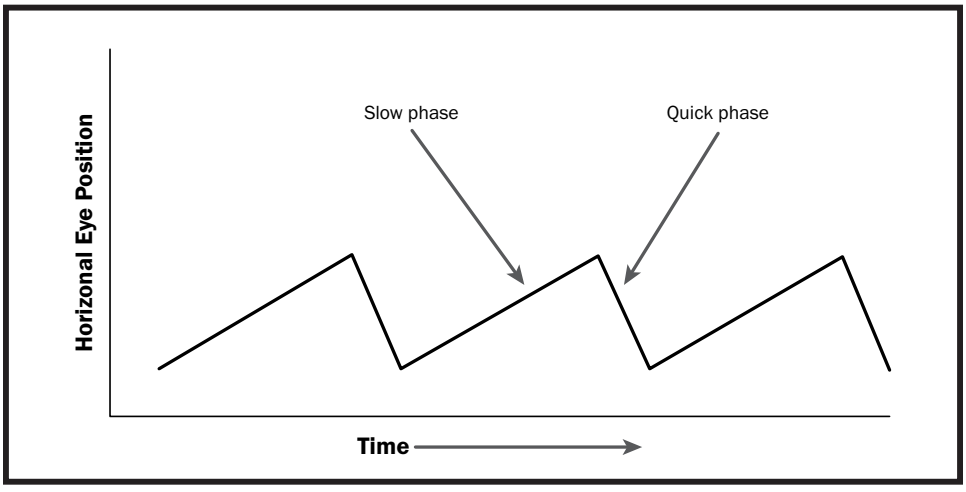


Figure 10.10

Vestibular nystagmus. Eye position has been recorded electrically during vestibular nystagmus revealing the repeated slow (VOR) and quick phases.

the world is spinning. This state is ***vertigo***. Damage to higher levels of the vestibular system can also lead to disordered perceptions of the orientations of external objects.

An individual suffering from vertigo may undergo testing to identify the possible source of the symptoms. The ***caloric test*** can be used to determine if the vestibular canals are functioning properly. In this test, a patient lies supine with the head tilted forward approximately 30 degrees, bringing the horizontal canals to a near vertical orientation. Then, either warm or cold water is injected into to the ear, and the physician will observe the patient's eyes, which should demonstrate induced vestibular nystagmus. Nystagmus occurs because the temperature gradient causes flow of the endolymph within the canals so that the patient experiences a feeling of rotation, with an accompanying VOR. If nystagmus does not occur, then there is reason to believe the vestibular apparatus itself is damaged, which may have caused the vertigo. If everything seems normal, however, then a more central cause would be indicated.

Terms and Techniques

utricle	saccule	semicircular canal
endolymph	perilymph	temporal bone
Scarpa's ganglion	otolithic membrane	hair cell
stereocilium	kinocilium	ampulla
cupula	crista ampullaris	macula utricularis
velocity	vertigo	disequilibrium
vestibular nystagmus	quick phase	slow phase
angular acceleration	linear acceleration	push-pull
vestibulospinal tract	reticulospinal tract	MLF
vestibulo-ocular reflex	gaze direction	visual axis
utriculopetal	utriculofugal otolith	otolith organ

The Body in Space: A Constructed Modality 11

Table of Contents

We are intuitively aware of the positions and motions of our bodies and the locations of objects around us. These capacities require the integration of information from virtually all the sensory modalities, as well as the engagement of memory storage mechanisms. The term *proprioception* generally refers to our sense of the positions of our limbs, trunk and head, and the receptors serving this function are called *proprioceptors*. These receptors include muscle spindles, Golgi tendon organs, skin/joint mechanoreceptors, and, importantly, the vestibular system, which encodes the position and motion of the head.

These sensory receptors have other functions in addition to that of mediating position sense. Other sensory systems supplement these signals from within the body. The visual system can monitor limb position and movement, as well as changes in the retinal image due to motion of the body through space. The auditory system locates sound sources from differences in the time and intensity of acoustic signals reaching the two ears. The brain must then synthesize all of this information to create the "dynamic body image in context," i.e. where we are, where we are going and where objects around us are located.

The brain processes proprioceptive information in two distinct systems: conscious and unconscious proprioception. To test conscious awareness of body location, one can ask the patient to report the position of a limb. This type of test probes the integrity of the dorsal column-medial lemniscus (DC/ML) system (see Figure 4.1). This system mediates conscious position sense. Damage to the DC/ML system not only causes deficits in conscious position sense but can also disturb locomotion, indicating that conscious position sense is also an important source of information for the motor system. The general term for incoordination of movement is *ataxia*, and that stemming from impaired function of the DC/ML system is called a *sensory ataxia*.

Unconscious proprioception involves a different target of proprioceptive information: the cerebellum receives a constant flow of proprioceptive information from the vestibular apparatus, muscles, joints and skin. Incoordination due to cerebellar damage is called ***cerebellar ataxia***, and this condition differs from sensory ataxia in important ways. Cerebellar damage causes no impairment of conscious position sense, so the proprioceptive function of the cerebellum is somewhat loosely called 'unconscious proprioception'. Also, blindfolding a subject with sensory ataxia makes the ataxia worse, because the subject needs the visual system to tell where the legs and arms are located. Blindfolding has little or no affect on cerebellar ataxia, because the condition arises not from inadequate information about limb position but rather the inability of the brain to use the information.

Proprioception

The neural substrate of joint-position sense was at first misunderstood. It was observed that pulling on muscle tendons gave rise to no identifiable sensation, so the idea that muscle receptors could play a role in conscious position sense was discounted. Early reports that local anesthesia of finger joints in humans impaired position sense led to the erroneous view that joint receptors alone are responsible for conscious proprioception. Certain neurons located in the ventral posterior thalamic nuclei, which receive the dorsal column input, were observed to be exquisitely sensitive to the full range of joint positions (Figure 11.1), and this was originally attributed to input from joint receptors.

These views changed when it was discovered that joint receptors of the knee discharge only at the extremes of flexion and extension but do not provide information about intermediate joint angles. Joint receptors could not be responsible for the full range of responses evident in Figure 11.1. Humans are exquisitely sensitive to changes in joint angle and can precisely adjust the angle of one joint to match the angle of its contralateral counterpart. Furthermore, people who have entire joints replaced by prostheses (knee, hip, shoulder, etc.) do not lose position sense in these joints, even though there is much destruction of the joint capsule and its nerve supply. These collective findings suggested that joint receptors were not the substrate of conscious proprioception.

It is now known that muscle spindles are imperative in conscious proprioception. As noted above, pulling on a tendon does not give rise to an identifiable sensation, as would touching the skin or sticking it with a pin. However, when a vibrating stimulus is applied to a tendon (a stimulus known to activate the Ia afferents of muscle spindles at high rates), subjects were asked to match the position of the stimulated limb with the contralateral limb. Based on this test, they found that subjects perceive that the stimulated limb has changed position. These results indicate that muscle receptor signals are used by the CNS to monitor joint angles. Consistent with this finding, humans with conditions that selectively affect large diameter muscle afferents lose their appreciation of joint angle. It has also been demonstrated that certain 1a afferents project via the DC/ML system to area 3a of sensory cortex, which is at the

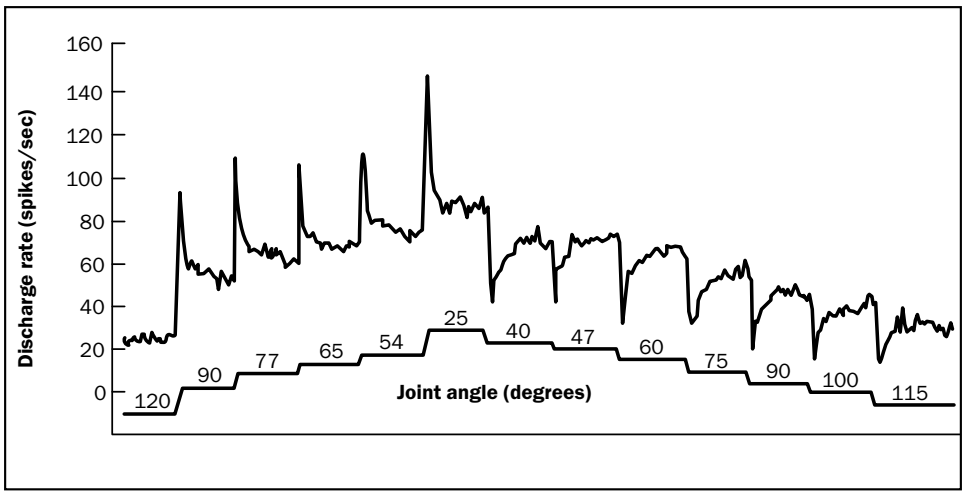


Figure 11.1

Discharge frequency of a neuron in the ventral posterior nucleus of the thalamus as the angle of the monkey's knee was rotated. Observe the great sensitivity of the phasic response, indicating sensitivity to changing joint angle. Adapted and redrawn from Mountcastle et al. in Gerard, R., editor, *Information Processing in the Nervous System*, Amsterdam, 1964, Excerpta Medica Foundation, 1964.

border between sensory and motor strips. It would make sense for proprioceptive afferents to project to an integrative location, such as the convergence of somatosensory and motor cortex.

Experiments devised to assess the contributions of skin, joint and muscle afferents indicate that these contributions vary at different joints and velocities of joint rotation. In the experiment of Figure 11.2, human subjects were tested for their ability to detect movements of the terminal joints of the middle finger. The figure plots the threshold movement (in degrees) that could be detected on 70% of trials. When muscle receptor contributions were eliminated by positioning the fingers in a particular way (inset), there was a significant loss of sensitivity at low velocities. Additionally, suppressing joint and skin afferents by local anesthesia but leaving muscle sense intact also had a significant effect, indicating that joint-position sense normally depends on an integration of all these afferent inputs.

Efference Copy

Sensory afferents do not account for all knowledge of joint or limb position. Although dorsal root section (***dorsal rhizotomy***) has a devastating effect on limb function, animals with bilateral dorsal rhizotomies can still use both limbs. Furthermore, following unilateral dorsal rhizotomy, an animal will use the deafferented limb with considerable skill and power if the normal limb is immobilized. Surprisingly, it was also found that a deafferented limb can be used with some accuracy by a ***blindfolded*** animal. Moreover, the animal appears to know the location of the deafferented limb without being able to see it. These results suggest that

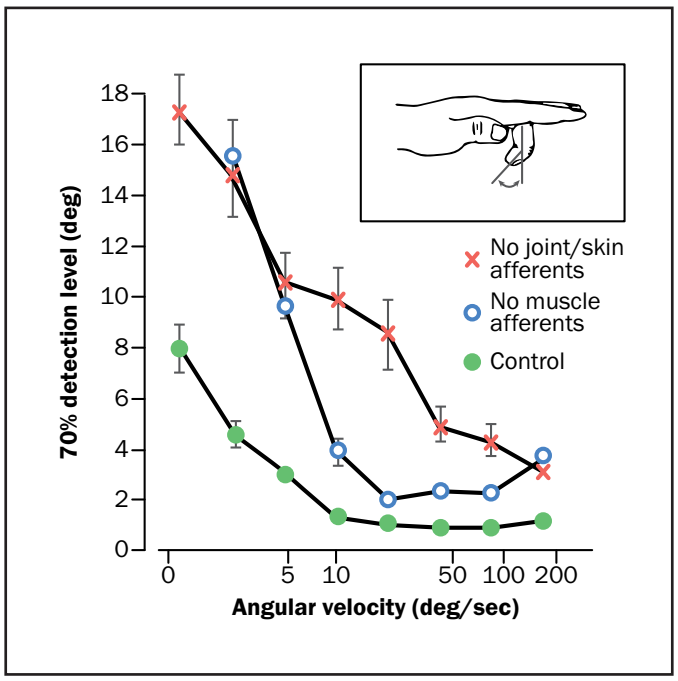


Figure 11.2

Detection thresholds for movements of the terminal joint of the middle finger. Filled circles: control. Open circles: muscle afferents disengaged. Crosses: joint and skin afferents anesthetized, muscle afferents functioning (adapted and redrawn from Gandevia et al. ***J. Physiol***. 355:507-517,1983). Inset: This position of the fingers disengages the muscles that move the distal digit of the middle finger (adapted and redrawn from Ferrell et al. ***J. Physiol***. 386:63-72, 1987)

the deficit following unilateral deafferentation is a form of learned uselessness—the limb is so clumsy that the subject ceases to use it in favor of the intact limb. Some proprioceptive ability must still be intact.

This result led to the hypothesis that, as each new motor command is issued, a copy of the command (***efference copy***) is available in the central nervous system. This system keeps track of intended limb position, and an estimate of position is available to the brain in the absence of sensory feedback. For example, if you place your arm straight out, you know the arm's starting position. If the brain sends a motor command to move it 90 degrees, you don't necessarily need proprioceptive feedback because your brain knows the starting position and how much that changed (the motor command). There is, in fact, considerable evidence that this occurs in both the oculomotor and somatic motor systems, as will be discussed in later chapters.

Role of the Parietal Lobes and the Dorsal or Parietal Stream

While damage to the primary somatic sensory pathways or to the post-central gyrus (S1) can impair limb position sense and aspects of somatic sensation, damage to posterior parietal areas 5 and 7 causes a much more global disturbance of spatial processing. Consider first our ability to reach in a pocket and pick a quarter from a collection of miscellaneous coins.

The ability to recognize the three-dimensional shape and the surface properties of the quarter requires the integration of proprioceptive information from joint, skin and muscle receptors with cutaneous information from mechanoreceptors and thermoreceptors. Because active touch is involved, motor control signals (efference copy) are also available. From this information we construct an image of the object we cannot see, and then we can compare this image to a remembered object in order to identify it. In fact, we can close our eyes and conjure up a visual image of the object and, if we have some previous experience with it, we can give it a name. This ability, which is much more complex than it would appear at first, is generally referred to as stereognosis and its absence as astereognosis. A lesion in the parietal cortex may impair stereognosis even when primary tactile and joint sense remains intact. Thus, the lesion would impair not the primary modality information but rather some interpretive function.

Damage to the parietal lobe, particularly on the right, also leads to a more global deficit called ***neglect syndrome***. This disorder often involves a denial of injury or deficit on the contralateral side. The patient may fail to recognize his or her own paralyzed limb and even claim that it belongs to someone else. The patient ignores stimulation of the body surface contralateral to the lesion. In mild cases, sensory inattention can be revealed by asking the patient to attend to (report the presence of) stimuli presented simultaneously on both sides of the midline, such as tactile stimulation of both cheeks. In such tests of ***double simultaneous stimulation*** the patient will ignore the stimulus on the affected side even though both stimuli are perceived when presented one at a time.

Disturbances of perception due to parietal lobe lesions are not restricted to the patient's body but extend into the surrounding peripersonal space as well. There is inattention to visual and auditory stimuli arising on the side opposite the lesion. Such patients will often ignore individuals speaking to them from the affected side.

Because spatial coordinate systems can be disturbed following parietal damage, tasks that require building or assembling multipart objects can be severely impaired. This pathology is called ***constructional apraxia***. (Apraxia means an inability to carry out a skilled movement or task that the patient had once been able to perform, even though there is no overt paralysis or primary sensory deficit). Even a task like dressing oneself can be very difficult for apraxic patients, as the spatial orientation of items of clothing must be arranged with respect to the body. The intimate relationship between the parietal lobe and the body is vividly seen in the phenomenon of ***hemiatrophy***, which afflicts children suffering from parietal damage. Here, the affected and often hemiparetic (weak) side does not grow normally, so that the body parts, bones as well as soft tissue, on one side appear to belong to a smaller person.

Lesions of the posterior parietal cortex clearly interfere with the "body image," or the high-level neural representation of the body, its parts and objects in peri-personal space. Much evidence indicates that the posterior parietal cortex performs this function as part of a system called the ***dorsal*** or ***parietal stream*** that is concerned primarily with spatial information. The dorsal stream complements the ***ventral*** or ***temporal stream*** that carries information about

object shape and color to the temporal lobe.

Coordinate System Transformations

An object's location must always be specified with respect to a coordinate system based on some reference point and direction. The various types of information that the brain uses in spatial processing arise from sources having very different coordinate systems. Visual information is encoded in coordinates that are fixed with respect to the eye (oculocentric/retinocentric) but move in extrapersonal space as the eye, head and body move.

Vestibular and auditory information is encoded in head-centered coordinates, which move with the head and body, but are independent of the position of the eyes. To perform many of its functions, the brain must be able to transfer the information from one coordinate system to another. There is considerable evidence that these coordinate transformations depend on the circuitry of the posterior parietal cortex (Figure 11.3).

To illustrate the need for such transformations from one system to another, imagine that you are reaching for a cup of coffee that is sitting to the left of your computer's keyboard. You look to the left and move your hand toward the cup. The hand movement is made in a coordinate system based, let us say, on the shoulder, but the image of the cup is encoded in retinal coordinates that are displaced with respect to the head, which itself is turned with respect to the trunk and shoulder. Thus, the retinocentric information must be corrected for eye position, and head position to specify the location of the cup outside the body. This information must then be converted to a motor sequence that takes into account the starting positions of the hand and arm.

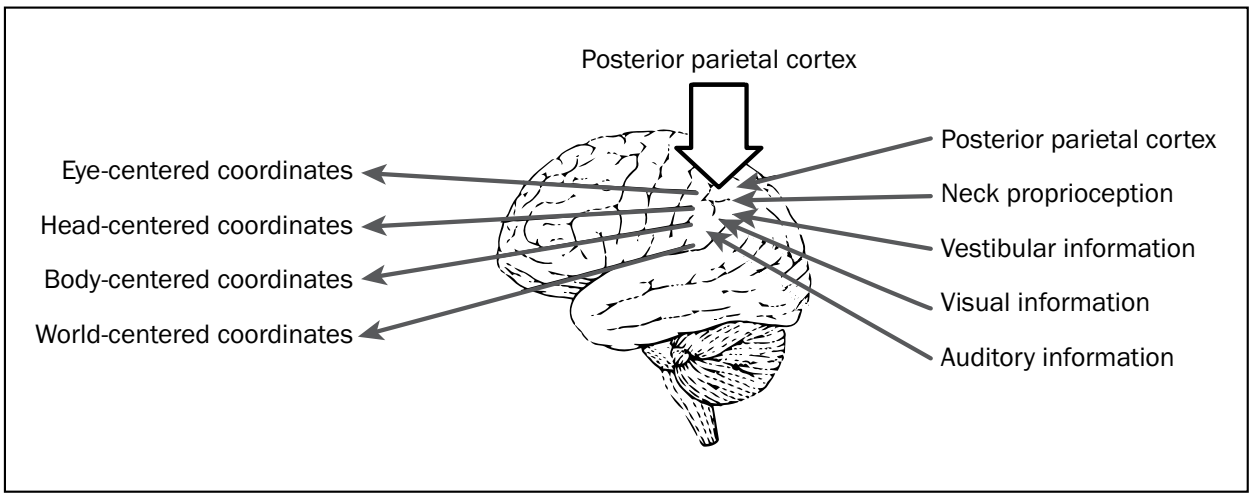


Figure 11.3

Coordinate system transformations in the posterior parietal cortex.

What kind of evidence would implicate the parietal cortex in these coordinate system transformations? To start with, information encoded in sensory coordinates should show signs of being modified by motor signals. An example of this was found in an experiment by Andersen and Mountcastle published several years ago (Figure 11.4). A monkey was trained to direct its gaze at different points on a screen. Responses of a neuron in the posterior parietal cortex were recorded when a visual stimulus appeared just above the fixation point. The location of the stimulus was at the same location on the retinal image, but as you can see, the responses were quite different in the two conditions. The sensory response was being affected by information from another source, presumably the oculomotor signals controlling eye position. In other words, the receptive field of this neuron moved with the eye, as though it were fixed in an oculocentric coordinate system, but the response was highly modulated by the position of the eyes. Because the visual response was now influenced by the direction of gaze, the signals contained information about the location of the stimulus with respect to the head. This experiment was one of the first to suggest a role of the parietal cortex in transferring sensory information between coordinate systems.

Later work has revealed evidence for other transformations taking place in posterior parietal cortex. For example, certain cells respond to a stimulus at a particular spot in the visual field if the stimulus is the target of either a saccade or a reaching movement of the arm. This type of cell appears to participate in a system representing behavioral goals, regardless of what

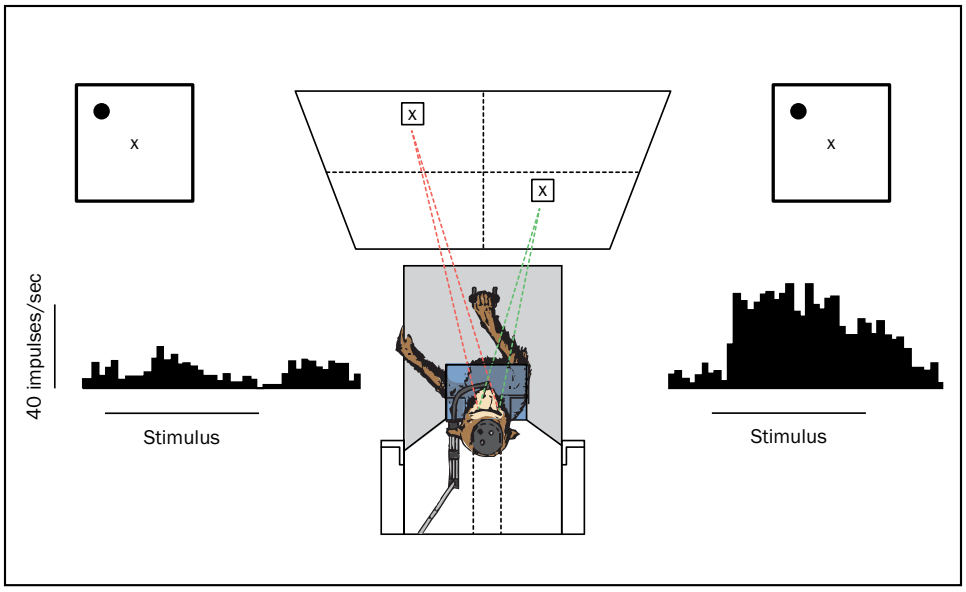


Figure 11.4

Modulation of visual responses of posterior parietal cell by changes in gaze angle. Responses to identical retinal stimulation at two different gaze angle are shown as histograms. Stimuli were much more effective when the direction of gaze and the cell's receptive field (small square) were positioned to the right of the midline. Adapted from R.A. Andersen and V.B. Mountcastle. ***J. Neurosci***. 3:532-548, 1983.

kind of behavior is involved in reaching that goal. One would expect this kind of high-level representation in a system that provides instruction to the frontal lobes, cerebellum and basal ganglia. In primates, recent evidence suggests that an eye-centered coordinate system may provide a common reference frame through which multiple sensory-motor systems communicate.

Terms and Techniques

dorsal column nuclei	DC/ML system	ataxia
cerebellar ataxia	sensory ataxia	apraxia
constructional apraxia	stereognosis	astereognosis
dorsal stream	parietal stream	ventral stream
temporal stream	postcentral gyrus	cortical areas S1 and S2
proprioception	efference copy	neglect syndrome
dorsal rhizotomy		

Chemosensation: Taste and Smell

12

Table of Contents

Our most primitive senses detect the chemical makeup of our environment. This chapter is mainly concerned with our sense of smell (olfaction) and our sense of taste (gustation). However, it should be noted that there are many other chemical receptors in our bodies, such as those in our circulatory system. Chemicals are detected by the olfactory and gustatory systems as they interact with receptors in the nasal cavity and on the tongue, pharynx and epiglottis (Figure 12.1). These senses are important in many basic behaviors such as mating, locating food and avoiding harmful chemicals.

Taste

The tongue has the highest concentration of taste receptors. These are located in small protrusions called papillae, which contain one to several hundred taste buds. The taste buds contain 50-150 taste cells, which have tiny microvilli that protrude into the pore of the taste bud. These microvilli contain the taste-receptor proteins that detect the “basic tastes”: sweet, sour, salt, bitter, and umami (glutamate). Note that the term ‘taste receptor’ is often used to refer to both the taste cells and the taste-receptor proteins located on these cells. Taste cells have about a 10-day lifespan and are replaced by differentiating basal cells also located in the taste buds. This cellular genesis explains why the sense of taste returns after burning our tongue on hot food.

The taste bud is innervated by gustatory afferents. These are the first neurons in the taste pathway, and their firing rate is modulated by the release of neurotransmitter from the taste cells. The taste buds from different parts of the tongue are innervated by different cranial nerves depending on their location (Figure 12.2). The anterior $^2/_3$ of the tongue is innervated by the facial nerve (VII), the posterior $^1/_3$ is innervated by the glossopharyngeal nerve (IX)

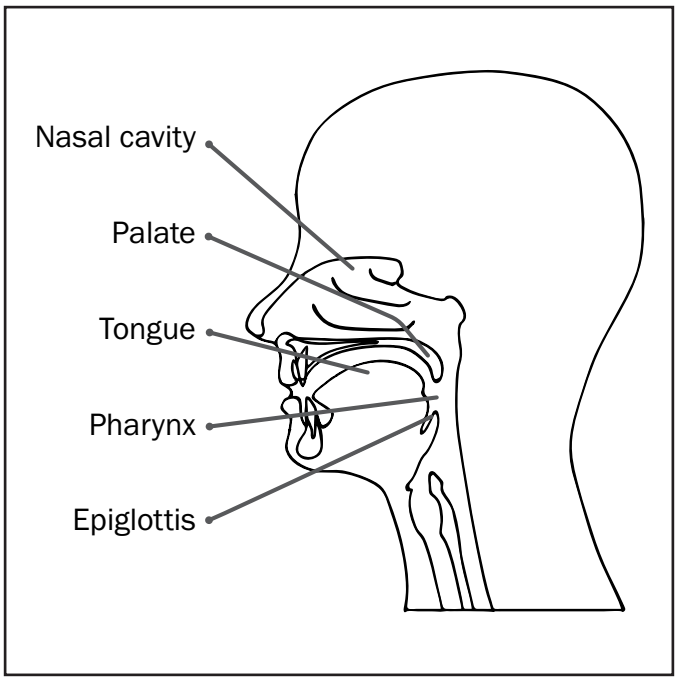


Figure 12.1

The gross anatomy of the mouth, throat and nasal cavity. The upper part of the pharynx is referred to as the nasopharynx and the lower part as the oropharynx.

and the pharynx, epiglottis and other regions surrounding the throat are innervated by the vagus nerve (X). Although they travel in different cranial nerves, all gustatory axons project to the ipsilateral gustatory nucleus (nucleus of the solitary tract) in the medulla. From there, the projection synapses in a special region of the ventral posterior medial (VPM) nucleus of the thalamus before continuing to the primary gustatory cortex. Primary gustatory cortex is located in insular cortex, but there is a gustatory cortical area near the somatic representation of the tongue that is also involved in taste perception. Interruption of the cranial nerves innervating the taste receptors leads to the loss of taste or ***ageusia***.

The transduction mechanisms differ for the basic tastes (Figure 12.3) and employ either ionotropic (salt, sour) or metabotropic (sweet, bitter, umami) receptors. However, each mechanism shares certain properties. When activated by its appropriate substrate, all taste cells will be depolarized, which opens voltage sensitive Ca^{2+} channels and leads to the release of neurotransmitter onto gustatory axons. Though the particular neurotransmitter used for all tastes is not known, it has been shown that taste cell activation can lead to both increases and decreases in the firing rate of primary gustatory afferents. While not strictly classified as neurons (i.e. derived embryologically from neural ectoderm), taste receptor cells do have voltage-gated sodium and potassium channels and can fire action potentials. These currents can be blocked by TTX (a sodium channel blocker), and the amplitude of their response is reduced in the presence of TEA (a potassium channel blocker).

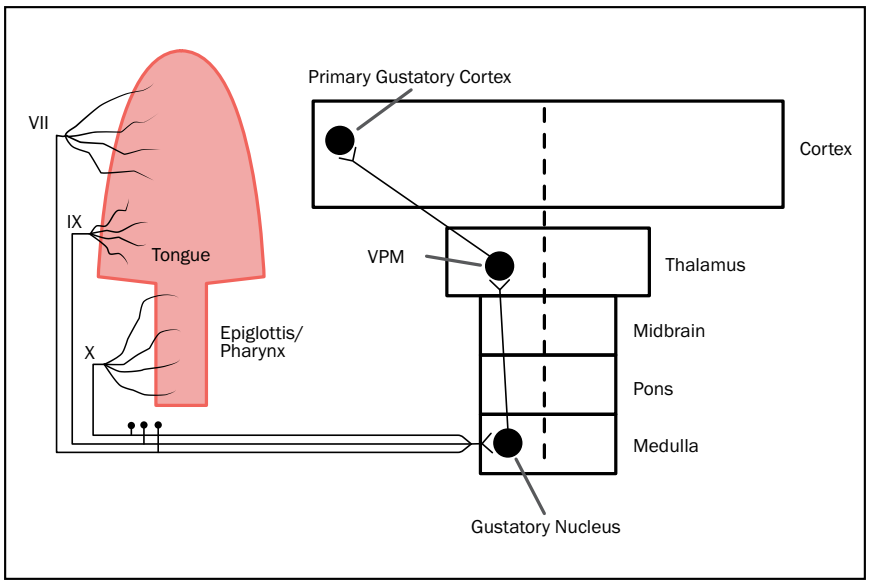


Figure 12.2

Cranial nerves VII, IX and X carry gustatory information from different parts of the tongue and throat to the gustatory nucleus in the medulla. Note that these projections are ***ipsilateral***. The projection continues to the VPM of the thalamus and then to primary gustatory cortex.

A Labeled Line Theory of Taste Coding

While researchers previously believed that perception of each basic taste is detected preferentially in a distinct region of the tongue, recent research has shown that the receptors for different tastants show overlapping patterns of expression on the tongue. Furthermore, while studies based on recordings from gustatory afferents led researchers to believe that most taste cells responded to two or more basic tastes, knockout mouse studies and protein expression assays have confirmed that each taste receptor cell only expresses one type of taste receptor and is thus responsive to a single basic taste.

Given the new results that suggest that each taste cell expresses only one type of receptor, there are two possibilities for downstream effects. In one scenario, the afferent fibers contain inputs from a mix of receptor cell types, while in the other scenario, each afferent fiber receives input from receptor cells expressing the same receptor types, thus carrying information related to just one of the five basic tastes. Again, the findings of previous research have been overturned by new experimental techniques.

The old research, based on electrophysiological recordings from primary gustatory afferent axons, showed a variety of responses to different taste stimuli – a broad response, rather than the narrow tuning that would be suggestive of a labeled line system. However, very recent research (2011) using two-photon calcium imaging in the cortex, shows that there are distinct regions of gustatory cortex that respond just to a single tastant. In order for this kind of distinct topography to exist in cortex, the input from the different taste receptors must remain separate. This is known as a “labeled line” code, because we could “label” the pathway with just the single response.

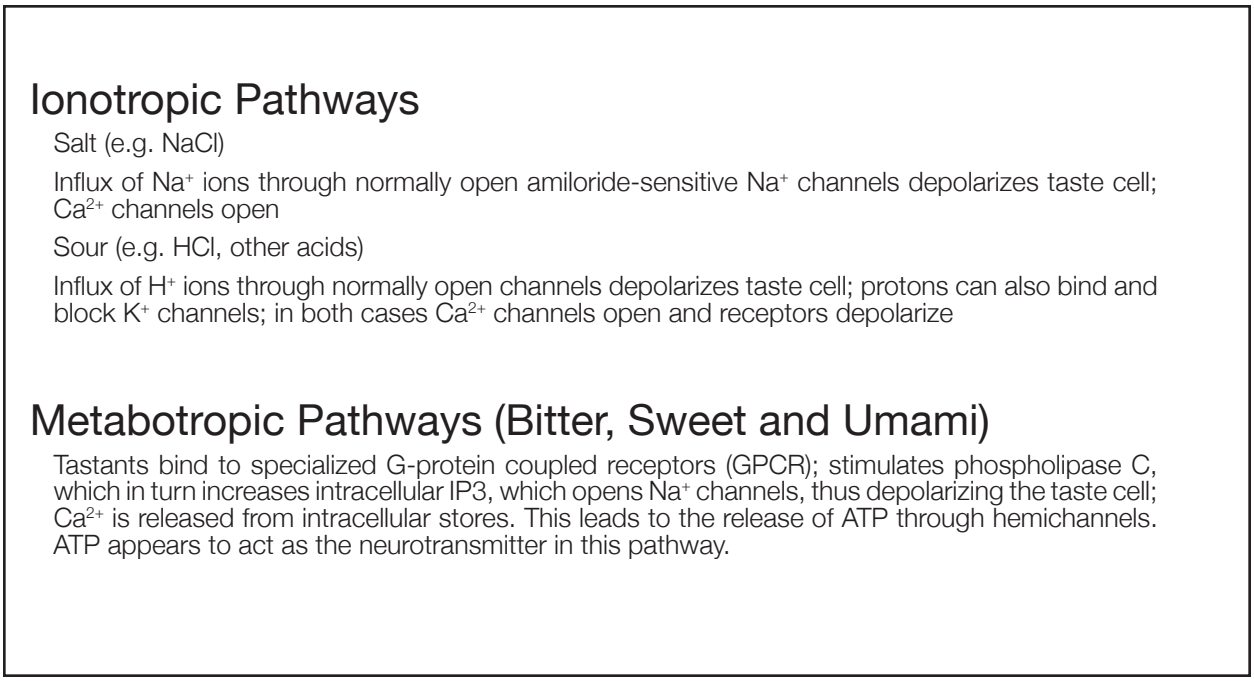


Figure 12.3

Transduction mechanisms of taste cells sensitive to basic tastes. Both ionotropic and metabotropic (G-protein coupled) pathways exist. In all cases, Ca^{2+} causes the release of neurotransmitter onto primary gustatory afferents.

Olfaction

The chemical receptors of our olfactory system are contained in the olfactory epithelium within the nasal cavity. Here, primary olfactory neurons extend hair-like cilia into a mucus membrane in which chemicals in the air are dissolved. These primary olfactory neurons are true neurons and produce action potentials. They are also the only neurons directly exposed to the external environment (covered only by a layer of mucus), which makes them susceptible to damage. Luckily, the olfactory epithelium is one place in the central nervous system where neuronal regeneration occurs in adults. Like taste cells, olfactory receptor neurons are replaced through the differentiation of basal cells. However, severe damage can lead to the loss of olfactory perception, which is called anosmia. For example, head trauma can result in anosmia due to shearing of the primary olfactory receptor axons traveling through the passageways of the cribriform plate.

It is interesting to note that the first complaint of someone suffering from anosmia is often that food does not taste the same. This phenomenon demonstrates the importance of the olfactory system in the perception of flavor, as chemicals from the oral cavity waft around the soft palate into the nasal cavity. Anosmia can also have more severe behavioral consequences. Spoiled food and fire, for example, have important olfactory signatures, which are used to avoid dangerous events.

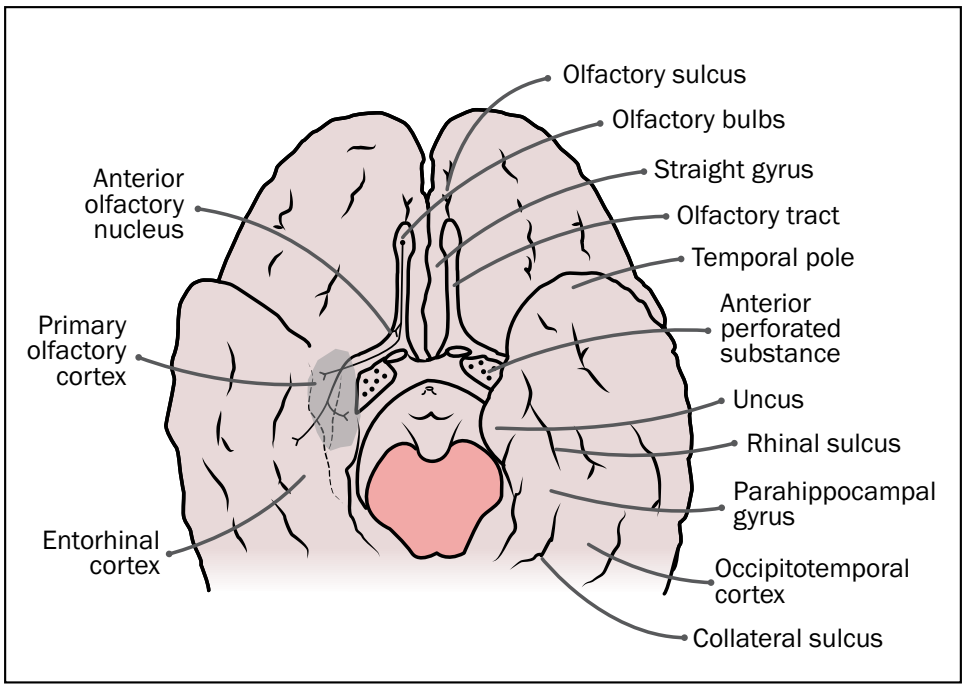


Figure 12.4

The ventral surface of the human brain contains the primary olfactory cortical areas. See text for details.

The axons of the primary olfactory neurons form the olfactory nerve (cranial nerve I), pass through the cribriform plate and synapse in the olfactory bulb. These synapses are clustered in anatomical structures called ***glomeruli*** (singular is glomerulus). There are approximately 2000 glomeruli in each olfactory bulb, and each is innervated by one of the projection neurons of the olfactory bulb, known as mitral and tufted cells. There are also inhibitory interneurons that form connections between glomeruli and adjacent ***mitral*** and ***tufted cells***. These lateral connections are formed by granule cells and periglomerular cells and likely play a similar role in sensory encoding as they do in other sensory systems.

From the olfactory bulb, the olfactory tracts carry signals to many parts of the brain. Interestingly, the olfactory system is the only sensory system that does not first pass information through the thalamus before cortical processing. The olfactory tracts form direct pathways to a variety of primitive cortical regions (paleocortex, which has 3 layers as opposed to 6-layered neocortex). Fibers in the olfactory tract terminate in the olfactory tubercles as well as the amygdaloid body, the uncus, and the anterior portion of the parahippocampal gyrus. The latter two regions form the primary olfactory cortex (Figure 12.4). Secondary cortical projections are directed to entorhinal cortex and orbitofrontal cortex, areas involved in the conscious perception of smell and taste. There are also projections to the medial dorsal nucleus of the

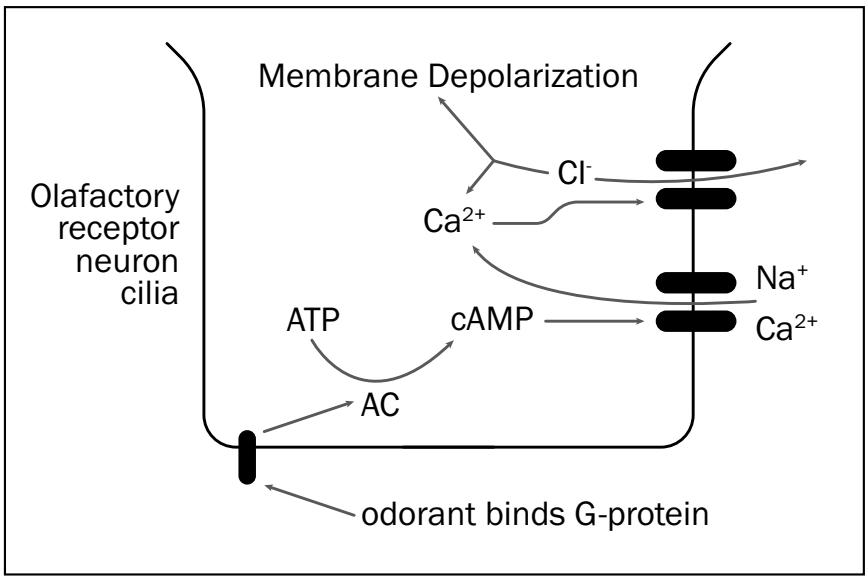


Figure 12.5

The transduction pathway for olfactory receptor neurons. See text for details.

thalamus. So while olfactory information passes directly from glomeruli to cortex, it can be processed in the thalamus as well.

In contrast to the variety of transduction mechanisms of taste cells, olfactory receptor neurons likely use a single mechanism (Figure 12.5). Odorant molecules are dissolved in the olfactory mucosa and bind to transmembrane olfactory receptors. This binding activates a special class of G-protein called G_{olf} proteins, which serve a similar function, as do transducin molecules in photoreceptors. The activated G-protein activates the enzyme adenylyl cyclase, which converts ATP to cAMP. cAMP binds to and opens cation specific ion channels. An inward flux of Ca^{2+} and Na^{+} depolarizes the cell. Additionally, Ca^{2+}-activated Cl^{-} channels are opened. Since olfactory neurons have an unusually high internal Cl^{-} concentration, Cl^{-} ions exit the cell, further depolarizing the membrane. The resulting depolarization triggers an action potential that propagates to the olfactory bulb.

The transduction cascade of the G_{olf} proteins has some interesting similarities to the transduction of light in the retina. In both cases, the appropriate stimulus activates a G-protein, which in turn activates an intermediate enzyme that regulates the level of cyclic nucleotides. These cyclic nucleotides bind to and open the aptly named ***cyclic nucleotide gated channels***. In the case of olfactory receptor neurons, the intermediate enzyme is adenylyl cyclase (AC), which converts ATP to cAMP. The increase in cAMP leads to the opening of the cyclic nucleotide gated channels and depolarization of the membrane. Similarly, in photoreceptors the intermediate enzyme is a phosphodiesterase, which breaks down cGMP into GMP. The decrease in cGMP causes the cyclic nucleotide gated channels to close, shutting off the "dark current" and hyperpolarizing the cell.

Ensemble Coding of Smell

In humans, there are approximately 1000 olfactory receptor genes, distributed across all but two of our chromosomes, and each olfactory receptor neuron expresses only one of these. It is interesting that, unlike in the mouse, which possesses some 1200 olfactory receptor genes, a large proportion of the human olfactory genes have accumulated mutations that make them functionally inactive. It is also important to note that olfactory receptor neurons, unlike taste cells, respond to a variety of odorants. The responses of olfactory receptor neurons to different scents would show a broad tuning response. This broad response profile partially stems from the many chemicals in typical smells such as almond or peppermint, which activate a variety of receptors. Furthermore, each odorant receptor can bind many structurally similar molecules, and each odorant may activate numerous odorant receptors. In short, unlike the labeled line coding in the gustatory system, the olfactory system uses a strategy of distributed/ensemble coding.

It does appear that receptor cells expressing a given receptor protein may be confined to one of four zones in the olfactory epithelium. These cells then project to a corresponding zone of the olfactory bulb. Within a given zone, the various receptor proteins are expressed randomly, i.e. in no particular spatial pattern. Remarkably, every receptor neuron that expresses a given receptor protein sends its axon to one or two specific glomeruli of the olfactory bulb (Figure 12.6). Thus, activation of a glomerulus indicates the presence of odorants that bind to a specific receptor protein. Because each odorant has a complex chemical makeup and binds to more than one receptor protein, a given odorant will activate multiple glomeruli.

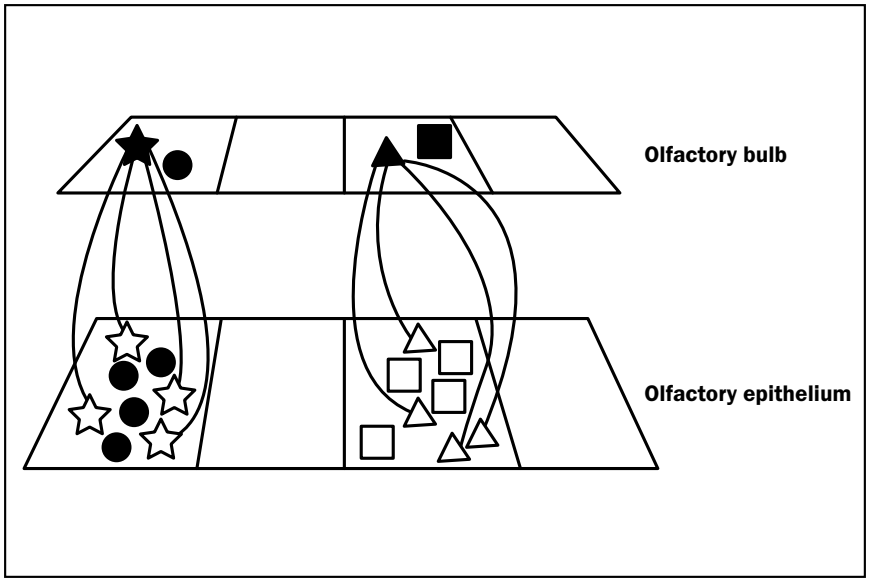


Figure 12.6

Interconnections between the olfactory epithelium and olfactory bulb. Odorant receptor neurons (outlines) expressing the same odorant receptor protein (represented by shape) project to the same glomerulus (filled) of the olfactory bulb.

