CHAPTER

16

ELECTRONIC COCHLEA

Richard F. Lyon Carver Mead

When we understand how hearing works, we will be able to build amazing machines with brainlike abilities to interpret the world through sounds—that is, to *hear*. As part of our endeavor to decipher the auditory nervous system, we can use models that incorporate current ideas of how that system works to engineer simple *electronic* systems that hear in simple ways. The relative success of these *engineered* systems then helps us to evaluate our knowledge about hearing, and helps to motivate further research.

As a first step in building machines that hear, we have implemented an analog electronic cochlea that incorporates much of the current state of knowledge about cochlear structure and function. The biological **cochlea** (inner ear) is a complex three-dimensional fluid-dynamic system, illustrated schematically in Figure 16.1. In the process of designing, building, and testing the electronic cochlea, we have had to put together a coherent view of the function of the biological cochlea from the diverse ideas in the literature. This view and the resulting design are the subjects of this chapter.

We hear through the sound-analyzing action of the cochlea and of the auditory centers of the brain. As does vision, hearing provides a representation of events and objects in the world that are relevant to survival.

Portions reprinted with permission from *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 36, No. 7, July 1988.

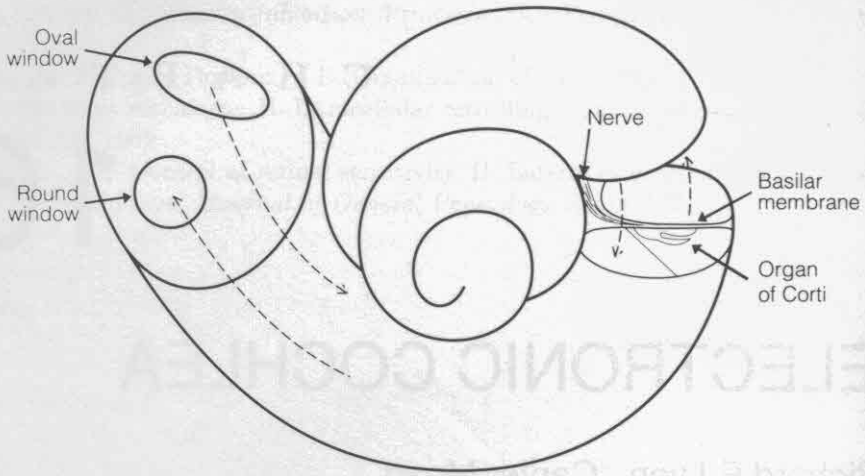

FIGURE 16.1 Artist's conception of the cochlea, with cutaway showing a cross section of a cochlear duct. The bending of the cochlear partition causes a shearing between it and the tectorial membrane, which can be both sensed and amplified by hair cells in the organ of Corti. The dashed lines indicate fluid paths from the input at the oval window and back to the round window for pressure relief.

Just as the natural environment of light rays is cluttered, so is that of sound waves. Hearing systems therefore have evolved to exploit many cues to separate out complex sounds. The same systems that have evolved to help cats to catch mice and to warn rabbits of wolves also serve to let humans speak with other humans. Except for the highest level of the brain (the auditory cortex), the hearing systems of these animals are essentially identical.

BASIC MECHANISMS OF HEARING

The cochlea consists of a coiled fluid-filled tube (see Figure 16.1) with a stiff **cochlear partition** (the **basilar membrane** and associated structures) separating the tube lengthwise into two chambers (called ducts, or **scalae**). At one end of the tube, called the **basal end** or simply the **base**, a pair of flexible membranes called **windows** connect the cochlea acoustically to the middle-ear cavity. A trio of small bones called **middle-ear ossicles** couple sound from the **tympanic membrane** (eardrum) into one of the windows, called the **oval window**. When the oval window is pushed in by a sound wave, the fluid in the cochlea moves, the partition between the ducts distorts, and the fluid bulges back out through the **round window**. Transducers called **hair cells** sit along the edge of the partition in a structure known as the **organ of Corti**, and are arranged to couple with the partition motion and the fluid flow that goes with that motion.

The distortion of the cochlear partition by sounds takes the form of a traveling wave, starting at the base and propagating toward the far end of the cochlear ducts, known as the **apical end**, or **apex**. As the wave propagates, a filtering action occurs in which high frequencies are strongly attenuated; the cutoff frequency gradually lowers with distance from base to apex. The **inner hair cells**, of which there are several thousand, detect the fluid velocity of the wave as it passes. At low sound levels, some frequencies are amplified as they propagate, due to the energy added by the **outer hair cells**, of which there are about three times as many as inner hair cells. As the sound level changes, the effective gain provided by the outer hair cells is adjusted to keep the mechanical response within a more limited range of amplitude.