The pattern of activation of glomeruli can be visualized using optical recording methods (see Technical Appendix of Chapter 4). These experiments have shown that the presence of particular odors leads to a reproducible activation of a specific set of glomeruli. Thus, the specificity in the projection from the olfactory epithelium to the olfactory bulb leads to the presence of an olfactory "odotopic" map. Unlike maps of space, which constrain the organization of sensory receptors in vision and somatosensation, and the frequency map found in the auditory system, this map is not based on any obvious continuous property of the chemical environment. Even so, the brain has apparently developed a way to read the olfactory map and use it to discriminate between thousands of different odorants.

The distributed activity across the olfactory map, representing the activation of specific receptors and their respective glomeruli, does not appear to be the only way information is encoded in the olfactory system. For instance, the pattern of activation in the olfactory bulb can change over the time course of stimulus presentation in a predictable fashion.

This is an example of ***temporal coding***. Temporal coding also manifests when it comes to timing of action potentials; individual neurons of the central olfactory system respond with specific spiking patterns to different stimuli. The importance of the temporal code was demonstrated with pharmacological agents that maintained all normal olfactory properties but disrupted synchronization of neural firing patterns. The experimenters found that animals seem to depend on the precise timing of spikes to discriminate between similar odors. Because normal olfactory sensory processing depends on active respiration, it is not so surprising that the timing of neuronal events is crucial in this modality.

Terms and Techniques

papilla	taste bud	basal cell
taste pore	gustation	pharynx
microvillus	facial nerve	glossopharyngeal nerve
vagus nerve	trigeminal nerve	VPM
tastant	primary gustatory cortex	ageusia
olfactory bulb	odorant	anosmia
cribriform plate	adenylyl cyclase	olfactory glomeruli
olfactory tubercle	olfactory cortex	G_{olf}
medial dorsal nucleus	population code	entorhinal cortex
labeled line	distributed code	ensemble code
cyclic nucleotide gated channels	nucleus of the solitary tract (gustatory nucleus)	odotopic map

Pain: More Than a Modality

13

Table of Contents

Ancient tradition has it that there are five senses: touch, sight, hearing, taste and smell. Note that pain is not included, nor is pleasure. Pain is somehow different. Sometimes it is experienced much like other cutaneous sensations when, say a mosquito stings your arm. Other times, it is a much more impactful experience than this.

We cannot conjure up an experience of pain in the same way that we can close our eyes and visualize the familiar face of a relative or 'hear' the words and tune of a familiar song. We remember having pain, and we recognize a pain that recurs, but we can't voluntarily recreate the sensation itself. Nor does pain occur in our dreams unless it is really present, as when we are awakened by a stomach or leg cramp that was incorporated into a dream. (Pain shares this last characteristic with smell, which gives us a clue that the mechanisms of pain are intimately related to very old structures of the brain). The effect of pain on a person highly depends on context. A soldier seriously injured in the heat of battle may feel no pain until back in the hospital, but the slightest pain may be intolerable to someone who is frightened, depressed or in a bad mood.

Pain clearly provides a protective function by alerting us to potential or present injury, but either too much pain or the inability to experience pain can be life threatening. As part of its protective function, pain tends to immobilize all or part of the body, which is maladaptive when one needs to escape the cause of the pain, such as the bite of a predator or a fire. On the other hand, individuals who are congenitally insensitive to painful stimuli accumulate injuries that shorten their lives.

Pain is also a unique modality because it can be modified by substances that have no known analgesic properties themselves (the placebo effect), but it often cannot be eliminated by cutting the nerves to the painful part. Thus, there are many aspects of pain that make

it more of an experience than a sensation, and it is somewhat artificial to treat pain as just another submodality of somatic sensation. Nonetheless, there are clearly components of the somatic sensory system that play crucial roles in the experience of pain and shape some of its properties. We will begin by discussing these.

The receptors responsive to noxious stimulation are the ***free nerve endings*** of two classes of afferent axons, the thinly myelinated A (Group III) and the unmyelinated C fibers (Group IV). In humans there are at least two types of myelinated A fibers that are nociceptive. One class responds only to mechanical stimuli and the other to both mechanical and thermal stimuli. They both may be sensitive to chemical stimulation. 75% of afferent fibers in a peripheral nerve are C fibers and in primate cutaneous nerves over 90% of these are nociceptive. All C fibers observed in transcutaneous recording studies of human nerves have been nociceptive. The major type of unmyelinated nociceptive neuron is the ***polymodal receptor*** that responds to mechanical, thermal and chemical stimulation. There is increasing evidence pointing to a class of nociceptors, referred to as silent or sleepy nociceptors, that are normally unresponsive to noxious stimulation but become active when the tissue they innervate becomes inflamed.

Two qualities of pain sensation can be correlated with the two major classes of afferent fibers. When a finger is burned by hot water, there is an initial sharp pain and the sensation of hot, both of which decay after the finger is removed from the water. This is ***fast pain*** (sometimes referred to as ***first pain***). A second or so later there emerges a diffuse, poorly localized, long lasting pain that causes more suffering than the initial sharp pain. This is ***slow pain*** (sometimes referred to as ***second pain***). Electrical stimulation of an exposed human cutaneous nerve while recording the compound action potential shows that the fast, spatially discriminable component of pain is mediated by Aδ afferents, while the slow, more excruciating and poorly localized pain is mediated by C fibers.

When a noxious stimulus (e.g. bumping your knee) causes no tissue damage, the activity of the A and C fibers subsides after a while and the perception of pain disappears. If tissue is actually damaged, with some cell destruction, a new stage of pain production sets in (Figure 13.1). Substances released by the damaged cells and by the terminals of the nociceptive neurons initiate the phenomenon of ***inflammation***. Among these substances are potassium, peptides such as bradykinin and substance P, serotonin (5-HT), histamine and prostaglandins. These chemicals cause dilation of blood vessels, increased leakage of fluid from these vessels and increased activity and sensitivity of the nociceptive afferents. The result is the classic Latin quartet of inflammation: ***rubor*** (redness), ***tumor*** (swelling) and ***dolor*** (pain) and ***calor*** (heat).

The redness or flare is of interest, as it spreads beyond the site of injury by way of an ***axon reflex***. Here, impulses from the injured zone, in addition to traveling centrally, send impulses backwards (antidromically) down the other branches of the activated sensory fibers to neighboring skin regions (Figure 13.1). These antidromic action potentials release some substance from the sensory terminal causing vasodilatation. These effects sensitize the adjacent nociceptors, lowering their thresholds/increasing sensitivity. This cascade may

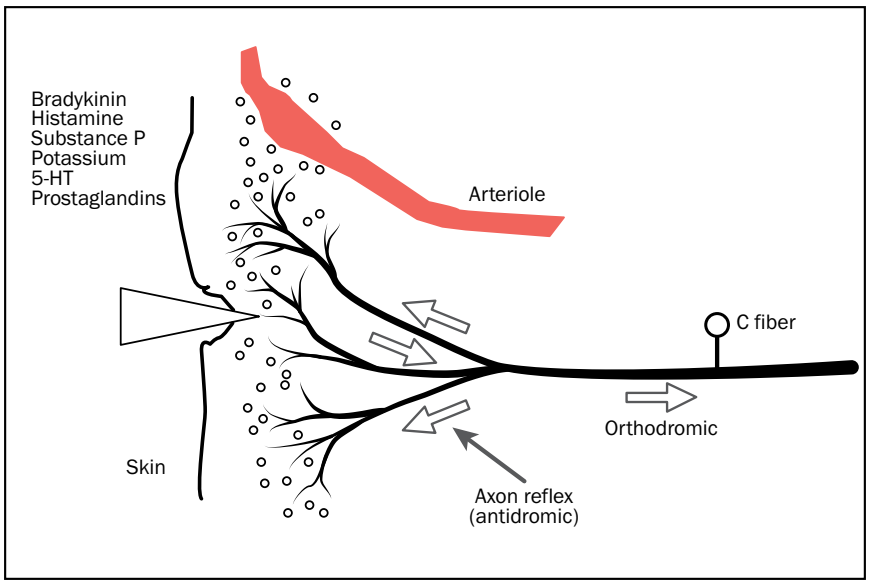


Figure 13.1

The axon reflex and inflammation. (see text)

also recruit some of the 'silent' nociceptors mentioned earlier. The region of lowered pain threshold immediately surrounding the injured site is called the ***zone of primary hyperalgesia***. Surrounding this region of primary hyperalgesia and extending beyond the flare is an area that has not been directly involved in the injury but also exhibits a lowered threshold for evoking pain by gentle mechanical stimulation. This is the zone of ***secondary hyperalgesia*** and its presence occurs from changes in the central nervous system, as will be discussed later.

The Sensation of Hot

The neural mechanisms mediating pain play an interesting role in the sensation of hot. Temperature sensitivity depends on two types of peripheral receptors generally known as cold and warm receptors. These two receptor systems are activated respectively when the skin temperature falls below or rises above a temperature of about 32ºC, which is experienced as neutral. However, when the skin temperature rises above 45ºC, the cold receptors once again become active, a phenomenon thought to be responsible for the ***paradoxical cold*** sensation sometimes experienced when one steps into a hot shower (Figure 13.2). At around this temperature, nociceptor discharge also begins, and when all three receptors (warm, cold and nociceptor) are activitated, the sensation is reported as 'hot'.

Itch

For many years itch and pain were considered together, as early evidence indicated that our sensation of itch is also carried by slowly conducting C fibers. The behavioral response

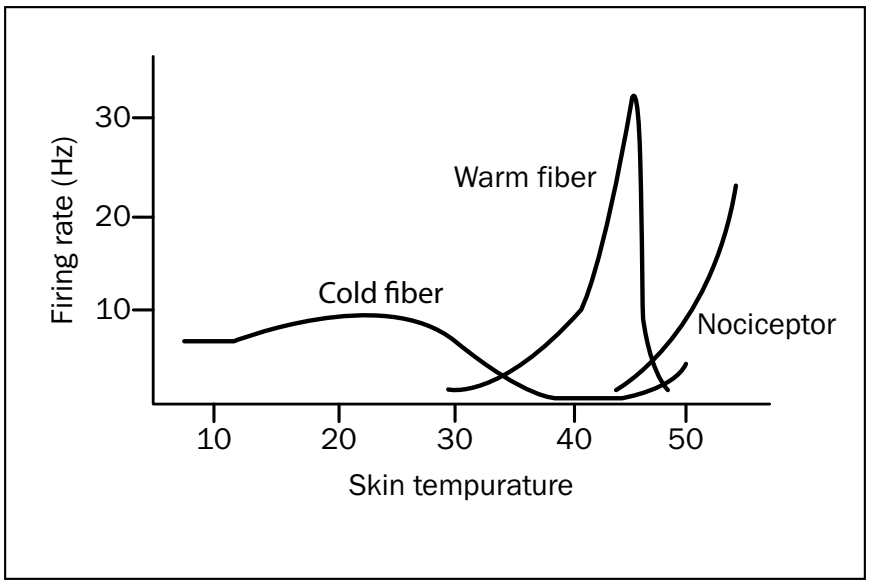


Figure 13.2

Thermal sensitivity of cutaneous cold, warm and nociceptor afferents. Note that the proper interpretation of temperature requires depends on ensemble codings, reminiscent of wavelength encoding in the retina (Figure 7.2).

to itch, however, is quite distinct from that of pain. Itch induces a scratch reflex, whereas pain evokes a withdrawal reflex. Recent physiological recordings from the peripheral nerves of humans provides a further basis for distinguishing the two, as researchers have found that a small, specialized subset of C fibers seem particularly sensitive to histamines, and, when activated, cause the ***pruritic*** sensation we call itch.

Referred Pain

When certain internal organs are damaged, the pain is often felt somewhere else, i.e. it is ***referred*** to a part of the body where we have felt pain in the past (Figure 13.3). For example, insufficient oxygen to the heart muscle can cause pain in the left arm, and an inflamed gallbladder causes pain in the back. People with degenerative disease of the temporomandibular joint (where the lower jaw or mandible articulates with the temporal bone of the skull) will often believe that they have a toothache or an earache.

There appear to be several possible mechanisms responsible for referred pain. In general, the affected organ and the place where the pain is felt are innervated from the same spinal or brainstem levels. Thus, nociceptive information arriving at the cervical spinal cord from the damaged heart muscle enters pain pathways normally associated with the arm. A similar explanation applies to the initial pain of appendicitis, which is referred to the area of the navel, rather than to the right lower quadrant of the abdomen (which begins to hurt when the lining of the abdominal cavity, the peritoneum, becomes inflamed). Although the appendix ends up far from the navel in adults, during embryological development they are close together and

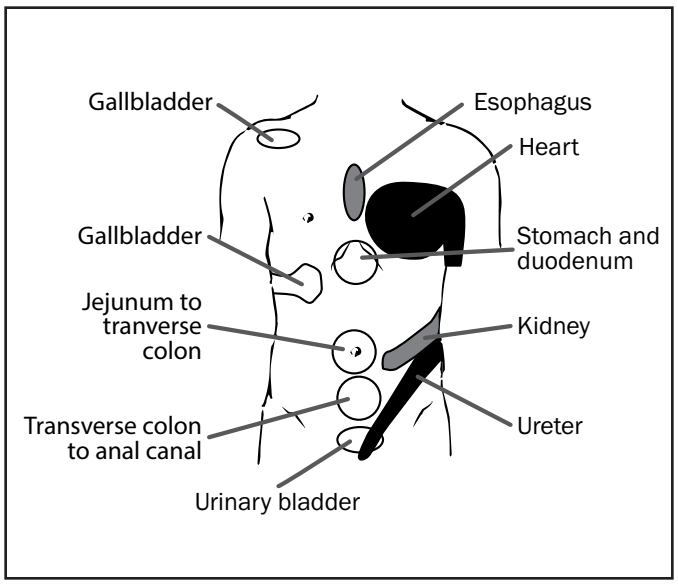


Figure 13.3

Typical locations of referred pain due to disease of the internal organs.

receive their innervation from the same spinal segments. Also, damage to other organs may give rise to nociceptive impulses causing muscle spasm in the body, and this spasm then gives rise to the perceived pain.

A related phenomenon, called ***allodynia*** (meaning "other pain") refers to a condition wherein normally innocuous stimuli cause pain or when pain occurs away from the area actually stimulated. This latter form of location transfer differs from referred pain because referred locations are regions that are innervated by the same nerves during development. Instead, a location allodynia generally occurs in a body part close to the stimulated area. An example of touch allodynia is pain from the touch of clothing. Thermal allodynia occurs from a draft of warm or cold air on the skin. One case, presented in the popular press, described a patient who dreaded the breeze from a ceiling fan because it felt like razors cutting his flesh.

Central Pathways of Somatic and Cutaneous Nociception

After entering the spinal cord, A and C fibers synapse almost immediately on at least two types of neurons in the dorsal horn. ***Nociceptor-specific*** neurons are activated only by noxious stimulation. The second variety, the ***wide dynamic range neurons***, respond to both noxious and innocuous input. Additional synapses may intervene before the fibers cross the midline and rise in the spinothalamic tract (Figure 13.4). Some of these axons terminate before the thalamus in the reticular formation and other components of the brainstem. Because not all of

these projection fibers reach the thalamus, a better name for the pathway is the ***anterolateral system***, reflecting the position of the fibers in the anterolateral column or funiculus of the spinal cord. (It should be noted that the anterolateral system also contains axons involved in crude touch, i.e. touch that lacks the fine discrimination of that mediated by the dorsal column/medial lemniscal system.)

The thalamic targets of the anterolateral system include the ventral posterior nuclei that also receive the input of the DC/ML system. In addition, the anterolateral system also contacts cells in the intralaminar nuclei and a complex of nuclei called the posterior group. Because of these multiple thalamic relays, the pain (and temperature) information traveling in the anterolateral system reaches more than one cortical area, and there is no single specific cortical area for pain. Neurons responding to noxious stimulation are found in area S1 as well as in a second somatic sensory area, S2, which contains cells with receptive fields on both sides of the body. The organization of the cortical representation of pain is poorly understood.

Visceral Nociception

Pain sensation from our internal organs is diverse and distinct from somatic pain of the skin, muscles, joints, and bones. The extent to which visceral pain is experienced depends very much on the particular organ that is damaged. Many of our "hollow" organs (gut, bladder, uterus) are lined with nociceptors and are extremely sensitive to even the smallest lesion, presumably because these come in contact with agents from the outside world. On the other

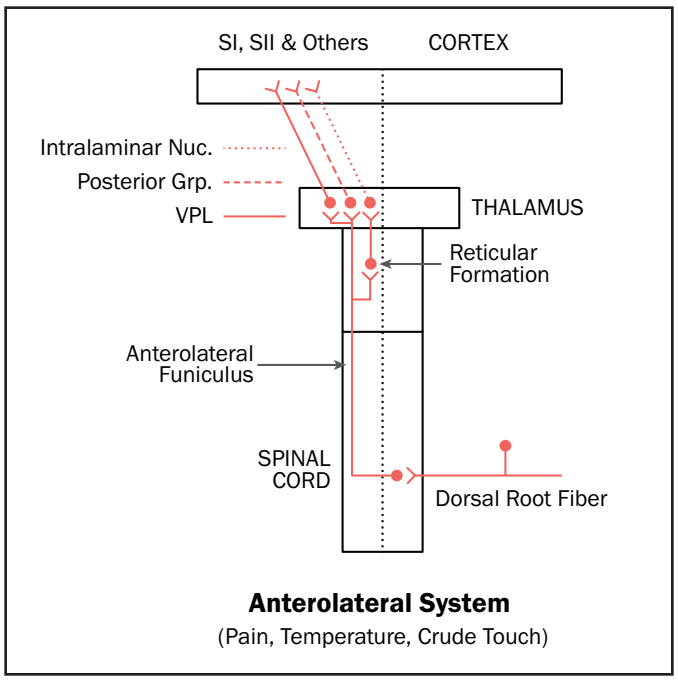


Figure 13.4

The spinothalamic or anterolateral system.

hand, the liver and kidneys are virtually insensitive to pain.

Visceral nociceptive afferents travel with both sympathetic and parasympathetic efferent nerves. Some of these afferents' cell bodies lie in the dorsal root ganglion (like somatic pain fibers) and send signals up the spinal cord after synapsing in the dorsal horn. Others reach the brainstem through autonomic nerve pathways. The overall number of visceral afferents is quite small compared to cutaneous nociceptors (less than 10%); however, in the spinal cord a surprisingly large proportion of cells respond to visceral pain, because incoming visceral pain signals significantly diverge and synapse on many areas in the spinal cord.

Observations Indicating the Anomalous Status of Pain as a Sensation

Various evidence suggests that pain perceptions can be influenced by processes occurring in the central nervous system. It appears, for example, that the phenomenon of secondary hyperalgesia depends on mechanisms that allow the normally non-painful signals in the A fibers to activate the central pain pathways that begin in the dorsal horn. As evidence for this, injection of a local anesthetic into the zone of primary injury or cutting the nerves to the damaged area eliminates the zone of secondary hyperalgesia immediately.

More compelling evidence comes from experiments by Torebjörk and colleagues who stimulated single A fibers in the lower leg. This stimulation normally evoked a non-noxious sensation projected to a small region on the dorsum of the foot. However, when this same region of the foot was injected with capsaicin, a substance that sensitizes pain fibers (and makes peppers hot), activation of the single A fiber now evoked a painful sensation. Because the electrical stimulation was applied to the axon well away from the site of injection, the stimulation bypasses the receptor, eliminating the possibility that the pain stemmed from increased sensitivity at the axon's terminals. Therefore, the new sensation of pain evoked by the electrical stimulation must have resulted from tonic input from the site of injection that caused the message in the artificially stimulated A fiber to activate the central pain pathway. A possible neural substrate for this will be discussed below.

Another set of observations led Patrick Wall, an English physiologist, to argue that the psychological experience of pain is not simply equivalent to the discharge of nociceptive afferents. Experiments have found that C fiber discharge did not always correspond with pain threshold, intensity, location and duration. For example, C fibers begin firing at 41^{o}C whereas the thermal pain threshold is 45°C. Additionally, psychological estimates of pain intensity do not always correlate well with the discharge rates of C fibers. The stimulus area also produces a curious effect: at 43°C, a 1 mm diameter stimulus probe produces pricking pain, a 2-5 mm probe produces stinging pain and a 20 mm probe produces pleasant warmth. Finally, pain can continue to increase after C fibers decrease their firing rates. These observations suggest that processes in the central nervous system are capable of maintaining the neural activity subserving the psychological perception of pain, even when the nociceptor input has died out.

The ability to experience pain in the absence of explicitly noxious stimulation is demonstrated by the phenomenon known as the ***thermal grid illusion***. In 1896 Thunberg showed that the burning pain associated with contact with very cold objects (e.g. dry ice) can be produced by touching the skin with interlacing bars that were alternately warm and cold. Making all the bars warm or cold produced normal sensations of warmth or cold, respectively, but no painful sensations. The painful experience evoked by the alternating warm and cold bars clearly shows that pain can be evoked by a particular spatial pattern of non-noxious stimuli. The neural mechanism responsible for this phenomenon is unknown.

Another example of central control of pain is the persistent pain in the absence of obvious noxious stimulation, or ***pathological pain***. These conditions arise after damage to peripheral nerves or to parts of the brain involved in somatic sensation. For example, amputation of a limb may be followed by development of a painful ***phantom limb***. Lesions of the thalamus caused by disease are often associated with unbearable pain in a part of the body that is, by sensory testing, completely anesthetic, a phenomenon called ***anesthesia dolorosa***.

Central Regulation of Nociceptive Inputs

A variety of mechanisms have now been identified as modulating nociceptive signals traveling through the central nervous system. As mentioned before, the body can be immobilized by pain, rendering the organism unable to fight or flee. It therefore makes sense for the CNS to provide some mechanism to govern such a necessary but potentially dangerous neural state. At one end of the spectrum there are mechanisms to increase the sensitivity of the nociceptive system (primary and secondary hyperalgesia) and at the other, a way to limit the central effects of nociceptive input.

The first attempt to neurally explain the modifiability of signals in the pain pathways was the ***gate control theory*** of Melzack and Wall, who were interested in the so-called ***counter-irritation effect***. When we hurt ourselves, we often instinctively rub the skin near an injured point to relieve the pain to some degree. These two investigators postulated that pain impulses reaching the spinal cord must pass through a control point in part of the dorsal horn called the ***substantia gelatinosa***. They proposed that rubbing and activating large diameter cutaneous fibers closed a 'pain gate' and prevented the nociceptive signals from ascending to the cerebrum. One hypothetical circuit that accounts for the counter-irritation effect and several other observations is illustrated schematically in Figure 13.5.

Neuron P in Figure 13.5 projects into the spinothalamic or anterolateral system and its activity is associated with the perception of pain. This neuron is under some tonic inhibitory control by the 'gate keeper cell' (synapse 2), an interneuron in the substantia gelatinosa. When the nociceptive afferents are stimulated, they inhibit the gatekeeper (synapse 1) and activate the P cell (synapse 3), with pain resulting. Rubbing the leg activates the A_{β} fibers, which contribute excitation to the gatekeeper cell (synapse 4) and weak excitation to the P cell (synapse 5). The excitatory effect on the gate keeper dominates, so the net inhibition on

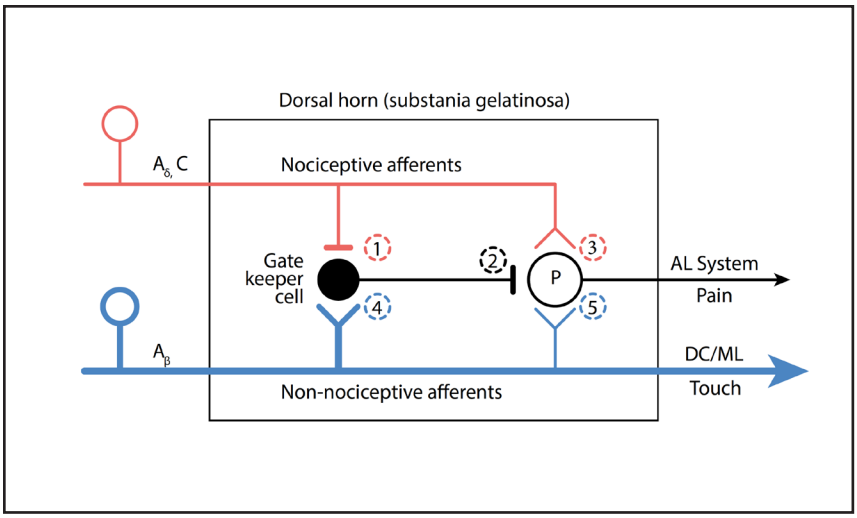


Figure 13.5

A possible neural circuit for the gate control mechanism in the dorsal horn. See text for details.

the P cell increases. That is, exciting the inhibitory interneuron inhibits the P cell, reducing the level of pain. This model predicts that electrical stimulation of the dorsal columns should also reduce pain by antidromically activating the non-nociceptive afferent axon and synapse 4. Indeed, this actually happens, and some patients with chronic pain have been equipped with dorsal-column stimulators. It is not known, however, whether this works through the mechanism proposed in Figure 13.5.

When there is minor injury, activity in the A_β fibers can overcome the C fiber inhibition via their excitation of the gatekeeper cell. If there is a severe injury with intense activity in the nociceptive afferents, however, the inhibition of the gatekeeper cell (at synapse 1) will be profound. In the face of powerful input from the nociceptive afferents, the excitatory A_β-gatekeeper synapse will be overpowered. Now, the weaker input via synapse 5 onto the P cell causes the A_β activity to excite the P cell, which evokes pain. Such a mechanism, which essentially 'switches' the A_β activity into the pain pathway, could account for secondary hyperalgesia in regions around the primary site of injury. The secondary hyperalgesia would be eliminated immediately by anesthetizing the nociceptive afferents, as this would remove the inhibition on the gatekeeper cell, and this is what is observed. Note that there is a cross-over of A_γ/C fibers from neighboring skin areas at the dorsal horn.

As discussed above, itch and pain seem to have distinct neural pathways. Interactions between these experiences share some similarities with the interactions between pain and tactile sensation. Pain, for example, can override or inhibit itch, but the painful stimulation can

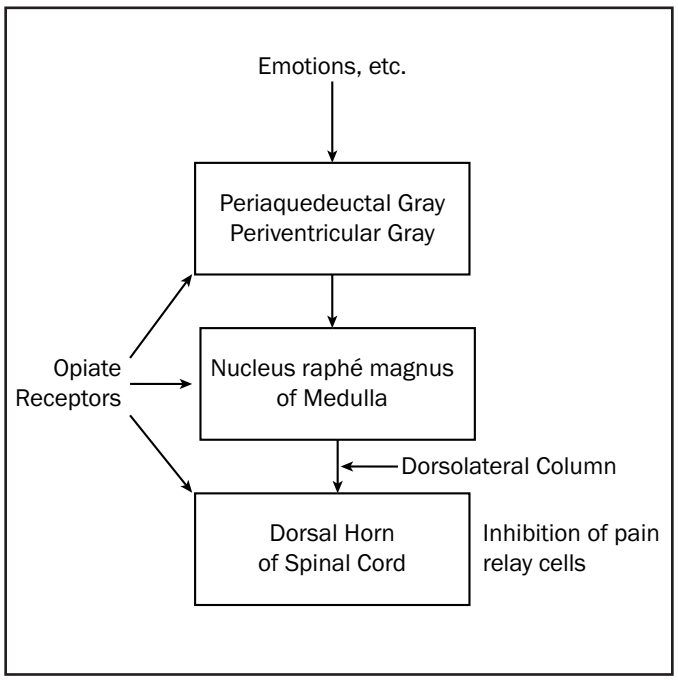


Figure 13.6

Descending pain control pathway

be many centimeters from the site of itch, suggesting that central mechanisms in the spinal cord may play a modulatory role in this interaction as well.

The gate control mechanism can also be activated from supraspinal structures (Figure 13.6). Emotions and other psychological factors provoke activity in certain neurons of the *periventricular gray* (PVG) and *periaqueductal gray* (PAG) matter of the upper brainstem and diencephalon, which in turn activate the *raphé nuclei* of the medulla and certain cell groups in the dorsolateral pons. Fibers from these brainstem nuclei project via the dorsolateral columns of the spinal cord to the dorsal horn. This descending pain-control pathway is also activated by nociceptive impulses ascending in the anterolateral system, indicating that there is a negative feedback component to the system; i.e. pain-related impulses engage the mechanism of pain control. Presumably this occurs via the collateral projections of the spinothalamic system into the brainstem. Some of the disturbances associated with pathological pain may be related to imbalances in this 'automatic gain control' of the pain pathways, such that the pain mechanisms are released to discharge spontaneously.

Narcotic analgesics likely act on certain stages of this pathway. Microinjections (25 ng) of opiates into PAG and PVG cause analgesia. Such injections also cause inhibition of lumbosacral nociceptive neurons, while neurons responding to non-noxious stimulation are unaffected. Cells in PAG and PVG are activated by intravenous morphine, and these cells also increase their activity when painful stimuli are applied, indicating that they are involved

in a system of negative feedback that regulates the flow of nociceptive impulses in the CNS. Electrical stimulation of PVG, PAG, or nucleus raphé magnus (NRM) produces long lasting analgesia. Importantly, all of these changes observed after opiate injections are blocked by opiate antagonist Naloxone, suggesting that these effects are all mediated by the opiates.

Other studies have also implicated the PVG, PAG and raphe nucleus in pain control. Electrical stimulation of this system inhibits nociceptive neurons in the lumbosacral cord and the inhibitory effects are specific for nociceptors and do not affect other modalities. Also, transection of only the dorsolateral funiculus of the cord blocks the effect of electrical stimulation of PAG, PVG, NRM and of microinjections of morphine.

A considerable number of humans suffering from intractable pain, usually due to metastases, have had electrodes implanted in and near the PAG and PVG. The patients are provided with stimulators so that they can activate their own electrodes. This ***deep brain stimulation*** has provided almost all patients with complete relief from their pain mediated by opiates, as the effects of stimulation are also blocked by Naloxone. Unfortunately for these patients, continuous stimulation results in refractoriness and a return of the pain. When the electrical stimulation no longer works, neither do narcotics. Return to intermittent stimulation results in a return of sensitivity to both electrical stimulation and to narcotics. The effects of a stimulus train may last for hours. The pain relief is not associated with increased threshold to pain, a characteristic also of opiate analgesia.

Among the least well-understood phenomena related to regulation of pain is the ***placebo effect***, in which a treatment with no intrinsic pharmacological effect may still produce analgesia. Evidence that placebos can modulate pain transmission has recently been demonstrated in fMRI experiments, which found that placebo analgesia was related to activity in particular brain areas associated with pain processing (thalamus, insula, anterior cingulate). In addition, Naloxone blocks the analgesic placebo effect, suggesting that this phenomenon is also mediated by endogenous opioids.

Terms and Techniques

nociceptor	inflammation	tumor
rubor	calor	dolor
slow pain	fast pain	phantom pain
anesthesia dolorosa	referred pain	allodynia
free nerve ending	nociceptor	silent nociceptor
A-delta axons	C-fibers	Group III axons
Group IV axons	anterolateral system	spino-thalamic tract
zone of Lissauer	substantia gelatinosa	thermal grid illusion
intralaminar nuclei	primary hyperalgesia	secondary hyperalgesia
axon reflex	analgesia	placebo effect
periaqueductal gray matter	periventricular gray matter	raphé nuclei
Naloxone	opioid receptors	pruritus (itch)
gate control theory	counter irritation effect	

Control Systems: Gain, Feedback and Set-points

14

Content Disclosure: The regulation of body weight is discussed in this chapter.

Table of Contents

When attempting to understand a neural system, it is often useful to divide it into separate functions or operations and put them in 'black boxes.' One does not need to know exactly how a particular function is implemented by a particular box, only that the function is carried out. Because each box represents an operation, the resulting diagram is said to be an ***operational model*** of the system. Such models help simplify complex systems, so it is important to get accustomed to this approach, to know its limits, and to be able to apply it to neural systems of various kinds. Here we will introduce some of the formal concepts of control systems by discussing the familiar problem of regulating the temperature of a room. We will then extend the example to the problems of regulating body temperature and body weight. In later chapters these ideas will be applied to motor control mechanisms.

Figure 14.1A shows the simplest of all system models, a black box with an input and an output. The G in the box indicates that this box performs some kind of multiplication, making the output equal to (G=1), larger than (G>1) or smaller than (G<1) the input. G can be negative or positive. If G has the value of -1, it simply changes the sign of the input. G symbolizes the ***gain*** of the system, defined as ***the ratio of output magnitude to input magnitude***. If both the input and output magnitudes have the same units (e.g. volts, centimeters, etc.), G will be a dimensionless number. If the units of input and output are different, this difference will be reflected in the units of G (e.g. volts/centimeter, spikes/decibel).

Varying the input signal over a certain range causes the output signal to vary. If the box is an audio amplifier (G>1), the input is a microphone, and the output controls a loudspeaker, speaking into the microphone will reproduce the sounds at a higher volume. The overall gain of such a system can be expressed as the ratio of the output sound level to the input sound level or using the decibels, as discussed in Chapter 8.

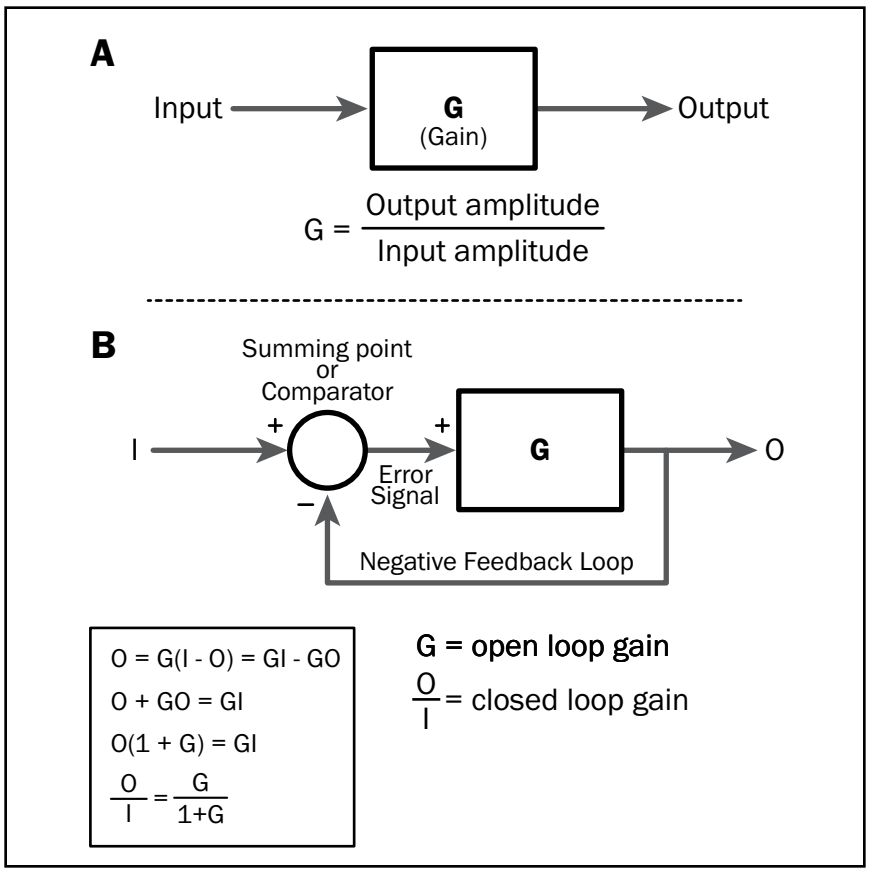


Figure 14.1

Operational system models. A. The gain of a 'black box'. B. A system with negative feedback. The feedback loop makes the overall gain (the closed-loop gain) of the system different from G (the open-loop gain). The calculation used to derive the closed-loop gain is given in the box.

Figure 14.1B adds a ***feedback loop*** to the circuit of Figure 14.1A. The output goes not only to some device, such as a loudspeaker, but is also sent back to a ***summing point*** where it is combined with the input. Many control systems are designed to compare the feedback signal to the input, so the summing point can also be called a ***comparator***. In Figure 14.1B the signal is in fact summed with a minus sign, so the feedback is negative. The signal sent to the black box is the difference between the input and the output and is sometimes referred to as an ***error signal***.

A system such as that of Figure 14.1B can be used in two ways. If the intention is to vary the input signal and modulate the output, the system is acting as a ***controller***. For example, the system might be used to constantly vary the position of a video camera monitoring the interior of a bank. An appropriately varying input signal to the controller would create the output signal needed to drive the motors on the camera's turntable. On the other hand, a ***regulator*** system applies a constant input signal to prevent the output signal from varying due to perturbations imposed from outside the system. Observe that if the input is constant, the feedback loop will insure that the output remains constant. If the feedback signal drops

below the input, a positive error signal will automatically increase the output to compensate. The reverse will happen if the feedback signal exceeds the input signal.

An example of a regulator system is a thermostat, which keeps the air temperature in a room constant. The thermostat compares the actual room temperature to the desired room temperature set on the thermostat. If there is a difference, an error signal is sent to the appropriate effector mechanism to raise or lower the temperature of the room (Figure 14.2). In this case, the reference input signal is referred to as a ***set-point***. More will be said about this function below, but first we need to point out another advantage of using a feedback loop such as that of Figure 14.1B.

As noted above, the gain of the system in Figure 14.1A is simply the gain of the mechanism in the black box. However, with the feedback loop connected, as in Figure 14.1B, the gain of the system (i.e. output/input) is no longer G but $G/(1+G)$. The calculation proving this is enclosed in the box in Figure 14.1B. (This result depends on the assumption that the input and output appear at the comparator at the same time, i.e. the system imposes no delay.) $G/(1+G)$ is the ***closed loop gain*** of the system. Interrupting the feedback loop (i.e. opening the loop) restores the system gain to G, the ***open loop gain***.

Note that if G is very large, the closed loop gain is very nearly 1, meaning the output will be controlled by the input and will have about the same magnitude. Making G even larger will never make the closed loop gain greater than 1. This configuration confers a great advantage on a control system. It allows the system to employ components of very high gain, but keeps the overall system gain close to 1. Also, if G is very high, small variations in G have little effect on the overall gain of the system. This makes it very stable in the face of changes in the properties of the components in the black box.

Feedback and Thermoregulation

The thermoregulatory system of the body is an example of a neural system that uses feedback to maintain body temperature within the physiological range. Not all animals have such systems. ***Poikilotherms*** are able to survive with highly variable internal temperatures. ***Homeotherms***, such as humans, maintain a stable internal temperature through physiological thermoregulatory mechanisms. Hibernators are homeotherms some of the time and poikilotherms at other times. During REM sleep humans briefly become poikilotherms.

Why is thermoregulation important? Our chemical systems work best at a particular temperature. The Q10, which is the increase in the rate of an enzymatic reaction for every 10°C rise in temperature, is 2 or 3 for many enzymatic reactions in the range around body temperature. However, many tissues rapidly denature (self destruct) above about 44°C, and the enzymatic denaturation process may have a Q10 of around 200! Thus, humans can survive only a few degrees below lethal temperature.

Which temperature is under control? If one monitors the major thermoregulatory mechanisms of the body, one will see that their activity varies continuously while the temperature of the brain changes hardly at all. This finding offers the misleading impression that brain temperature has nothing to do with thermoregulatory mechanisms. However, one must remember that the purpose of a regulatory system is to keep a ***controlled variable*** at some desired value.

One can ascertain the importance of the controlled variable by manipulating it. In one experiment demonstrating the importance of brain temperature, subjects held ice in their mouths, cooling the blood passing through the carotid arteries and, in turn, cooling regions of the brain. Even if the air temperature was warm, physiological mechanisms were invoked to conserve and generate body heat, as though the environment was cold. Thus, ***brain temperature is the controlled variable***; it is 'defended' by the thermoregulatory mechanisms of the body and not allowed to vary outside a narrow range.

Neural Mechanisms in Temperature Regulation

Temperature in the body is controlled by a similar mechanism as a thermostat's maintenance of room temperature by means of air conditioning and heating systems (Figure 14.2). When the room temperature is perturbed by some external factor, such as opening a window, an error signal appears on the output of the thermostat. The heating (or cooling) system kicks in to return the air temperature at the location of the thermostat to the desired temperature (or set-point).

The body's temperature control system relies on neurons whose activity is exquisitely sensitive to the local temperature. Such cells exist in various parts of the CNS including the spinal cord, but a particularly critical site is located in the ***preoptic region of the hypothalamus***. Here the cells monitor the temperature of the hypothalamus, which is determined by the temperature of the arterial blood reaching the brain (Figure 14.2).

Before arriving at the brain, this blood has been warmed or cooled as it circulated through other parts of the body. In short, the brain controls body temperature in order to control the temperature of the carotid blood pumped to it from the heart.

The 'set-point' of the hypothalamic thermostat is an abstraction that captures an essential ***property*** of the system and does not correspond to any particular component or process. At a given moment there is some blood temperature with which the hypothalamus calls for neither heating or cooling. When the blood temperature deviates from this value, the set-point mechanism produces an error signal that results in an action to counteract the deviation.

When the hypothalamic temperature is too low, its circuits initiate heat production and conservation through such mechanisms as shivering, constriction of the skin's blood vessels and, most importantly, appropriate behaviors such as seeking shelter, donning warm garments etc. The rising body temperature warms the blood entering the hypothalamus and returns it to its desired temperature, the set-point. When the hypothalamus is too warm, heat loss

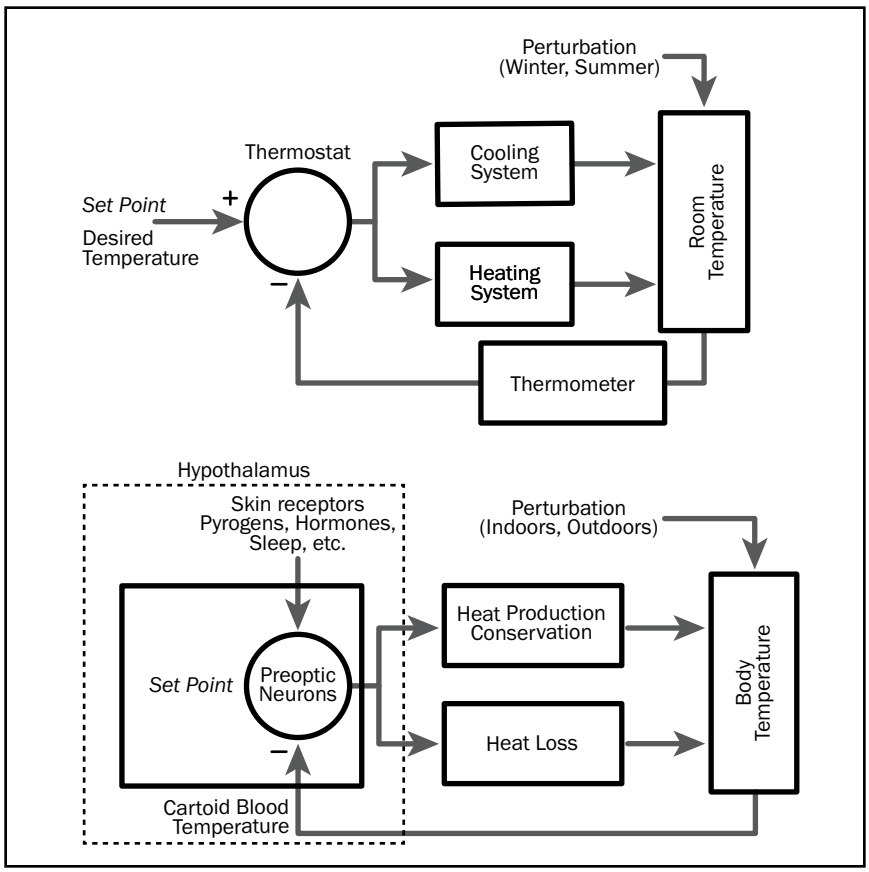


Figure 14.2

Thermoregulatory systems. A. System for controlling the temperature of a room. B. Thermoregulatory system of the mammalian body. (see text for details)

mechanisms are invoked to cool the blood reaching the brain from the rest of the body. This cooling is accomplished by sweating (evaporative heat loss), peripheral vasodilatation and, again most importantly, appropriate behaviors such as decreased activity and seeking shade.

The set-point temperature is determined by factors such as the properties of the cells' temperature-sensitive membrane conductances, neural signals reaching the cells from other parts of the CNS and the skin, and the hormones and other substances making up the chemical environment of the cells.

As just suggested, the temperature sensitivity of the hypothalamic neurons can be modified by chemical and neural factors, meaning the set-point can correspond to different temperatures at different times. One important neural signal occurs when cold receptors of the skin initiate heat production and conservation in anticipation of a decrease in core temperature. Cutaneous warm receptors also play a role, but the effect is of lower magnitude. *Pyrogens*, chemicals produced by the body or by micro-organisms during infections, effect a fever by raising the body temperature by altering the set-point upwards. Circadian rhythms in hormones or neural activity also affect the hypothalamic set-point for temperature. During sleep, it seems that the

set-point is lowered, resulting in sweating and vasodilatation just after sleep state is entered. During REM sleep the set-point mechanism seems to go off-line, so no thermoregulatory responses occur, and body temperature sinks even lower. Upon awakening, the set-point is raised, but because it takes a while for the control mechanisms to respond, there is sometimes shivering and a feeling of chilliness.

Regulation of Feeding Behavior

A model analogous to that just discussed for thermoregulation can be usefully applied to the regulation of feeding behavior. While it is natural to think that body weight is under the control of feeding behavior, the causal relationship is actually the reverse. Body weight (or more exactly, energy stored in the form of fat) appears to be the ***controlled variable*** and feeding behavior can be manipulated by artificially changing body weight. A study published in 1972 points out that an average woman gains 24 pounds between the ages of 25 and 65. This corresponds to an excess daily energy intake of .025% of her total needs. In other words, women have an average excess energy intake equivalent to 350 mg of food per day over 40 years, a period during which she consumes 20 tons. If one considers the tremendous variety of eating patterns, which occur over such a period, the long-term stability of weight is remarkable. Experimental animals deprived of food (or force-fed), will eat (or starve) their way back to 'normal' weight, if allowed ad lib access to food. Thus, body weight appears to have a 'desired' value that is defended by the regulatory system.

Lesions of the hypothalamus permanently alter this regulatory process. A lesion of the ventral medial nucleus of the rat's hypothalamus results in ***hyperphagia*** (overeating) and obesity in laboratory animals. A lesion in the lateral hypothalamus will cause an animal to cease eating and drinking, but if given highly palatable food, the rat will reduce its weight to a lower than normal level and keep it there. These and other experimental findings support the idea that the hypothalamic lesions disturb some kind of set-point mechanism that regulates body weight. When signals reflecting body weight are higher than the set-point, feeding behavior is inhibited. When they drop below the set-point, feeding behavior is initiated to regain the normal body weight.

When weight is gained or lost, the change largely occurs in body fat or adipose tissue. Thus, it has long been assumed that some blood chemical signals the size of the body's fat mass and the brain regulates feeding to keep this signal and, consequently, body fat constant. ***Lipids*** are a major constituent of fat cells so this idea is called the ***lipostatic hypothesis*** of feeding control. In recent years evidence has accumulated that an adipocyte (fat cell) derived hormone called ***leptin*** is a circulating signal linking fat mass to the brain's control of energy balance. This protein is missing in certain genetically obese mice and was isolated in experiments designed to find the responsible gene, now known as the ***ob*** gene. Leptin injection into the cerebral ventricles of normal and genetically obese mice causes a long-lasting decline in feeding and leads to a reduction in body weight. Blood levels of leptin in humans are proportional to body

weight, and leptin receptors have been identified in the choroid plexus and hypothalamus. Recent evidence also suggests that leptin levels are particularly important during development.

While leptin seems like a regulatory panacea, it is but one player in a complex network of hormones and peptides involved in the hypothalamic control of feeding. Others include insulin, neuropeptide Y, melanocyte stimulating hormone, cholecystokinin, and a recently recognized class of peptides called orexins or hypocretins, which are also involved in the regulation of sleep. Recent evidence even suggests that even our bones play a role in weight regulation by secreting hormones that increase insulin production, which in turn can reduce fat stores.

How all of these factors interact in the control of weight and feeding behavior is a matter of intense research, but the control system model of Figure 14.3 schematizes the operations that appear to be involved. In this model, body weight, as signaled through levels of various circulating compounds such as leptin, is compared with a 'set-point' by a mechanism located in the hypothalamus. Like with thermoregulation, the conceptual 'set-point' is in reality a complex relationship among a number of factors that determine whether weight-related hypothalamic cells will issue commands to feed or fast. It is easy to see how disturbances in this set-point mechanism could lead to obesity or extreme thinness. Also, normal variations in its function may be related to the observed variability of normal body weights among humans. The challenge for neuroscience is to understand the role that each of the factors plays and how they interact.

Short Term Regulation of Feeding: Control of Meal Size and Frequency

Acute changes in the blood levels of various nutrient substances can initiate or inhibit feeding, probably by acting through the same hypothalamic mechanisms schematized in Figure 14.3. The substance most thoroughly investigated is glucose. The brain needs glucose for its normal function, so one might expect changes in blood glucose levels to have large effects on feeding behavior. Indeed, intravenous infusion of glucose inhibits feeding, while substances that suddenly lower glucose levels stimulate feeding. Such findings supported the hypothesis that the brain uses feeding behavior to keep its supply of glucose within normal limits, the so-called ***glucostatic hypothesis***.

However, changes in blood glucose have to be rather heroic to evoke or inhibit feeding; the behaviors are sensitive to glucose levels, but not that sensitive. Studies of monkeys eating a variety of diets show that minute-to-minute control of feeding is relatively independent of blood glucose levels. Similar findings have emerged from studies of other factors such as temperature, critical amino acids, and various salts and vitamins.

Over the last decade, though, there has been significant interest in a hormone called ***ghrelin***, whose concentration in the blood seems to peak right at meal time, and then dip in the minutes just following feeding. Although the mechanisms controlling its secretion by the stomach and intestinal tract are not well understood, it is of great interest in the world of

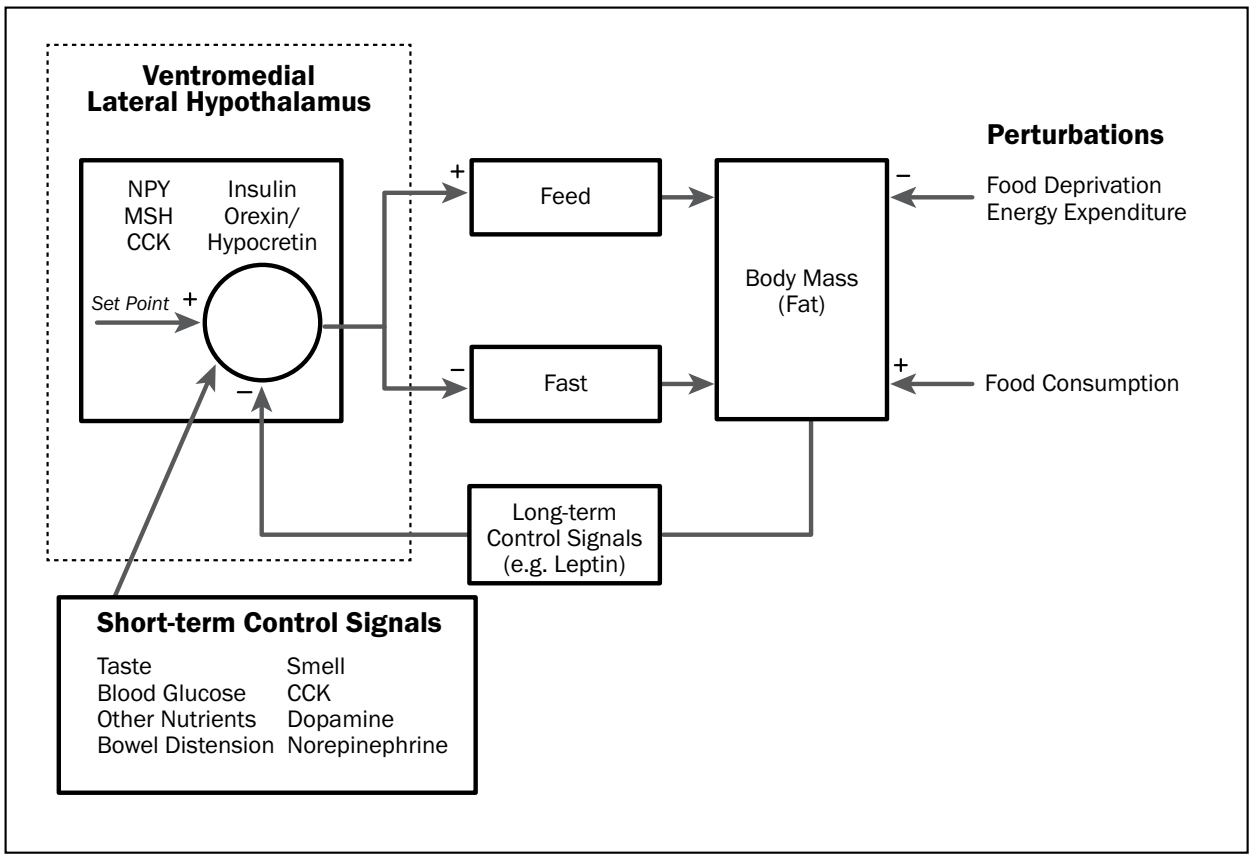


Figure 14.3

Schematic diagram of feedback system for the control of feeding behavior.

obesity control treatments.

Feeding can also be directly inhibited by suddenly expanding a balloon in the esophagus, stomach or bowel. This finding led the development of adjustable gastric bands (lap bands) placed around the top of the stomach. Because the volume of the stomach is artificially constrained, a small amount of food will force the walls of the stomach to expand toward the cuff, and effectively trick the brain into believing the whole stomach is full. This restrictive band is inserted during a laparoscopic surgery and has the advantage that it can be removed in case of complications.

Recent studies lend support to the idea that taste and smell are critical factors in initiating and halting food intake. This ***hedonic*** (pleasure-related) regulation demonstrates what Walter B. Cannon called the 'wisdom of the body.' Comfort is decoupled from the physiological regulatory mechanisms of vital parameters like body temperature and energy stores. By allowing hedonic factors to govern behaviors related to these vital parameters, the parameters themselves never drift into the danger zone that requires emergency action. A feeding system based on blood glucose, for example, would only activate feeding once the glucose levels

were dangerously low. On the other hand, the pleasure of eating prompts us to eat before we are on the brink of starvation. This hedonic regulation also applies to thermoregulation: consider how walking into a cool room prompts us to put on a sweater and turn up the heat even before our body temperature changes. In this way the brain anticipates a cold threat and keeps body temperature within normal limits.

The same strategy applies to regulating not just when we eat but how much we eat. When appropriate food is available, its taste and smell 'reward' feeding behavior and sustain consumption up to a point. After some quantity of the food has been consumed, the brain appears to modify the reinforcing properties of taste and smell, presumably by switching these sensory inputs into circuits that 'punish' the feeding behavior. This mechanism will clearly regulate meal size without there ever being a threat to, say, blood sugar levels or the capacity of the gastrointestinal tract to contain and process the food. It will ensure that eating takes place when tasty food is available and that eating stops before the animal compromises its digestive system or is so full that it can't flee predators.

Terms and Techniques

operational model	negative feedback	positive feedback
comparator	error signal	controller
summing point	hedonic mechanism	hypothalamus
gain	closed loop gain	open loop gain
regulator	controlled variable	set-point
pyrogen	poikilotherm	homeotherm
hyperphagia	leptin	glucostatic hypothesis
lipostatic hypothesis		

Oculomotor Control, Part 1

15

Table of Contents

Eye movements involve varying patterns of contraction of the six extraocular muscles of each eye. Thus, oculomotor control is a somewhat stripped down version of the more complex somatic motor system with its hundreds of muscles acting around a myriad of joints. Because of its relative simplicity, the oculomotor system has illuminated a number of the brain's motor control strategies, such as the use of higher-level neurons to activate entire sub-systems of neurons to create packaged movements. In the oculomotor system we also encounter the use of feedback to 'compute' certain signals that can be used to guide movements.

We begin our study of motor systems with the phenomenology of eye movements and the neuromuscular apparatus subserving them. Then we introduce some of the models that have been advanced to account for the properties of the oculomotor system.

Positions and Motions of the Eyes

In animals with lateral eyes, such as fish, the directions of horizontal movements of one eye are named for the part of the visual field toward which the visual axis is displaced. Thus, the eye is said to move ***nasally*** or ***temporally*** in the horizontal plane. Upward movements ***elevate*** the eye, downward movements ***depress*** it.

Elevation and depression also describe the corresponding movements in frontal-eyed animals, such as humans (Figure 15.1), but horizontal movements are named according to whether the visual axis (i.e. the line of sight) is carried toward or away from the midline. Movement toward the midline is called ***adduction*** (ad meaning "towards" the midline) and movement away is called ***abduction*** (ab meaning "away" from the midline). Torsional (twisting) rotations about the visual axis are named for the direction in which the top of the eye moves. ***Intorsion*** rotates this point toward the nose, ***extorsion*** rotates it away from the nose.

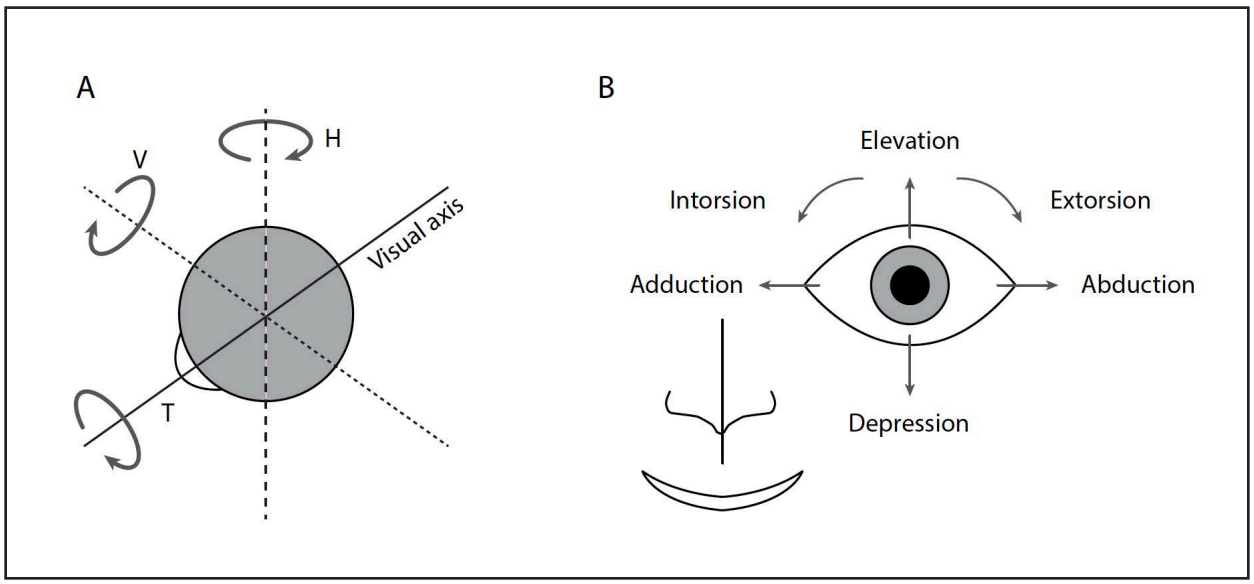


Figure 15.1

A. Rotations of the human eye are expressed in angular units and occur around three main axes (Vertical, Horizontal, Torsional) passing through the center of the globe. The reference axis for ocular rotations is a line from the fovea through the center of the pupil called the ***visual axis*** or ***line of sight***. B. Elevation, depression, abduction and adduction refer to displacements of the visual axis. Rotations about this axis are called torsional movements.

When no rotation has occurred around any of the major axes, the eye is said to be in ***primary position***. Purely horizontal or vertical displacement puts it in a ***secondary position***. Any combination of horizontal and vertical displacement results in an oblique movement that places the eye in a ***tertiary position***.

A separate set of terms is used to describe combined movements of the two eyes (Figure 15.2). These, too, refer to the directions in which the visual axes are displaced. ***Conjugate movements*** displace the visual axes of both eyes in the same direction, as if they were yoked together. Conjugate movements can be vertical, horizontal and oblique. ***Vergence movements*** move the visual axes in opposite directions, either toward (***convergence***) or away (***divergence***) from each other.

Recording Eye Movements

Several methods are used to measure eye movements in humans and laboratory animals, some of great sensitivity and precision. The most commonly used clinical technique, ***electro-oculography***, exploits a standing potential of about 6 mV that exists between the cornea

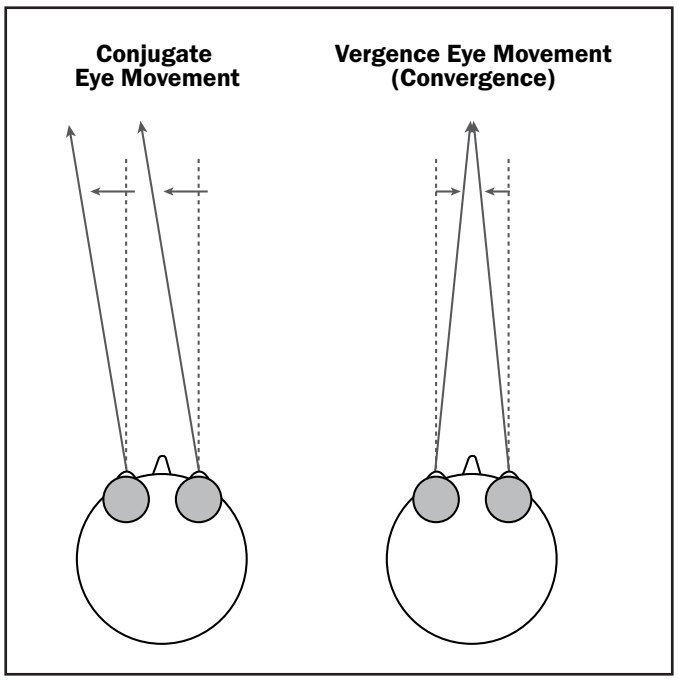


Figure 15.2

Conjugate eye movements (left) displace the visual axes in the same direction. Vergence movements move the axes toward (convergence) (right) or away from (divergence) each other.

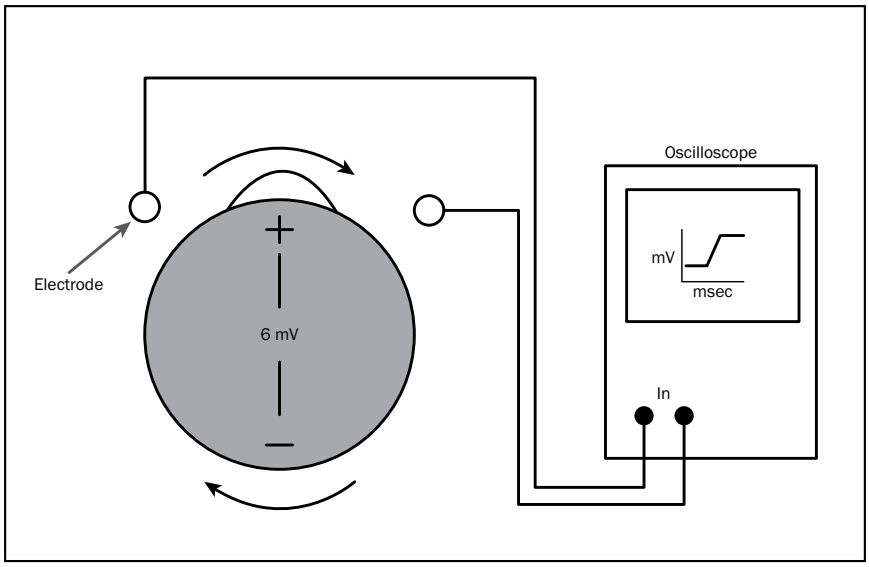


Figure 15.3

Method for recording the electro-oculogram or EOG.

and the back of the eye, the cornea being relatively positive (Figure 15.3). Electrodes placed on either side of the eye will detect horizontal rotation of the eyes because the positive pole swings toward one electrode and the negative pole toward the other. Similarly, vertical movements can be detected with electrodes above and below the eye. With this technique one can measure eye movements as small as 1-2 degrees in amplitude.

Another approach tracks the image of the pupillary border using infrared illumination and a suitable recording apparatus. Often used in human studies, this non-invasive technique can measure movements of less than a degree of arc.

Very small eye movements can be recorded using a contact lens bearing a small mirror (Figure 15.4A). A light beam projected onto the mirror is reflected into a recording apparatus such as a camera or an array of photodetectors. As the eye moves, the angles of incidence and reflection of the light change, so that the ***light lever*** moves through twice the angle of the eye movement. The amplification achieved depends on the distance from the mirror to the recording apparatus, which can be very large. This method can resolve movements on the order of seconds of arc.

A recording method widely used in oculomotor research exploits the fact that an alternating magnetic field induces a current in a conductive wire. In this technique a coil of fine wire, the search coil, is attached directly to the eye or to a contact lens worn by a subject (Figure 15.4B). When the subject's head is placed in an alternating magnetic field generated by the field coils, the current induced in the search coil varies in amplitude with eye position. Vertical and horizontal components of eye movements can be resolved simultaneously by adding a second set of ***field coils*** and appropriate signal processing apparatus. The search-coil method provides a signal of eye position over a wide range and has a maximal sensitivity of about 0.25°.

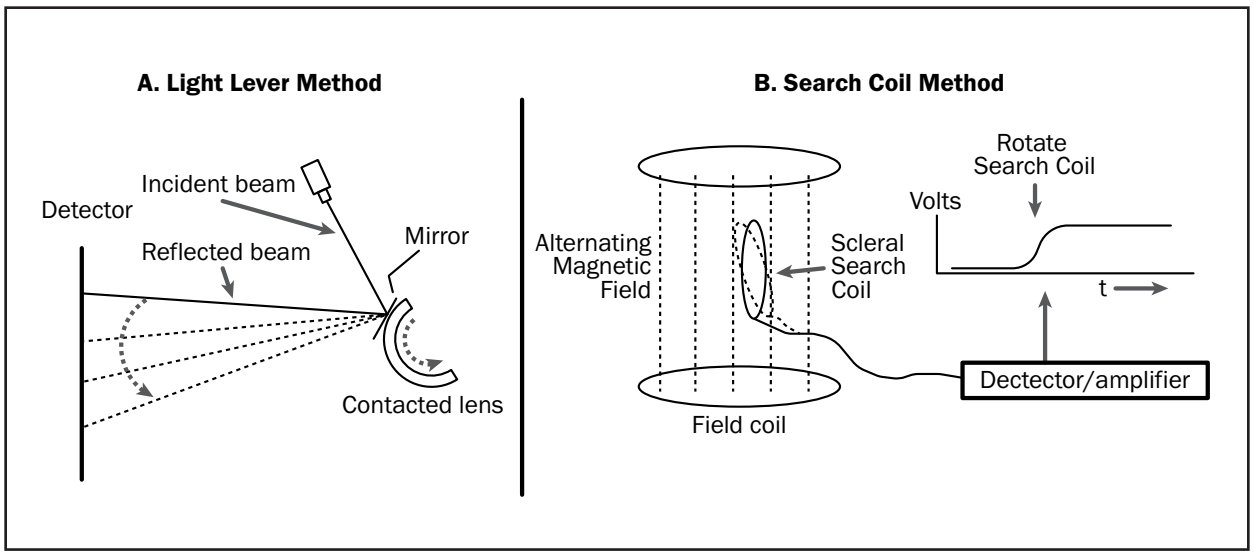


Figure 15.4

Experimental methods for recording movements of the eye. A. Light lever method. B. Scleral search coil method. See text for details.

Muscles That Move the Eyes

In most vertebrates, each eye is moved by the activity of ***extraocular*** muscles, which are attached to the bony orbit at one end and to the eyeball at the other. The geometry of the muscles and their attachments varies considerably among species. The six extraocular muscles of the human left eye are illustrated in Figure 15.5, which also shows the location of the axes about which vertical and torsional movements are made. The vertical axis (for horizontal movements) is normal to the page.

The actions of the various muscles can be deduced by considering their relationship to the various axes of rotation when the eye is in primary position. Thus, contraction of the ***medial rectus*** rotates the front of the eye toward the nose around the vertical axis, the action of adduction. Similarly the ***lateral rectus***, abducts the eye around the vertical axis, rotating the visual axis away from the nose.

Acting around the horizontal axis, the ***superior rectus*** elevates and the ***inferior rectus*** depresses the globe.

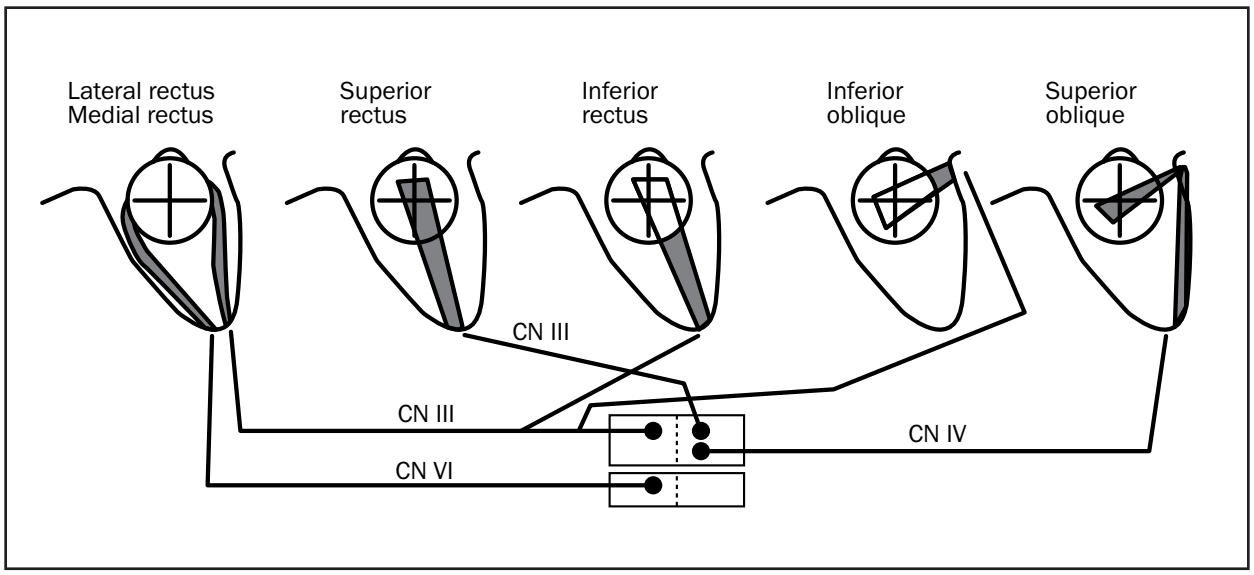


Figure 15.5

Geometry of extraocular muscles of the human left eye viewed from above and the corresponding cranial nerve innervation. The eye is in primary position, i.e. the visual axis is directed straight ahead. The arms of the cross in the center of the eye represent the rotational axes of vertical and torsional movements. The vertical axis about which horizontal movements occur is normal to the page. Filled circles below show the targets of the three cranial nerves controlling the extraocular muscles. Note that CN IV (trochlear) is unique as it is entirely contralateral (and is also unique because it is the only CN to arise from the dorsal aspect of the brainstem.)

The ***superior oblique*** passes through a small ring of bone, the ***trochlea***, on the medial wall of the orbit and has major actions around all three axes. It depresses the line of sight by lifting the back of the eye upward around the horizontal axis. It intorts or pulls the top of the eye toward the nose around the tortional axis and abducts the eye around the vertical axis. The ***inferior oblique*** muscle extends from the medial wall of the orbit and inserts on the underside of the eye. An analysis of its actions around the three axes reveals that it extorts, elevates and abducts the eye.

These muscles are all innervated by cranial nerves. The lateral rectus is innervated by the ipsilateral VIth cranial nerve or ***abducens*** nerve and the superior oblique by the contralateral IVth cranial nerve or trochlear nerve. The superior, inferior and medial recti and the inferior oblique are innervated by CN III, the oculomotor nerve, with the superior rectus innervated by the contralateral nerve, and the rest innervated by the ipsilateral nerve.

Coordination of Extraocular Muscle Contraction

The extraocular muscles must be very precisely controlled to coordinate the movements of the two eyes and to maintain fixation once it is established. The nervous system has therefore automated the muscles in subsystems that are under voluntary control yet do not require conscious thought about how they are working. The difference between the oculomotor and somatic motor control systems may be appreciated from the degree to which they permit independent movements. For instance, we can wiggle our right little finger at will but we cannot voluntarily wiggle only our right eye. Nor can we move one eye up and the other eye down at the same time as chameleons and some fish can do. When the two eyes are not properly aligned and one constantly deviates in or out relative to the other, it is called ***strabismus***. This often results in the person having diplopia or double vision (seeing two of everything).

An example of a low-level component of the coordinating circuitry is provided by the ***medial longitudinal fasciculus*** or MLF, a large myelinated tract of the brainstem. From the diagram of Figure 15.6 we see that a conjugate movement to the left requires that the lateral rectus of the left eye and the medial rectus of the right contract together, meaning that the left abducens (VI) nucleus and the medial rectus component of the right oculomotor (III) nucleus must be activated simultaneously. This synergism is accomplished through a special projection from the abducens nucleus on one side to the oculomotor nucleus on the other (Figure 15.6). ***Internuclear neurons*** with their cell bodies in the nucleus of VI send their axons across the midline into the medial longitudinal fasciculus. These ascend to the nucleus of III and excite the motor neurons innervating the medial rectus. Thus, activation of the left VIth nucleus produces contraction of the left lateral rectus and, indirectly, the right medial rectus, resulting in a conjugate deviation to the left. All conjugate horizontal eye movements employ this circuitry.

Vertical and vergence eye movements require muscles innervated by cranial nerves III and IV whose nuclei are located in the mesencephalon. It is not surprising, then, that the circuits that organize the premotor commands for vergence and vertical movements are also located

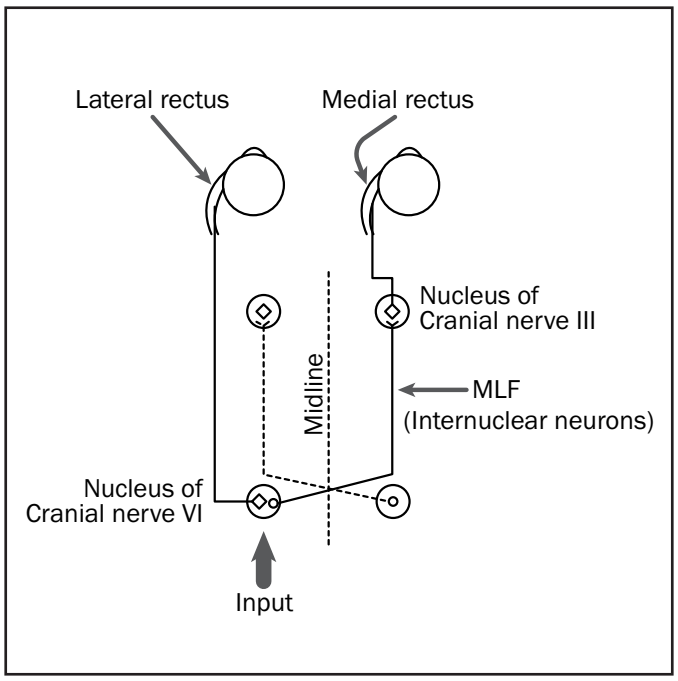


Figure 15.6

Connections of the medial longitudinal fasciculus or MLF involved in horizontal conjugate movement of the eyes. This diagram shows connections for a leftward conjugate eye movement.

in the mesencephalon and that lesions there interfere with vertical eye movements.

Terms and Techniques

fixation	abduction	adduction
depression	intorsion	extorsion
visual axis	line of sight	primary position
azimuth	elevation	eccentricity
secondary position	tertiary position	vergence
convergence	divergence	conjugate
strabismus	superior rectus	inferior rectus
EOG	superior oblique	inferior oblique
MLF	trochlear nerve	abducens nerve
conjugate	vergence	search coil
dipolopia	oculomotor range	light lever
internuclear neurons	oculomotor nerve	IIIrd cranial nerve
IVth cranial nerve	VIth cranial nerve	medial rectus
lateral rectus		

Oculomotor Control, Part II

16

Table of Contents

Quick and Slow Eye Movements

The oculomotor systems of vertebrate brains commonly generate two important subclasses of movements: ***quick*** and ***slow eye*** movements (Figure 16.1). The principal function of quick eye movements is to ***displace*** an image from one part of the retina to another. The principal function of slow eye movements is to prevent an image from moving on the retina, i.e. to ***stabilize*** the image, to stick it to the retina. These basic movements are incorporated into a variety of oculomotor programs.

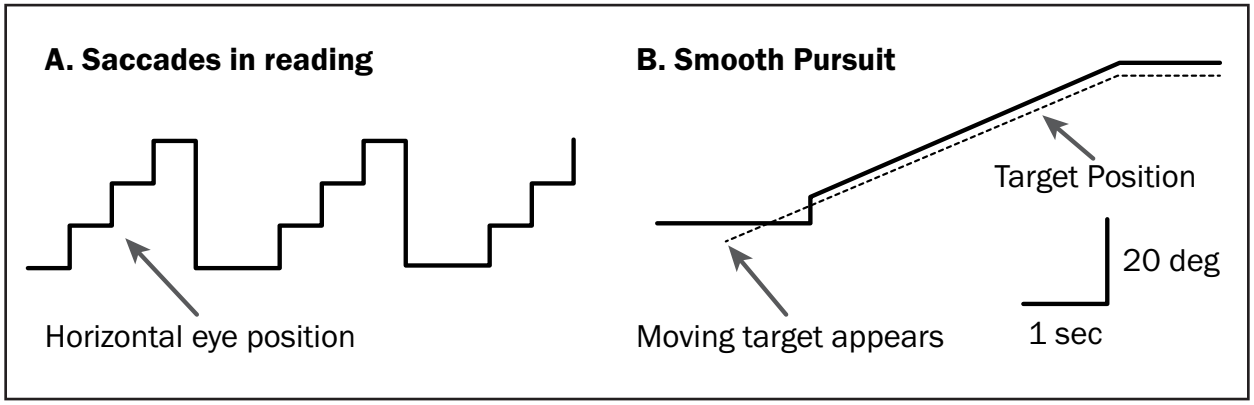


Figure 16.1

Quick and slow eye movements. A. Schematic representation of saccadic eye movements of one eye during reading. B. Position of the eye as it pursues a target. Note the latency to pursuit onset and the initial saccade to fixate the stimulus.

An example of a quick movement is the ***saccade*** which is executed voluntarily during visual exploration and reading (Figure 16.1A). A novel stimulus in the peripheral visual field may elicit a reflex saccade that brings the image of the stimulus to the fovea, known as the 'visual grasp reflex.' Quick eye movements are always conjugate.

An example of a slow movement is ***smooth pursuit*** in which the eyes track a moving target in a smooth, conjugate deviation (Figure 16.1B). A person who is fixating an airplane passing high overhead is executing a smooth pursuit movement. Pursuit movements cannot ordinarily be performed voluntarily in the absence of a moving target.

Conjugate movements may be slow (pursuit) or quick (saccades). Vergence movements are usually slow but they may be assisted by quick conjugate movements of unequal size. The basic patterns occur in both voluntary and reflex eye movements.

There is considerable evidence that quick and slow eye movements are controlled by different brain systems. When a subject gets sleepy, quick movements disappear before slow movements. Administering a barbiturate (depressant) eliminates slow movements before suppressing saccades. Electrical stimulation of the superior colliculus evokes saccades but not slow movements, whereas stimulation of the cerebellum can elicit slow movements alone. Recordings from the nuclei of cranial nerves III, IV and VI indicate that the same oculomotor neurons are employed for both quick and slow eye movements. This means that the quick and slow systems are separate in the higher, premotor parts of the system and converge at the final common pathway, the oculomotor neurons themselves.

The Vestibulo-Ocular Reflex or VOR

The vestibulo-ocular reflex or VOR is a powerful mechanism that automatically produces a slow eye movement to counteract motion of the retinal image due to head movement (Figure 16.2). As discussed earlier in the course (Chapter 10), the vestibular apparatus has two components, the otolith organs and the semicircular canals. The otolith organs are sensitive to linear accelerations such as gravity and that experienced when taking off in an airplane. Tilt of the head results in a signal from the otolith organs to the central nervous system that produces a partially compensatory torsion of the eyes. While the otolith organs are connected with the eyes, the most important signal in terms of oculomotor effects comes from the semicircular canals, which sense angular rotation of the head. During normal movements of the head this signal represents the velocity with which the head is turning and produces a compensatory movement of the eyes in the opposite direction.

The *gain* of the VOR is the *ratio of angular eye velocity to angular head velocity*. Under most conditions, this gain is close to -1, meaning that the reflex compensates nearly perfectly for the head rotation. VOR gain may be modified experimentally by having a subject wear special goggles that, for example, double the apparent displacement of an object for a given rotation of the head. For instance, if the head moves 1 degree, but the eye's view of the

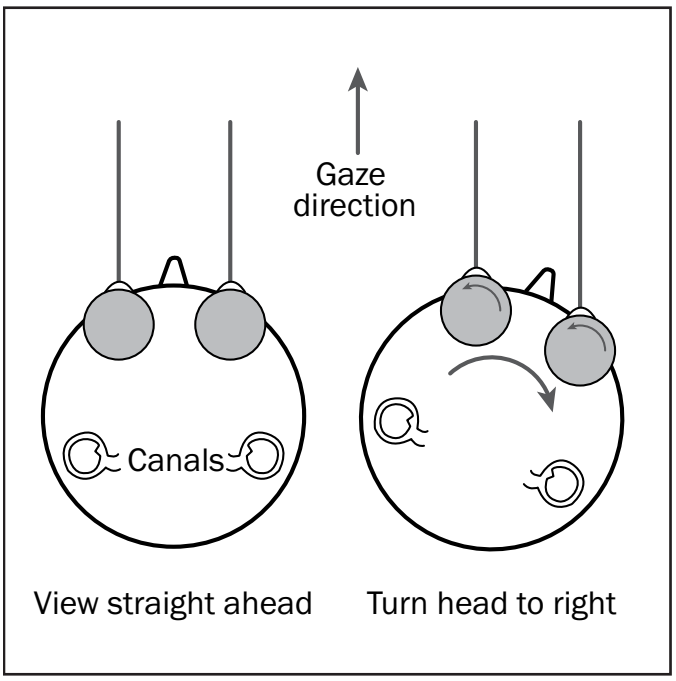


Figure 16.2

The vestibulo-ocular reflex. Turning the head to the right produces a conjugate, corrective rotation of both eyes to the left, stabilizing the direction of gaze in space.

outside world is caused to shift -1/2 degree, the eyes will not rotate enough to compensate if the VOR gain is only -1. Wearing such lenses for a time results in a gradual change in VOR gain to compensate for the mismatch between eye movement and image movement.

Such plasticity in the VOR proves useful because stabilization of the retinal image must be kept nearly perfect over the course of an animal's life, during which the size of the eyes and distance between them changes and the nervous system suffers various insults and normal cell death. Thus, the VOR must be constantly monitored and recalibrated to preserve its capacity to stabilize the retinal image. This requires that the connections between the vestibular system and the oculomotor control system must be capable of modification.

The cerebellum also plays an important role in modifying the gain of the VOR. Destruction of the cerebellar flocculus eliminates the adaptive capacity, or plasticity of the VOR. The cerebellum receives all the information required to adjust the gain properly, since it is the target of both vestibular and visual projections. The cerebellum has been characterized as a 'repair shop' for the VOR, maintaining just the right conditions for optimal image stabilization during head rotation. A major focus of current research is to identify the locations within the cerebellum and/or brainstem where the plastic changes occur.

Optokinetic Movements

Displacements of the entire visual scene evoke slow, ***optokinetic*** eye movements that tend

to oppose the resulting motion of the retinal image (Figure 16.3A). The moving visual scene essentially 'drags' the eyes along with it. For instance, when we move our heads with our eyes open, the entire retinal image is displaced. This type of motion will lead to optokinetic eye movements that will stabilize the image (Figure 16.3B). Importantly, this response is dependent on visual input.

The motion of the head also evokes the VOR, which keeps the eyes aligned on the external scene. Be careful not to confuse the OKR and VOR. The OKR is ***opto***kinetic and depends on vision, while the ***vestibulo***ocular reflex depends on head position and will occur whether the eyes are open or not. The optokinetic movements work ***synergistically*** with the VOR during natural movements of the head to stabilize the retinal image of the background, allowing the visual system to detect potentially significant movement of prey or predator moving against the background. This stabilizing action on the ***entire*** background image is to be contrasted with the action of smooth pursuit, which is designed to stabilize only part of the retinal image, normally the image on the fovea. This important function is considered later.

Vestibular and Optokinetic Nystagmus

If the head is turned through a large angle, the eye rotates in the opposite direction but cannot continue to rotate because of mechanical limits on its motion. If head rotation continues, the eyes cannot keep rotating without rolling into the back of the head. There is therefore a

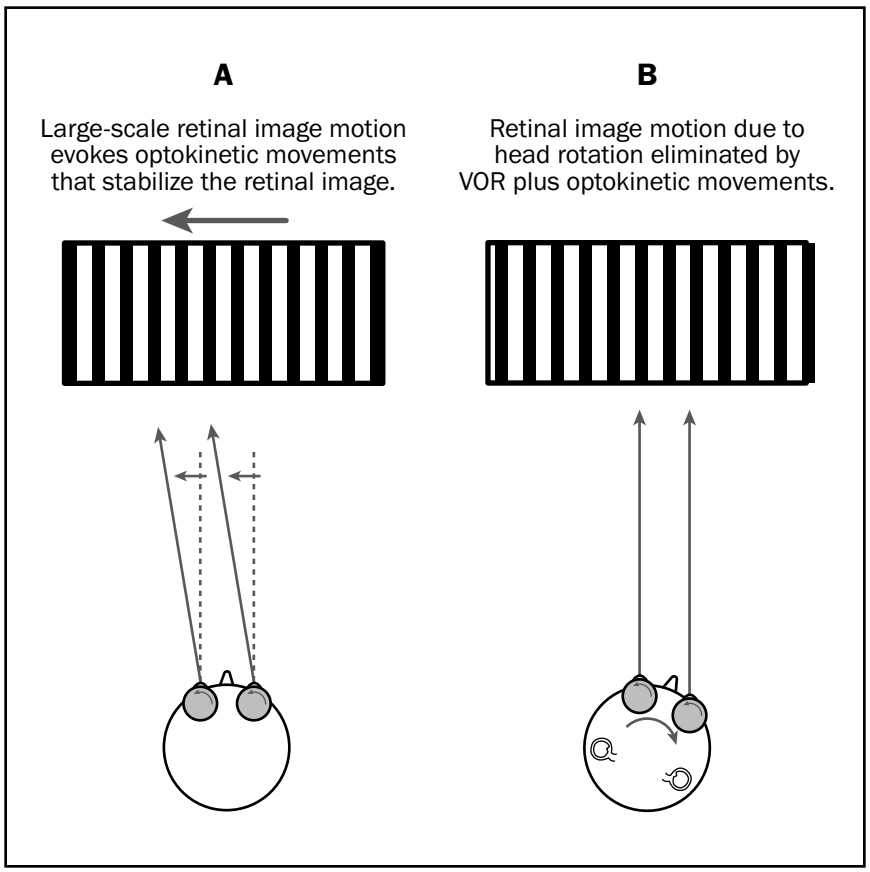


Figure 16.3

A. Optokinetic movements are elicited by motions of the entire visual scene that cause image motion across large areas of the retina. B. Rotation of the head will cause the retinal image of a stationary background to move, evoking optokinetic movements of the eye. These will be combined with movements in the same direction due to activation of the VOR.

quick eye movement during the VOR, in which the eye jerks back toward the primary position. This jerking movement of the eyes, composed of a slow phase and a quick phase, is called *nystagmus* (Figure 16.4). When the movements are triggered by the vestibular system, such as in the case of sitting in a spinning chair, it is called *vestibular nystagmus*.

Optokinetic eye movements made in response to motion of the retinal image can also take the eye nearly to the limits of its movement range, with the result that a quick phase is added in the opposite direction. This is called *optokinetic nystagmus*. A good example of this phenomenon is 'railroad nystagmus.' When a person gazes from the window of a moving train, the eyes are irresistibly driven in the same direction as the rapidly passing telegraph poles or trees at the side of the track and a series of slow and quick movements follow in succession. It is important to remember that the slow phase stabilizes the retinal image, while the quick phase simply returns the eye to a position from which a new slow movement can begin. The mechanisms responsible for optokinetic movements involve widely dispersed regions of the brain including certain nuclei of the pretectum and even the visual cortex.

Physiological Nystagmus

When very sensitive measuring devices are used, such as the light-lever method described in Chapter 15, it is found that the human eye is never still. Even during intense voluntary fixation the eye continues to exhibit small drifts, a tremor and periodic microsaccades. The microsaccades tend to direct the fixated image back toward the center of the fovea.

One can learn to suppress the microsaccades, but not the other two types of movement.

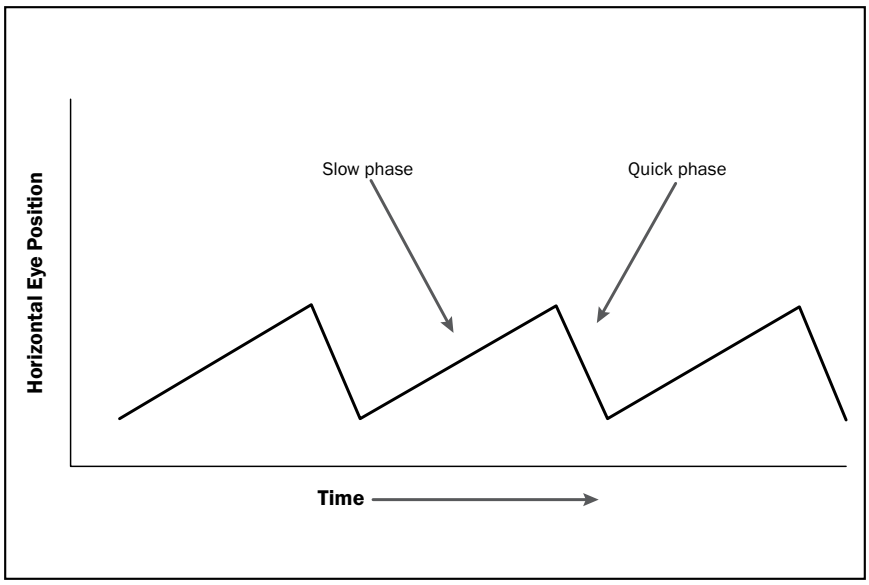


Figure 16.4

Nystagmus of the vestibular or optokinetic type. These movements comprise a slow and quick phase. The slow phase is the compensatory component designed to stabilize the retinal image. The quick component simply resets the eye's position in the orbit.