The propagation of sound energy as hydrodynamic waves in the fluid and partition system is essentially a distributed low-pass filter. The membrane velocity detected at each hair cell is essentially a bandpass-filtered version of the original sound. Analyzing and modeling the function of these lowpass and bandpass mechanisms is the key to understanding the first stage of hearing. Because analog circuits cannot easily be made precise, they must be made self-adjusting; if the circuits must adjust for their own long-term offsets and drifts, they might as well also adjust to the signal. The use of adaptation to optimize the system response seems to be a pervasive principle in perception systems, which are built from sensitive but imprecise components. **Lateral inhibition** is one term often applied to physiological systems that self-adjust [von Békésy, 1967].

Traveling Waves

As sounds push on the oval window, a pressure wave is initiated between the ducts. The fluid is incompressible, and the bone around the cochlea is incompressible, so as the pressure wave moves, it displaces the basilar membrane. When the eardrum is tapped, the middle-ear ossicles tap on the oval window, and a pressure pulse travels down the length of the cochlea. In propagating, the pressure pulse deforms the basilar membrane upward; if we could watch, we would see a little bump traveling along the basilar membrane. There is a well-defined velocity at which a signal will travel on such a structure, depending on the physical parameters of the membrane.

If the basilar membrane is thick and stiff, the wave will travel very quickly along the cochlea; if it is thin and flexible, the wave will travel very slowly. The changing properties of the basilar membrane control the velocity of propagation of a wave. Hearing starts by spreading out the sound along this continuous traveling-wave structure. The velocity of propagation is a nearly exponential function of the distance along the membrane. The signal starts near the oval window with very *high* velocity of propagation, because the basilar membrane is very thick and stiff at the basal end. As the signal travels away from the oval window, toward the apical end, the basilar membrane becomes thinner, wider, and more flexible, so the velocity of propagation *decreases*—it changes by a factor of about 100 along the length of the cochlea.

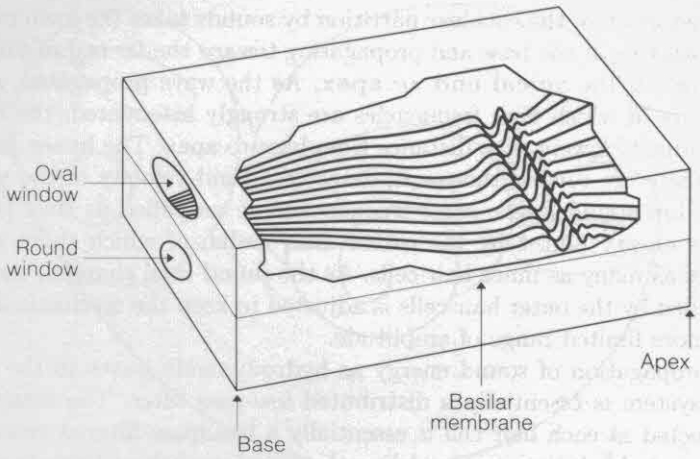

FIGURE 16.2 A sinusoidal traveling wave on the basilar membrane, in a simplified rectangular-box model. The basilar membrane (the flexible part of the cochlear partition between the bony shelves) starts out stiff and narrow at the base and becomes more compliant and wider toward the apex. Thus, a wave's propagation velocity and wavelength decrease—the wavenumber, or spatial frequency, increases—as the wave travels from base to apex.

Sine-Wave Response

When the sound is a sine wave of a given frequency, it vibrates the basilar membrane sinusoidally. The wave travels very quickly at the basal end, so it has a long wavelength. A constant energy per unit time is being put in, but, as the wave slows down and the wavelength gets shorter, the energy per unit length builds up, so the basilar membrane gets more and more stretched by the wave. For any given frequency, there is a region beyond which the wave can no longer propagate efficiently. The energy is dissipated in the membrane and its associated detection machinery. Past this region, the wave amplitude decreases rapidly (faster than exponentially). Figure 16.2 is an artist's conception of what the membrane deflection might look like for a given sine wave.