This residual jitter of the eye is called ***physiological nystagmus*** and it is vital to visual function. When the position of the retinal image is decoupled from eye movement, for instance by mounting a tiny projector on a contact lens, the stabilized image fades rapidly from perception. This is why we don't see the shadows cast by our retinal blood vessels unless we use some technique to cause these shadows to move on the retina.

Voluntary Smooth Pursuit or Tracking Movements

The initiation of smooth pursuit normally requires the presence of a moving visual stimulus. The most important aspect of the stimulus for smooth pursuit is ***retinal slip***, i.e. an actual movement of the image on the retina expressed as a velocity (slip velocity or retinal image velocity) with units of millimeters/second or degrees/second. Both the velocity and acceleration of retinal slip are detected by the control system and affect the eye movement. A second important parameter is ***position error***, the static displacement of the image from the fovea.

Note that smooth pursuit allows the eye to stabilize the image of a moving target (e.g. an airplane) while the image of the background (e.g. the clouds) also moves on the retina. This phenomenon presents a unique problem: the eye tracks the target in one direction, but because the background also changes, there should be background-stabilizing eye movements in the opposite direction. The control system must therefore disable mechanisms that would normally oppose movement of the background image.

For instance, when the entire background moves on the retina, the optokinetic reflex drives the eyes to move the opposite direction. The retinal control system will therefore turn off this reflex. Similarly, the VOR might also be inhibited because head movement towards the pursued object would drive the eye away from the target. For the same reason, the optokinetic reflex and the VOR must also be 'turned off' during quick voluntary eye and head saccades used to redirect gaze.

Figure 16.5A shows a model of the pursuit system that would use retinal slip to produce an eye movement to track a moving target. In this model, the difference between target and eye velocity produces a signal of retinal slip or retinal image velocity in the afferent pathways from the eye. Central processing mechanisms convert this error signal to an eye velocity command in the efferent pathways. This command drives the motor pathways which move the eye with a particular velocity. (Note in Figure 16.5A that the signals marked Eye Velocity and Target Velocity in the dashed box are not representations in the brain, but rather represent real motion of the eye and target). Imagine now what would happen if this system generated an eye velocity that exactly matched target velocity. The retinal image velocity signal and the eye velocity command would drop to zero and the eye would stop. This does not happen, so this particular model cannot be correct.

Also unexplained by the model of Figure 16.5A is the fact that the eye continues to move even when the target disappears briefly (see Figure 16.6). Moreover, eye velocity can momentarily

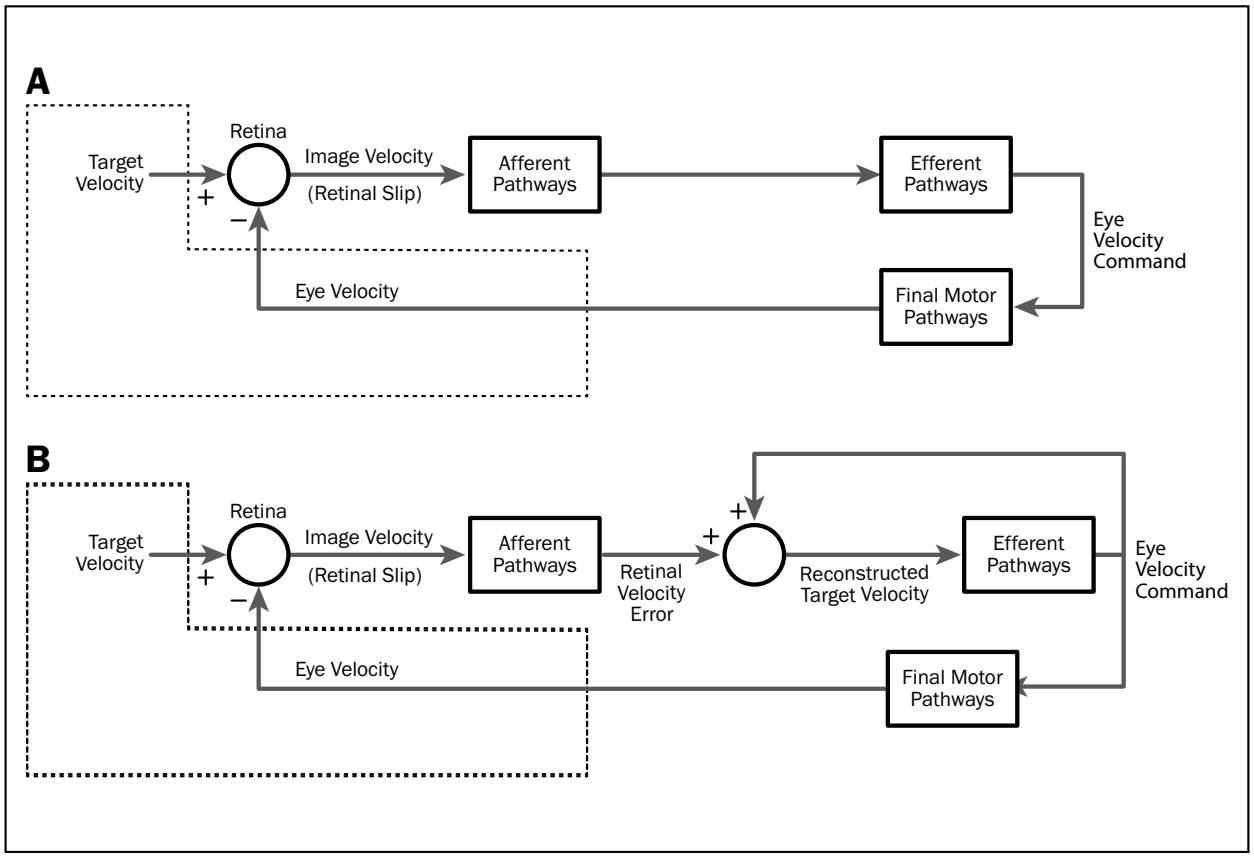


Figure 16.5

Models of the pursuit system. A. Retinal slip drives the efferent pathways directly. Note that in this model would not work because the eye would stop moving when eye velocity and target velocity are the same. B. A feedback connection is used to compute a neural replica of target velocity and this drives the efferent pathways. Dashed outlines identify processes occurring outside the CNS.

exceed stimulus velocity, producing a slip in the opposite direction, without the movement stopping or reversing direction. A perfect match between eye and stimulus velocity would stabilize the image on the fovea, eliminating retinal slip, but the eye moves nonetheless when the image is artificially stabilized for an instant. These observations indicate that the pursuit system uses a predictive strategy, generating an internal 'neural replica' of target velocity and matching eye velocity to this over a certain period of time.

The model of Figure 16.5B shows how the brain could create such an internal replica of target velocity by means of a feedback loop. Note that:

Target Velocity = Eye Velocity + Retinal Image Velocity (Retinal Slip)

Thus, if there were a way to sum retinal slip with eye velocity, the brain could reconstruct

the velocity of the target. The retinal slip signal is available from the visual system, and the brain produces a motor signal to move the eye at a particular velocity. So, by feeding back a neural replica of the pursuit velocity command and summing it with retinal velocity error, the brain could reconstruct a signal proportional to target velocity.

The neural replica theory of target velocity addresses some of the problems with the model in 16.5A. For instance, if the target disappears momentarily and there is no retinal slip signal, a neural replica of the target velocity could keep driving the efferent pathways so that the eyes keep moving. Also, if eye movement and target movement were perfectly matched and retinal velocity error dropped to zero, the eye velocity commands still feed positively into the efferent pathways and keep the eyes moving. Said another way, when actual eye velocity becomes equal to target velocity, retinal slip is zero, but the internal replica of eye velocity continues to drive the efferent pathways through the loop at the right of Figure 16.5B.

According to this model, the system operates on the assumption that the current and future velocity of the target will not be different from its immediate past velocity. The model does not account for why the eye stops moving when the target disappears permanently, but more complete models take care of this by limiting the amount of time the feedback signal continues in the absence of a real retinal slip signal.

The model of Figure 16.5 implies that there should be cells in the brain that carry an internal signal of target velocity, even when the target disappears momentarily. Figure 16.6 compares the responses of two cells, one in cortical area MT and the other in area MST of the temporal lobe. Both are activated when a stimulus begins to move in their receptive fields and smooth

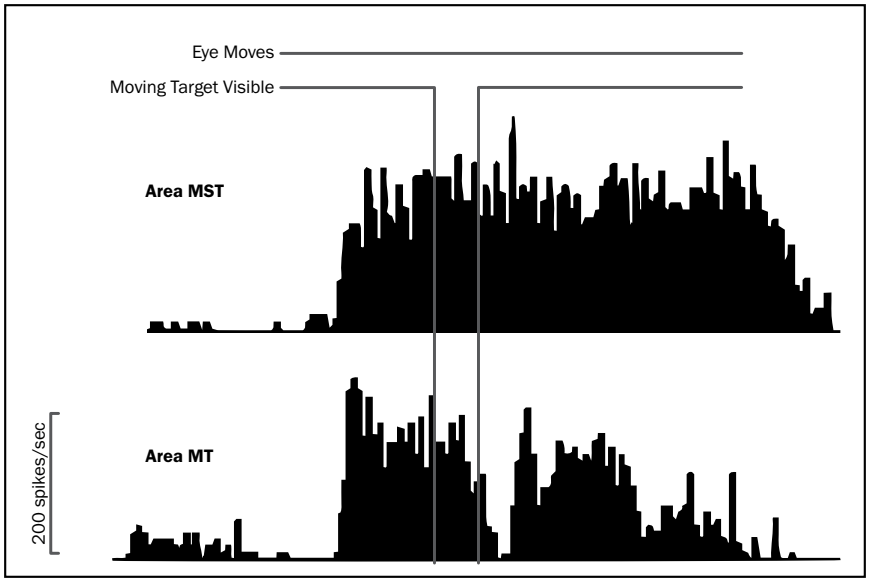


Figure 16.6

Responses of cells in areas MT and MST as the eye pursues a moving stimulus. When the stimulus is blinked off (vertical lines), the cell in MT ceases to fire while the discharge of the cell in MST is maintained as the eye continues to move. The response of the cell in MST is not dependent on the continued presence of the visual stimulus. (Data from Newsome et al., *J. Neurophysiol.* 1988)

pursuit is initiated. When the stimulus is blinked off, the cell in MT ceases to discharge, as though it were dependent on the visual input, while the cell in area MST carries on as though the moving stimulus were still present. The discharge of this cell might therefore embody the internal representation (or 'memory') of stimulus velocity postulated in the model of Figure 16.5.

While MST seems to encode the neural memory of target velocity, studies have implicated many other brain areas in smooth pursuit. Lesions of area MT in the temporal lobe interfere with pursuit in the area of visual field affected by the lesion. Lesions in frontal cortex modify but do not eliminate smooth pursuit. The posterior parietal cortex contains cells that respond during smooth pursuit. Unilateral lesions in the striate cortex of monkeys irreversibly eliminate pursuit of objects in the 'blind' hemifield, implying that subcortical systems alone cannot sustain this function. The pontine nuclei play a major role in smooth pursuit, relaying signals from cortical areas to the cerebellum. Following cerebellectomy, there is total loss of smooth pursuit. The cerebellar region most clearly involved is the flocculus, where electrical stimulation elicits smooth conjugate movements and where neurons carry a signal of gaze velocity and eye velocity. Cells in several nearby brainstem nuclei, including the vestibular nuclei, discharge during pursuit and carry a signal proportional to eye velocity. These nuclei probably make up the final premotor substrate for pursuit.

Generation of Saccadic Eye Movements

As noted above, the function of quick eye movements is to displace an image from one place on the retina to another. Saccades are one form of quick eye movement that generally move an image from the periphery of the retina to the fovea. Such movements are always used when we read or to scan the environment. The separation of an image from the fovea is called the ***retinal error*** of the stimulus (e_r in Figure 16.7), a quantity analogous to that of position error discussed above.

The task of the oculomotor system is to create a saccadic eye movement equal to the retinal error in amplitude and direction. The oculomotor system could presumably use two strategies to do this. It could create a signal equal to retinal error and use this to drive the eye toward the target. Or, it could combine the retinal error signal and an internal signal of current eye position to create a signal of the target's position with respect to the head, then issue a command to move the eye to the correct position in the orbit (Figure 16.7). Under most circumstances, the result would be the same, but the underlying strategies require very different computations by the oculomotor system.

The distinction between these two strategies is made clear in an experiment by Sparks and colleagues, which is illustrated schematically in Figure 16.8. A monkey is asked to make saccades to a target (asterisk) that appears briefly right above the fixation point (F) and then disappears. Before the eye can move upward to where the target appeared, the eye is displaced horizontally to the left by electrical stimulation of the superior colliculus. The monkey then makes a second saccade to the target.

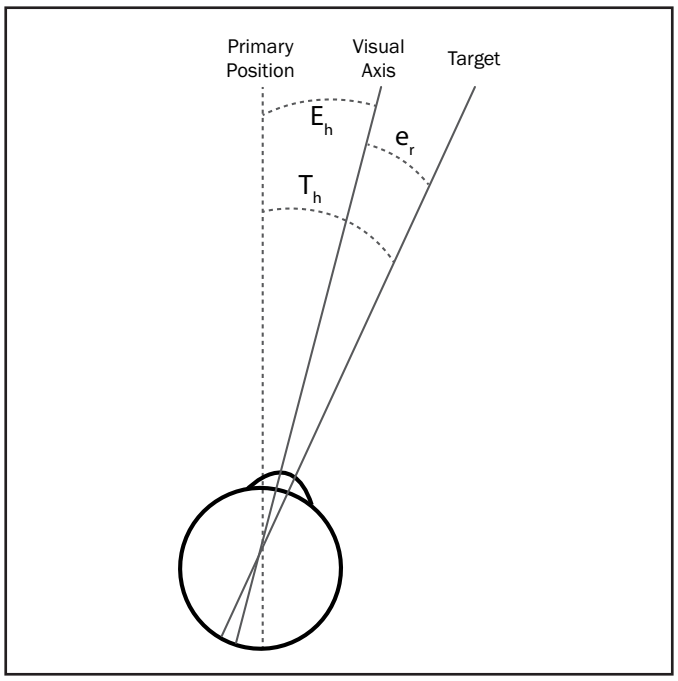


Figure 16.7

Critical angles defining target location. The eye is deviated by angle E_h (eye in the head) from the primary position. The target's image is displaced by angle e_r (retinal error) from the fovea. The sum of e_r and E_h is the target's position with respect to the head, T_h.

The retinal error signal (e_r) is purely vertical. If this error were translated directly into a movement vector, the result should be a saccade that is also purely vertical as represented by e_r. However, if the control system computes the actual position of the target with respect to the head (T_h) and positions the eye accordingly, the saccade should be oblique. The saccades in fact are oblique, indicating that the system has taken into account the electrically-evoked change in eye position and generated a signal corresponding to target position, not retinal error. This implies a computational complexity in the circuits responsible for saccade generation that has a parallel in the predictive character of smooth pursuit described above.

Most current theories of saccade control are based on a seminal model of David A. Robinson. Though extensively modified in recent years, his theory incorporates the ideas just described and will serve to illustrate the model (Figure 16.9). In this model the oculomotor system uses feedback to generate an internal representation of target direction with respect to the head. The system drives the eye until this signal has the same value as an internal representation of eye position. Interestingly, the entire sequence of events is divorced from what is happening in the real world and is done solely computationally.

One might wonder why internal representations of target direction and eye position must be used instead of visual signals and proprioceptive feedback from the eye muscles. The answer is that the saccade takes place so quickly that it is over before visual and proprioceptive signals could reach the brain and be used in programming the movement. For example, a 10 degree saccade may have a duration of 25 milliseconds whereas the latency of a visual

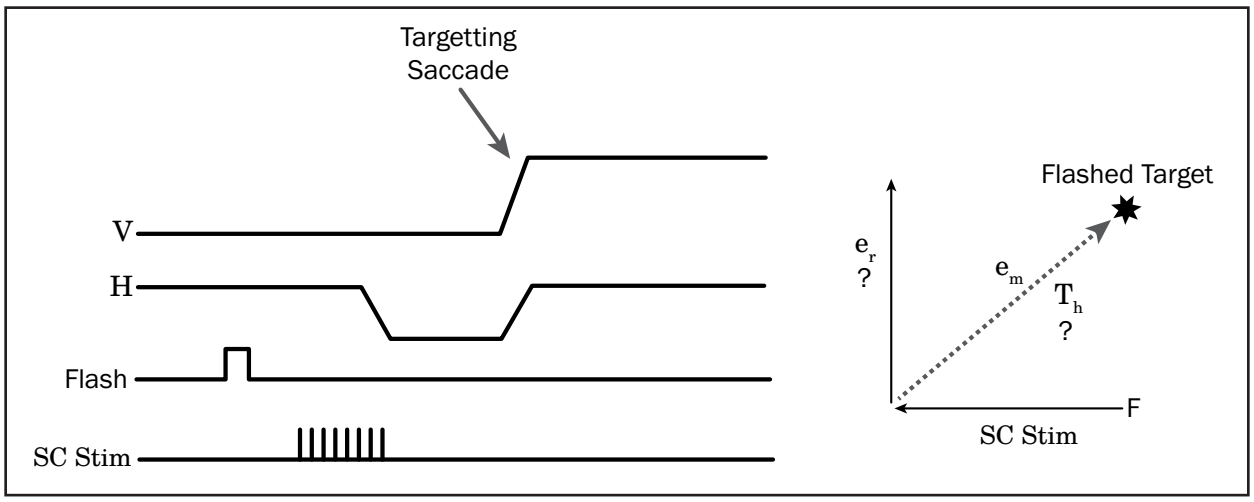


Figure 16.8

Schematic representation of an experiment showing that the saccadic control system computes target position with respect to the head. Left: Timing of the events. Right: Paths of the direction of gaze. V, H, vertical and horizontal eye position. Asterisk, flashed saccade target. F, fixation point. SC stim., electrical stimulation of the superior colliculus. e_r, retinal error. e_m, motor error. T_h target with respect to the head. See text for details.

response in the optic nerve can be longer than this and depends on a number of factors such as the intensity of the stimulus. Similarly, a saccadic eye movement is complete long before proprioceptive information could reach the brain to stop it.

The Robinson model assumes that there exists within the circuits of the brainstem a signal E_h proportional to current eye position. This assumption is reasonable because the oculomotor neurons require such a signal in order to maintain the eye in its current position. Signals with this property are found in nuclei of the oculomotor system, including the vestibular nuclei, which have direct access to the oculomotor neurons.

In the model, the internal replica of current eye position (E_h) is added to the signal of retinal error (e_r), derived from the visual input, to reconstruct the position or direction of the target with respect to the head ($T_h = E_h + e_r$; see Figure 16.7). Then, after a delay that buffers or stores this sum for a brief period, the updated E_h is subtracted from the stored T_h to produce a signal called ***motor error*** (e_m), which is the immediate input to the saccade generating machinery of the brainstem. When a saccade is triggered, the eye movement causes E_h to change in the direction of T_h. Movement continues until these two signals are equal and e_m goes to zero, at which point the eye stops. The delay/buffer mechanism is required to prevent the signal T_h from changing at the right-hand summing point during the brief instant that the saccade is occurring and E_h is changing.

It is important to differentiate between e_r and e_m. In short, the former signal will not change

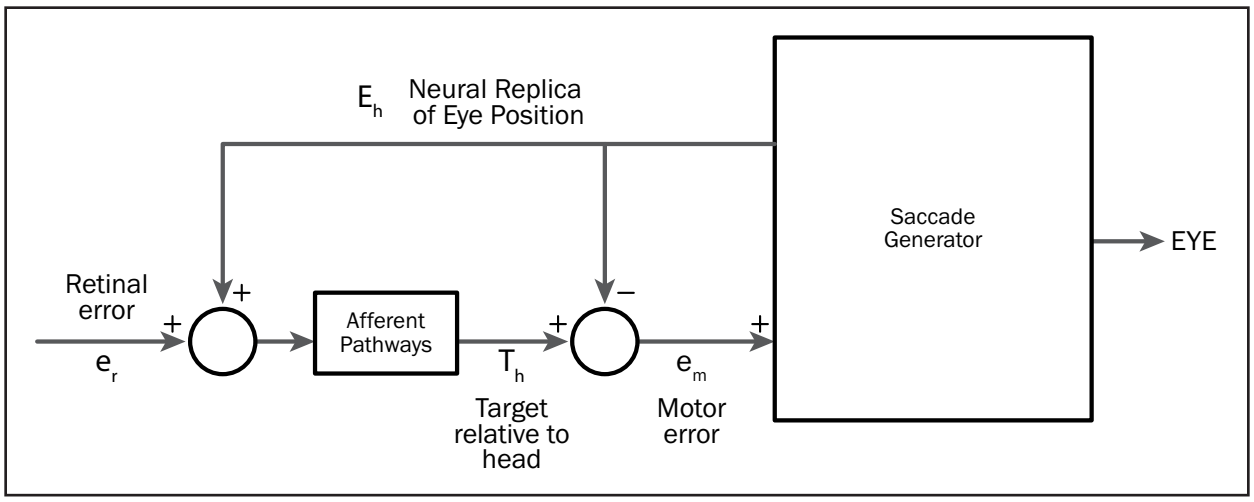


Figure 16.9

Robinson's model of local feedback control in the saccadic system. See text for details.

when the eye moves, while the latter value depends on E_h and will change during a saccade. The retinal error, e_r, is a static quantity that is ***sampled*** before the eye movement begins. This error signal is summed with current head position, E_h, and stored as T_h by the delay/buffer. On the other hand, motor error is a signal used to decide whether and how a saccade is carried out. Motor error necessarily depends on eye movement, as this signal compares the current position of the eyes to the desired destination of the eyes.

We can see this phenomenon mathematically: because $e_m = T_h - E_h$, and T_h is a constant, when eye position changes, so must e_m. The control system aims to eliminate motor error by driving E_h to match T_h – that is, to make the eyes reach the target. In terms we learned in Chapter 11, we can see that e_r is neurally coded differently than e_m. Retinal error is encoded in coordinates based on the retina, while e_m is encoded in coordinates based on the head. At the beginning of the saccade they will have the same magnitudes in terms of degrees, but during the saccade these magnitudes diverge.

To produce a rapid displacement of the eye to a new position, the oculomotor system must generate a high frequency burst of action potentials in the oculomotor neurons, followed by a constant level of activity appropriate to hold the eye at the new position (Figure 16.11). Figure 16.10 expands the box in Figure 16.9 labeled 'Saccade Generator' and shows how this pulse-step pattern of activation is created by the brain.

Neural elements in the brainstem feed to oculomotor neurons and are responsible for starting/stopping saccades and generating the pulse-step pattern of neural activity. The main driving force is a class of Burst Neurons (BN) that produce an intense burst of action

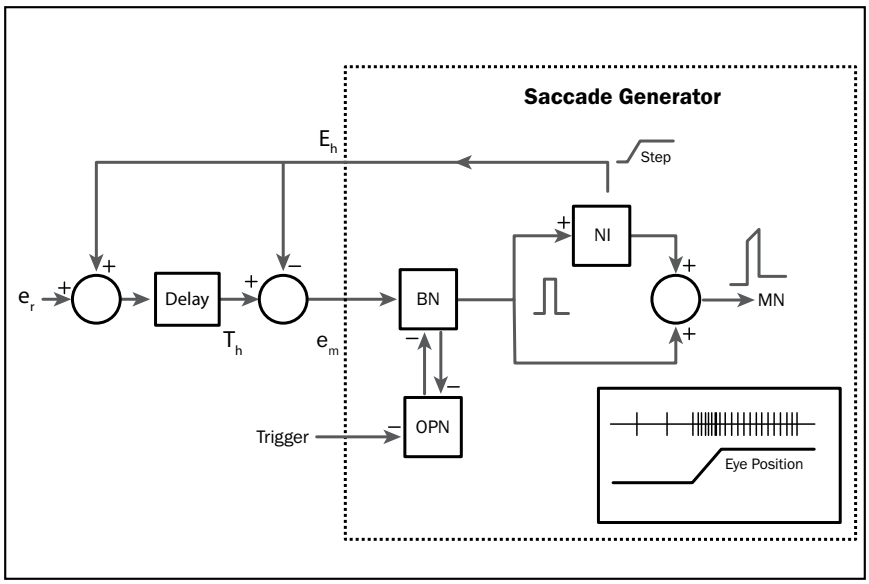


Figure 16.10

Expanded version of Robinson's model of the saccadic control system. The dashed box contains the neurons forming the saccade generator, whose activity results in the pulse-step pattern of activation of the oculomotor neurons during a saccade. This pattern is schematized in the inset enclosed by dotted lines. See text for further details.

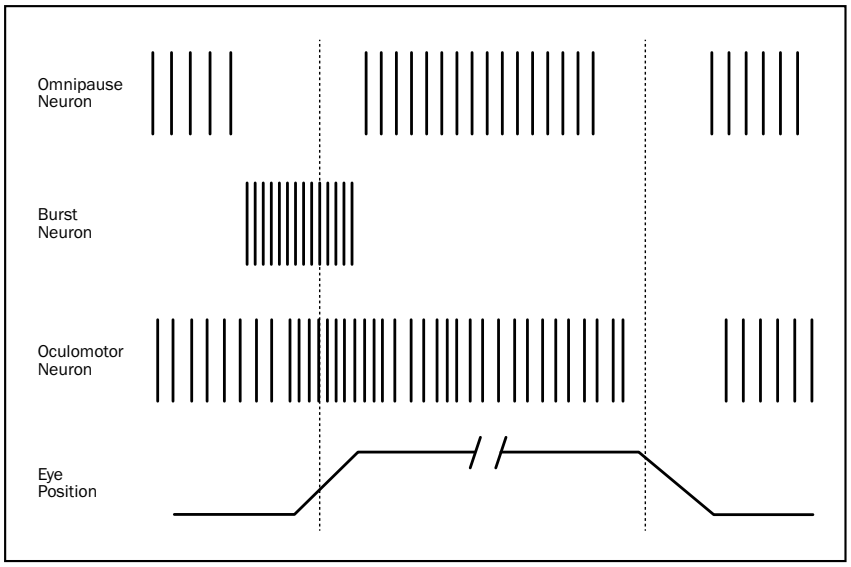


Figure 16.11

Discharge patterns of omnipause (OPN), burst (BN) and oculomotor neurons during two saccades in opposite directions. Note that the BN and oculomotor neuron have specific 'on' directions whereas the OPNs pause for all saccades.

potentials during the saccade (Figure 16.11). BNs associated with horizontal components of the saccade are found in the paramedian pontine reticular formation (PPRF) while those associated with vertical saccades are located in the rostral midbrain reticular formation (Figure 16.12). BNs are known to project directly to oculomotor neurons (MN).

The BNs are tonically inhibited by a class of neurons found in the caudal PPRF that are active continuously but pause during saccades (Figure 16.11). These Omnipause Neurons (OPN) or 'pausers' are switched off by the burst neurons themselves so that the burst neurons can respond to the motor error signal. While firing, burst neurons simultaneously inhibit the omnipause neurons, 'latching them off' for the duration of the saccade. The pulse of activity in the BNs is transmitted directly to the oculomotor neurons.

BN's also feed to a neural integrator circuit (NI). We call this circuit a "neural integrator" because it actually performs a calculus integration of signals. Notice how the burst neurons produce a single pulse, while the NI neurons produce a tonically increasing ramp. If you remember calculus, it is clear that the integration of the horizontal function will lead to a linear function that levels off when the burst ends.

It is important to integrate activity of BN's for two reasons. First, the integral of the pulse is added to the input to the motor neurons to complete the signal and hold the eye in its new position. The output of NI also provides an internal representation of current eye position, E_h, which feeds back to T_h (as in Figure 16.9) to yield motor error. When E_h equals T_h, motor error goes to zero and the BNs switch off, the OPNs are no longer inhibited and the saccade stops.

Cells forming the neural integrator (NI) circuit reside in the vestibular and neighboring nuclei. This location makes sense, given that the NI concerns proprioceptive sense of the eye in the head. We also know that the cerebellum is critical to the function of this integrator because, if it is damaged, the eye cannot be maintained in its new position after a saccade. It is not known where or how the computation of T_h and e_m occurs.

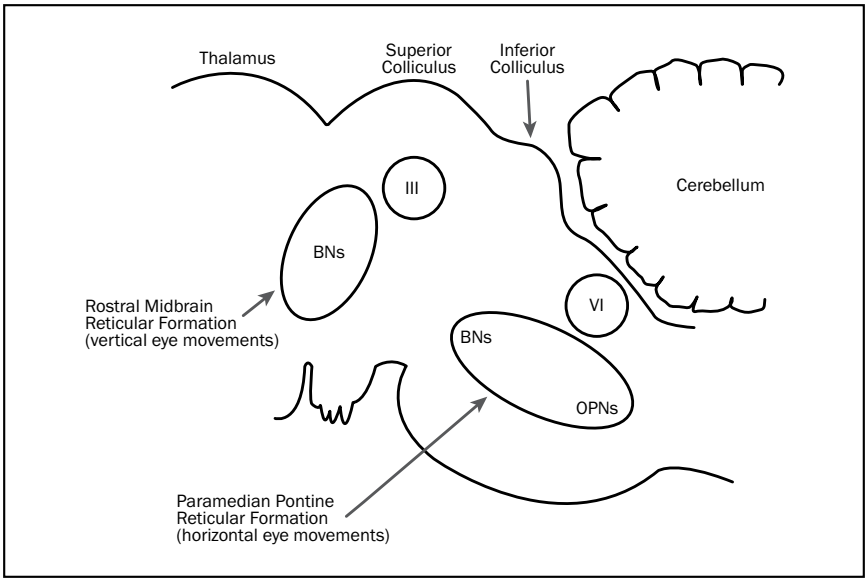


Figure 16.12

A midsagittal section of the brainstem showing the locations of the BNs and OPNs associated with saccade generation. Location of motor neurons of cranial nerves III and VI are also indicated.

While the Robinson model is useful in understanding saccades, it has been modified and extended in recent years to accommodate many complexities associated with the control of gaze. For instance, gaze-direction may be changed with both head and eye movements, so there must be important interactions between the oculomotor mechanisms and the vestibular system. Also, saccades of different size may be controlled in different ways and gaze shifts of comparable amplitude may be accomplished by varying combinations of head and eye movements. Analysis of the circuitry underlying these phenomena remains an active area of research.

Higher Level Systems Controlling Saccadic Eye Movements

Among subcortical areas, the most important for saccade control is the superior colliculus of the mesencephalon. Electrical stimulation of the SC will evoke saccades; inactivation interferes with the generation of saccades. Certain collicular neurons burst before saccades and may represent inputs to the BNs of Figure 16.10.

Many cortical areas also appear to be involved in the targeting and guidance of saccades. Electrical stimulation of visual cortex evokes saccades, and lesions there impair or eliminate visually guided saccades. Electrical activation of the frontal eye fields (FEF, Brodmann's area 8) causes conjugate saccades towards the opposite hemisphere. Cells in this area do not discharge before random/spontaneous saccades, but they do discharge prior to saccades to visual targets, suggesting the FEF are involved in planned saccades. The frontal eye fields were long thought to be the cortical 'motor areas' for voluntary eye movements, but this analogy with the primary motor cortex is not helpful and has largely been abandoned. A more apt analogy is with the role of a premotor area, Area 6, in somatic motor control. Recent studies have found neurons in areas 5 and 7 of the parietal lobe which discharge prior to eye movements.

Terms and Techniques

saccade	pursuit	PPRF
VOR	optokinetic	nystagmus
position error	retinal slip	motor error
VOR gain	internal replica	neural replica
efference copy	burst neuron	retinal error
neural integrator	Area MT	Area MST
frontal eye field (FEF)	superior colliculus	pontine nuclei
physiological nystagmus		

Grading Muscle Contraction

17

Table of Contents

Length-Tension Relationships of Muscle

The somatic motor system controls muscles involved in posture, locomotion and skilled movements of all kinds. In this section of the course, we will examine somatic motor mechanisms, beginning with the lowest level of organization and progressing to the highest. To understand the task faced by the nervous system we must first examine properties of muscle that are relevant to motor control.

The force developed by a muscle produces tension in its tendon, which in turn exerts a force on the load. At a constant level of neural activation, the force exerted (and the tension in its tendon) depends on the muscle's length. Experiments that measure this relationship produce what is traditionally called a ***length-tension diagram*** (Figure 17.1).

The basis for a relationship between length and tension lies in the sliding filament mechanism of muscle contraction spanning the sarcomere. Because the force is developed by the myosin cross bridges, it will depend on the number of such bridges that are forming at a particular moment. When the sarcomere is relatively short, the actin filaments overlap, interfering with the formation of cross bridges and thus decreasing tension. When the sarcomere reaches about 2 μm long, overlap of the actin and myosin filaments is optimal for cross bridge formation and maximum force develops. At longer lengths, however, overlap of actin and myosin decreases, and the amount of tension plateaus.

In the experiment illustrated in Figure 17.1, the muscle was held at a constant length as it developed force. In this situation, the muscle is said to contract ***isometrically***, that is, at a constant length. The term 'isometric contraction' is somewhat incorrect because the muscle does not actually shorten (i.e. contract), but it develops force nonetheless.

This force is generated by the molecular interactions of the thick and thin filaments of the

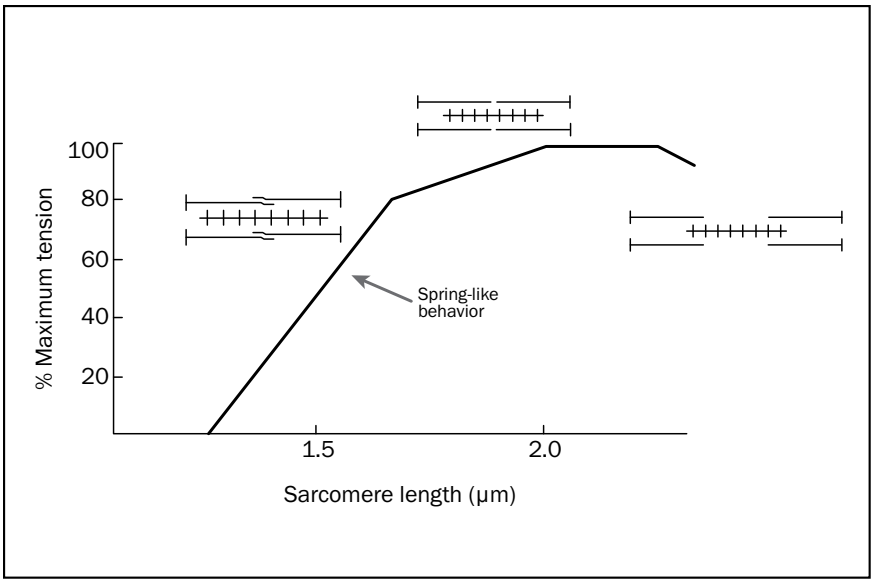


Figure 17.1

The length tension diagram of a maximally contracting frog muscle clamped at various starting lengths. Ordinate: average fraction of maximum isometric force developed. Abscissa: estimated length of a sarcomere in the test segment. Relative positions of actin and myosin are shown. (see text)

muscle and requires the expenditure of energy. Thus, a muscle can develop force without changing its overall length (just as a runner on a treadmill can run without moving anywhere). In fact, if muscle couldn't do this, it could never gather enough force to change its length. When you lift a weight a certain distance and hold it there, your muscles are contracting isometrically.

If the muscle is allowed to shorten, it is said to contract ***isotonically***, implying that a constant tension is developed in the tendon or a constant force is exerted on the load. This situation never actually occurs, however, because the amount of tension developed at a constant level of neural activation depends on the length of the muscle, as just noted. In any case, it is clear that, in the apparently simple act of lifting an object, the nervous system must adjust the neural input to the active muscles dynamically to compensate for the variation in force developed as the muscles shorten.

The length-tension behavior of muscle (Figure 17.1) illustrates why some models of motor control invoke the principle of a spring. Figure 17.2 schematizes the behavior of two mechanical springs that obey Hooke's law. Over a certain range of muscle lengths, we can model the muscle as a spring and use this law. Hooke's law states that the amount of force in a spring increases in direct proportion to the length of the spring. Or, in equation form:

$$F = kx$$

Where F is force, x is distance and k is the spring constant. The spring constant is the slope of this linear equation, and its magnitude relates to the ***stiffness*** of the spring: the higher the

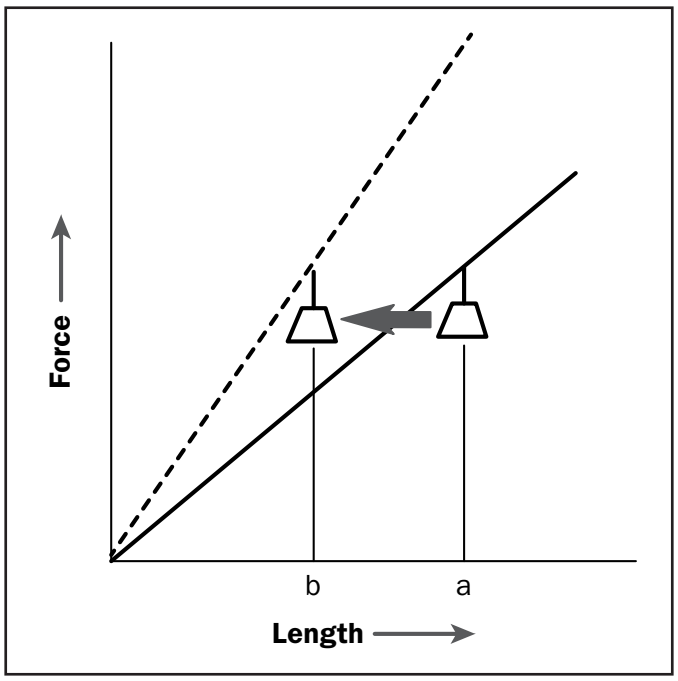


Figure 17.2

Behavior of an ideal spring. Solid and dashed lines: the relationship between length and force developed by two springs obeying Hooke's law. When a given load (symbolized by small weight) is attached to the weaker spring, the spring settles at length a. If the stiffness of the spring is 'magically' changed to that corresponding to the dashed line, the force exerted suddenly increases and the spring shortens to length b (filled arrow), at which the upward force exerted by the spring is again equal to the force of gravity acting on the load.

constant, the stiffer the spring. This notion of stiffness is often used in models of the behavior of muscle as its motor innervation changes, so it is useful to understand this concept.

Observe what would happen if there were some way suddenly to increase the stiffness "k" of a spring from which an object is suspended. Initially, the load (symbolized by the little weight in Figure 17.2) has come to rest at a point where the weaker spring is pulling up as hard as gravity is pulling down (length a). If the spring constant is suddenly changed to that of the stiffer spring (dashed line), the equation must still hold true. Because k has increased, and F stays the same, x must decrease. Indeed, in the diagram the spring shortens (arrow) to a new length (b) at which the lifting force again exactly matches the weight of the object.

In Figure 17.1 it can be seen that the muscle behaves more or less like a spring at sarcomere lengths from about 1.3 μm to 2 μm. In that experiment the muscle was activated maximally. At lower levels of activation the amount of force developed at a given length would be less. Thus, one can think of the level of activation of the muscle, i.e. the level of alpha motor neuron input, as varying the spring constant or stiffness of the muscle. Increasing the spring constant or stiffness of the muscle would be analogous to shifting the length-force plots of Figure 17.2 from the solid to the dashed line.

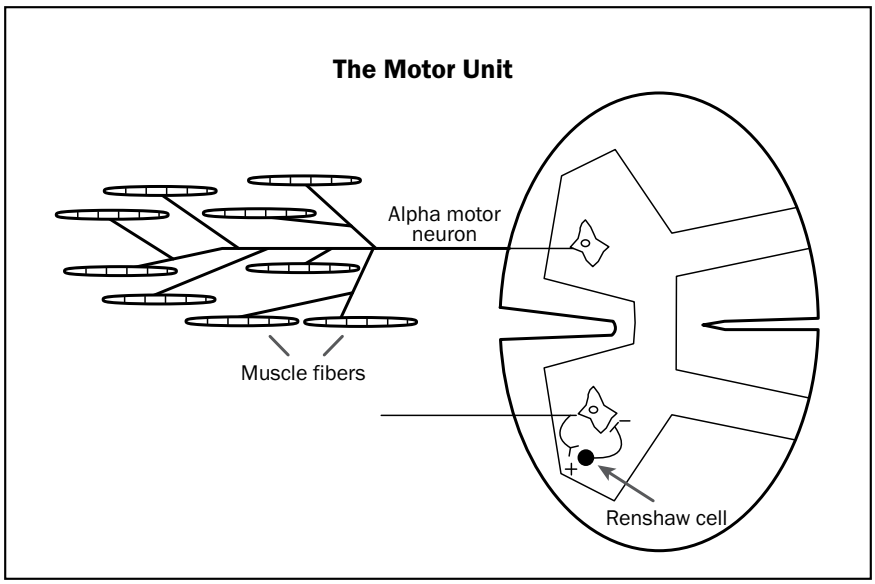


Figure 17.3

The motor unit is defined as an alpha motor neuron and all the muscle fibers it innervates. The firing rates of alpha motor neurons are controlled in part by feedback inhibition from nearby Renshaw cells. These are activated by axon collaterals of the very motor neurons they inhibit.

The Motor Unit

The CNS can exert motor control by modulating the firing rates of specific alpha motor neurons involved in a given movement. Thus, the smallest functional unit controlled by the CNS is the ***motor unit***: one alpha motor neuron and all the muscle fibers that it innervates (Figure 17.3).

Motor units are controlled by different inputs. In the spinal cord, circuits contained within one spinal segment or from other spinal segments exert control on motor units. There also exists higher-level descending influences from the brainstem and forebrain. The cerebral cortex even exerts direct control over certain spinal motor neurons via the corticospinal tract.

Motor unit size correlates with the size of the alpha motor neuron and that of the innervated muscle fibers. Larger motor neurons generally innervate many large muscle fibers, whereas small motor neurons innervate relatively fewer and smaller muscle fibers. In mammalian muscle each muscle fiber generally receives innervation from only one motor neuron. There are some exceptions to this. For instance, certain fibers in the extraocular muscles are multiply innervated.

One can estimate the average size of the motor units in a given muscle by dividing the number of muscle cells in a muscle by the number of innervating alpha motor neurons. Muscles with small motor units lend themselves to fine control. For example, an extraocular muscle may have one axon for every three muscle fibers, whereas in the large postural muscles of the leg a single axon may innervate over a thousand muscle cells.

Why do a motor neuron and its target muscle fibers form a unit? There is a very high safety factor in synaptic transmission at the neuromuscular junction, so if an action potential occurs in the motor neuron, all of the innervated muscle fibers contract. This concerted contraction is necessary for smooth motion. The formation of units also means that all the muscle fibers innervated by a motor neuron atrophy when that neuron dies or if its axon is cut. Motor neuron diseases include ALS (Lou Gehrig's disease), genetically inherited premature cell death, and infectious diseases such as polio.

Slow vs. Fast Muscle

The various muscle types are distinguished by their mechanical and biochemical properties. Table 17.1 compares the major functional properties of the two commonest types. Fast muscle relies on anaerobic metabolism and is suited to short but powerful contractions. Slow muscle relies on aerobic metabolism and is suited to sustained contraction. A given muscle may have a predominance of one type fiber (soleus has all slow) or may be mixed (gastrocnemius has both fast and slow). All fibers in a given motor unit are of the same type

Fast (White) Muscle	**Slow (Red) Muscle**
Large motor neurons	Small motor neurons
Axons conduct rapidly	Axons conduct slowly
Burst at high frequencies	Tonically active during movement
Fast rise in tension	Slow rise in tension
Fast decay in tension (Fast twitch)	Slow decay in tension (Slow twitch)
Rapid fatigue	Slow fatigue
Large fibers, many in motor unit	Small fibers, few in motor unit
Much stored glycogen	Little stored glycogen
High levels of glycolytic enzymes	Low levels of glycolytic enzymes
Low lipid, low oxidative enzymes	High lipid, high oxidative enzymes
Sparse mitochondria	Plentiful mitochondria
Poor capillary supply	Rich capillary supply

Table 17.4

A comparison of fast and slow muscle.

with respect to biochemistry, contraction speed and resistance to fatigue (rule of the motor unit). In a mixed muscle, such as gastrocnemius, the motor units are dispersed so that fibers of different motor units are intermingled.

The motor neuron has a role in determining the mechanical and biochemical properties of all the muscle fibers it supplies. When a motor nerve is cut, all muscle fibers atrophy and become alike in their histochemistry. When motor nerves to different types of muscle are switched, the reinnervated muscle takes on the characteristic dictated by the nerve (motor neuron), although this conversion is never complete. Both of these findings suggest that the motor neuron somehow dictates part of the chemical makeup of the fibers.

Modulation of Muscle Force

The motor unit is the smallest functional unit that the CNS can manipulate in motor function. Therefore, to grade muscular contraction, the CNS must somehow vary the activity of motor units. The CNS controls both the number of active motor units (***recruitment***) and the firing rates of motor units already recruited (***rate coding***). Both mechanisms can change the force developed by the muscle.

The firing rate of individual motor neurons depends on both peripheral and central pathways that change the membrane potentials of the cells. These systems will be discussed in detail later. A single action potential in a motor neuron produces a so-called twitch of the innervated muscle (Figure 17.4). This twitch is of much longer duration than the action potential, and, if evoked at short intervals, the twitches will fuse into a sustained contraction or tetanus. A

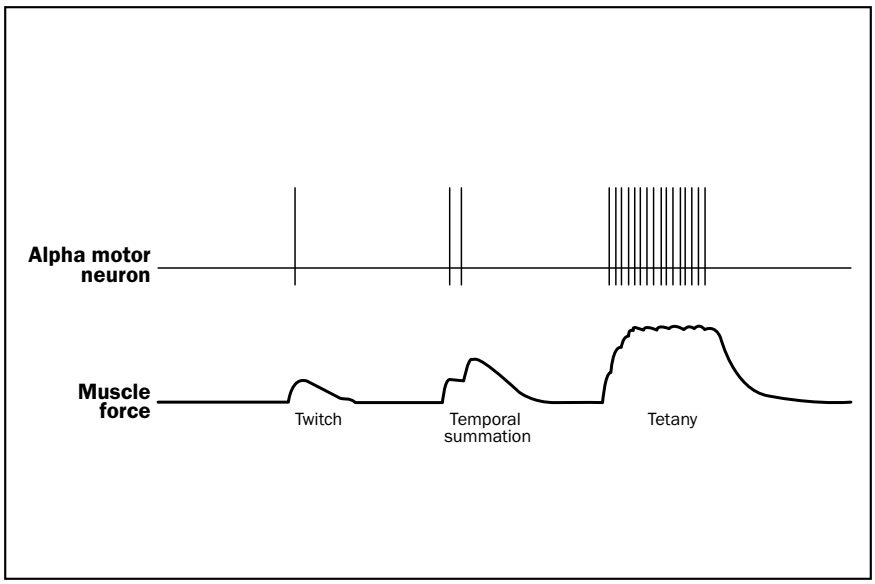


Figure 17.4

Summation of muscle force. Single action potentials in the motor nerve produce brief contractions called twitches. The duration of a twitch is longer than that of the action potential, so repeated action potentials lead to force summation. High frequency activation of the nerve causes a sustained contraction or tetany.

tetanus occurs when the axon fires at about 40 times a second.

Although motor neurons can be excited electrically to fire at over 800/sec., most almost never reach their highest possible firing rates. This limitation arises partially from collaterals of the motor axons which excite local inhibitory cells (called ***Renshaw cells***) which in turn inhibit those same motor neurons (Figure 17.3). Thus, the harder the motor neurons fire, the more they inhibit themselves. This feedback not only regulates the average firing rate but prevents the motor neurons from producing surges of force in the contracting muscle that would make the contraction jerky. Note that the activity of the Renshaw cell represents negative feedback.

During a voluntary contraction, small motor units that produce weak forces are recruited first; larger motor units that produce larger forces (because they engage the largest muscle fibers and many of them) are recruited last. This phenomenon is known as the ***size principle*** of motor neuron recruitment. Small motor neurons are activated first because, compared to larger motor neurons, it takes less synaptic current to depolarize their membranes to the threshold for action potential generation. Smooth movements are maintained even when larger units come in, because their force is added to an ever increasing baseline of force, so that percentage increase in force is kept relatively constant. Also, the muscle fibers innervated by one motor neuron are scattered within a muscle, an arrangement that contributes to smooth action of the muscle because the action potentials traveling in the axon terminals do not all reach their muscle fibers simultaneously. This fiber diversity also means that loss of a single motor unit (as in early stages of motor neuron disease) will not seriously affect control of a particular muscle.

Terms and Techniques

myosin	cross bridge	actin
thick filament	thin filament	sarcomere
isometric	isotonic	spring constant
stiffness	motor unit	fast muscle
slow muscle	recruitment	size principle
twitch	tetany	Renshaw cell

Muscle Receptors and Motor Control 18

Table of Contents

Efficient execution of movement depends critically on sensory information, not only about the goal of the movement but also about the positions of the limbs, the direction of motion of the body and the status of the muscles involved. Thus, it will come as no surprise that sensory nerves compose most of a muscle's innervation. For example, Table 18.1 lists the kinds of nerves that innervate the cat's soleus muscle, a large muscle of the leg, most of which are sensory. Whereas there are 150 fibers for direct action (alpha motor neurons), there are 240 for control (Ia, Ib, II, gamma motor neurons).

	#	Fiber Type	Target
Afferent	50	Ia	50 spindle primary endings
	40	Ib	40 Golgi tendon organs
	50	II	50 spindle secondary endings
	200-400	IV	pain and vasculature
Efferent	150	alpha motor neurons	25,000 extrafusal muscle fibers
	100	gamma motor neurons	300 intrafusal fibers in 50 spindles

Table 18.1

Sensory innervation of the soleus muscle.

There are two major classes of muscle receptors involved in the control of muscle contraction and proprioceptive sense. These are the Golgi tendon organ and the muscle spindle. In short, the Golgi tendon organ signals the tension developed by the muscle and the spindle transduces several aspects of muscle length.

The Golgi Tendon Organ

Golgi tendon organs (GTOs) are small encapsulated structures lying near the junction of tendon and muscle. Each is innervated by a single Ib muscle afferent, whose branches spread widely within the capsule and intertwine with the collagen fibrils of the tendon. When the tendon is placed under tension, either by pulling on it or by contraction of the muscle, the mechanosensitive nerve endings are squeezed by the fibrils of the tendon. This squeezing effect causes depolarization of the terminal branches and activation of the Ib afferent. Because anything that stretches the muscle also stretches the tendon and potentially activates the Ib fibers, the GTO is said to be mechanically ***in series*** with the muscle by analogy with the series resistors of Figure 2.6.

Pulling on the tendon is not an efficient way of activating the GTO because the visco-elastic elements of muscle lengthen and dissipate most of the stretch. However, when the muscle actively contracts against a load, the tendon is pulled in both directions, a stimulus to which the GTO is exquisitely sensitive. In fact, the tendon organ will respond briskly to the contraction of a single motor unit. The Ib fibers therefore send a signal to the spinal cord related to the amount of force being developed by the muscle.

A major target of this input is an interneuron that inhibits alpha motor neurons innervating the muscle from which the Ib fiber arises. Thus, the circuit provides a negative feedback pathway: contraction of the muscle increases tension in the tendon, which inhibits the alpha motor neurons via the GTO and Ib afferent and opposes muscle contraction. Note, however, that the inhibitory Ib interneuron is also regulated by many other inputs, and its role in the spatio-temporal control of muscle contraction is considerably more complex than this simple explanation suggests. Similarly, the signals arriving over the Ib fiber are dispersed to a number of circuits that appear to be involved in the timing of muscle action during stepping and other movements.

The Muscle Spindle

Muscle spindles are intramuscular receptors that transduce various aspects of muscle length. They regulate muscle length under changing loads and affect the dynamic control of muscle contraction. Spindles are about 4-10 mm long and are scattered throughout a muscle, occurring in relatively higher numbers in muscles used for fine movements. For example, control of head position is very precise and one muscle involved in this is the ***obliquus capitis superior***. In one study on human cadavers, the muscle was found to weigh 3.3 grams on

average and to contain 141 spindles, for a spindle density of 42.7 spindles per gram. By contrast, a muscle acting about the shoulder, *latissimus dorsi*, weighed 246 grams and contained 368 spindles, for a spindle density of 1.4 spindles per gram. Intrinsic muscles of the hand also have high densities of spindles, as one might expect.

Muscle spindles are innervated by at least two kinds of sensory afferents and receive their own motor innervation that is different from that of regular muscle fibers. Because the spindle has different types of afferents than the regular muscle, the CNS can selectively control the spindles.

A muscle spindle consists of several modified muscle fibers enclosed in a fibrous capsule. These small, thin muscle fibers are called *intrafusal* fibers, named after the fusiform shape of the spindle capsule. They are to be distinguished from the larger, power generating *extrafusal* fibers that form the main mass of the muscle. The intrafusal fibers are connected at each end to the extrafusal fibers and are thus said to lie in *parallel* with them. When the whole muscle is stretched, so are the spindles. When the extrafusal fibers contract, neural mechanisms ensure that the intrafusal fibers shorten along with them. This fact has major implications for motor control and will be discussed in greater detail later.

Terminals of the afferent fibers are positioned near the middle of the intrafusal fibers in the *equatorial zones*. These zones importantly contain few contractile elements. Stretch of the muscle tends to stretch this region primarily, deforming the sensory endings and activating stretch-sensitive channels in the nerve membrane. The spindle's afferents can send information to the CNS about two major aspects of the muscles: its length and the rate at which length is changing at a particular moment. These are often referred to respectively as the *static*

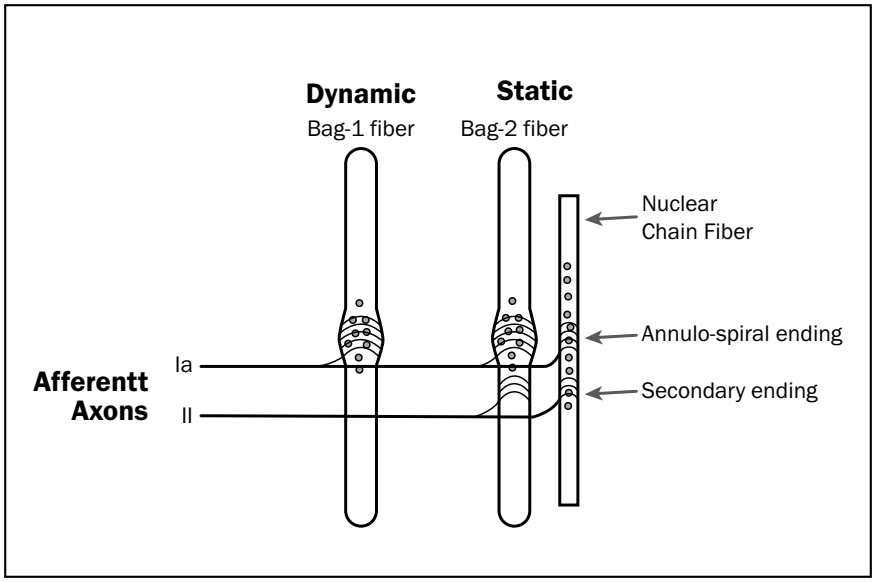


Figure 18.1

Nuclear chain and nuclear bag fibers and their innervation. Observe that all the intrafusal fibers receive Ia contacts but only those transducing static aspects of stretch have group II endings. A given spindle contains several of each fiber type.

and *dynamic* aspects of length. The components of the spindle organ are specialized to discern these two features and, as we will see, the properties of the spindle can be modified to emphasize either static or dynamic information.

The differential sensitivity of the spindle to dynamic and static aspects of stretch apparently arises from the mechanical properties of its intrafusal fibers. There are two types of intrafusal fibers named for the nature of their nuclei: ***nuclear chain*** and ***nuclear bag fibers*** (Figure 18.1). A given spindle may contain several intrafusal fibers of each type. Generally, the nuclear chain fibers transduce static aspects of a stretch. In other words, if a stretch is applied and maintained, the afferent terminals on the nuclear chain fiber are subjected to a sustained deformation and continue to signal as long as the stretch is maintained. Behavior of the nuclear bag fibers is more complicated.

Recent work has shown that there are two types of nuclear bag fiber. The ***bag-1*** fiber transduces dynamic aspects of stretch, while the ***bag-2*** fiber transduces static aspects. It is not clear why two subsystems are concerned with static stretch. Perhaps there are subtle but important differences in the connections and functions of these two systems.

The bag-1 fibers' capacity to somewhat selectively transduce dynamic aspects of stretch appears to depend on the visco-elastic elements in the equatorial region of these fibers. An applied stretch produces an immediate deformation of the afferent nerve ending located in this region. As the stretch is maintained, the polar regions of the fiber begin to lengthen, thus reducing the stretch of the equatorial region and relieving the distortion of the primary ending there. This reduces the firing rate of the ending, so the overall pattern of the discharge is phasic. Discharge accelerates during the stretching phase and then adapts during the maintained phase.

Spindle Afferents

Each spindle is supplied with a single Ia afferent fiber forming the ***primary or annulospiral endings on the equatorial zones*** of all the intrafusal fibers contained in the spindle capsule (Figure 18.1). Smaller diameter group II fibers contribute ***secondary*** endings that terminate primarily on the bag-2 and nuclear chain fibers. From this it follows that the Group II afferents carry information about static aspects of muscle length. Because the Ia fiber innervates both nuclear chain and nuclear bag fibers, its discharge contains a mixture of static and dynamic information.

Figure 18.2 illustrates these facts schematically. If the muscle is subject to a linear increase in length, the group II fiber gradually accelerates its discharge, which ends up at a higher rate at the new, longer length. In contrast, the Ia fiber discharges most intensely during the dynamic phase of the stretch, when the muscle length is actually changing. The Ia signal also contains a static component, as is evident from the difference in discharge rate between the end and beginning of the stretch.

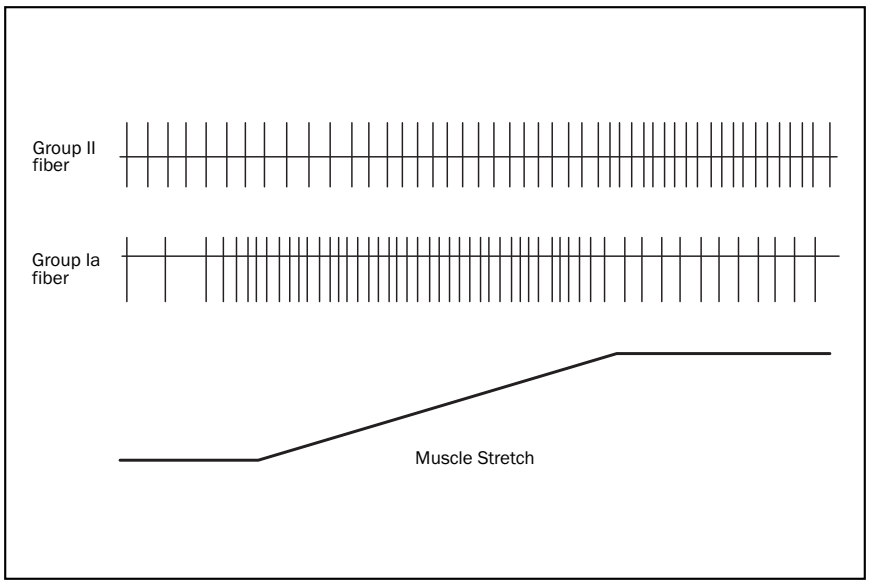


Figure 18.2

Signaling static and dynamic aspects of stretch. See text for details.

Motor Innervation of the Muscle Spindle

The intrafusal muscle fibers can contract, but they do not contribute directly to the force developed by the muscle. The extrafusal fibers, innervated by alpha motor neurons, are the force-generating elements. Unlike the extrafusal fibers, the intrafusal fibers do not discharge action potentials or exhibit contractile twitches. Rather, they contract in a graded fashion dictated by the sum of synaptic depolarizations (called ***end-plate potentials*** at the neuromuscular junction) produced by their motor innervation. This innervation is supplied by a class of small motor neurons called ***fusimotor neurons***, located in the anterior horn of the spinal cord, of which two major types, ***gamma*** and ***beta***, have been identified. The gamma motor neurons are the best understood, so we will mostly discuss this fusimotor neuron type.

Activation of the gamma motor neurons causes the contractile poles of the intrafusal fibers to shorten. The equatorial region, where the annulospiral ending is located, contains a lower concentration of contractile proteins and tends to be stretched by this contraction of the polar regions (Figure 18.3C). This action modulates the tension on the equatorial region and consequently the excitability of the sensory ending. The gamma innervation is said to ***bias*** the spindle, increasing or decreasing its sensitivity or modifying its transducing properties.

During contraction of the extrafusal fibers, shortening of the ***contractile poles*** can decrease the overall length of the intrafusal fibers, but the stretching force on the annulospiral ending can be maintained, or even increased, by accelerated discharge of the gamma motor neurons. Thus, intense activation of the gamma motor neurons can drive the Ia fiber to increased discharge even when the muscle is shortening.

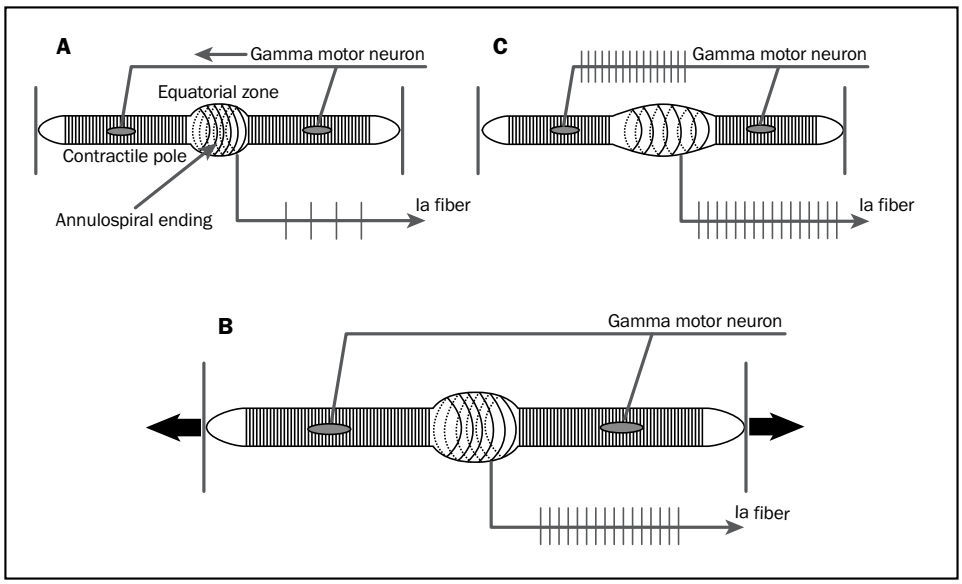


Figure 18.3

Action of the gamma motor neurons on the intrafusal fiber and Ia afferent. A. This schematic nuclear bag intrafusal fiber is fixed at both ends (vertical lines) as it would be in a real muscle. B. If the entire muscle is stretched (filled arrows), the intrafusal fiber and its annulospiral ending are stretched and the Ia activated. C. Because there are few if any contractile elements in the equatorial zone, the annulospiral ending will also be stretched if activity in the gamma motor neurons causes the contractile poles to shorten (open arrows), pulling each side of the equatorial zone towards the fixed ends of the intrafusal fiber.

Gamma motor neurons can be divided into two functional groups based on how they affect spindle sensitivity. Activity of the ***dynamic*** gammas enhances the response of the spindle to the rate of stretch, whereas activation of the static gammas enhances the response of the spindle to maintained stretch. The ***dynamic gamma*** innervates the nuclear bag-1 fiber and the ***static gamma*** innervates both the nuclear chain fiber and the bag-2 fiber (Figure 18.4). Thus, the Ia fiber, which supplies both nuclear bag and nuclear chain fibers, can be influenced by both the dynamic and static gamma motor neurons.

The effect of this differential gamma innervation on the Ia discharge may be seen in Figure 18.5. Stretch of the spindle without any gamma bias results in a Ia discharge which is greatest as length is changing (dynamic) but which also reflects the longer length at the end of the stretch (the static component). When the static gamma motor neuron is stimulated it causes the Ia discharge to reflect largely the static aspect of the stretch. Conversely, dynamic gamma activation enhances the phasic nature of the Ia discharge. It is known that the CNS can selectively modulate the discharge rates in static and dynamic gamma motor neurons and thereby "dial up" the relative proportions of dynamic and static information desired in the Ia discharge. It is uncertain whether such differential action on the different types of fusimotor neurons is used in the normal control of movement.

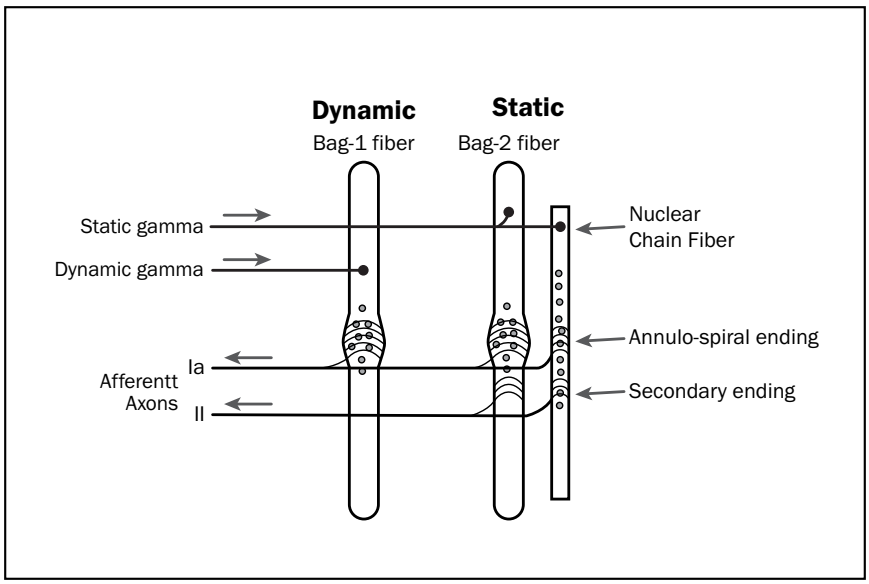


Figure 18.4

Static and dynamic gamma innervation of the intrafusal fibers.

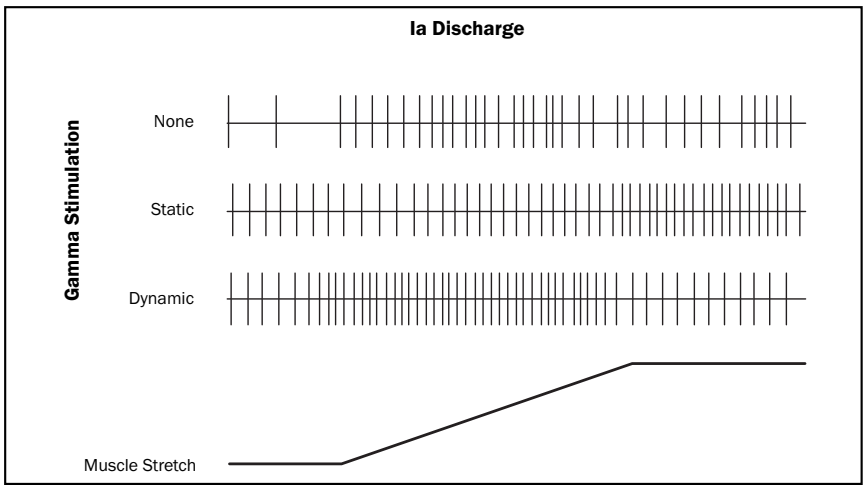


Figure 18.5

Effects of dynamic and static gamma bias on the Ia response to muscle stretch.

Significance of the Spindle Innervation

If the muscle spindle is to supply the CNS continuously with information about muscle length, it must be able to function as a stretch receptor as the muscle assumes different lengths. Thus, the overall length of the spindle must be adjusted to permit the afferents to 'stay on the air' as the gross muscle contracts and relaxes. Fusimotor innervation helps achieve such modulation. During active muscle contraction, the alpha and gamma motor neurons are activated simultaneously, allowing the contractile poles of the intrafusal fibers to contract along with the shortening of the extrafusal fibers (Figure 18.6). This ***alpha-gamma co-activation*** allows the sensory endings to monitor muscle length continuously and detect

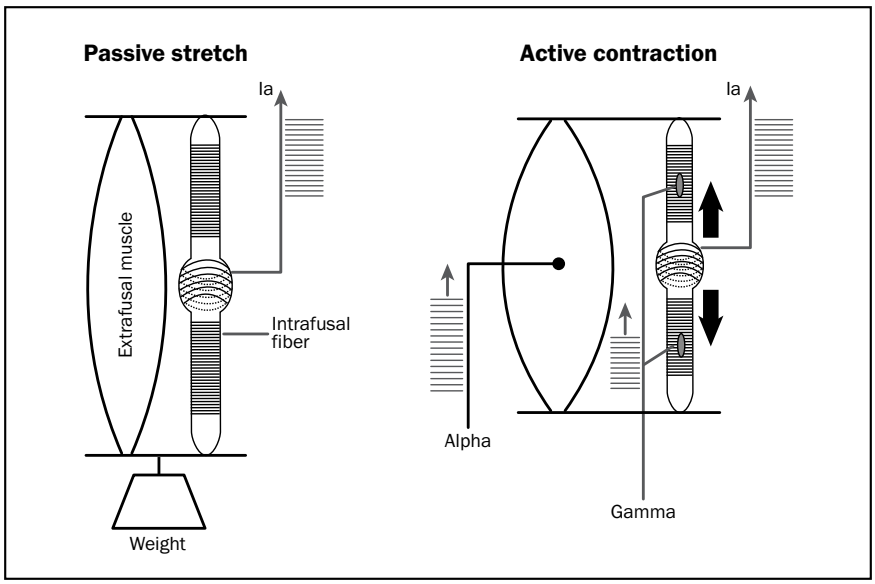


Figure 18.6

Gamma motor neuron modulation of spindle length to accommodate change in extrafusal length. By causing the contractile poles to shorten (filled arrows), the gamma activity maintains tension on the annulospiral ending and allows the spindle to continue to function as a stretch receptor.

any perturbations that might occur.

From this it will be clear that the discharge of the spindle afferents does not convey information about ***absolute*** muscle length or rate-of-change of length. To interpret the afferent activity, the CNS must take into account the gamma bias operating on the spindle and it appears to do this. The ability of the CNS to use motor output in this fashion to interpret sensory input is an example of a principle called ***efference copy***, which was discussed earlier in relation to proprioception.

The Myotatic Reflex

The Ia afferents from the muscle spindle participate in a unique circuit that plays several fundamental roles in the control of posture and locomotion. Sir Charles Sherrington was the first to observe that if one pulls on a muscle, it pulls back. He called this the ***stretch*** or ***myotatic reflex***. Stretch of a given muscle (the ***agonist***) excites primary endings on spindles, which results in a volley of action potentials to the CNS via Ia axons. These axons make monosynaptic excitatory connections with alpha motor neurons innervating the muscle from which the volley arose, and, via an interneuron, inhibitory connections with the motor neurons of ***antagonist***, or opposing, muscles. This is an example of ***reciprocal innervation***.

When a new load is added, the stretch reflex tends to oppose lengthening of the muscle

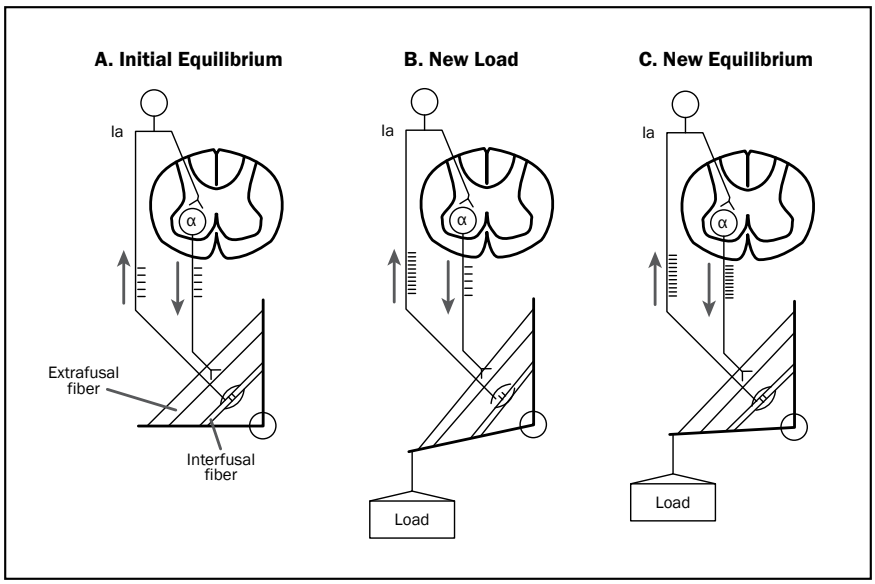


Figure 18.7

Load compensation by the myotatic reflex. A. Initial condition - the alpha motor neuron activity is sufficient to hold the elbow flexed at nearly a right angle. B. When a load is imposed on the arm, the elbow extends, the spindle is stretched and the Ia discharges intensely. C. The Ia activity activates the alpha motor neurons, the extrafusal fibers shorten until Ia input and alpha motor neuron activity are sufficient to hold the elbow at the new equilibrium position.

by adjusting the force of contraction to compensate for the new load (Figure 18.7). This circuitry is very prominent in muscle systems that face constantly changing loads, such as the ***antigravity*** (postural) muscles. Muscles of mastication also face changing loads during the grinding and chewing of foods of different consistency, so they also exhibit powerful stretch reflexes. Muscles that do not have to deal with variable loads, such as the muscles of facial expression and the extraocular muscles, do not exhibit stretch reflexes. The myotatic reflex by itself is not sufficient to support the weight of the organism by causing contraction of antigravity muscles. Other inputs to the motor neurons are required and these descend from the brainstem.

Figure 18.8 presents an operational model of the myotatic reflex as a regulator of muscle length in the face of changing loads. A control signal provided by the CNS imposes a certain level of activation on the alpha motor neurons. Monosynaptic input from the Ia afferents also feeds to alpha motor neurons, with the result that a certain amount of muscular force is developed to hold the load in a certain position and the muscle at a certain length. Increasing the input control signal will increase the muscular force, and the muscle will shorten and move the load to another position. If an additional weight is unexpectedly added to the load ("external force" in the figure), the muscle will stretch a little, stretching the annulospiral ending of the Ia afferent and increasing the input to the alpha motor neuron to compensate for the additional load and reduce muscle length.

Conversely, if the load is lightened, the muscle will shorten a little, reducing the Ia input, decreasing reflex activation of the alpha motor neurons and holding the muscle near its original length.

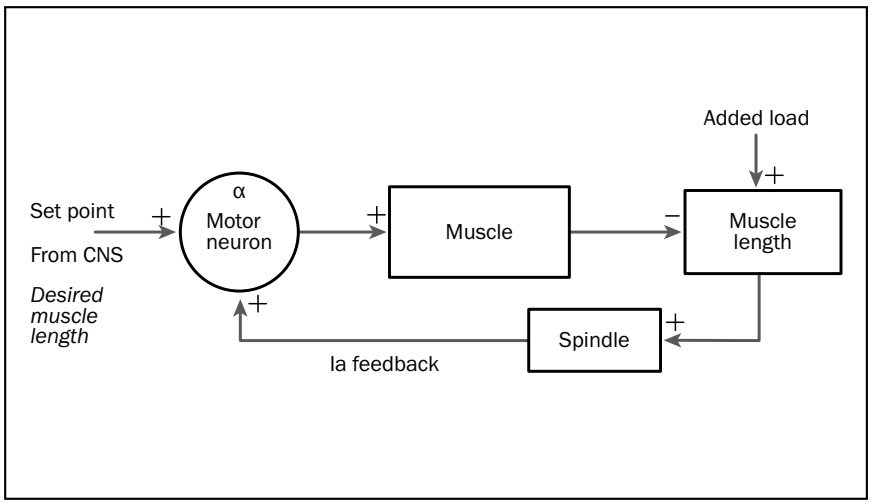


Figure 18.8

The myotatic circuitry as a regulator of muscle length. Note that with respect to muscle length, the Ia feedback is negative because an increase in spindle output results in a decrease in muscle length.

Even though the Ia feedback is positive to the alpha motor neuron, the overall effect of the loop is negative because the effect of increasing muscle force is to ***reduce*** muscle length. In other words, the effect of increasing muscle force negates the effect of increasing the load, hence the negative sign following the muscle in the loop of Figure 18.8. This reflex circuitry allows the muscle to automatically oppose any change in length caused by a change in the load.

It is useful to think of the myotatic reflex as contributing more ***stiffness*** to the muscle than that inherent in the muscle itself, as reflected in the length-tension diagram. Stiffness is the increment of force required to produce a given increment in length of the muscle. Similarly, as fusimotor bias increases, the sensitivity of the spindle increases and through the Ia connections produces a stronger opposition to change in muscle length. Thus, the gamma motor neurons can be thought of as controlling the ***stiffness of the myotatic reflex***.

The usefulness of such a control strategy is easiest to see if one considers not just one muscle but all the muscles acting across a joint. The rate at which the biceps, for instance, can flex the elbow depends not only on how hard it is contracting but on how much the triceps (an elbow extensor) is resisting, i.e. how stiff the antagonist is. Also, the stability of the knee joint increases with the stiffness of the opposing muscles acting across it.

It is evident that any CNS pathology that causes abnormally high levels of gamma motor neuron activity would be characterized by increased resting muscle tone or stiffness as well as increased sensitivity of the myotatic reflex to muscle stretch. Patients with such lesions in fact exhibit ***hypertonia*** (increased resting tone), as revealed by increased resistance to passive movement. They also exhibit ***hyperreflexia***, increased reflex excitability of the stretch reflex. A tap on the patellar tendon of the affected side will evoke a much brisker knee-jerk than that

elicited from the normal side.

The combination of hypertonia and hyperreflexia is called ***spasticity***. Animal models demonstrated that the myotatic reflex is abnormally active in individuals with spasticity: When experimenters cut the dorsal roots, the reflex pathway was interrupted, and the spasticity disappeared. Damage to the CNS that causes spasticity can also cause paralysis of the affected limb. This condition of ***spastic paralysis*** is to be distinguished from the ***flaccid paralysis*** caused by damage to the motor nerves or the muscles themselves.

Terms and Techniques

agonist	antagonist	synergist
reciprocal innervation	reciprocal inhibition	stretch reflex
myotatic reflex	muscle spindle	Golgi tendon organ
Ia, Ib, Group II fibers	intrafusal fiber	extrafusal fiber
nuclear bag fiber	nuclear chain fiber	annulospiral ending
primary ending	secondary ending	fusimotor neuron
contractile pole	equatorial region	gamma bias
static gamma	dynamic gamma	interneuron pool
series connection	parallel connection	alpha-gamma coactivation
efference copy	monosynaptic	polysynaptic
stiffness of the myotatic reflex	spasticity	hyperreflexia
hypertonia	flaccid paralysis	spastic paralysis

Spinal Reflexes and Motor Control

19

Table of Contents

An Expanded Role for the Myotatic Reflex

We have discussed the role of the myotatic reflex in regulating muscle length or stiffness. Perturbations in a muscle's load change the Ia input to the motor neurons to compensate for the perturbation. By 'cinching up' the intrafusal fibers as the extrafusal fibers contract, activity in the gamma efferents allows the spindle to 'stay on the air' at different muscle lengths so that the reflex can function in this ***regulatory mode***. While this idea was introduced in terms of a stationary load, it also applies to dynamic situations like a muscle shortening to lift an object. Because the spindle is allowed to function continuously as a stretch receptor, momentary variations in load or resistance are compensated virtually instantaneously via the monosynaptic connection between Ia afferents and alpha motor neurons.

Ordinarily one would think of the input to the alpha motor neurons as being the control signal for muscle length or force, but the gamma input to the intrafusal fibers could also supply such a signal. It occurred to Merton that voluntary movement might be accomplished simply by using the gamma motor neurons to contract the spindle and drive the Ia afferents, which in turn would activate the alpha motor neurons to excite the muscle. This is the ***gamma loop:*** gamma motor neuron » spindle » Ia afferent » alpha motor neuron » muscle. Such a system would resemble a ***servomechanism*** in which a low-power device (the spindle) is used to control a high-power device (the muscle) through a feedback loop. Note that the tiny, weak intrafusal fibers add no force directly to that developed by the extrafusal fibers. They serve only to vary the tension on the sensory nerve endings of the spindle.

Figure 19.1 illustrates Merton's basic idea. In the initial condition, a load on, say, the elbow flexor muscle (***biceps brachii***) stretches the muscle spindle just enough to maintain alpha motor neuron activity at a level sufficiently high to balance the load. There is some ***gamma***

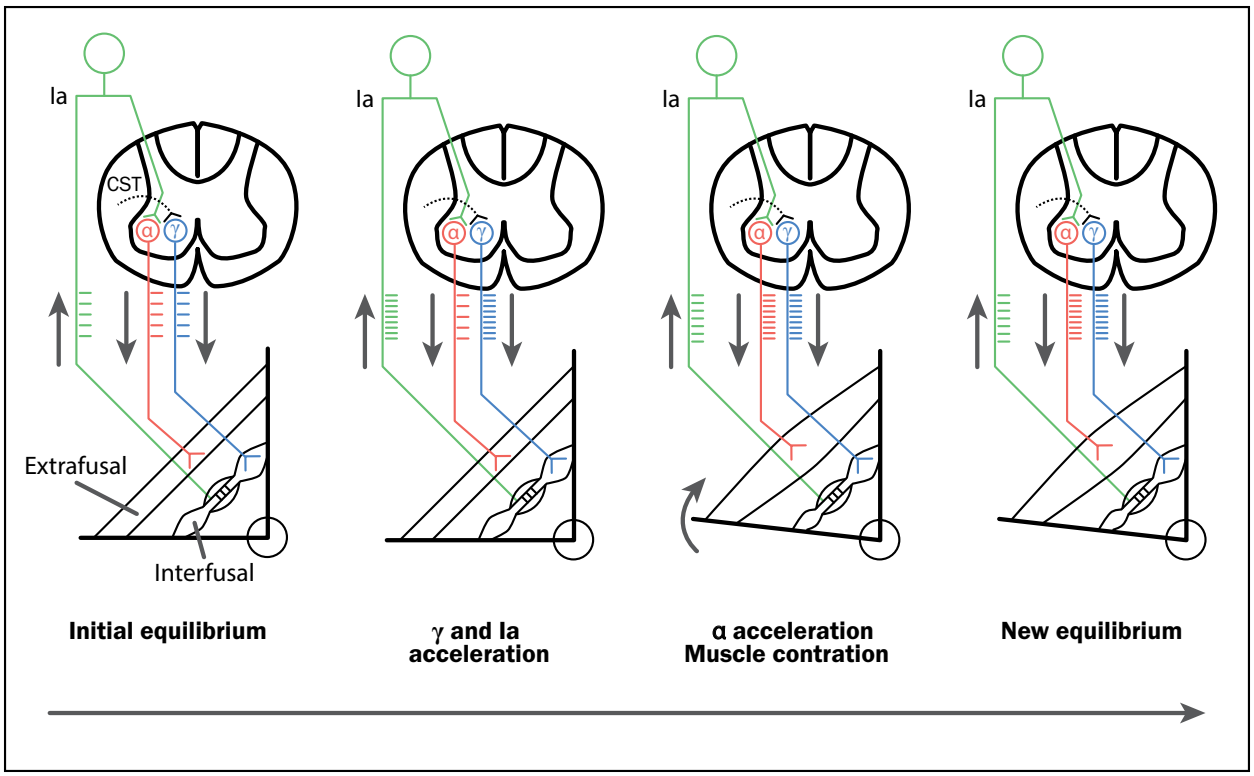


Figure 19.1

Merton's follow-up-length-servo hypothesis. CST-corticospinal tract. Notice that the CST is *assumed* to contact only the gamma motor neurons. See text for details.

traffic to keep the spindle on the air. The CNS now increases the input to the gamma motor neurons from, say, the corticospinal tract (via the dotted connection in Figure 19.1). This causes the contractile poles of the intrafusal fibers to shorten, stretching the annulospiral endings, which in turn leads to increased Ia and alpha motor neuron activity, and the muscle shortens. As the muscle shortens, the stretch on the spindle is relieved (the spindle is ***unloaded***); the Ia input declines to some point at which it is just sufficient to sustain the muscle at the new equilibrium length. Merton called this system a ***follow-up-length servo***: the command for a new length goes out via the low-power mechanism (the gamma motor neurons) to the high-power mechanism (Ia to alpha motor neurons and muscle) which ***follows up*** to achieve the new length. Here one might say that the gamma innervation is functioning in ***controller mode***, i.e. it is doing more than 'keeping the spindle on the air'.

This hypothesis was tested experimentally by Vallbo using percutaneous microelectrode recordings in human volunteers to record Ia activity from an identified muscle. It is difficult to measure alpha and gamma motor neuron activity directly but one can record the electromyogram (EMG) of the target muscle (reflecting alpha activity) and compare it with

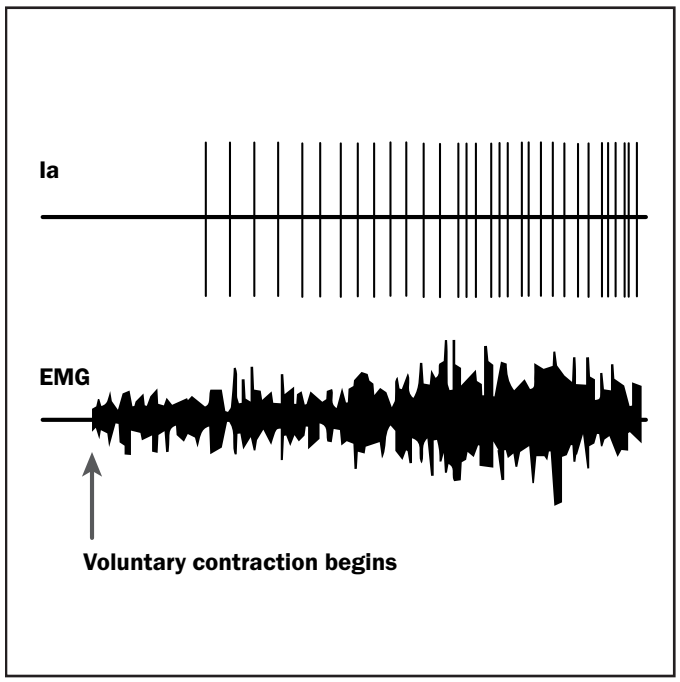


Figure 19.2

Schematic illustration of Vallbo's test of Merton's hypothesis. The upper trace in the pair represents impulses from a single spindle Ia and the lower trace is the electromyographic signal from the same muscle. Observe that the EMG activity begins before the Ia activity. After Vallbo, *J. Physiol.* 218:405-431, 1974.

the activity of Ia afferents (reflecting gamma activity). Merton's hypothesis predicts that, ***during voluntary movement, the gamma motor neurons should become active before the alpha motor neurons to the contracting muscle***—that is, Ia activation will precede EMG activation.

Vallbo tested this hypothesis directly and the result of his experiment is illustrated schematically in Figure 19.2. The EMG activity (lower trace) preceded the Ia activity (upper trace), disproving Merton's hypothesis. When Vallbo took into account the slower conduction velocity of the gamma motor neurons' axons, he concluded that on average the alphas and gammas were activated simultaneously. In a word, during voluntary movement there is alpha-gamma coactivation.

This observation led to the notion that the co-activation of alphas and gammas allows the CNS to exploit the gamma loop to assist in depolarizing the alpha motor neurons, boosting the input to the muscles. This is called ***servo-assistance*** (Figure 19.3). Figure 19.4 illustrates schematically an experiment in which Vallbo demonstrated the phenomenon of servo-assistance. In this study the muscle was allowed to shorten (isotonic contraction), which might be expected to reduce Ia activity by unloading the spindle, or at least leave the Ia activity unchanged. In reality, the Ia activity actually increased as the muscle shortened, rising well above its level at the beginning of the contraction. This finding shows that the intrafusal fibers are being activated so vigorously that the spindle not only stays 'on the air' but actually augments its input to the alpha motor neurons to assist in the voluntary contraction.

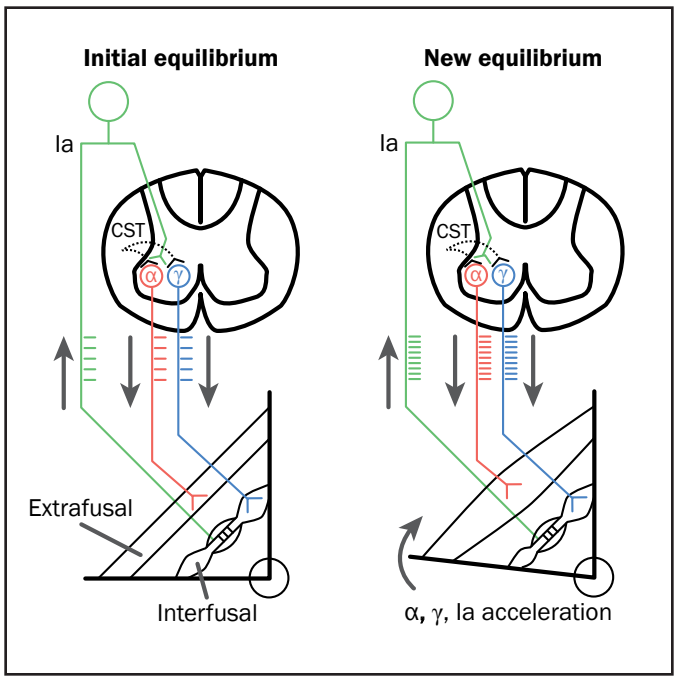


Figure 19.3

Servo-assisted movement. Intense activation of the gamma motor neurons by the corticospinal tract (CST) and other inputs allows the gamma loop to assist in depolarizing the alpha motor neurons to the contracting muscle.