Neural Machinery

The auditory-nerve signal comes out of the machinery in the organ of Corti, along one side of the basilar membrane. A protrusion called the **tectorial membrane** is located just above the basilar membrane. We can think about the tectorial membrane's relation to the organ of Corti this way: As the pressure wave travels along the cochlea, it bends the basilar membrane up and down; when the basilar membrane is pushed up, the tectorial membrane moves to the right relative to the basilar membrane. This linkage arrangement is a way of converting up-and-down motion on the basilar membrane to shearing motion between it and the tectorial membrane.

Mounted on the top of the basilar membrane is a row of inner hair cells. The single row of 3500 (in humans) inner hair cells that runs along the length of the membrane is the only source of the nerve pulses that travel to the cochlear nucleus and on up into the brain. All the auditory information is carried by about 28,000 (in humans) nerve fibers. Everything we hear is dependent on that set of hair cells and nerve cells. Fine hairs, called **stereocilia**, protrude from the end of the inner hair cells; they detect the shearing motion of the membranes and act as the *transducers* that convert deflection to an ion current.

Let us consider the behavior of the auditory system for loud signals, in which situation the outer hair cells are not needed. Signals are propagating down the cochlea, so there are bumps traveling along the basilar membrane. The bumps travel quickly at first, and then they slow down. For any given frequency, there is a point at which the displacement is maximum; this is the point of maximum velocity of the vibration of hair cells with respect to the tectorial membrane.

On the scale of a hair cell's stereocilia—a small fraction of a micron—the viscosity of the fluid is high. The fans of cilia sticking out of the hair cells have enough resistance to the motion of the fluid that they bend. When the cilia are bent one way, the hair cells stimulate the primary auditory neurons to fire. When the cilia are bent the other way, no pulses are generated. When the cilia are bent to some position at which the neurons fire, and then are left undisturbed for some time, the neurons stop firing. From then on, bending the cilia in the preferred direction away from that new position causes firings. So the inner hair cells act as auto-zeroing half-wave rectifiers for the velocity of the motion of the fluid.

Outer Hair Cells

The large structure of the organ of Corti on top of the membrane absorbs energy from the traveling wave. In the absence of intervention from the outer hair cells, the membrane response is reasonably damped. The ear was designed to hear transients, and therefore the basilar membrane itself is not a highly resonant structure.

What then is the purpose of the outer hair cells? There are only a few slow afferent (*to* the brain) nerve fibers coming into the auditory nerve from the outer hair cells. On the other hand, there are a large number of slow fibers coming *down* from higher places in the brain into the cochlea, that synapse onto these outer hair cells. The outer hair cells are *not* used primarily as receptors—they are used as *muscles*. If they are not inhibited by the efferent (*from* the brain) fibers, they provide *positive* feedback into the membrane. If they are bent, they push even harder in the same direction. They can put enough energy back into the basilar membrane that it will actually oscillate under some conditions; the resulting ringing in the ears is called **tinnitus**.

Tinnitus is not caused by an out-of-control sensory neuron as one might suppose. It is a *mechanical oscillation* in the cochlea that is driven by the outer hair cells. The cells pump energy back into the oscillations of the basilar membrane, and they can pump enough energy to make the traveling-wave structure

unstable, so it creates an oscillatory wave that propagates back out through the eardrum into the air. In 1981, Zurek and Clark [Zurek et al., 1981] reported spontaneous acoustic emission from a chinchilla that made such a squeal that it could be heard from several meters away by a human's unaided ear. An excellent and insightful overview of the role of outer hair cells for active gain control in the auditory system is given by Kim. In a classic monument of understatement, he comments on the chinchilla results, "It is highly implausible that such an intense and sustained acoustic signal could emanate from a passive mechanical system" [Kim, 1984, p. 251].

So, the outer hair cells are used as muscles and their function is to reduce the damping of the basilar membrane when the sound input would be otherwise too weak to hear. This arrangement provides not just gain, but also control of the gain, it controls gain by a factor of 100 in amplitude (10,000 in energy). When the signal is small, the outer hair cells are not inhibited and they feedback energy. This system of *automatic gain control* (AGC) works for sound power levels within a few decades of the bottom end of our hearing range by making the structure slightly more resonant and thereby much higher gain—by reducing the damping until it is negative in some regions. We will discuss the details of gain control after we have developed the basilar membrane model.

Based on the threshold of hearing and linear extrapolations from observations on loud signals, researchers once estimated that a displacement of the cilia by less than one-thousandth of an angstrom (10^{-3} angstrom or 10^{-13} meter) would be large enough to give a reasonable probability of evoking a nerve pulse. Because the detectability of sounds near the threshold of hearing involves active mechanical amplification, the actual motion sensed probably is on the order of 1 angstrom of basilar-membrane displacement or hair-cell bending.

WAVE PROPAGATION

As we have discussed, sounds entering the cochlea initiate traveling waves of fluid pressure and cochlear-partition motion that propagate from the base toward the apex. The fluid-mechanical system of ducts separated by a flexible partition is like a waveguide, in which wavelength and propagation velocity depend on the frequency of a wave and on the physical properties of the waveguide. In the cochlea, the physical properties of the partition are not constant with x, but instead change radically from base to apex. The changing parameters lead to the desirable behavior of sorting out sounds by their frequencies or time scales; unfortunately, the parameter variation makes the wave analysis a bit more complex. In this section, we will discuss the mathematics of waves in uniform and nonuniform media, including the cochlea.

The instantaneous value W of the pressure or displacement of a wave propagating in a one-dimensional uniform medium due to a sine-wave input can be expressed as

$$W(x,t) = A(x)\cos(kx - \omega t)$$

If frequency ω and wavenumber k (spatial frequency) are positive and real and A is constant, W is a wave propagating to the right (toward $+x$) at a phase velocity $c = \omega/k$ with no change in amplitude. If the wavenumber k is complex, the wave will either grow or die out exponentially with distance, depending on the sign of the imaginary part of k.

Differential equations for W can be derived from some approximation to the physics of the system. We can then convert the differential equations to algebraic equations involving ω, k, and parameters of the system by a generalization of the technique described in Chapter 8, noting that W can be factored out of its derivatives when $A(x)$ is constant (or if A is assumed constant when it is nearly so). These algebraic equations relating ω and k are referred to as **dispersion relations**. Pairs of ω and k that satisfy the dispersion relations represent waves compatible with the physical system.

From the dispersion relations, we can calculate the velocity of the wave. If the velocity is independent of ω, all frequencies travel at the same speed and the medium is said to be **nondispersive**. In the cochlea, higher frequencies are known to propagate more slowly than do lower frequencies, and the basilar membrane is therefore **dispersive**. In a dispersive medium, we distinguish two different velocities. The **phase velocity** $c = \omega/k$ is the velocity at which any given crest or valley of the wave propagates. The **group velocity** $U = d\omega/dk$ is the speed at which the wave envelope and energy propagate.

Dispersion relations generally have symmetric solutions, such that any wave traveling in one direction has a corresponding solution, of the same frequency, traveling with the same speed in the opposite direction. If, for a given real value of ω, the solution for k is complex, then the equations imply a wave amplitude that is growing or diminishing exponentially with distance x. If the imaginary part k_i of $k = k_r + jk_i$ is positive (for the complex exponential wave conventions we have adopted), the wave diminishes toward the right ($+x$); in any dissipative system, a wave diminishes in the direction that it travels. The wave may thus be written as the damped sinusoid

$$W(x,t) = A\cos(k_r x - \omega t)e^{-k_i x}$$

Fluid Mechanics of the Cochlea

In the cochlea, finding the relations between ω and k is more complex than is determining them for a one-dimensional wave system, such as a vibrating string. We must first work out the fluid-flow problem in two or three dimensions; ultimately, we can represent displacement, velocity, and pressure waves on the cochlear partition in one dimension as the relation between k and ω changes with x (conventionally referred to as the cochlear **place** dimension).

It is highly likely that the lowest-order loss mechanism in the cochlea is the viscous drag of fluid moving in a boundary layer near the basilar membrane and through the small spaces of the organ of Corti. The sensitive cilia of the inner hair cells that detect motion are moved by viscous drag. The outer hair cells also interact with the fluid and membranes in the organ of Corti, and are known to

be a source of energy, rather than a sink. At low sound levels, the outer hair cells can supply more than enough energy to make up for the energy lost to viscous drag. We therefore model the effects of the outer hair cells as a negative damping.