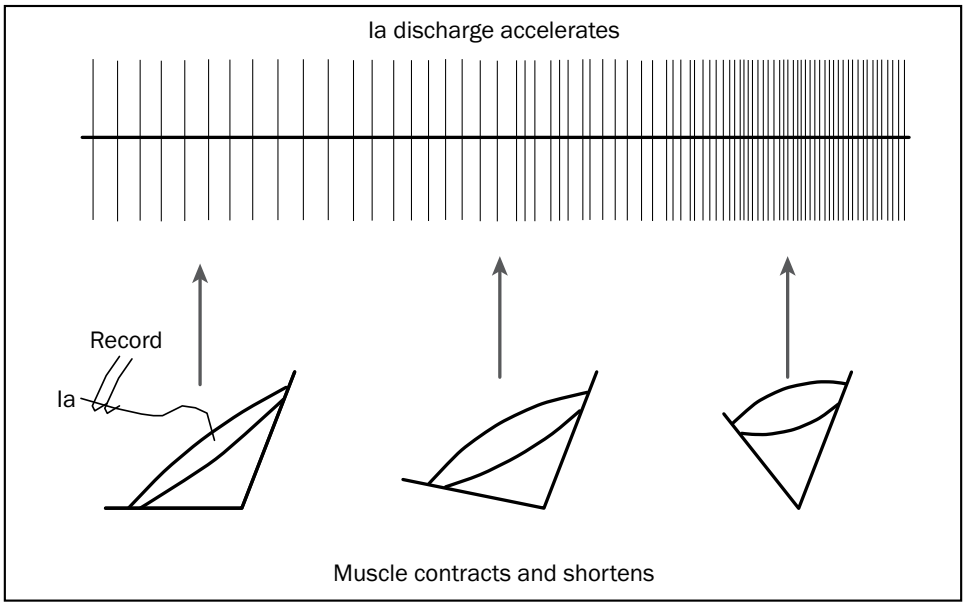


Figure 19.4

Evidence of servo-assistance during voluntary muscle contraction. The Ia discharge increases even though the muscle is shortening. This boosts the excitation to alpha motor neurons serving the muscle.

The servo-assistance mechanism has another use, namely the control of muscle contraction. Recall from the length-tension diagram of a muscle that, over physiological lengths and if neural input is constant, the muscle generates less force as it shortens (Figure 17.1). Thus, to maintain a constant force as it contracts, the muscle must be activated more vigorously as it shortens, and this can be accomplished by increasing the alpha motor neuron activity via the gamma loop.

Models of the Myotatic Reflex

Figure 19.5 diagrams the myotatic reflex machinery as a servomechanism for controlling muscle length. There are two input signals, one directly via the alpha motor neurons and the other indirectly via the gammas. By controlling the sensitivity of the spindle, the gammas can modulate the gain of the entire system. As activity in both alpha and gamma motor neurons increases, the load is lifted and a new equilibrium length established. Even though the Ia input to the alpha motor neurons is positive, the net effect of the loop is negative because an increase in muscle force lifts the load and results in a decrease in muscle length. This is signaled by the presence of the minus sign in the pathway between the boxes for muscle and the load.

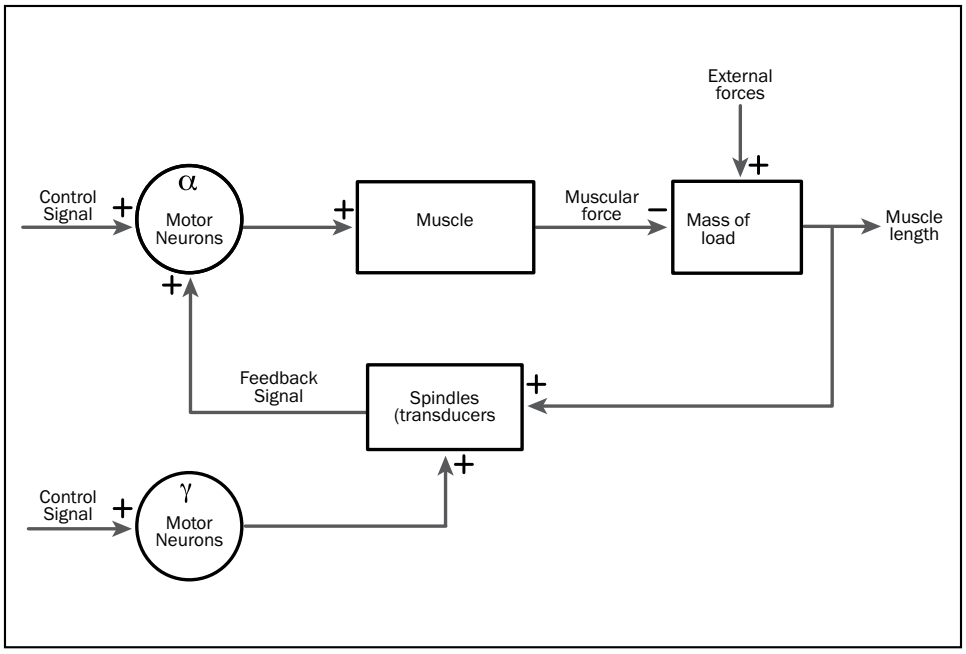


Figure 19.5

The myotatic reflex modeled as a servomechanism. See text for discussion

In earlier discussions we pointed out the parallel between the mechanical properties of a muscle and a spring. This notion can be extended to the myotatic reflex, which causes a muscle to exert more force as it is stretched. Observe in Figure 19.5 that the input to the alpha motor neuron is the sum of the control signal and the Ia discharge from the spindle, which is a function of the muscle's length.

Now, imagine that the spindle has some gain B whose contribution to the motor neuron is proportional to B*L, where L is muscle length. Then, let's impose a stretch ΔL on the muscle (and assume that the control signal is constant). When including gain B, we see that the input to the muscle has been incremented by BΔL. The muscle also adds a gain G ($^{force}/_{input}$) to the system, yielding an increment in force of BGΔL.

Like that of a spring, the stiffness of the myotatic reflex be defined as change in force divided by change in length, or $^{\Delta F}/_{\Delta L}$. Substituting in the above calculations, we see that:

$$\text{Stiffness} = \frac{\Delta F}{\Delta L} = \frac{BG\Delta L}{\Delta L} = BG$$

This equation mathematically models that the stiffness of the myotatic reflex is a function of both the gain of the spindle and the gain of the muscle. Because the gamma motor neurons can change the gain of the spindle, they can control the stiffness of the reflex.

Equilibrium-Point Models of Joint-Angle Control

The concept of the myotatic reflex as a spring with variable stiffness can be extended to the control of agonist and antagonist muscles acting across a joint. Consider Figure 19.6 in which two springs (muscles) with constants (or stiffnesses) K_L and K_R swing a load around a joint. In the initial condition, $K_L = K_R$ and the tensions developed by each spring just balance each other. The load is therefore supported when it hangs straight down. Now, if the ratio of K_L and K_R suddenly changes so that K_R is higher, the right-hand spring's tension increases and will no longer be in balance with the load and with the force exerted by the left-hand spring. As a result the right hand spring will shorten and the load will be pulled to the right until all the forces are once more in equilibrium. The reverse will occur if K_L becomes larger than K_R.

Note that if K_R is greater than K_L, when you displace the load to the left and let it go, it will swing rightwards. The load will rest at the position at which the force exerted by the left-hand spring will exactly equal that of the right-hand spring, meaning the combinations of length and tension of the two springs are in equilibrium. If the load is displaced to the right past this position and let go, it will swing to the left until the equilibrium point is reached. If the load is not at the equilibrium position, the imbalance between the two springs can be thought of as a sort of ***position error*** that is eliminated when the load is allowed to reach its equilibrium position from any direction.

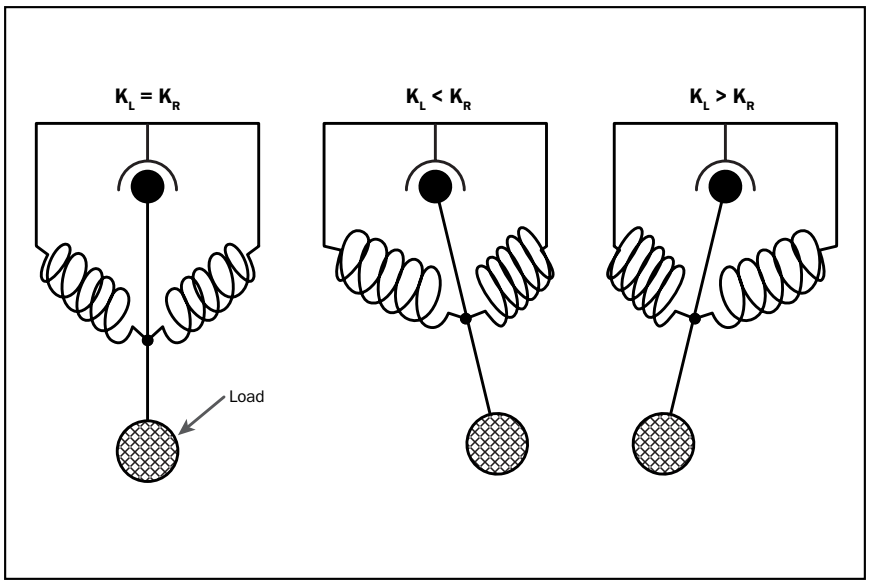


Figure 19.6

Spring model of joint angle control. KR and KL are the spring constants or stiffnesses of the right and left springs, respectively. See text for discussion.

This idea has given rise to a family of so-called ***equilibrium point*** models of motor control which assume that, when directing movement, the nervous system specifies a joint position at which the net forces acting on the joint are zero. Displacement away from that position creates a net force to return the joint to its equilibrium position. These forces are generated by the spring-like action of the myotatic reflex circuitry or, in some models, by the muscles themselves. In neural terms, the equilibrium point might be a specific pattern of 'desired' activity among the several motor neurons, sensory afferents and interneurons involved in controlling the limb musculature. The difference between this 'set point' and the current configuration of activity in these neurons would constitute an error signal, and the limb would move until this signal is eliminated. Dynamic movement of the joint (or limb) can be thought of as caused by a signal that continuously varies the equilibrium point along some trajectory until the limb has carried out the intended action.

Bizzi has provided some evidence for this hypothesis by stimulating the spinal cord locally in frogs (Figure 19.7). Microstimulation of a given location among the spinal interneurons causes the limb muscles to develop forces that move the limb to a particular position. When he stimulated the same site but changed the initial position of the limb, he found that the distribution of evoked muscle activity changed, but the limb went to the same place.

The study suggests that the activity of the interneurons specifies the equilibrium position of the limb as a certain pattern of stiffness in the relevant muscles (realized through the alpha and gamma motor neuron activity). If these stiffnesses are not in equilibrium, as in the spring model of Figure 19.6, the limb will move until this equilibrium is achieved. The pathway the limb takes to get to equilibrium will depend on the initial position of the limb, but the end position

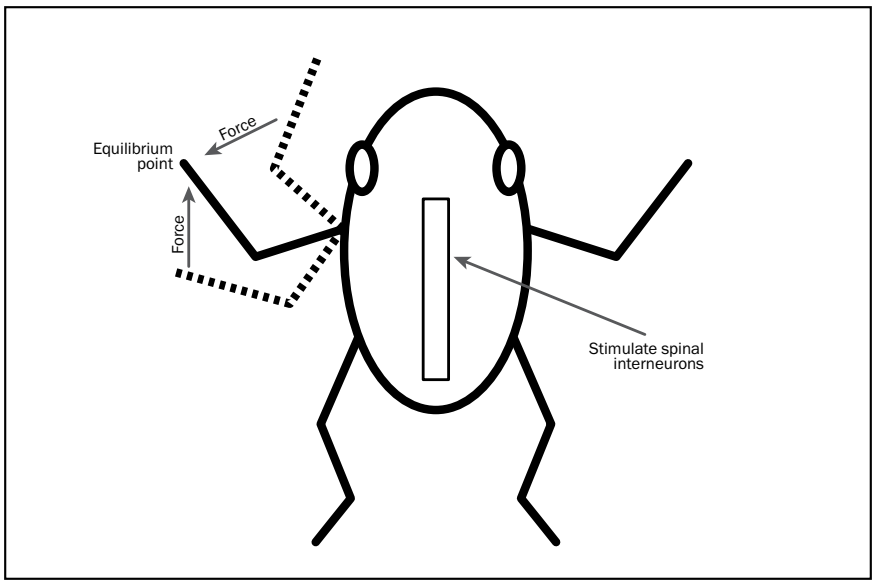


Figure 19.7

Bizzi's experiment supporting an equilibrium point process. When the limb is displaced away from the equilibrium position (i.e. to the positions indicated by dotted lines), muscle forces in response to focal electrical stimulation of the spinal cord are such as to return the limb to the equilibrium position. See text for details.

will be the same no matter where the limb is first positioned. This hypothesis illustrates how the brain may exploit the mechanism of the myotatic reflex in the control of movement. The same concepts may apply to other built-in circuitry of the spinal cord, as we shall now discuss.

Multi-Segmental Connectional Patterns in Spinal Motor Circuits

Until now we have considered only the monosynaptic myotatic reflex, and it should be clear that even this low-level circuitry supports control processes of great sophistication. The myotatic reflex is said to be ***segmental*** because the complete circuit can be found in a segment of the spinal cord. More elaborate connections link multiple segments, and their workings may be seen in certain stereotyped responses to strong stimulation. These circuits control the patterning of movements of head and limbs during normal posture and locomotion. The scope of their action is broader than that of the myotatic reflex. Nonetheless, they are confined to the spinal cord and also operate under the influence of descending inputs as well as afferent projections from skin, muscles and joints. The ideas of servo-assistance and equilibrium points can be applied to their function, as well. Here it will suffice to identify the more important circuits as revealed by the reflexes they mediate.

An important example is the ***withdrawal reflex***, sometimes called the ***flexor reflex***. A painful (noxious) stimulus to the skin of, say, the foot leads to excitation of flexors and inhibition of extensors of the leg so that the foot is withdrawn from the offending stimulus. Note that there are reciprocal effects on agonist and antagonist muscles. This reflex is said to be ***prepotent***, meaning that it will override other reflexes, such as the myotatic, which would otherwise keep

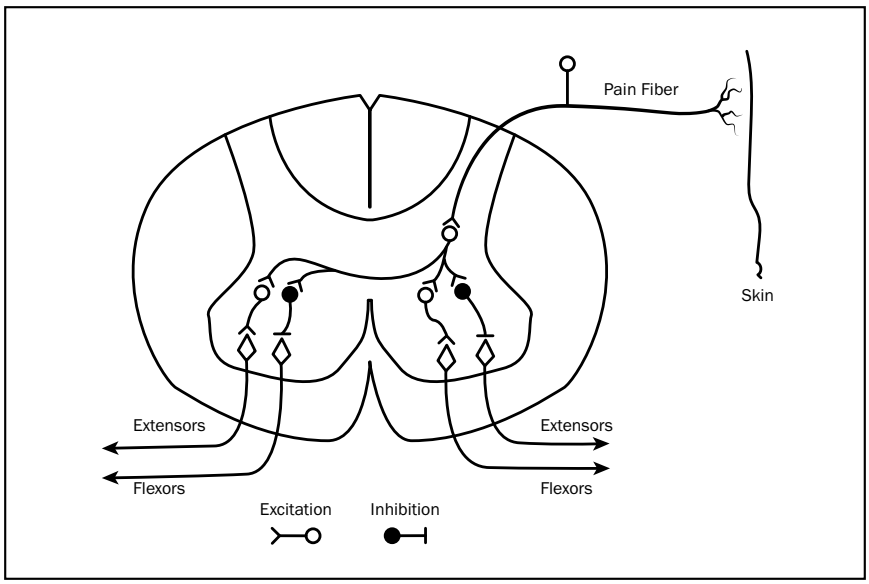


Figure 19.8

Basic segmental circuitry of the flexor-crossed extensor reflex. Interneurons organize the combination of ipsilateral flexion and contralateral extension in response to a noxious stimulus.

the leg extended. The reflex also shows ***local sign***, i.e. the pattern of contraction changes as the locus of the noxious stimulus changes. The movement is always appropriate to produce an effective withdrawal from the stimulus. Thus, if one places the palm of ones hand on something hot, the reflex produces flexion about the elbow. However, if the back of the hand is lifted up against something hot, the reflex produces elbow extension to distance the hand from the source of stimulation.

Thus, withdrawal reflex is a generally more accurate term than flexor reflex. The reflex is graded such that more muscles become involved as the intensity of the stimulus is increased. This recruitment based on stimulus intensity means that the movements may vary from a small flick of the hand to a concerted jump away from the stimulus. The neural pathways of the withdrawal reflex involve interneurons at all levels for both excitation and inhibition.

The withdrawal reflex may involve both sides of the body if a stimulus is very intense. If the right foot is affected and the right leg flexes, the extensors of the left leg must stiffen to support the body. Hence, the withdrawal reflex is sometimes referred to as the ***flexor/crossed extensor reflex***. Here we have an example of ***reciprocal innervation*** that is crossed, i.e. the pattern of innervation of flexors and extensors on one side is the converse of that on the other side of the body.

Note that by switching this pattern from one side to the other one obtains the alternating pattern of extension and flexion involved in locomotion. Thus, the spinal machinery mediating the alternating flexor and extensor activation of the withdrawal reflex is without doubt the same as that employed in voluntary locomotion, both bipedal and quadrupedal. The pools

of spinal interneurons mediating these connections are accessible from higher motor control areas. Thus, to produce a hopping or galloping mode of ambulation the brain must switch off the connections responsible for the flexor/crossed extensor pattern and allow both legs to flex and extend at the same time. These examples illustrate the critical pattern-generating role of the interneuron pools.

Normal body movements also require coordination of muscles controlling the position of the head, arms and legs. Some of this coordination is mediated by ***propriospinal neurons*** that lie entirely inside the spinal cord and may extend for long distances. These neurons are responsible for the ***long spinal reflexes***, which can be evoked by activating the proprioceptors of the neck and vertebrae or the receptors of the vestibular system (Figure 19.9). For example, extending the neck as though one is looking upward results in increased extensor tone in both arms and increased flexor tone in the legs, allowing the whole body to adjust to the new direction of gaze. As the head moves to look down, to the right or to the left, appropriate adjustments are made in the posture of the limbs and trunk. Such reflexes are less prominent in humans but may appear in exaggerated form when there is brain damage.

Technical Appendix: The EMG

Fine needles inserted into a muscle can be used to record the voltages generated by extracellular currents associated with the action potentials of adjacent muscle cells. Some typical records are illustrated in Figure 19.10. Remember, when an alpha motor neuron fires, all the muscle cells with which it synapses also generate action potentials and a mechanical

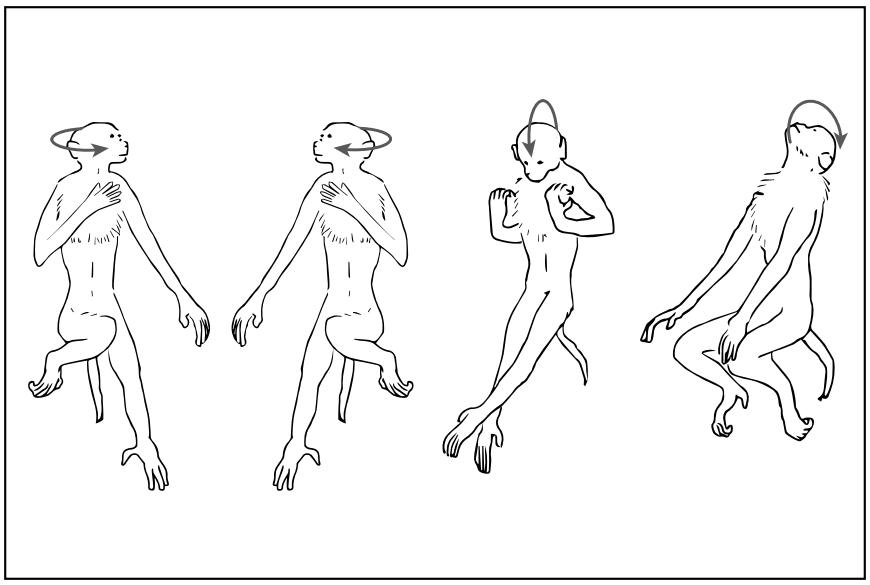


Figure 19.9

Long spinal reflexes elicited by inputs from the vestibular system and receptors in neck muscles and the vertebral column. Modified from Twitchell, T.E. ***J. Amer. Phys. Ther. Assoc.*** 45:413, 1965.

twitch. Thus, in the normal EMG the voltage transients are due to the activation of entire motor units, not just single muscle fibers.

However, because the muscle cells in a motor unit are scattered throughout an area of the muscle, some may contribute more to the recorded potentials than others. When a motor neuron is diseased, it often fires spontaneously, producing a sporadic series of twitches called ***fasciculations***, because all of the muscle cells in the fascicle are activated more or less simultaneously. The resulting contraction can often be seen under the skin and the series of accompanying motor unit discharges are evident in the EMG.

When a muscle fiber is denervated, it becomes hypersensitive to circulating acetylcholine and to other stimuli. The resulting asynchronous contraction of the muscle fibers, called ***fibrillation***, often cannot be detected by visual inspection but can be seen in the EMG.

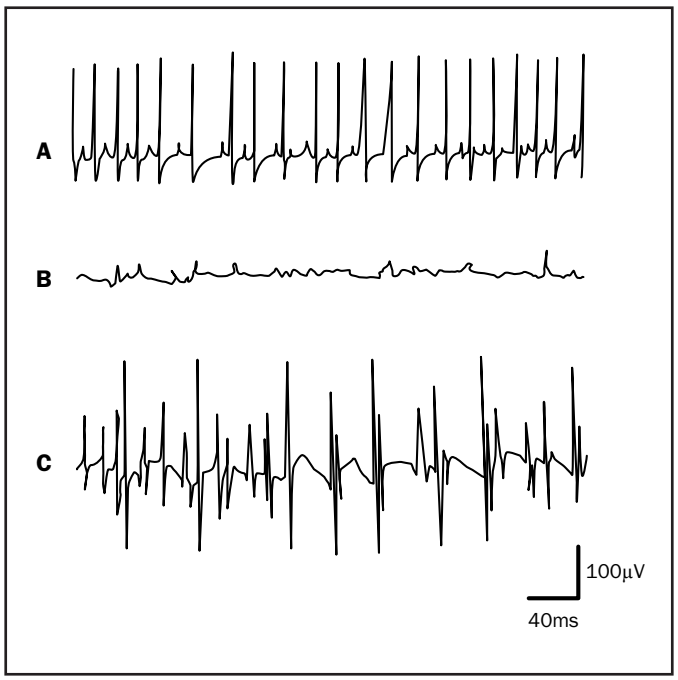


Figure 19.10

A. A needle electrode in the biceps of an alert monkey shows the action potentials of one large motor unit and one or two smaller ones during active contraction.

B. Spontaneous fibrillation potentials in a denervated gastrocnemius muscle.

C. Spontaneous fasciculation and fibrillation in a case of polio, six months after onset of paralysis. Modified from Gelfan, "Muscle" in Fulton, ***Textbook of Physiology***, W.B. Saunders, 1955

Terms and Techniques

regulator mode	controller mode	servo mechanism
follow-up-length servo	servo-assistance	equilibrium point model
segmental reflex	multisegmental reflex	withdrawal reflex
flexor reflex	crossed extensor reflex	local sign
electromyogram	fibrillation	fasciculation
long spinal reflexes	propriospinal neurons	gamma loop

Supraspinal Motor Control Systems

20

Table of Contents

An Overview of Somatic Motor Control Circuitry

The final common pathway for all motor acts is the alpha motor neuron and its associated muscle fibers. Alpha motor neurons are directly modulated by ***interneuron pools***, which provide ready-made circuits for a range of coordinated movement patterns. For example, the flexor crossed extensor reflex is a full motor pattern mediated by spinal interneurons. This reflex, as well as many other spinal reflexes not covered in this course, represents low-level movement patterns carried out purely by spinal motor commands.

Because the spinal reflexes limit us to only a subset of movements, in order to achieve the vast diversity of movement of which humans are capable, descending input is required to modulate the low-level reflexes. This pattern modulation stems from ***supraspinal*** (above the spinal cord) inputs from both cortical and subcortical regions of the brain. For instance, the flexor-crossed extensor reflex is useful in a task like walking, but not when shooting a basketball, which requires simultaneous bending at both knees. The flexor-crossed extensor reflex must therefore be inhibited via supraspinal control systems. This supraspinal control of interneurons and motor neurons involves a number of structures, some of which are identified in the overview scheme of Figure 20.1.

In the next few chapters, we will be discussing the different supraspinal systems that control motor output at the spinal level. First, we will discuss the descending pathways that take information from the brainstem/cortex and synapse directly onto interneurons in the spinal cord. After this discussion, we will look at sensorimotor cortex, where neurons selectively respond during the planning and execution specific motor tasks. Finally, we will study systems that do not directly project to the spinal cord—the basal ganglia and cerebellar loops. These

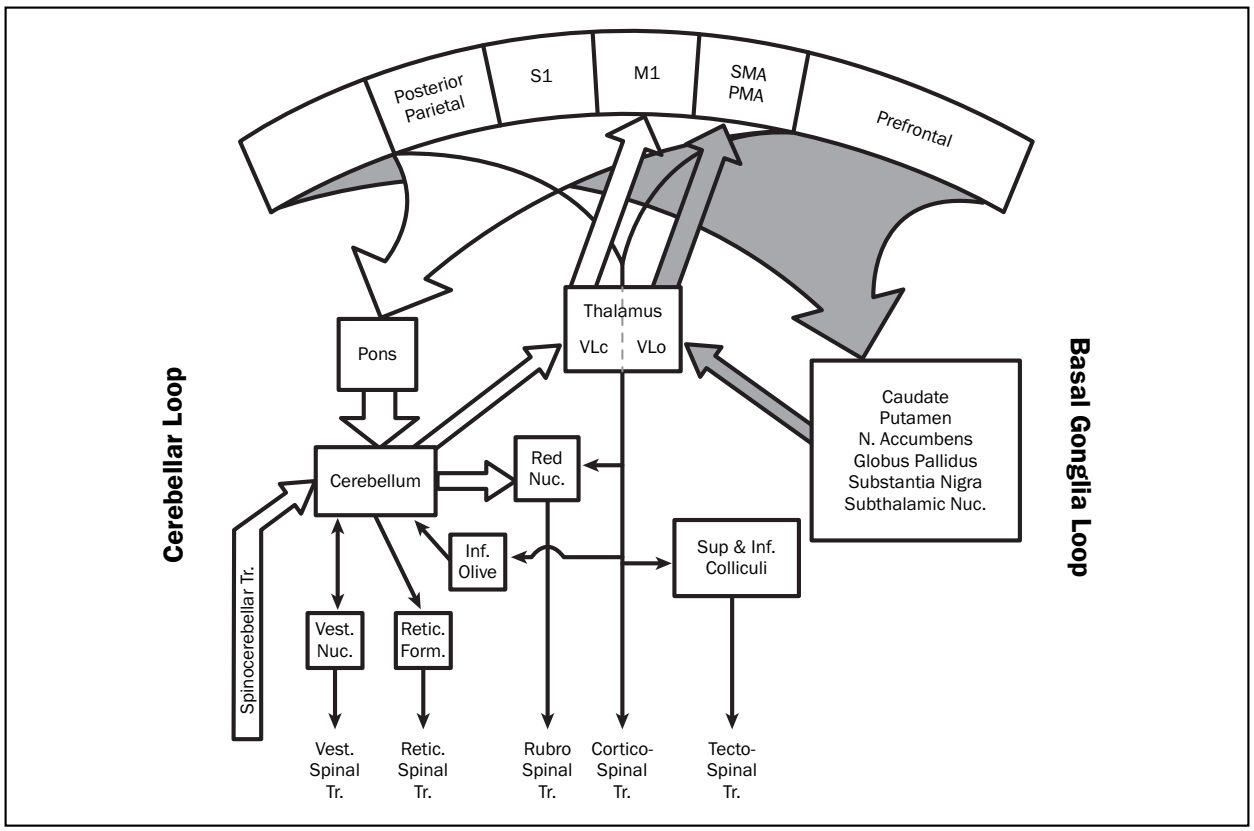


Figure 20.1

Schematic overview of supraspinal motor control systems. VLo, ventral lateral nucleus, pars oralis; VLc ventral lateral nucleus, pars caudalis. The superior and inferior colliculi of the tectum receive inputs for visual and auditory cortices as well as from sensorimotor cortex.

two systems process motor information related to planning and execution of movement, respectively, and will be covered in the last two chapters.

These different motor systems are hugely interconnected. Observe how the basal ganglia exclusively relates with the cerebral cortex, whereas the cerebellum influences and is influenced by both cortical and brainstem motor systems. The basal ganglia project via the thalamus to premotor and prefrontal cortex, so that they influence a large part of the frontal lobe, whereas the cerebellum has its major effects on the primary motor cortex. This suggests that the cerebellum is mostly concerned with the execution of movement, once the orders have been issued, whereas the basal ganglia are more concerned with planning, programming and higher cognitive aspects of movement. Note also, that cortical neurons projecting into the corticospinal or pyramidal tract are directly and indirectly subject to cerebellar and basal ganglia influence.

We will now explore in some detail the descending pathways, in which the cerebral cortex and associated brainstem structures influence the motor circuitry of the spinal cord.

Descending Motor Pathways

The most useful conceptualization of the organization of descending control comes from a series of experiments carried out more than 50 years ago by D. Lawrence and H. Kuypers. Their experiments, published in *Brain* (91:1-36, 1968), are worth reading in detail. Here we will summarize their approach and their findings.

The basic strategy was to interrupt a particular descending pathway and compare the resulting behavioral effects with that pathway's synaptic targets in the spinal cord. The projection patterns were determined by cutting the pathways and allowing the axons to degenerate distal to the cut (so-called Wallerian or anterograde degeneration—see appendix at end of Chapter 7 and this chapter). Using special stains, they could determine the location of the degenerating synaptic terminals.

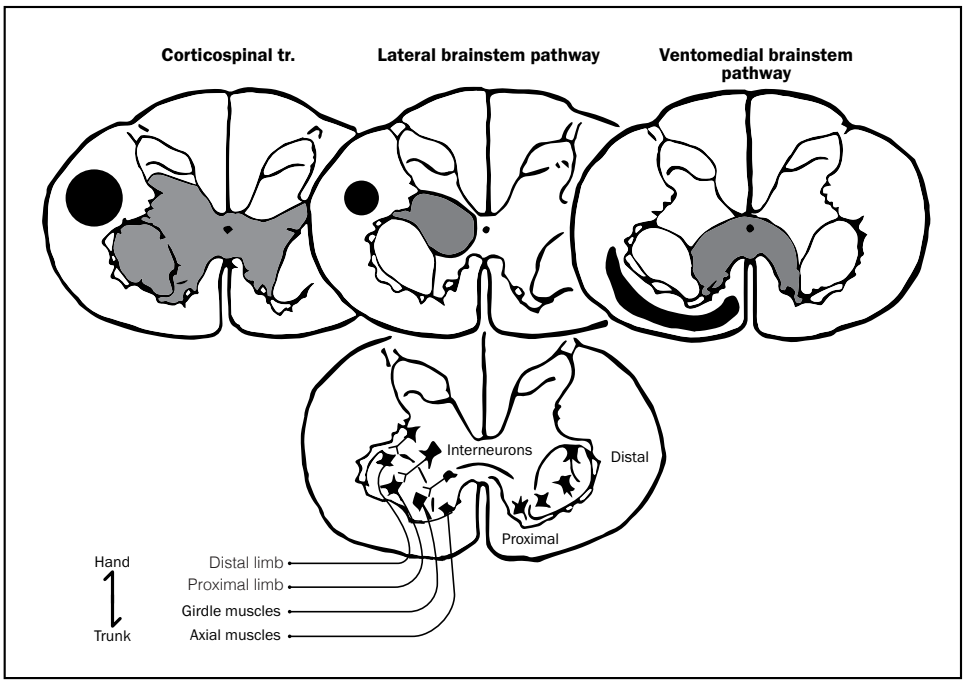


Figure 20.2

Descending motor control pathways. Sections above: location of degenerating synapses (gray) following unilateral section of the three major descending pathways. The approximate locations of the pathways at this level are in black. Section below: topography of motor neurons and interneurons. (Redrawn from Lawrence and Kuypers, ***Brain***, 91:1-36, 1968).

Lawrence and Kuypers knew that the alpha motor neurons of the spinal cord's anterior horn are arranged topographically: those innervating axial (trunk) and proximal limb muscles are located medially, while distal (hand) muscles are located laterally. The interneurons associated with these groupings are distributed in a similar fashion (Figure 20.2). Thus, Lawrence and Kuypers could compare how heavily a pathway projects onto this topographic motor circuitry with how intense the behavior deficits are when the pathway is interrupted.

Lawrence and Kuypers distinguished three descending systems in their study, one originating from cells of the cerebral cortex, the ***corticospinal tract***, and two arising from nuclei in the brainstem. Cells giving rise to the ***rubrospinal tract*** reside in the red nucleus (nucleus ruber) of the midbrain and, after crossing the midline, their axons travel in the lateral column of the spinal cord near those of the corticospinal tract.

Some texts group the corticospinal and rubrospinal tracts together as the "lateral pathway," but because the rubrospinal tract starts in the brainstem, it can be called the ***lateral brainstem pathway***, shown in Figure 20.2 and the paper of Lawrence and Kuypers. Also, following Lawrence and Kuypers, the ***ventromedial brainstem pathway*** comprises the ***reticulospinal tract***, the ***vestibulospinal tract*** and the ***tectospinal tract***, which travel near one another in the ventromedial part of the spinal cord. They are grouped together because it is very difficult to cut one of them without damaging the others.

Lawrence and Kuypers first studied the corticospinal tract, a pathway that arises from layer 5 of cortex. Specifically, the tract starts in pyramidal cells in several areas of the cerebral cortex adjacent to the central sulcus, and it descends through the internal capsule into the brainstem.

Once in the medulla the axons are organized into a discrete bundle called the ***pyramidal tract***, which can be observed on the ventral surface of the medulla where it forms the ***medullary pyramids***. The pyramidal tract fibers decussate in the caudal medulla and enter the lateral column of the spinal cord as the lateral corticospinal tract. A small component descends ipsilaterally in the anterior columns of the cord. Many axons of cortical origin also terminate in the brainstem and are called ***corticobulbar*** fibers.

After interrupting the corticospinal tract above its decussation in the medulla (see Figure 20.3), Lawrence and Kuypers found degenerating synaptic terminals distributed over the entire contralateral ventral horn, including some spread to the ipsilateral side (Figure 20.2). Most of the terminals were in the interneuron region but many were also located in the lateral zone on motor neurons innervating the hands.

Following interruption of the corticospinal fibers, individual limb and total body movements were preserved, but speed, agility and power were diminished in all muscle groups. Effects were seen on axial and proximal muscles, but the deficits were most pronounced on distal ones, particularly those moving the fingers and hand. This finding makes sense, as it has since been shown that distal muscles are innervated by motor neurons that receive monosynaptic excitatory input from the cortex, but the other pathways do not control these muscles. Skilled

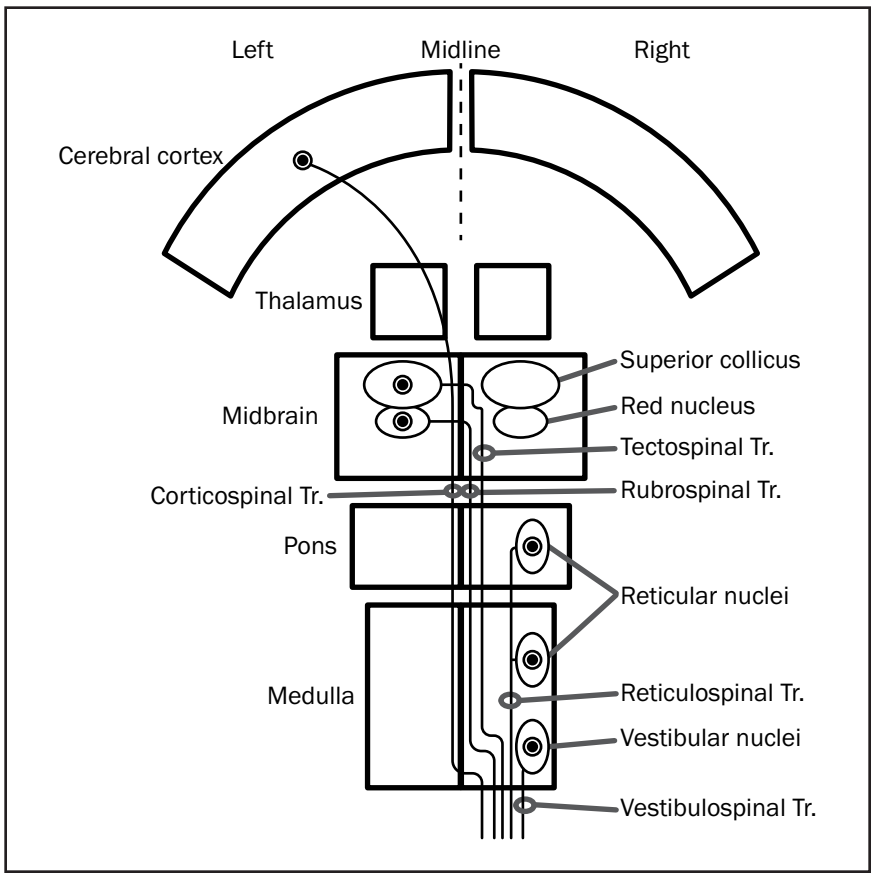


Figure 20.3

Decussation patterns of the descending motor pathways. Note that those pathways arising above the pons decussate, while those arising from the pons and medulla do not, at least for the greater part of their descending trajectories.

movements that require fractionated contraction of hand muscles were permanently lost (e.g. the precision grip of thumb against index finger). The ***sign of Babinski*** appeared (dorsiflexion and fanning of the toes on stroking the bottom of the foot).

The behavioral results indicated that this pathway enables relatively independent movements of muscles used in skilled activities such as piano playing and writing, and adds a measure of power and control to all motor acts. The widespread pattern of deficit, with its major appearance in hand control, was consistent with the distribution of corticospinal terminals (Figure 20.2).

Because the effects of corticospinal tract interruption were so widespread, Lawrence and Kuypers decided to study the effects of lesions of the two other pathways in the absence of corticospinal function. Therefore, they sectioned the corticospinal tract in some monkeys and allowed the animals to recover. They reasoned that motor control capacities remaining in these animals had to be mediated by either the ventromedial pathway or the rubrospinal tract. They then interrupted one or the other of these two pathways and observed the results.

The rubrospinal pathway arises mainly from the red nucleus of the midbrain, decussates in the midbrain near the nucleus (Figure 20.3) and innervates the spinal cord on the opposite side. Lawrence and Kuypers found its terminals distributed mostly in the lateral part of spinal gray, synapsing upon interneurons concerned with proximal and distal muscles, especially flexors (Figure 20.2). As might be expected from this pattern of termination, following interruption of the pathway there was severe impairment of independent distal extremity and hand movements. (Independent movements of the digits were already gone because of the previous corticospinal tract lesion). The results indicated that the lateral brainstem pathway has an important role in voluntary limb and hand movements such as reaching and grasping.

As noted above, Lawrence and Kuypers lumped together the reticulospinal, vestibulospinal and tectospinal tracts into the ventromedial brainstem pathway because these systems descend in that part of the spinal white matter. While the tectospinal pathway is crossed, the reticulospinal and vestibulospinal pathways descend ipsilaterally from their cells of origin to their targets in the spinal cord. However, as they terminate, many send some collaterals across the midline.

In the experiment of Lawrence and Kuypers, unilateral interruption of these pathways was followed by the appearance of bilateral degenerating synaptic terminals in medial parts of the spinal gray on spinal interneurons concerned with trunk, girdle, and proximal limb muscles (Figure 20.2). Behaviorally, this lesion caused the trunk to flex toward the side of the lesion (a flexion bias) because of the severe impairment of axial and proximal extremity movements. Independent distal extremity movements were preserved (except for the hand because of the existing corticospinal tract lesion). These observations indicated that the ventromedial brainstem pathway is important for control of balance, posture, and orienting movements using ***axial musculature***.

While it was difficult to parse out specific traits of the vestibulospinal, reticulospinal and tectospinal tracts, some is known about the effects of interrupting each path. For instance, the reticulospinal pathways have both excitatory and inhibitory effects on spinal motor neurons, mediated by the interneurons on which they terminate, but the major effect is excitatory to antigravity muscles which are mostly extensors. Lawrence and Kuypers found that a lesion of the pathway shifted the balance away from extensors and caused a flexion bias of the trunk, as mentioned above. Other studies have shown that the the tectospinal tract affects muscles of the neck and is primarily involved in controlling the position of the head.

Damage to the Descending Motor Tracts in Humans

Strokes and other pathological processes often damage the descending motor pathways in humans. Because these natural lesions are never localized exclusively to one pathway or

another, the resulting clinical deficits are more complex than those described by Lawrence and Kuypers. Clinicians have identified a number of characteristic signs and symptoms associated with damage to these ***long motor tracts*** that descend from brain to spinal cord.

When diagnosing long motor tracts pathologies, clinicians distinguish between two types of motor neuron. Cells with axons in the long motor tracts that do not directly cause movement are somewhat archaically called ***upper motor neurons***. The term ***lower motor neuron*** refers to an alpha motor neuron that directly innervates a muscle.

Damage to the lower motor neurons results in ***flaccid paralysis*** of the muscle, in which patients experience weakness and decreased muscle tone. ***Spastic paralysis***, on the other hand, can occur when lesions affect the central motor pathways. Patients with spastic paralysis often exhibit increased resting tone (***hypertonia***) and exaggerated myotatic reflexes (***hyperreflexia***). If both the arm and leg on one side of the body are paralyzed and exhibit hyperreflexia and hypertonia, the patient is said to have a ***spastic hemiplegia*** (-plegia referring to the complete paralysis). If the limbs are only weak, they are said to be ***paretic*** rather than paralyzed, and the patient would exhibit a spastic ***hemiparesis***. If the two legs, but not the arms are paralyzed, there is ***paraplegia***. If all four limbs are involved, the condition is called ***quadriplegia***. Palsy is another term often associated with weakness due to nervous system damage. Nowadays, this term is usually applied to weakness or paralysis of the facial muscles or extraocular muscles, e.g. VIIth nerve palsy (also called Bell's palsy), IIIrd nerve palsy, etc.

Another phenomenon called ***Clonus*** (G. ***klonos***, a tumult) is commonly seen in a spastic extremity. Clonus in the ankle, for example, would mean a rhythmical contraction of the calf muscles following a sudden passive dorsal flexion of the foot, the leg being semiflexed. This leads to flexion-extension beating movements. Based on the frequency of flexion and extension, ***beats*** are often counted to describe how quickly this effect diminishes or is 'damped'.

It is interesting that sectioning only the corticospinal tract does not produce spasticity in humans or monkeys, nor does a lesion restricted to the motor cortex. Instead there is a profound weakening or paralysis of the distal muscles and loss of precise control of the intrinsic muscles of the hand. Mild cases can appear as a 'clumsy hand syndrome.' Because the system is so interconnected, it is unclear which of the other descending pathways must be interrupted to produce spasticity. In any case, one can conclude that spasticity is not a clinical sign of pyramidal tract disease, but rather indicates a ***long tract sign***, or disturbance in some of the descending motor control pathways.