The analysis of the hydrodynamic system of the cochlea yields a relation among frequency, place, and complex wavenumber. The analyses we used are based on three approximations commonly employed to make the hydrodynamics problem relatively simple:

1. We assume that the cochlear fluids have essentially zero viscosity, so the sound energy is not dissipated in the bulk of the fluid, but rather is transferred into motion of the organ of Corti
2. We assume that the fluid is incompressible, or equivalently that the velocity of sound in the fluid is large compared to the velocities of the waves on the cochlear partition
3. We assume the fluid motions to be small, so we can neglect second-order motion terms; for sound levels within the normal range of hearing, this is a good approximation

The details of the hydrodynamic analysis and reasoning about physical approximations are too lengthy to include in this chapter [Lyon et al., 1988], but we can summarize the results by the short-wave dispersion relation (with complex k), which is

$$\omega^2 \rho = \pm k[S - j\omega(\beta + k^2\gamma)] \tag{16.1}$$

where β is a low-order (viscous) loss coefficient (which may be negative in the actively undamped case), γ is a high-order (bending) loss coefficient, S is the membrane stiffness, and ρ is the mass density of the fluid. Parameters S and γ are functions of place, but at any given place are fixed for all time. The active negative damping term β also is a function of place, and changes slowly with time as the system adapts to changes in incoming sound level. The sign of the right-hand side of Equation 16.1 is taken as positive for positive ω, and negative for negative ω.

Because the physical parameters in Equation 16.1 are changing with x, a closed-form solution is not possible except under specific restrictions of form. Nevertheless, excellent approximate solutions for wave propagation in such nonuniform media are well known, and correspond to a wave propagating locally according to local wavenumber solutions. Any small section of the medium, of length Δx, over which the properties change slowly behaves just as would a small section in a uniform medium: It contributes a phase shift, $k_r \Delta x$, and a log gain, $-k_i \Delta x$. We also may need to adjust the amplitude $A(x)$ to conserve energy as energy-storage parameters such as the membrane stiffness change, even in a lossless medium.

In the cochlea, the amplitude of the pressure wave remains nearly constant as the wave propagates, but the amplitude of the velocity or displacement wave grows to conserve energy as the stiffness decreases.

Scaling

A wave medium is said to **scale** (or be **scale-invariant**) if the response properties at any point are just like those at any other point, with a change in time scale. The response properties of the cochleas of all known animals are approximately scale-invariant over most of the length of the basilar membrane. This scaling property is achieved by an exponential slowing of the wave-propagation velocity with the distance x. In a system that scales, the response for all places can be specified as a single transfer function $H(f)$,

$$f = \omega/\omega_N$$

where f is a nondimensional normalized frequency, and ω_N is any conveniently defined natural frequency that depends on the place—for example, the peak in the frequency response. Because of the assumed exponential form for the variation of parameters, we can write ω_N as

$$\omega_N = \omega_0 e^{-x/d_\omega}$$

where ω_0 is the natural frequency at $x = 0$ (at the base), and d_ω is the characteristic distance in which the velocity, and therefore ω_N, decreases by a factor of e.

Changing to a log-frequency scale in terms of $l_f = \log f$, we define the function $G(l_f) = H(f)$, which we can write as

$$G(l_f) = G\left(\log \frac{\omega}{\omega_N}\right) = G\left(\log \frac{\omega}{\omega_0} + \frac{x}{d_\omega}\right)$$

This equation shows that the transfer function G expressed as a function of log frequency l_f is identical to the transfer function for a particular frequency ω expressed as a function of place x, for an appropriate offset and place scaling. Thus, we can label the horizontal axes of transfer-function plots interchangeably with either place or log-frequency units, for a particular frequency or place respectively.

In the cochlea, the function G will be lowpass. Above a certain cutoff frequency, depending on the place, the magnitude of the response will quickly approach zero; equivalently, beyond a certain place, depending on frequency, the response will quickly approach zero.

The stiffness is the most important parameter of the cochlear partition that changes from base to apex, and it has the effect of changing the characteristic frequency scale with place (ω_N varies as the square root of stiffness if other parameters, such as duct size, are constant). Over much of the x dimension of real cochleas, the stiffness varies approximately exponentially [Dallos, 1978]. The scaling assumption simply allows us the convenience of summarizing the response of the entire system by a single function G, and does not prevent us from adopting more realistic parameter variations in the region where the variation is not exponential.

Approximate Wavenumber Behavior

It is instructive to look at approximate solutions to Equation 16.1 that make clear the dependence of the wavenumber on the parameters, because it is this dependence that relates the fluid dynamics to the circuit model we will describe later. In practice, more exact solutions for k may be achieved by Newton's method, starting from simple approximations such as those we discuss here.