Decerebrate and Decorticate Rigidity

Extensive damage to the descending motor tracts can lead to conditions known as ***decorticate rigidity*** and ***decerebrate rigidity***. These are best understood in terms of the overall effects of damage at different levels of the brain on the muscle groups that are targets of the descending pathways. These conditions are summarized in highly schematic form in

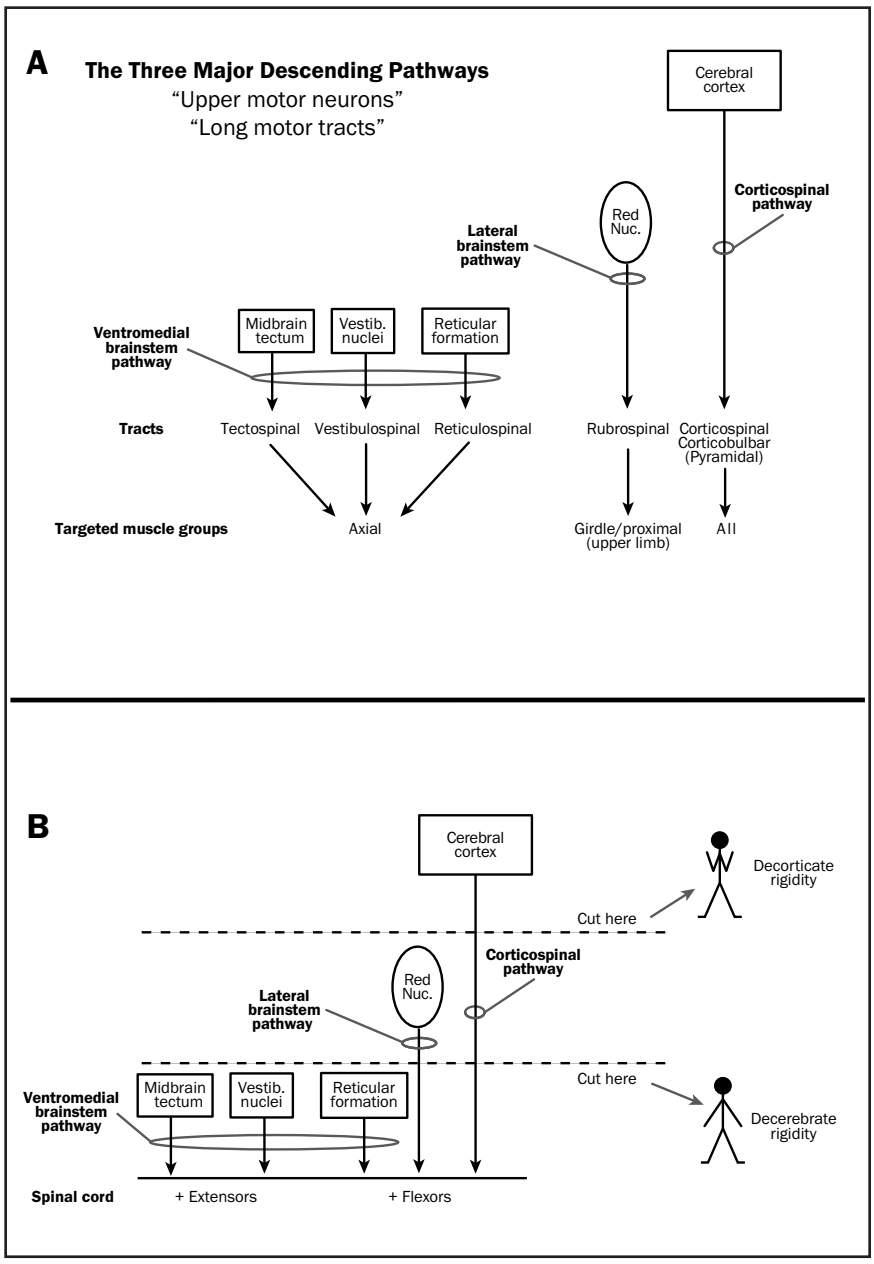


Figure 20.4

A. Schematic overview of the major descending motor pathways (long motor tracts, upper motor neurons). B. Effects of extensive damage to long motor tracts at different levels of the CNS. Damage below the midbrain produces the pattern called decerebrate rigidity. Damage above the midbrain leads to a pattern of decorticate rigidity.

Figure 20.4A, which ignores many subtleties about these projections.

Observe in Figure 20.4B that the ventromedial brainstem pathway's major influence is on extensor muscles, consistent with their role in opposing the pull of gravity on the body. The lateral brainstem pathway and the corticospinal pathway have their major influence on flexor muscles, which are important in moving the body or limbs from one place to another. Thus, when the brainstem is damaged below the red nucleus, the ventromedial brainstem pathway

acts unopposed and muscle tone in the arms and legs shifts toward extension, i.e. that of exaggerated anti-gravity posture. This pattern is called decerebrate rigidity.

In contrast, large cortical lesions or decortication produce tonic arm flexion and leg extension (Figure 20.4B). The difference in behavior deficit occurs because the cortical lesion leaves intact the lateral brainstem (rubrospinal) pathway, which provides a net flexor drive to the upper, but not the lower, limbs.

Technical Appendix: Tracing Neural Pathways

Anatomical Methods

When an axon is cut, the distal part (between the cut and the synaptic terminals) degenerates. This holds true for axons in both the CNS and the periphery. The degeneration occurs because the distal axon depends on the biochemical machinery of the neuron's soma for the production of structural proteins, transmitter molecules among other things. By interrupting the normal anterograde axonal transport mechanisms, the transection essentially starves the axon distal to the cut. The phenomenon was first described in detail by Waller, hence the name ***Wallerian degeneration.***

One can take advantage of the machinery behind this degeneration process to localize cells. During a short window of time after a cut, the degenerating synapses will take up various stains that can trace the location of cells. One of the most famous of these stains was developed by Walle Nauta, and is called the Nauta stain. This method is no longer used to trace pathways.

Modern anatomical track-tracing methods use a variety of chemical and biological tracers that are carried along axons by anterograde or retrograde transport mechanisms. Injection of a retrograde tracer into a region of nerve terminals results in labeling of the cell bodies from which the terminating axons arose. Injection of an anterograde tracer results in labeling of the axon terminals of those cells (Figure 20.5). One of the most widely used tracers is ***horseradish peroxidase***, which is transported both anterogradely and retrogradely.

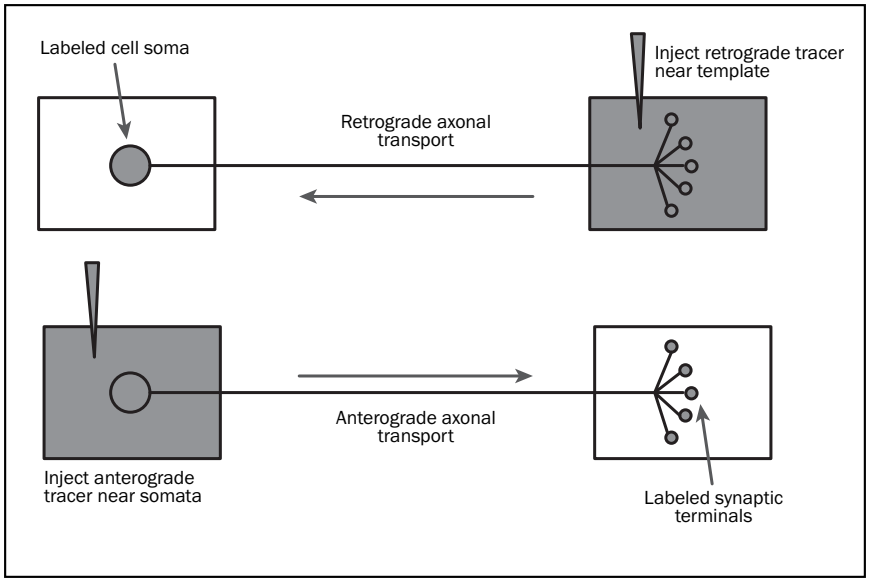


Figure 20.5

Pathway or track tracing with substances carried by anterograde and retrograde axonal transport mechanisms.

Electrophysiological Methods: Antidromic Activation

In addition to tracers, one can also use several electrophysiological methods to establish the connectivity of different brain regions. Here we will describe the method of antidromic activation and the collision test. Other methods will be discussed in the following chapter.

Neurons often send their axons long distances, and knowing the target of these projects informs our analysis of neural systems. For example, one might wish to show that a neuron in primary motor cortex sends its axon to the spinal cord. One way to assess this is to activate the axon, say, in the spinal cord while recording from the cell body. Axons normally conduct action potentials away from the cell body, in the ***orthodromic*** direction, but suprathreshold electrical stimulation of an axon produces action potentials traveling in both directions away from the stimulus site, i.e. orthodromically and ***antidromically*** (Figure 20.6A). If one can demonstrate that an antidromic action potential reaches the cell body from the distant site of electrical stimulation, then the cell has to have an axon at that site.

Because electrical stimulation may also excite other cells or axons that eventually synapse on neurons at the recording site, one must be able to distinguish antidromic action potentials

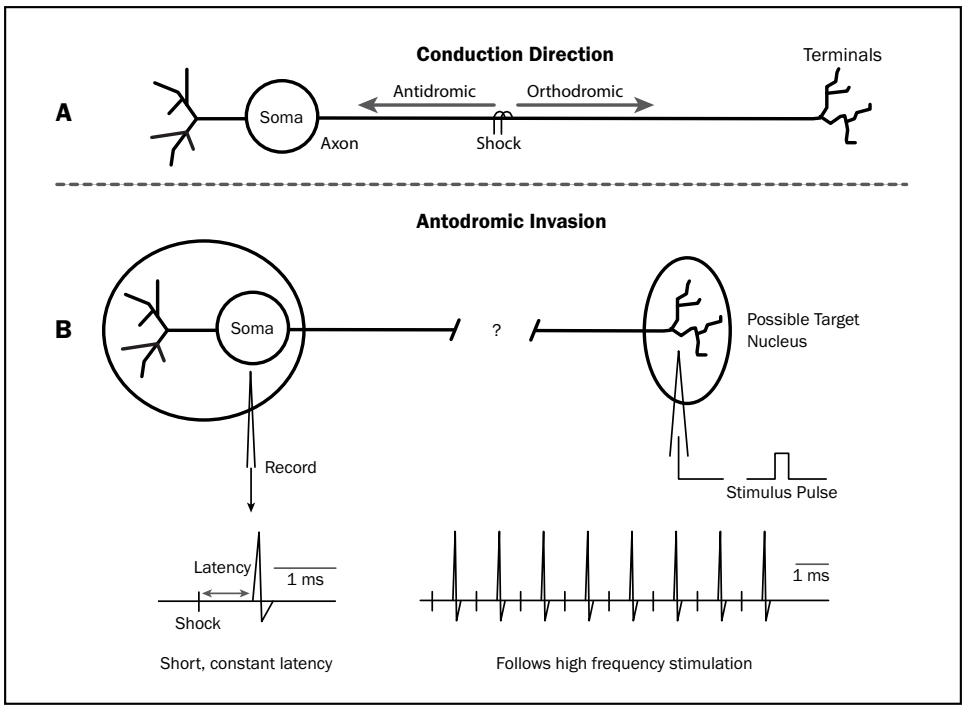


Figure 20.6

Using antidromic invasion to determine if a cell soma has an axon in a distant location. (see text). A. Anti- vs orthodromic conduction. B. Two tests of antidromicity: short, fixed latency; high frequency following of the antidromic stimulus. (see text).

from synaptically evoked spikes. A number of criteria are used for this. First, the latency from electrical stimulation to the appearance of an antidromic spike will be very constant because it is not subject to the normal latency fluctuations inherent in a synaptic relay. The latency will also be so short that one can usually exclude a synaptic delay.

Second, during high frequency electrical stimulation (e.g. 300-500 Hz) the antidromic action potential will faithfully occur with each stimulus because the rate at which the axon will fire is limited only by the refractory period of less than a millisecond (Figure 20.6B). Synaptically driven spikes, on the other hand, do not 'follow' such stimulation because repeated activation of the synapse at short intervals results in transmitter depletion.

While these criteria can help identify an antidromically evoked action potential, but they are not foolproof. Some researchers therefore use another test is used that is considered the "gold standard" for proving antidromic invasion: the ***collision test***.

Because axons experience a period of electrical inexcitability (the absolutely refractory period) just after the passage of an action potential, one can block the passage of an antidromic action potential by having it collide with one traveling in the orthodromic direction (Figure 20.7). If the two action potentials do indeed collide, it the two stimulation spots reside on the same cell.

When doing the collision test, experimenters can place a recording electrode at the soma and a stimulating electrode at the axonal location that they believe is connected to the soma. When action potentials are recorded at the cell body, the apparatus automatically triggers stimulation of action potentials at the distant site. The apparatus can vary the delay between the action potentials at the soma and the stimulation of action potentials at the putative axon. The orthodromic spike appears on the oscilloscope.

Trace 1 in Figure 20.7 shows how this experiment appears on the oscilloscope. The arrowhead marks the time of electrical stimulation of the distant axon, i.e. the antidromic (A/D) stimulus. In trace 2, the A/D stimulus has been moved forward in time by decreasing the delay between the orthodromic spike and the activation of the stimulator. In trace 3, the electrically evoked spike begins to 'fail,', showing up on some traces and not others. This disappearance of the antidromic spike occurs because the orthodromic spike has traveled past the A/D stimulus site, so the axon is still relatively refractory. Some stimuli are effective, others are not. Moving the A/D stimulus even closer in time to the onset of the orthodromic spike results in collision every time.

Another way to carry out this test is to determine the critical interval in the timing of the spikes recorded at the cell body. This measurement takes advantage of the fact that a collision will occur only if the antidromic spike is initiated before the orthodromic spike reaches the distant stimulus site. If the orthodromic spike has already passed the site of stimulation, the antidromic spike can travel up the axon to the cell body without collision.

In order to understand this concept, let's assume that the distant stimulus evokes an antidromic spike that takes 4 ms to travel the distance from the stimulus site to the cell body.

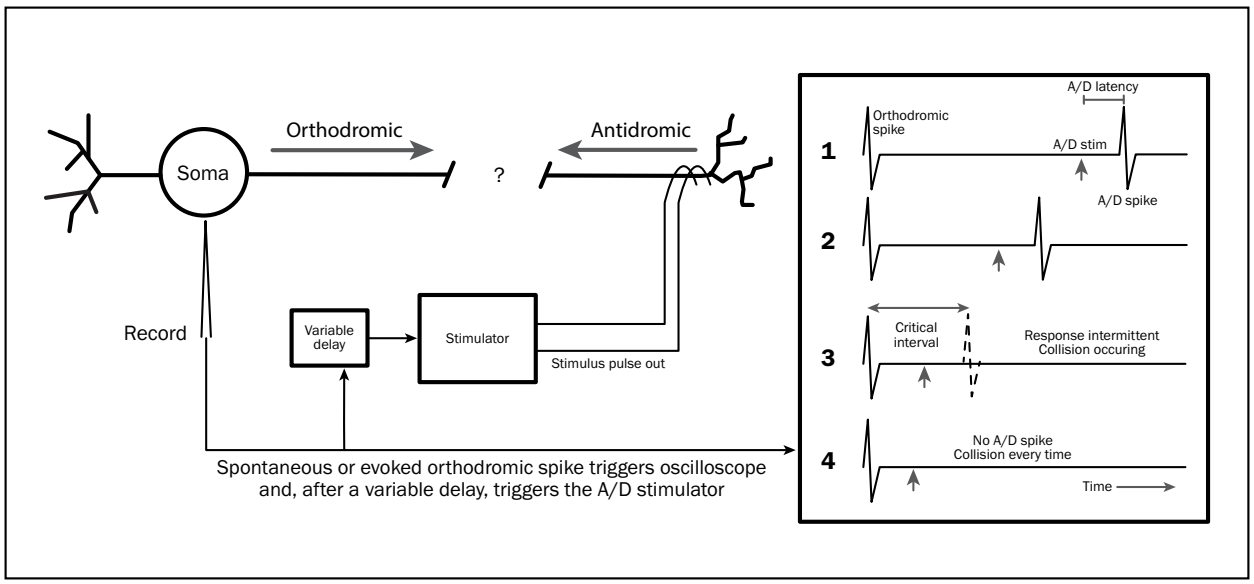


Figure 20.7

The collision test. A/D - antidromic. See text for details.

Because the speed of conduction is the same for both the orthodromic and the antidromic spikes, it also takes the orthodromic spike 4 ms to go from the cell body to the distant site of stimulation. So, if an antidromic spike is evoked about 4 ms after the orthodromic spike is recorded at the cell body, the orthodromic spike will have gone by, the refractory period will have passed. There will be no collision, and the antidromic spike will be recorded at the cell body 4 ms later. The shortest interval between the ortho- and antidromic spikes *observable at the cell body* will therefore be 8 ms, i.e. time for the orthodromic spike to go all the way past the distant electrode plus the time for the antidromic spike to travel from that electrode to the cell body (trace 3).

This time period is called the ***critical interval*** and is exactly equal to twice the antidromic latency plus the refractory period of the axon. The refractory period is very short, tenths of a millisecond perhaps, so a good estimate of the critical interval is twice the antidromic latency. Decreasing the delay between the appearance of the orthodromic spike and the initiation of the antidromic spike will result in complete disappearance of the antidromic spike as collision occurs every time (trace 4).

It is important to realize that one does not directly observe the collision taking place between the orthodromic and antidromic spikes. This occurs somewhere along the length of the axon. What one sees is the failure of the antidromic spike to appear at the recording site when the interval between the initiation of orthodromic and antidromic spikes is sufficiently short.

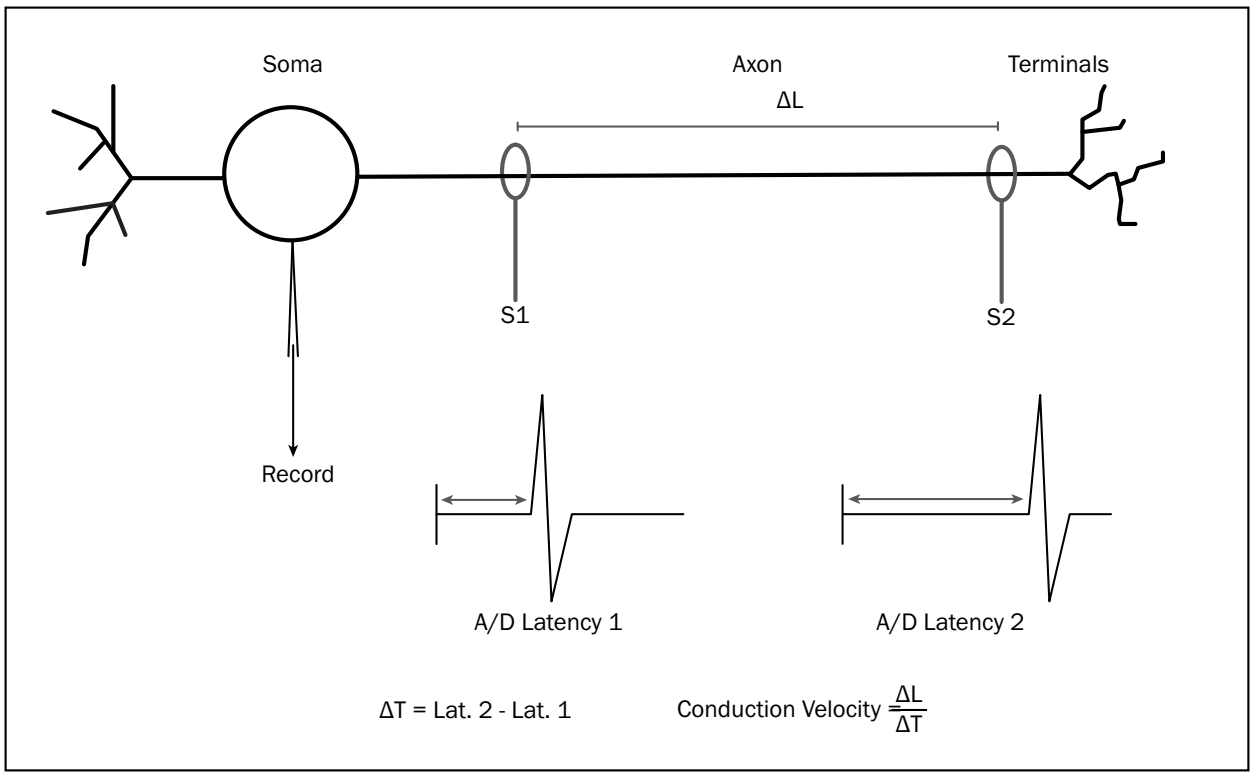


Figure 20.8

Determining conduction velocity from antidromic latencies. (see text).

Using Antidromic Stimulation to Estimate Axonal Conduction Velocity

In many cases, cells serving different functions have axons of different size. Since conduction velocity varies with axon diameter, the conduction velocity of the axon can be a useful indicator of cell class. For example, alpha motor neurons have large, rapidly conducting axons, while gamma motor neurons have fine, slowly conducting axons.

There are several ways of estimating axonal conduction velocity, but one of the most commonly used is illustrated in Figure 20.8. With a recording electrode in or near the cell body, two stimulating electrodes, S1 and S2, are positioned at different distances away. Each is used to evoke an antidromic spike in the recorded cell. The latencies from the stimulus to the appearance of the spike at the cell's soma will depend on the distance from the electrode to the cell. If this length is known, it can be used to estimate conduction velocity.

Because it is often hard to estimate the actual conduction distance to the soma, experimenters often use another approach. With two electrodes positioned a known distance apart (L), antidromic latencies are determined for stimulation at each electrode. The latency from S2 is the sum of the time it takes the spike to travel from S2 to S1 and the time from S1 to the cell. Thus, the difference between the two latencies is the time it took the action potential to cover the distance from S2 to S1 and the conduction velocity of that stretch of axon is L/(Lat. 2 - Lat. 1). The units are conventionally meters/second.

Terms and Techniques

ipsilateral	proximal	contralateral
distal	decussation	somatotopy
synergist muscles	axial muscles	antagonist muscles
girdle muscles	proximal muscles	distal muscles
ventromedial brainstem pathway	lateral brainstem pathway	corticospinal tract
vestibulospinal tract	reticulospinal tract	pyramidal tract
corticobulbar fibers	tectospinal tract	rubrospinal tract
upper motor neuron	lower motor neuron	medullary pyramids
hemiparesis	paraparesis	quadriparesis
hemiplegia	paraplegia	quadriplegia
Sign of Babinski	clonus	precision grip
spasticity	hyperreflexia	hypertonia
decerebrate rigidity	decorticate rigidity	propriospinal
supraspinal	basal ganglia loop	cerebellar loop
Wallerian degeneration	Nauta stain	anterograde transport
retrograde transport	horseradish peroxidase	orthodromic
antidromic	collision test	critical interval
antidromic latency	medullary pyramid	internal capsule

Sensory-motor Cortex 21

Table of Contents

Cortical Areas Involved in Motor Control

According to one long-standing tradition in neuroscience, the mammalian cerebral cortex can be subdivided into sensory, motor and association areas. This notion has its analog in the system of spinal reflexes, in which the sensory afferents receive information, transmit it to the association interneurons, which then process the information and relay it to the motor neurons.

While this tripartite system vaguely explains cortical circuitry, the inadequacy of such a scheme will become apparent as we explore the connectivity and workings of the supraspinal motor mechanisms.

We may begin by asking how the three-part system was conceived in the first place. When studies of the nervous system began, one of the few available tools were ablation/lesion studies. Experimenters could observe the effects of removing a part of the brain on the animal's behavior. A cortical area was labeled 'motor' when a restricted ablation significantly impaired some movements but did not seem to reduce the animal's ability to respond in some way to various sensory stimuli. In contrast, sensory areas were identified when the lesion did not impair movement, but the animals failed to respond to stimulation of one modality or another. Ablation of other regions seemed to have little sensory or motor effect, so these were called association areas, a default non-sensorimotor designation. (The ablation/lesion method is discussed in the appendix to this chapter).

The advent of electrical stimulation and recording methods produced results that reinforced this scheme. Thus, motor areas were defined operationally as those sites yielding movements at the lowest intensities of electrical stimulation. Sensory areas were identified by their production of ***evoked field potentials*** following application of low intensity sensory stimuli that did not

greatly affect other cortical areas (see appendix at end of this chapter). And finally, recording and stimulation of association areas generally produced little in the way of interpretable results. It should also be noted that much of this work was carried out under deep barbiturate anesthesia, which inevitably reduced the excitability of the brain's circuitry.

Even while using these crude experimental tools, evidence soon emerged that the sensory and motor areas of the cortex were not always segregated spatially. For example, the primary motor cortex of the rat, defined by the methods describe above, was found to overlap spatially with the hindlimb area of the primary somatic sensory cortex and some of the forelimb area as well. The overlap appears to be even more complete in the possum and hedgehog. When the activity of single cells in the primary motor cortex of monkeys was studied, the cells were found to have sensory receptive fields on the skin and comparable results were obtained on a rare occasion when the activity of single cells was recorded in human primary motor cortex during neurosurgery.

These findings demonstrated that, despite the impression given by the labels placed on different gyri adjacent to the central (Rolandic) sulcus, the distinction between sensory and motor processes is blurred in the CNS. Experimenters found that, among the various species studied, the spatial segregation of sensory and motor cortical areas becomes more pronounced as the dexterity and independent control associated with a given body part increase.

This blurring of function is demonstrated when looking at the contributions of different cortical areas to the pathways to the corticospinal and corticobulbar tracts (see Figure 18.1). These descending projections arise from layer V pyramidal cells in Brodmann's areas 4 and 6, but primary somatic sensory cortex (S1 or areas 3, 1 and 2) and areas 5 and 7 of the posterior parietal lobe also contribute to the tracts. In fact, while 60% of the corticospinal fibers arise from frontal cortex, about 40% come from parietal cortex. Because all of these areas, even those we traditionally consider "sensory" in nature, contribute to motor control, it is appropriate to refer to this *peri-Rolandic region* as *sensory-motor cortex*. We have also noted in the previous chapter that this region receives important projections from the basal ganglia and cerebellum via the ventral lateral nucleus of the thalamus (Figure 18.1).

Organization and Function of Primary Motor Cortex

M1, the primary motor cortex, is that region of the cerebral cortex where movement of the contralateral musculature can be evoked most easily by electrical stimulation (i.e. with the weakest stimuli), reflecting its especially close ties with the somatic musculature. In humans it is found on the *precentral gyrus* and is topographically organized with the foot and lower leg represented on the medial surface of the hemisphere and the rest of the body on the lateral surface. Because of its *somatotopic organization*, small lesions of this region can produce effects limited to a particular part of the body. The apparent simplicity of this map is deceptive for there is considerable uncertainty about what exactly is represented at a particular point on

the topographic map. Much ink has been spilled over the question of whether the discharge of cells at a given locus encodes a specific movement or activation of a specific muscle or something else.

This debate continues, but it is largely being displaced by knowledge that the transformations between cortical activity and movement are more complex and contingent than previously believed. For example, it is now known that the terminals of a single corticospinal axon may be distributed to the spinal motor neuron pools of several muscles located in several adjacent spinal segments. Obviously, this is inconsistent with the idea that a given cortical locus represents a single muscle. Also, whereas very weak stimulation of a given cortical site may cause contraction in a single muscle, that same muscle may be activated by similar stimulation of other areas.

When the activity of single cortical neurons is correlated with muscle action during voluntary movements, it becomes clear that the relationship of a cell's discharge to the contraction of a given muscle can be variable (Figure 21.1). This and other evidence indicates that the link between cortical and muscle activity is contingent on contexts that are not easily controlled by the experimenter. Suffice to say that the functional architecture of the motor map in M1 is not fully understood, and the 'homunculus' used to represent that map is a caricature that gives only the barest outline of the complex pattern of connectivity.

As discussed in the previous chapter, some axons of the corticospinal tract have synaptic terminals in the lateral region of the ventral horn of the spinal cord where the motor neuron pools of the distal muscles also sit. If these terminals are located on alpha motor neurons,

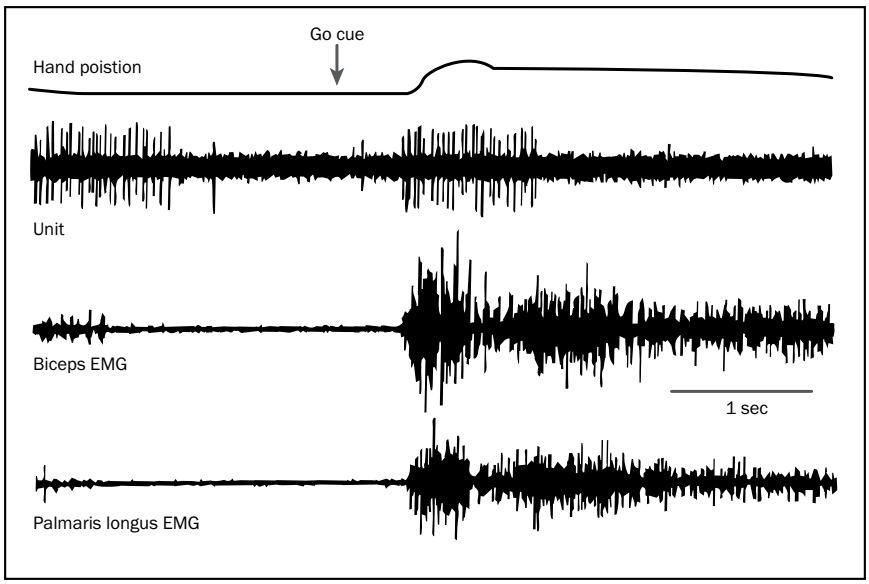


Figure 21.1

Discharge of a unit in monkey M1 during voluntary movement of the hand. Traces from the top down: hand position, unit discharge, biceps EMG, Palmaris longus EMG. The animal was trained to make the movement on the Go Cue. Note that the unit could fire actively during a little or a lot of muscle activity. Courtesy of John Donoghue.

they could give the cortex highly specific control over individual muscles and, as shown by Lawrence and Kuypers, these connections seem to be necessary for skilled, independent movements of the fingers and hands (see Figure 18.2).

To test the connectivity of cortex and motor neurons, Phillips and Porter, among others, have stimulated the corticospinal axons electrically while recording intracellularly from motor neurons that could be activated ***antidromically*** from specific muscles of the hand. The volley of impulses evoked in the corticospinal fibers produced EPSPs in the hand motor neurons that had latencies so short that only one synapse could have intervened between the descending axon and the motor neuron. This finding clearly established the existence of a cortico-motoneuronal projection capable of controlling certain distal musculature.

Using a technique called ***spike triggered averaging***, other investigators have been able to correlate the occurrence of an action potential in a specific cortical neuron with the appearance of an EPSP in a motor neuron in the hand area of the ventral horn (see appendix to this chapter). Although these findings show that the cortex can exert direct control over certain muscles, they do not prove that a given cortical cell or a given cortical locus is uniquely and exclusively related to a particular muscle.

Investigators have asked if the discharge of neurons in M1 is linked to one or more of the possible parameters that describe the movement of a limb or joint, such as direction, force and final position. The pioneering studies of this type were done by Evarts, who trained monkeys to flex and extend their wrists against a load of variable magnitude and direction (Figure 21.2).

Recording from antidromically identified corticospinal tract neurons, Evarts observed no strong correlation between wrist position and discharge, but there was a correlation between the cell's activity and the amount of force needed to flex or extend the wrist. The neuron in Figure 21.2 was clearly related to the act of flexing the wrist because it fired vigorously as the wrist moved from extension to flexion (middle trace). When flexion was made more difficult by applying a load opposing flexion (upper trace), the neuron fired even more intensely as the animal flexed the wrist. When the load pulled in the direction of flexion, making extension more difficult, the neuron was virtually silent (lower trace). In this situation the muscles did not have to work to flex the wrist. The discharge frequency also seemed to be correlated with the rate at which force was applied.

Evarts' results indicate that the discharge of neurons in M1 that project into the corticospinal tract does not specify the final position of a particular joint but rather encodes certain aspects of the muscle force needed to place the joint in that position. In interpreting results such as those of Evarts, it is useful to remember that to change the angle of a particular joint (e.g. the wrist) the brain must specify a pattern of activity in muscles acting as agonists and antagonists around that joint and perhaps other muscles acting around related joints (e.g. the elbow and shoulder).

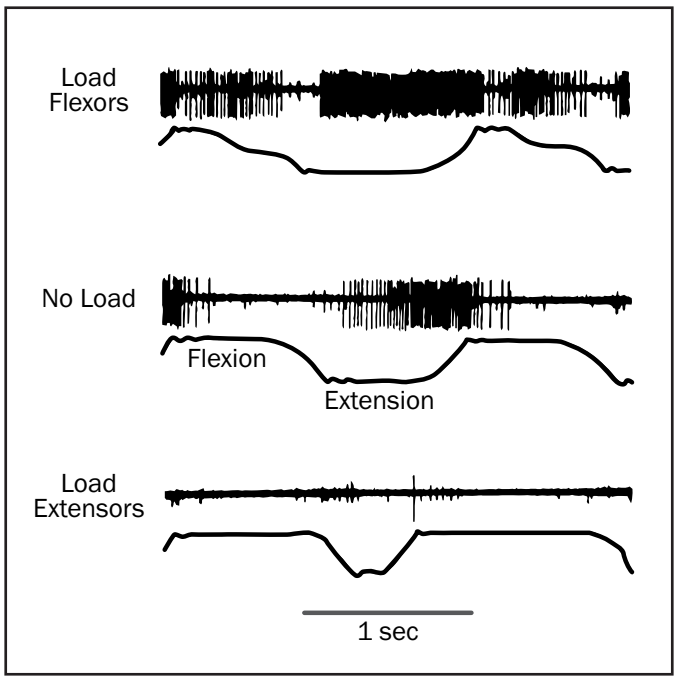


Figure 21.2

Set up for extracellular recordings. Behavior of a corticospinal neuron in the arm area of MI. The upper trace in each panel is the response of a neuron activated antidromically from the pyramidal tract. The lower trace in each set shows the position of the wrist with flexion upward. In the middle panel there is no load. The load opposes flexion in the top panel and extension in the bottom panel. Redrawn from Evarts, *J. Neurophysiol.* 31:14-27, 2168.

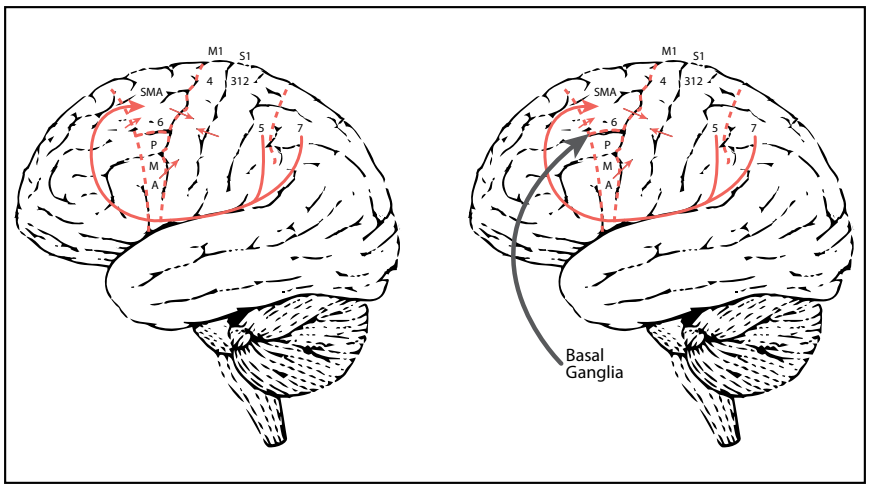


Figure 21.3

Key inputs to sensory-motor cortex. SMA-supplementary motor area; PMA-premotor area.

Motor Cortical Areas Rostral to M1

In addition to giving rise to part of the corticospinal tract, Brodmann's area 6 is an important source of input to primary motor cortex (Figure 21.3, left). Immediately rostral to M1 in the frontal lobe, this area has several distinct functional areas identified on the basis of their

projections and the behavior of their neurons during motor performance. Two of the most important are the ***Supplementary Motor Area*** (SMA) and the ***Premotor Area*** (PMA). We will not describe here the distinctions between these two regions or the other components of the areas rostral to M1 that are involved in motor control, but Table 21.1 compares certain characteristics of area 6 with those of the primary motor cortex.

All of these features suggest that area 6 is indeed "upstream" of M1 in that it represents a higher level of control than M1.

	Area 6	**Area 4, M1**
Lesions	• Complex motor acts defective • No weakness or paralysis	• Weakness or paralysis of limited muscle groups
Electrical stimulation	• High electrical thresholds • Complex movements evoked across several joints	• Low electrical threshold • Movements limited to small groups of muscles
Recorded neural activity	• Associated with 'motor set' • Contingent on whether generated by a stimulus or internally	• Timing highly correlated with muscle contraction. • Few muscle groups or joints involved • Clear relation to force
Blood flow increase	• 'Mental' movement alone	• Actual movement

Table 21.1

Comparison of area 6 and area 4, the primary motor cortex.

Further evidence of the 'higher' position of the SMA in motor control was provided by an experiment by Per Roland and colleagues (Figure 21.4). These experimenters measured blood flow in the cerebral cortex using the PET method and had the subject do various kinds of motor tasks. Voluntary finger flexion alone caused an increase in activity (i.e. blood flow) in the primary motor cortex. Sequential movement of the fingers activated both M1 and SMA. Of particular note is the finding that SMA was activated even when such movements were simply imagined but not actually performed (Figure 21.4).

PET demonstrates brain activity by measuring blood flow to a region of cortex, which is regulated by the metabolic needs of that area and rises and falls with its activity. In other words, the blood flow to cortex is ***autoregulated***. A related technique uses the method of

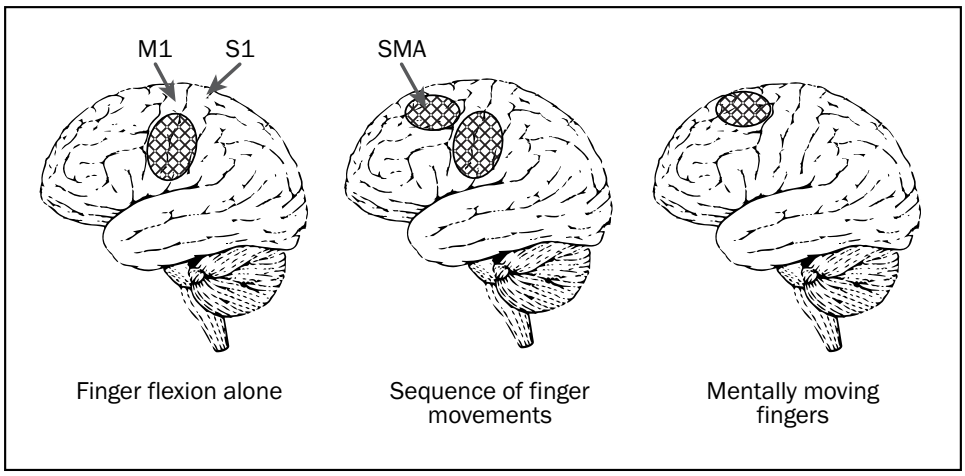


Figure 21.4

Summary of a PET study by Per Roland and colleagues. During active flexion of contralateral fingers, the location of the PET signals (stippled oval) indicated activity in the hand area of M1 and S1. When the subject executed a sequence of finger movements, the Supplementary Motor Area became active. When the subject was asked to imagine the sequence of movements, but not make them, only SMA was active.

magnetic resonance imaging or MRI. This method is enormously useful when looking for damaged regions of brain, but it can also be used for research. When doing MRI, one usually acquires one image when the subject is doing nothing (the control) and another when a task is being performed (the experiment). Subtracting the control image from the experimental image yields a measure of those areas selectively active during the task. This approach is called ***functional*** MRI or fMRI.

Certain areas of the parietal somatic sensory cortex are included in the sensorimotor region not only because they send some axons into the corticospinal tract, but also because they project heavily to motor areas of the frontal lobe. These parietal areas include the primary somatic sensory cortex (S1, Brodmann's areas 3, 1, 2 on the ***postcentral gyrus***) and the posterior parietal cortex (areas 5 and 7). These areas process cutaneous (skin) and deep (muscle) receptor information and relay it to M1 either directly or via area 6 (Figure 21.3).

The corticobulbar and corticospinal projections of the parietal areas particularly target sensory relay structures, such as the dorsal column nuclei and the posterior horn of the spinal cord, where they may control or modify the sensory information rising to the cortex. It is suspected that they have a 'gating' function because it is known that ascending sensory information can be gated into or out of motor areas. As we have discussed elsewhere, the parietal cortex forms part of the so-called ***dorsal stream***, a complex pathway that is crucial to the integration of spatial information from a number of sensory modalities.

Although antidromic activation can reveal the presence of a cell's axon at a site distant from the cell body, it does not necessarily prove that the axon makes synaptic contacts near the electrode. The electrode could be located anywhere along the length of the axon and excite it antidromically, so the synaptic terminals could be far away from the site of stimulation. Two other techniques have been used extensively to demonstrate functional connectivities and they will be discussed here.

Evoked Field Potentials

When large numbers of neurons are activated simultaneously, large currents flow in the extracellular space. As in the case of extracellular single cell recording, electrodes can detect the voltages produced by these currents. The resulting potentials are called ***evoked field potentials*** or, simply, ***evoked potentials***. Electrodes suitable for recording evoked field potentials can be large relative to those used in single cell recording and are sometimes referred to as gross electrodes or macro-electrodes (Figure 21.5).

The evoked field potentials may be very low amplitude and are virtually concealed by on-going activity in the local tissue, which produces somewhat random currents in the extracellular space. A common technique for revealing the evoked potential is to sum the recordings from a sequence of stimulus presentations in a process analogous to that used in averaging the responses of single cells to repeated trials (see Figure 1.4). The component of the record that results from the stimulation (i.e. the 'signal') is of constant form and time-locked to the stimulus,

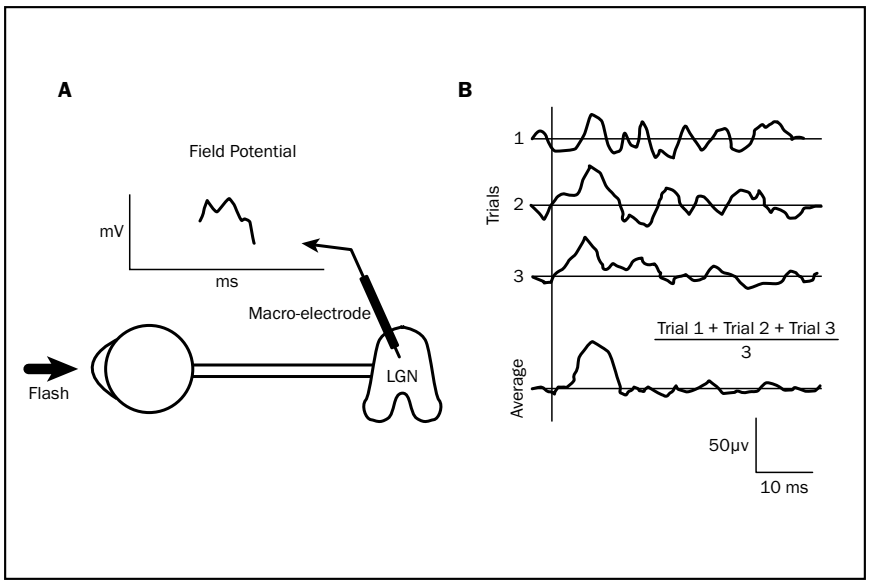


Figure 21.5

Evoked field potentials and signal averaging. A. Schematic of a field potential recorded from the lateral geniculate nucleus in response to light flashed in the eye. B. Averaging of 3 trials increases the signal-to-noise ratio in the recording.

but the random "spontaneous" voltage fluctuations (the 'noise') are as likely to be positive as negative at any instant. Thus, when the traces are summed, the random fluctuations tend to cancel while the successive evoked potentials add incrementally.

Summing the traces and then dividing by the number of stimulus presentations yields the average evoked response (Figure 21.5B). This process essentially increases the signal-to-noise ratio of the mesurement by averaging out the random, time-varying components of the signal.

Spike-Triggered Averaging

The technique of signal averaging can also be applied at the single-cell level to detect weak synaptic connections between two cells. Suppose we want to know if a given neuron in the primary motor cortex provides an excitatory input to a given motor neuron in the anterior horn of the spinal cord. For this we can use spike-triggered averaging.

In this method, experimenters allow a spontaneously occurring or stimulus-evoked spike recorded from the cortical neuron to trigger a device that averages any changes in the membrane potential of the spinal motor neuron (Figure 21.6). If there is a monosynaptic connection between the cortical cell and the motor neuron, even if it is weak, the synaptic potential should always occur with a fixed latency after the cortical spike and should be revealed by the averaging process, just as in the case of evoked field potentials.

The conduction velocity of the axon can be also estimated if the distance separating the two cells is known. If multiple synaptic relays are interposed between the two cells, the synaptic potentials in the recorded motor neuron become undetectable because they are too spread

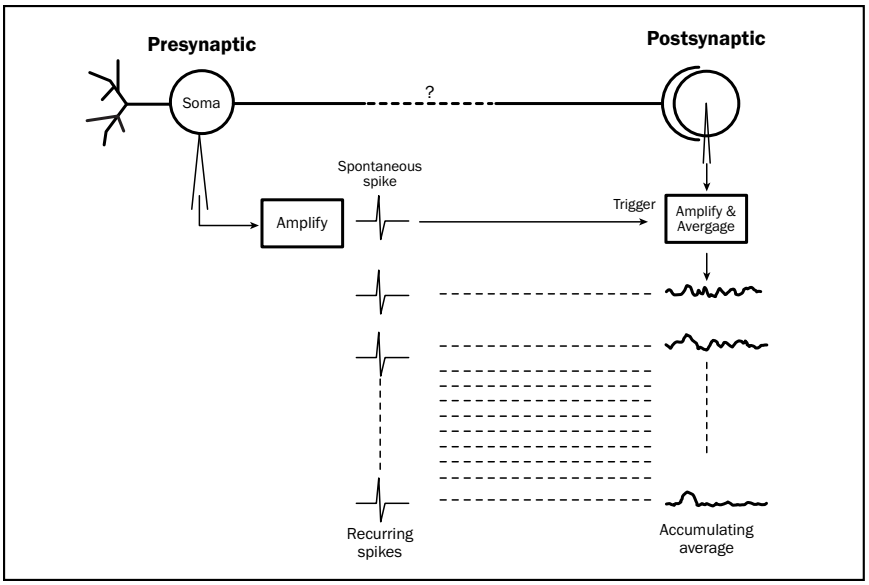


Figure 21.6

The method of spike-triggered averaging used to detect synaptic connections between two cells separated by a long distance. See text for details.

out in time.

The Ablation/Lesion Method

As mentioned earlier this chapter, experimenters first assessed the function of cortical areas by removing that area and observing the result on the behavior of the experimental animal. This method is easy to apply and has been used extensively, but it has major shortcomings. Isn't it curious that the ablation technique identifies a structure one wishes to understand, then removes it and throws it away? In other words, the behavior one observes after an ablation or lesion are the capacities of what is left, placing major constraints on the kinds of interpretations that can be made.

This limitation is demonstrated clearest in the case of an ablation that has no effect. When there is no observable deficit after ablation, one can only conclude that the ablated or lesioned structure was not *necessary* to the function observed. One cannot conclude that the structure, when present, plays no role in the function.

The limits of lesion studies were illustrated in the study of the superior colliculus and a region of frontal cortex called the ***frontal eye fields***. Both of these structures had been implicated in the control of eye movements on the basis of recording and stimulation studies, but ablation of either structure alone had little, if any, lasting effect on the ability of the monkey to move its eyes. However, removal of both structures rendered the animal incapable of making certain eye movements. As with many other systems, this task was supported by redundant mechanisms, which can substitute for each other. The presence of such redundancy makes the interpretation of ablation/lesion experiments very difficult.

Terms and Techniques

VL thalamic nucleus	internal capsule	reticular formation
SMA, PMA, M1, S1	prefrontal cortex	posterior parietal cortex
area 4	area 6	postcentral gyrus
peri-Rolandic cortex	motor strip	precentral gyrus
PET	fMRI	autoregulation
spike-triggered averaging	evoked field potentials	ablation/lesion method

Basal Ganglia and Motor Control

22

Table of Contents

The basal ganglia are among the most mysterious and confusing structures in the brain. This is a large area of the brain that's poorly understood. The list of structures included in the basal ganglia has undergone various additions and deletions over the years, which adds to the confusion. Also, the word 'ganglion' is usually restricted to structures located outside the CNS, so even the term 'basal ganglia' is anomalous. Now, these structures are often referred to as 'basal nuclei.'

The original grouping of these nuclei was based on the fact that they have a common developmental origin from the telencephalic vesicles at the rostral end of the neural tube. Early in the development of the brain, the ventral part of the vesicle wall thickens into a mass of tissue called the basal telencephalon and it is this region that gives rise to several of the major nuclei originally designated as the basal ganglia. These include the ***caudate nucleus, putamen, nucleus accumbens, globus pallidus (pars interna and pars externa), claustrum*** and ***amygdala*** (Figure 22.1).

The caudate and putamen constitute a single mass of tissue that is split by the axons of the internal capsule. The two nuclei remain connected by bands of cells that appear as striations across the internal capsule (visible in Figure 22.1-bottom panel). Because of this striated appearance, the caudate and putamen are together called the ***neostriatum***, or simply, the ***striatum***.

Continuous with and ventral to these two nuclei is the ***nucleus accumbens***, which is sometimes called the ***ventral striatum.*** The name derives from the way the nucleus rests against the septum (*cf*. incumbent - to lie down against). The lumens of the telencephalic vesicles eventually become the paired lateral ventricles of the brain, which serve in the adult brain as useful landmarks for the location of certain of the basal ganglia and other structures originating from the basal telencephalon, such as the hippocampus.

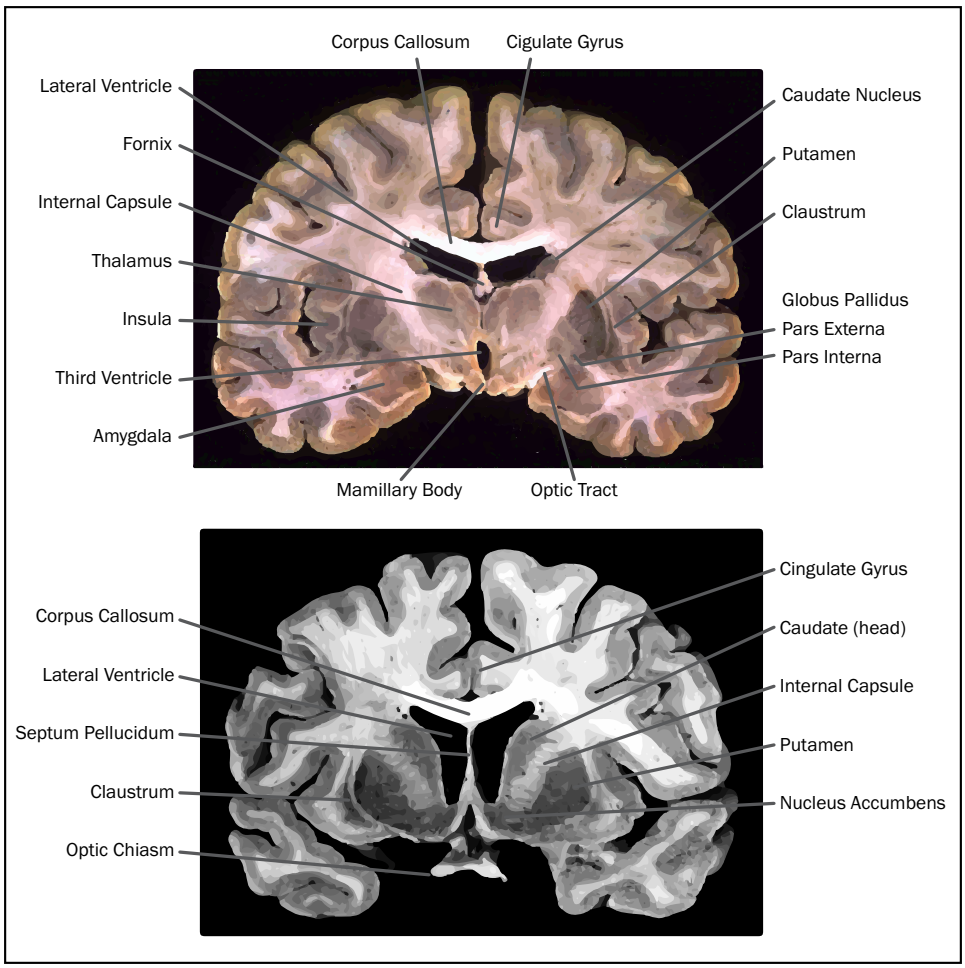


Figure 22.1

Nissl-stained, coronal sections of the human forebrain, illustrating the locations of certain basal ganglia. The bottom section is anterior to the top one. Observe how the internal capsule separates the caudate nucleus from the putamen and globus pallidus. In the anterior (bottom) section the nucleus accumbens can be seen 'leaning' against the septum. Were it not for the presence of the internal capsule, the entire striatum would form the lateral wall of the lateral ventricle, as predicted from its embryology.

In recent years the term 'basal ganglia' has been associated with a particular circuit involved in motor control, with the result that some additional structures have been added to the list and some of the original members have been dropped. When you hear or read the term these days, it is usually meant to include the caudate, putamen, n. accumbens, both components of the globus pallidus (pars externa and pars interna), the ***subthalamic nucleus*** and the ***substantia nigra*** (pars compacta and pars reticulata).

With this new classification of basal ganglia structures, it is no longer appropriate to think of the basal ganglia as purely telencephalic. For instance, the substantia nigra is part of the

midbrain and stems from embryonic mesencephalon. Also, the pars interna of the globus pallidus and the subthalamic nucleus emerge from the part of the neural tube that forms the diencephalon.

Two structures now removed by convention from the list of basal ganglia are the amygdala and the claustrum. The former is functionally related to the limbic system and is important in mediating emotion, especially fear. Nothing much is known about the function of the claustrum.

Though all basal ganglia processing involves connections from the cerebral cortex through the basal ganglia and back to the cortex, there are four different loops with different connectivities and distinct functions: a skeletomotor loop, an oculomotor loop, an association loop and a limbic loop. The basic architecture of each loop is schematized in Figure 22.2, the contents of the boxes differing somewhat for each loop. The pattern, however, is similar for each loop.

Information descends from several cortical areas to the entry point of the system, always some component of the striatum (see Figure 22.1). Information can then take two pathways: direct and indirect. The striatal input nuclei project directly to the output nuclei, which in turn project to various thalamic nuclei. In addition, an indirect pathway to the output nuclei goes by way of the external segment of the globus pallidus and subthalamic nucleus, which also have a feedback connection. The thalamic nuclei that receive the output of the basal ganglia project to the frontal lobe.

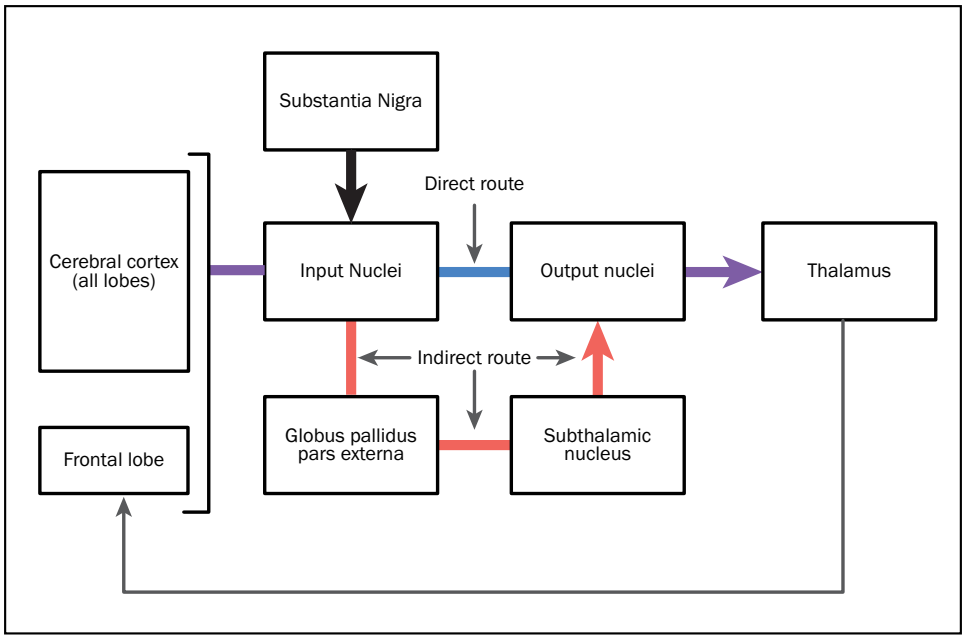


Figure 22.2

Basic circuitry of the basal ganglia. Note the direct and indirect routes to the output nuclei and thalamus and the fact that the projection is back to the cerebral cortex.

Of primary concern to us here is the skeletomotor loop, which is schematized in Figure 22.3. Observe that the basal ganglia form an inhibitory box; their only output inhibits the thalamus, specifically the VL and VA nuclei. The subthalamic nucleus is the only excitatory structure in either the direct or indirect pathway. The substantia nigra, which is not in either pathway, modulates input nuclei with dopamine, and the dopaminergic action excites some cells and inhibits others.

Note also that the net effect of the direct route is to disinhibit the thalamus – the putamen inhibits the output nuclei, which removes the inhibition of the thalamus. In contrast, the net effect of the indirect loop is to increase the inhibition of the thalamus. These connections shed light on certain clinical signs of disease in the basal nuclei.

Disorders of the basal ganglia cause disturbances of tone or ***dystonias*** and abnormalities of movement or ***dyskinesias***. The abnormalities disappear in sleep and, when produced in experimental animals, are eliminated by cutting the corticospinal tract, which is consistent with the anatomical connections schematized in Figure 22.3.

Because the motor components of the basal ganglia do not contribute axons to the corticospinal tract, they are often said to form the ***extrapyramidal motor system***, and the disorders associated with them are called ***extrapyramidal diseases***. The dystonias and

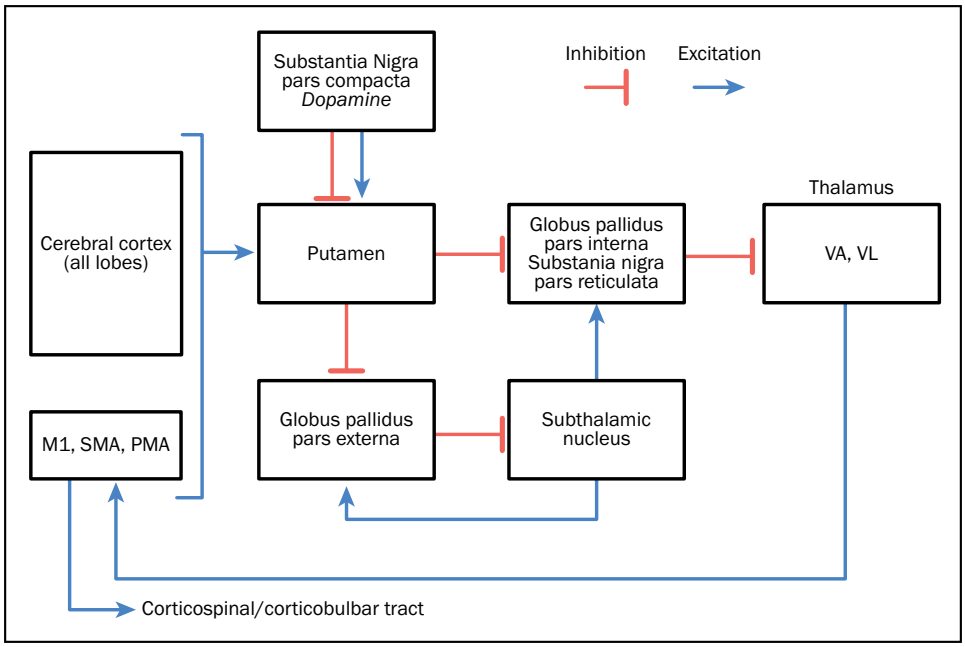


Figure 22.3

The skeletomotor loop through the basal ganglia. Synaptic inhibition is mediated largely by GABA and excitation by glutamate. Acetylcholine and numerous peptides are also involved but are not shown here.

dyskinesias appear to represent what the neurologists call ***release phenomena***, a disinhibition of normal processes leading to too much tone or too much movement.

As an example of this, when a lesion damages the subthalamic nucleus on one or both sides, the person exhibits wild flailing movements of the limbs when they attempt to move voluntarily. This is called ***ballismus*** or, in the case of unilateral disease, ***hemiballismus***. In essence, there is much more movement than is intended. Observe in Figure 22.3 that the net effect of a lesion of the subthalamic nucleus is to reduce the excitatory input to the output nuclei, thereby releasing the thalamus from inhibition, and indirectly the cortical motor areas as well. It is easy to imagine that, as a consequence, the motor commands issued by the cortex to the brainstem and spinal cord will be exaggerated, with large amplitude accidental movements superimposed on the desired movement.

Ballismus is perhaps the most dramatic of the dyskinesias associated with extrapyramidal disease, but slower movements of lower amplitude also occur in degenerative diseases of the striatum, such as Huntington's disease. When slow and writhing in character, these are called ***athetoid*** (noun is ***athetosis***). Quicker movements are called ***choreic*** or ***choreiform***, because they resemble the motions a dancer might make. Choreic movements can sometimes resemble ballismic movements, and it is customary in British neurology to use the term 'hemichorea' for hemiballismus. ***Tics*** are quick movements of the face and head seen in Tourette's syndrome, another disorder affecting the basal ganglia.

The movement pathologies probably represent a continuum with a common underlying neural substrate. All are thought to result from release of the corticospinal/corticobulbar pathways from the normal inhibitory control of the basal ganglia loop.

Another example of an extrapyramidal disease is Parkinsonism, named after the physician James Parkinson who first recognized and described it. Parkinsonism or Parkinson's disease is characterized by a marked slowing of movement (***bradykinesia***), poverty of movement (***hypokinesia***) and sometimes absence of movement (***akinesia***).

Patients can lose control over anything from facial expression to postural muscles. For example, the face is expressionless (***mask facies***) and there is infrequent blinking. Because the normal postural mechanisms are impaired, there is a tendency to fall when the body's center of gravity is suddenly displaced. A characteristic tremor affects the hand, jaws and head when no voluntary movements are underway. This ***resting tremor*** or ***tremor at rest*** inspired Parkinson to call the disorder the shaking palsy. There is also increased resistance to passive movement of the limbs reminiscent of the resistance to bending offered by a soft lead pipe. This is called ***lead-pipe rigidity***. Sometimes the resistance is momentarily interrupted leading to ratchet-like jerks of the limb: so-called ***cogwheel rigidity***.

The origin of these features is less easily explained in terms of the circuitry of Figure 22.3 because there is a mixture of release-like signs (tremor, rigidity) and other signs of motor inhibition (akinesia, mask facies). However, the rationale for the current methods of treating

Parkinsonism can be seen in Figure 22.3. The major change in the brains of parkinsonian patients is the loss of the dopaminergic cells of the substantia nigra, pars compacta, that innervate the striatum. Thus, an obvious therapeutic approach is to substitute external dopamine for that normally delivered to the striatum by the substantia nigra. Clinicians do so by administering dopamine precursor, L-dopa, which crosses the blood-brain barrier and can be converted to dopamine by the remaining dopaminergic cells of the brain. Various surgical approaches have attempted with mixed success to eliminate some of the motor symptoms of Parkinsonism by making selective lesions in the globus pallidus and ventral lateral nucleus of the thalamus. More recent and still experimental approaches involve the implantation directly into the brain of cells that manufacture dopamine.

The oculomotor, association and limbic loops through the basal ganglia have the same basic architecture as the skeletomotor loop. The major differences lie in details of the internal circuitry and, especially, in the cortical areas targeted by the related thalamic nuclei (Figure 22.4). It may seem curious that the basal ganglia also include a component of the limbic system, which we usually associate with the hippocampus, amygdala, hypothalamus and so forth. Recall, though, that the amygdala is embryologically one of the 'original' basal ganglia.

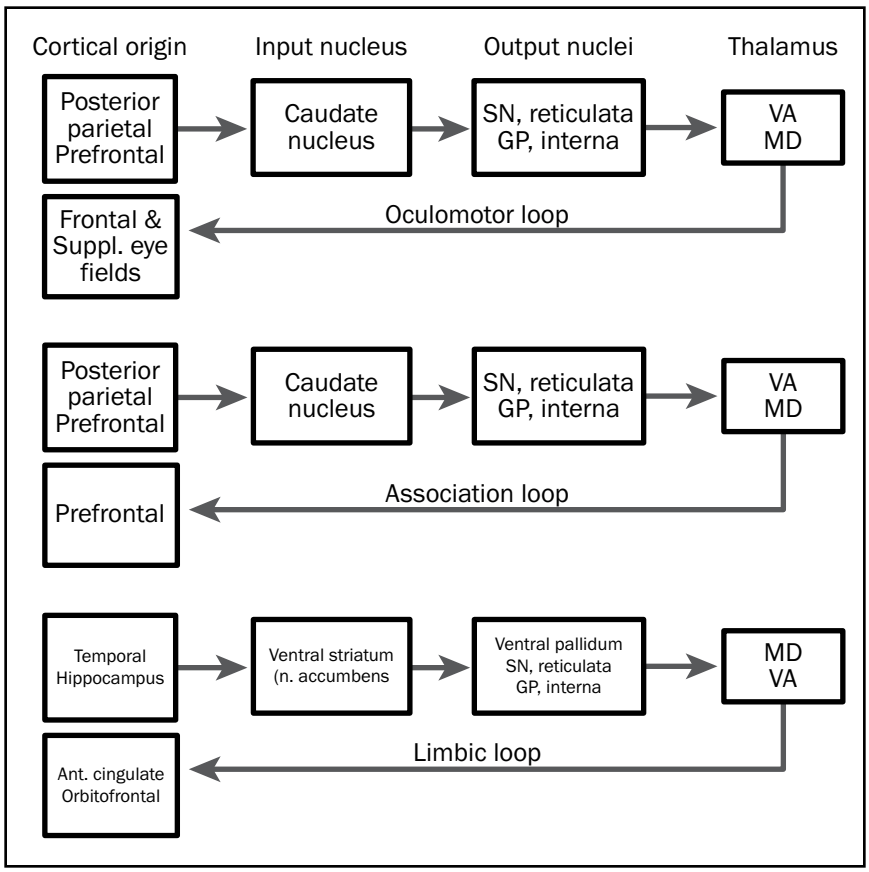


Figure 22.4

Reduced schematic of the oculomotor, association and limbic loops through the basal ganglia. The contributions of the indirect and modulatory components are not shown (see Figure 22.2).

The involvement of the striatum in mental function has long been suspected because the antipsychotic activity of the class of drugs knows as ***neuroleptics*** is highly correlated with their ability to bind to and block dopamine receptors. Amphetamines and cocaine, which can cause psychosis, block the reuptake of catecholamines, including dopamine, thus adding to the suspicion that some dopaminergic system has a role in mental and emotional function. Much recent work points to the nucleus accumbens (Figure 22.1 and Figure 22.4) as a possible site responsible for the effects of these drugs. Why, we might ask, should a nucleus that forms part of the basal ganglia have anything to do with emotion? Very possibly, the answer to this still unknown question is the same as that to the question: "what makes a movement start?"

A body of recent research suggests that two distinct neural processes are at work in those behaviors that ensure the survival of the individual and the species, one of which is mediated by the nucleus accumbens. The first, the ***appetitive-approach process*** (wanting) increases the probability of exploratory or foraging behavior that moves the animal toward an important goal (food, mate, etc). The ***consummatory-satisfaction process*** (liking) ensures that, once found, the goal will be effectively utilized (feeding, mating).

It is natural to think of processes like feeding as generating its own reward, but it is not immediately obvious why searching for food needs positive reinforcement. However, a little reflection makes it clear that something has to ensure that such behavior takes place in the absence of the goal that is being sought. Otherwise, the animal would never find the food or mate that it needs. It is not enough to say that hunger is what drives foraging behavior, because hunger is unpleasant and should punish (negatively reinforce) the animal for behavior associated with it.

Now, it is well known that rats will voluntarily stimulate their nucleus accumbens electrically over and over again, and that pharmacological manipulation of the nucleus affects the probabilities that certain kinds of behaviors will occur. Thus, the evidence for involvement of this nucleus in motivational reward is strong. Experiments have also shown that injection of dopamine receptor blockers into the nucleus accumbens, or GABA, which directly inhibits cells, has no effect on a rat's eating behavior when food was available, but reduces the animal's approach and foraging behavior significantly. In other words, these manipulations of the nucleus accumbens affect the 'wanting and seeking' part of the behavior but not the 'liking' part. Viewed from this perspective, it is less mysterious to find a part of the striatum involved in the process of increasing the probability of (i.e. rewarding) certain critical patterns of behavior that have no immediate external source of reinforcement.

An unfortunate consequence of the multiple functions subserved by dopamine in the brain is that drugs intended to affect one of the systems will also affect the others. This is seen in the side effects of the administration of neuroleptics, such as haloperidol, a mainstay in the pharmacological treatment of schizophrenia. After long-term treatment, these individuals develop a syndrome called ***tardive dyskinesia*** that comprises elements of Parkinsonism and other extrapyramidal disorders. Because several different kinds of dopamine receptors exist, it

is possible that more specific medications can be developed which can target certain systems and avoid the side effects associated with actions on systems not involved by the disease.

Terms and Techniques

extrapyramidal motor system	basal ganglia	striatum
caudate nucleus	putamen	ballism or ballismus
nucleus accumbens	direct pathway	indirect pathway
subthalamic nucleus	globus pallidus	substantia nigra
pars reticulata	pars compacta	dyskinesia
dystonia	lead pipe rigidity	cogwheel rigidity
chorea	athetosis	tic
Parkinson's disease	hypokinesia	tremor at rest
dopamine	dopaminergic	GABA
nigro-striatal pathway	neuroleptic	tardive diskinesia

The Cerebellum and Motor Coordination

23

Table of Contents

Our earliest clear indication of the function of the cerebellum came from British neurologist Gordon Holmes' systematic observations of soldiers injured in the First World War. This work demonstrated that the cerebellum is essential for the coordination of muscle action in the course of a movement. Disturbances in this function lead to incoordination or ***ataxia***. Ataxia manifests itself in various ways, all related to an inability to produce a smoothly blended sequence of muscle contractions. Some typical manifestations are:

- **dysmetria** - improperly scaled movements made in attempting a motor task; leads to overshooting and past-pointing.
- **decomposition of movement** - movements broken into isolated components
- **dysdiadokokinesis** - inability to perform rapid alternating movements such as slapping the hand against the thigh, first palm down, then palm up.
- **pendular knee jerks**
- **staccato speech**
- **megalographia** - large amplitude excursions in writing

The overshooting or past-pointing of cerebellar ataxia causes the limb to oscillate as it seeks to place the hand or fingers on a target and is known as ***intention tremor*** (though the phenomenon is not a true tremor as in a patient with Parkinsonism). Intimate connections between the cerebellum and the vestibular system explain the appearance of ***nystagmus*** with cerebellar lesions.

Cerebellar ataxia is not made worse by closing the eyes unlike ataxia due to interruption of the dorsal columns. A patient with dorsal column disease depends on the visual system

for knowledge of the position of the legs. If such a patient stands with legs together and closes the eyes, the body will sway and become unsteady. This is called ***Romberg's sign*** and indicates a sensory ataxia, rather than a cerebellar ataxia.

Although the cerebellum is the target of massive afferent input from muscles, joints and skin, cerebellar lesions cause no deficit in ***conscious sensation***. Patients with cerebellar lesions can accurately report the positions of their limbs in space because the mechanism of conscious proprioception depends entirely on the dorsal column/medial lemniscal system, which projects to the cerebral cortex through the thalamus.

Gross Anatomy of the Cerebellum

The cerebellum fills the posterior cranial fossa. Despite its small size, the cerebellum contains more than half the cells in the brain. It has an outer ***cortex***, white matter containing input and output fibers, and a collection of deep nuclei, which are aggregates of output cells. From medial to lateral these are the ***fastigius, interpositus*** and ***dentatus*** nuclei (Figure 23.1). In humans, the nucleus interpositus comprises two smaller nuclei, the ***globose*** and ***emboliform nuclei***. The cerebellum is connected to the brainstem by three large peduncles that contain axons carrying information to and from the cerebellum.

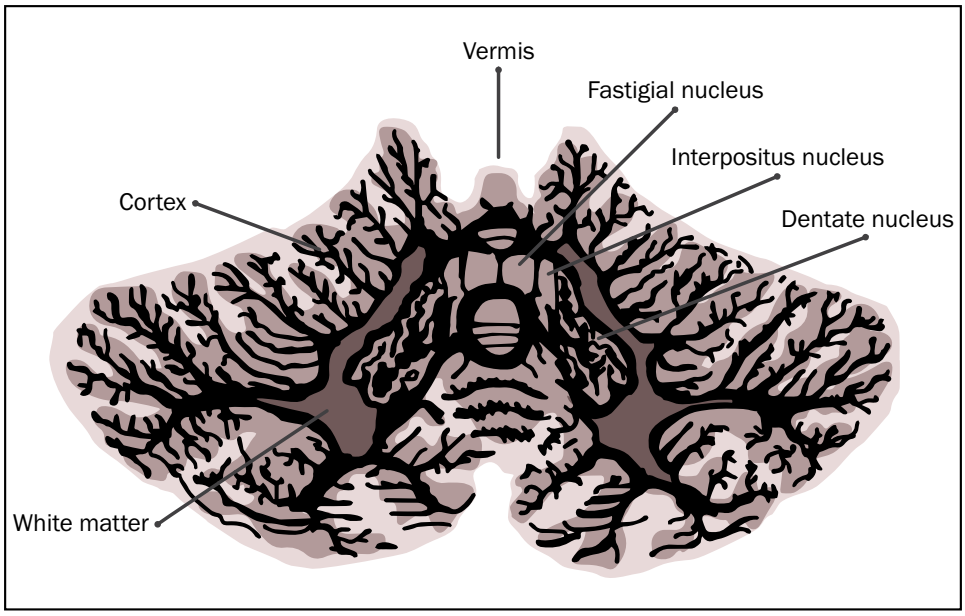


Figure 23.1

A horizontal section through the cerebellum showing the cortex, white matter and deep nuclei. Fibers are stained. Instead of gyri, the cerebellar cortex is formed into folia (Latin for leaves), which give it a tremendous surface area. If unfolded, the surface area of the cerebellum would equal that of one cerebral hemisphere.

Functional Organization of the Cerebellum

Anatomists have divided the cerebellum in a number of ways, but the most useful functional parcellation is illustrated in Figure 23.2. In this scheme the cerebellum is divided from medial to lateral into an unpaired midline ***vermal*** region, paired ***intermediate*** (or ***paramedian*** or ***paravermal***) regions and the paired ***cerebellar hemispheres***, plus a distinct element called the ***flocculo-nodular*** lobe.

These regions have different patterns of projection to the deep cerebellar nuclei and to the brainstem, as is illustrated in Figure 23.2. Their inputs also correspond functionally with their output projection patterns. It should be emphasized that the compartments are not as discrete as is implied in Figure 23.2 because the projections shift gradually from medial to lateral as one compartment changes gradually into the next.

The input to the cerebellar hemispheres largely comes from the cerebral cortex via the pontine nuclei. The hemispheres send their output back to the cerebral cortex via the dentate nucleus, which projects to the ventral lateral nucleus of the thalamus (Figure 23.2). This circuit is in some ways analogous to that of the basal ganglia. The cerebellar hemispheres appear to underly high-level coordination of fine voluntary movements, especially of the distal joints. Because the cerebellar hemispheres have a somewhat close relationship to the cerebral cortex, they are sometimes referred to as the ***cerebro-cerebellum***.

Vermal and intermediate zones are called the ***spino-cerebellum*** because of their primary

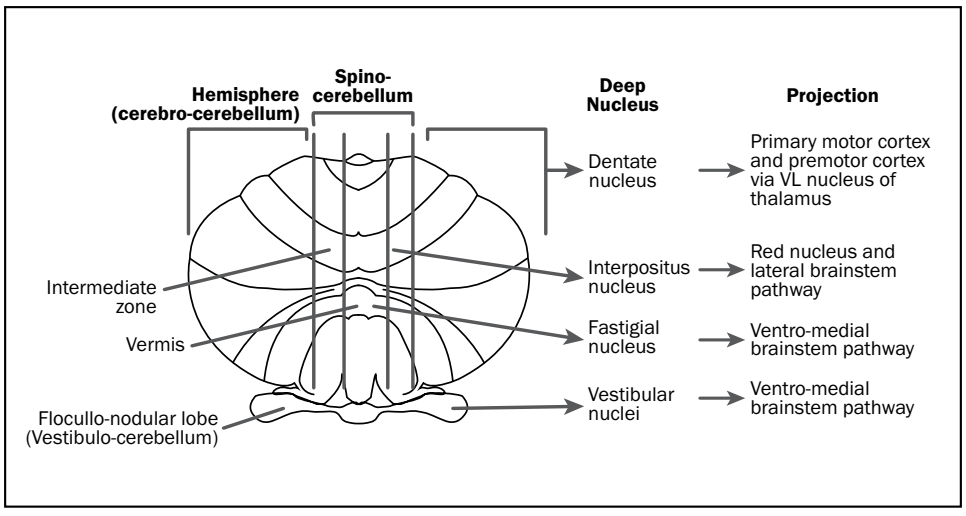


Figure 23.2

Organization of the cerebellum and deep nuclei. The medial-to-lateral organization is reminiscent of the organization of the anterior horn of the spinal cord where axial muscles are represented medially and distal muscles laterally.

relationship with motor and sensory systems of the spinal cord. Afferent information from the trunk and the vestibular system reaches the vermis, which influences axial muscles via projections through the fastigial nucleus to the vestibular nuclei and the reticular formation (the ventromedial brainstem pathway of Chapter 20).

Input to the intermediate zone comes from the spinal cord via the spinocerebellar pathways and also from the sensorimotor cortex by way of the pontine nuclei. Somatotopic relations are preserved in these projections but they are much less distinct than those of sensory cerebral cortex. The intermediate zone is concerned with girdle and proximal limb muscles and projects primarily by way of the interpositus nucleus to the red nucleus (origin of the rubrospinal tract) and via the thalamus to the motor cortex.

Spinocerebellar pathways carry information about body and limb position received from skin, joint, tendon organ and spindle receptors. Some of these signals concern specific muscles and joints, while others provide more global information about multiple muscles and joint. The spinocerebellar pathways also carry information about the status of interneuron pools of the spinal cord that are organizing the basic patterns of motor activity.

The flocculonodular lobe is the phylogenetically oldest part of the cerebellum and is closely related to the vestibular system, hence the name ***vestibulo-cerebellum***. While it shares with the rest of the cerebellum the pattern of inputs and outputs schematized in Figure 23.2 and Figure 23.4, it also has special connections not found in the other functional zones. The flocculonodular lobe receives some direct projections from the vestibular labyrinth, i.e. primary afferent fibers from Scarpa's ganglion that have not synapsed in the medullary vestibular nuclei. Also, certain cells in the cortex of the flocculonodular lobe project directly to the vestibular nuclei, without a relay in the deep cerebellar nuclei. The vestibular nuclei seem to serve as a 'deep cerebellar nucleus' for the flocculonodular lobe. This part of the cerebellum is clearly involved in balance and postural maintenance. It also plays a role in the control of eye position (see Chapter 15).

Laterality Of Cerebellar Connections

In contrast to damage to the cerebral hemispheres, unilateral damage to the cerebellum leads to signs on the same side of the body as the lesion. The pattern of connections underlying this fact can be confusing, so it is worth spending a little time on them (Figure 23.3).

The cerebral cortex on one side projects to the ipsilateral pontine nuclei, whose cells then send their axons across the midline into the middle cerebellar peduncle and reach the contralateral cerebellar hemisphere. The projection from that same cerebellar hemisphere exits via the superior cerebellar peduncle, descussates in the midbrain and ends in the contralateral thalamus, which completes the relay to the cerebral cortex. This double decussation pattern maintains the appropriate laterality in the relationship and the cerebral and cerebellar hemispheres concerned with one side of the body are connected together (Figure 23.3, left).

The afferent pathways are also organized appropriately. Thus, the dorsal column medial lemniscal system serving the left side of the body projects functionally to the right cerebral hemisphere. The right cerebral hemisphere is connected to the left cerebellar hemisphere, which receives an uncrossed spino-cerebellar projection carrying information about the left side of the body.

The corticospinal fibers from the right cerebral hemisphere also decussate to govern movement of the left side of the body. Similarly, the left side of the cerebellum projects across the midline to the right red nucleus (Figure 23.3, right) whose axons recross the midline to project into the left rubrospinal tract. The reticulospinal and vestibulospinal tracts are essentially uncrossed. In a word, the connections of the cerebellum with the cortex, red nucleus, vestibular nuclei and reticular formation ensure that its influence is on the ipsilateral side of the body.

Microcircuitry of the Cerebellum

Figure 23.4 schematizes the microcircuitry of a cerebellar folium, which is relentlessly repetitious across the entire structure. There are two major anatomical classes of input fiber:

> ***Mossy fibers*** constitute most of the input axons. Mossy fibers give off collaterals that excite cells in the deep cerebellar nuclei. These same mossy fibers then excite the ***granule cells***, the most numerous cells in the cerebellum and probably in the brain itself. The granule cells give rise to a single axon that rises toward the cortical surface and splits to form a ***parallel fiber***. The parallel fibers run parallel to the folia

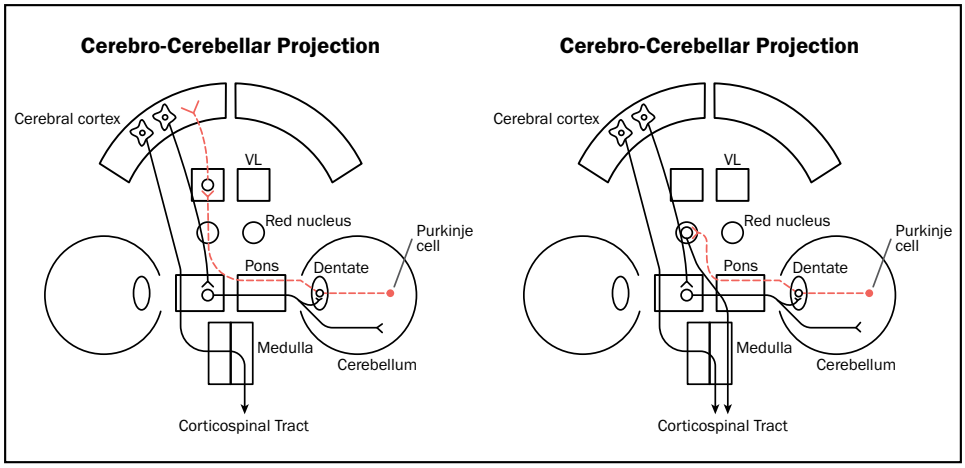


Figure 23.3

Double decussation connections of the cerebellum.

for long distances, contacting thousands of Purkinje cells whose elaborate, flattened dendritic trees stand at right angles to the parallel fibers. The parallel fibers make excitatory synaptic contracts on the Purkinje cell's dendrites.

Climbing fibers come from the inferior olivary nucleus and get their name from their unique pattern of termination. After giving off collaterals to the deep nuclei, the climbing fibers send tendrils out along the complex dendrites of the Purkinje cells and cover them with excitatory synapses. A given Purkinje cell receives input from only one climbing fiber, and a climbing fiber may excite as many as 10 Purkinje cells. An action potential in a climbing fiber strongly depolarizes the target Purkinje cell. The function of the climbing fibers, with their powerful excitatory effects, is not known.

The Purkinje cells provide the only output axons of the cerebellar cortex and most project to the deep nuclei. Purkinje cells make *inhibitory* contacts and use the neurotransmitter GABA.

With the exception of the granule cell, all cells whose cell bodies live in the cerebellar cortex are inhibitory. The deep nuclear cells make excitatory contacts wherever they project. As noted

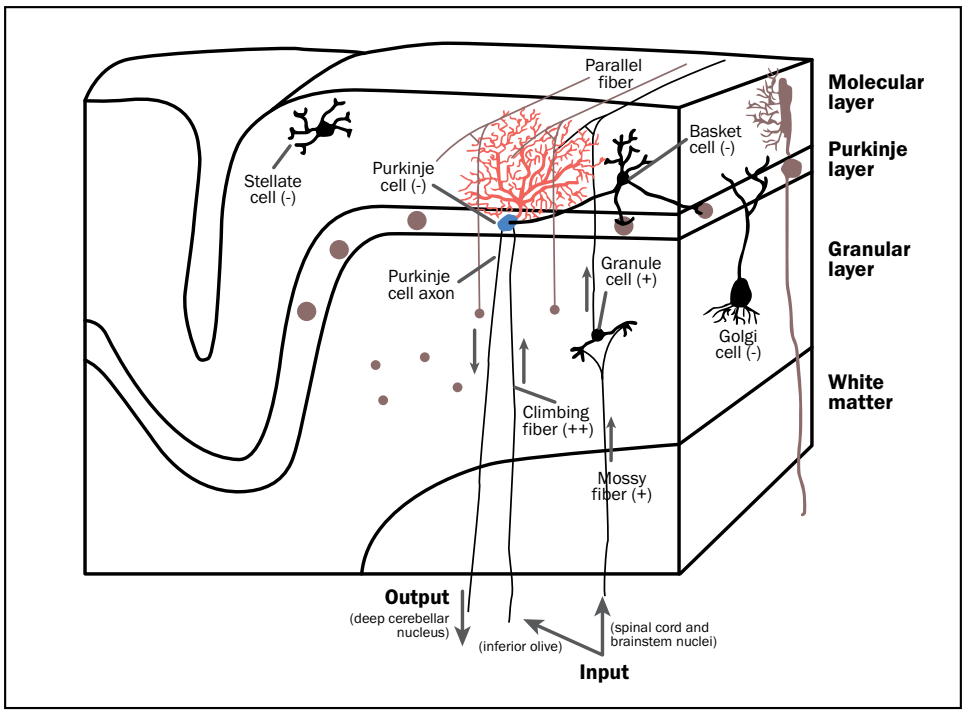


Figure 23.4

Cross section of a cerebellar folium. The granule cells are the only excitatory cells in the cerebellar cortex. Purkinje, stellate, basket and Golgi cells are all inhibitory. Adapted from J.H Martin, *Neuroanatomy: Text and Atlas*, 2nd, Appleton & Lange, New York, 1996.

earlier, the connections between the cerebellum and the vestibular system are somewhat different. Some cells in Scarpa's ganglion project directly to the flocculonodular lobe and vermis (as mossy fibers), and Purkinje cells in these areas project directly to the vestibular nuclei.

How Does the Cerebellum Coordinate Movement?

From Figure 23.4 it can be seen that the 'true' output of the cerebellum arises almost exclusively from the deep cerebellar nuclei. (The one exception, already noted, is that some Purkinje cells in the flocculo-nodular lobe send their axons to vestibular nuclei.) The only influence that the cerebellar cortex exercises over this output is GABA-ergic inhibition from the Purkinje cells.

One may ask how a system based on inhibition could be used to coordinate movement. This becomes less mysterious if one considers that a coordinated movement is a sequence of all of the right muscle actions and none of the wrong ones. Cerebellar coordination can be viewed as the process that simply ***vetoes*** all possible combinations of muscle action except those needed to provide a smooth, efficient movement.

Recall that collaterals of axons carrying signals from the cerebral cortex, brainstem and spinal cord to the cerebellar cortex create a background of excitation in the deep nuclei. This same information goes to the cerebellar cortex, which integrates it all and develops an appropriate pattern of activity in the sheet of Purkinje cells. These now impose a pattern of inhibition on the activity in the deep nuclei, which project out to the various components of the motor system (Figure 23.5). The pattern of inhibition acts like a dynamic ***stencil***, allowing only a particular sequence of excitation to flow from the deep nuclei to their targets in motor effector systems.

How is this pattern of Purkinje cell activity determined? To begin with, massive divergence and convergence occurs among the mossy fibers, which bring most of the relevant information to the cerebellum. A given mossy fiber may emit several collaterals within a given folium and also send branches to neighboring folia (Figure 23.4). These activate the granule cells and, consequently, the system of parallel fibers, which is clearly designed to distribute information to a population of Purkinje cells. The average length of a parallel fiber is 3 mm and it is estimated that in humans as many as 250,000 parallel fibers may contact the dendritic tree of a single Purkinje cell.

Because of the general somatotopy of the projections to the cerebellum, this lateral spread of information is well suited to link together information relevant to neighboring body parts. If the cerebellum is damaged, this coordinating function is impaired, and the cortical inhibitory mechanisms fail to constrain the pattern of muscle action to just the right ones. The result is movement of the wrong magnitude and timing, i.e. cerebellar ataxia.

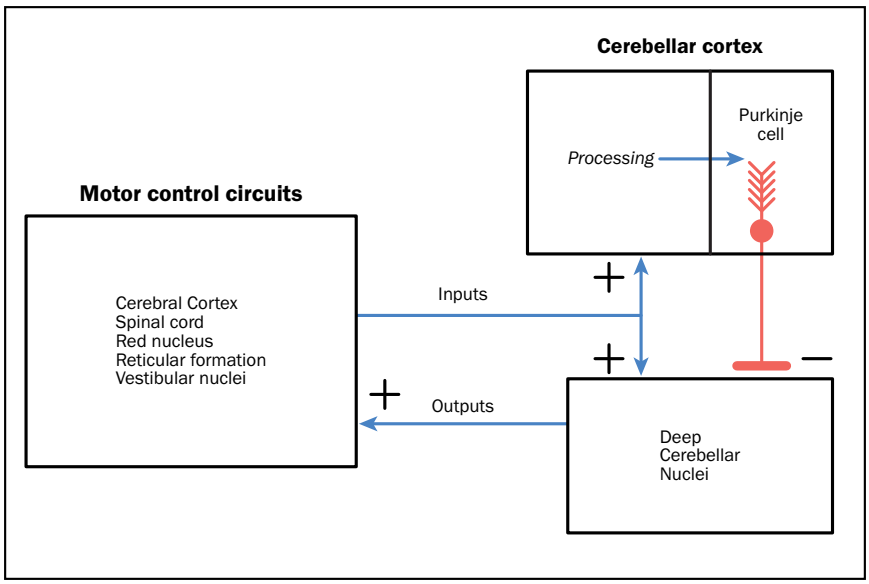


Figure 23.5

A box diagram of the connections of the primary motor circuits with the cerebellum. The cerebellar cortex and the deep nuclei get the same input, but the cortex imposes a pattern of inhibition on the deep nuclei and, consequently, the output of the cerebellum.

Terms and Techniques

spino-cerebellum	cerebro-cerebellum	cerebellar peduncle
granule cells	Purkinje cells	mossy fibers
climbing fibers	parallel fibers	deep cerebellar nuclei
fastigial nucleus	interpositus nucleus	dentate nucleus
inferior olivary nucleus	ataxia	dysmetria
dysdiadokokinesis	intention tremor	Romberg's sign