Starting with real k, we obtain a first approximation to the dispersion relation by setting the imaginary part of Equation 16.1 to zero. This lossless approximation is

$$k_r \approx \frac{\omega^2 \rho}{S}$$

Using this approximate solution for k_r, we can obtain a first approximation for the imaginary part k_i by solving for the imaginary part of Equation 16.1, assuming that k_i is much less than k_r and ignoring terms with k_i^2 and k_i^3:

$$\begin{aligned} k_i &\approx \frac{k_r \beta \omega}{S} + \frac{k_r^3 \gamma \omega}{S} \\ &\approx \frac{\beta \omega^3 \rho}{S^2} + \frac{\gamma \omega^7 \rho^3}{S^4} \end{aligned}$$

Because these relations are derived in the short-wavelength limit, they are not applicable at very low frequencies. However, they can give us an excellent representation for the behavior of the peak frequency and the characteristics of the high-frequency cutoff.

We can interpret these complex wavenumber approximations either as *frequency-dependent* at a constant place (constant S, β, and γ), or as *place-dependent* at a constant frequency (constant ω). Thus, a wave of frequency ω will propagate until the damping gets large; the loss per distance, k_i, grows ultimately as ω^7 or e^{7x/d_ω} (assuming the exponential dependence of S discussed earlier, and with γ proportional to ω_N and therefore to e^{-x/d_ω}). For a given frequency, the damping is near zero for small x and becomes dominant very quickly as x approaches a **cutoff place** x_C. Similarly, at a given place, low frequencies are propagated with little loss; as ω grows, however, the loss grows quickly, and waves above a **cutoff frequency** ω_C are heavily attenuated.

The cochlea is known to have sharp cutoff behavior, so it is reasonable to suppose that only the high-order γ loss term is significant in determining the cutoff points. We can estimate the cutoff frequency to be near the point where k_i becomes comparable to k_r. Based on the previous simple approximations,

$$\omega_C \approx (S^3/\gamma \rho^2)^{1/5}$$

If the system scales, cutoff frequencies and cutoff places are related exponentially:

$$\omega_C \approx \left(S_0^3/\gamma_0 \rho^2\right)^{1/5} e^{-x/d_\omega}$$

$$x_C \approx -d_\omega \log \frac{\omega}{(S_0^3/\gamma_0 \rho^2)^{1/5}}$$

where the subscript 0 refers to values at the base ($x = 0$). If the damping and stiffness coefficients do not scale exponentially, there is still a cutoff place as a function of frequency, but it is not a simple function of $\log \omega$.

The **best frequency**, or frequency of highest wave amplitude, will be somewhat less than the cutoff frequency, for any place. Because the cilia of the hair cells sense the velocity or displacement of the membrane, we should calculate the best frequency using the velocity rather than the displacement of the cochlear partition. Conversion from pressure to velocity involves a spatial differentiation, contributing another factor of k, or a tilt in the place response of about 12 decibels per octave. The velocity will peak within less than an octave of cutoff.

The effect of the variable damping on the cutoff points is not included in our approximation, due to the assumption that the higher-order loss mechanism mainly determines the sharp cutoff. According to this model, the best frequency should shift by nearly an octave (depending on parameters) as the damping is changed. The cutoff point, however, measured as the frequency where the response decreases with a particular high slope, changes relatively little. Experimentally, a shift in best frequency of up to 0.75 octave has been observed from healthy cochleas with active outer hair cells to traumatized cochleas where the outer hair cells could no longer provide active mechanical undamping [Cody, 1980]. In these experiments, the steep high-frequency side of the response was unchanged, as we would expect.

SILICON COCHLEA

All auditory processing starts with a cochlea. Silicon auditory processing must start with a *silicon cochlea*. The fundamental structure in a cochlea is the basilar membrane. The silicon basilar membrane is a transmission line with a velocity of propagation that can be tuned electrically. Output *taps*, where the signal can be observed, are located at intervals along the line. We can think about the taps as crude inner hair cells. Unfortunately, we cannot build a system with as many taps as living systems have hair cells. Human ears have about 3500; we will be lucky to have 1000. On the other hand, we can make many delay-line elements; the delay element we use in this delay line is the second-order section, described in Chapter 11. We expect this model delay line to be good enough to duplicate approximately the dynamics of the second-order system of fluid mass and membrane stiffness, including the active role of the outer hair cells.

Basilar-Membrane Delay Line

Our basilar-membrane model is fabricated with 480 sections in the boustrophedonic arrangement illustrated by the 100-section version in Figure 16.3. The only reason for using this serpentine structure instead of a straight line is that

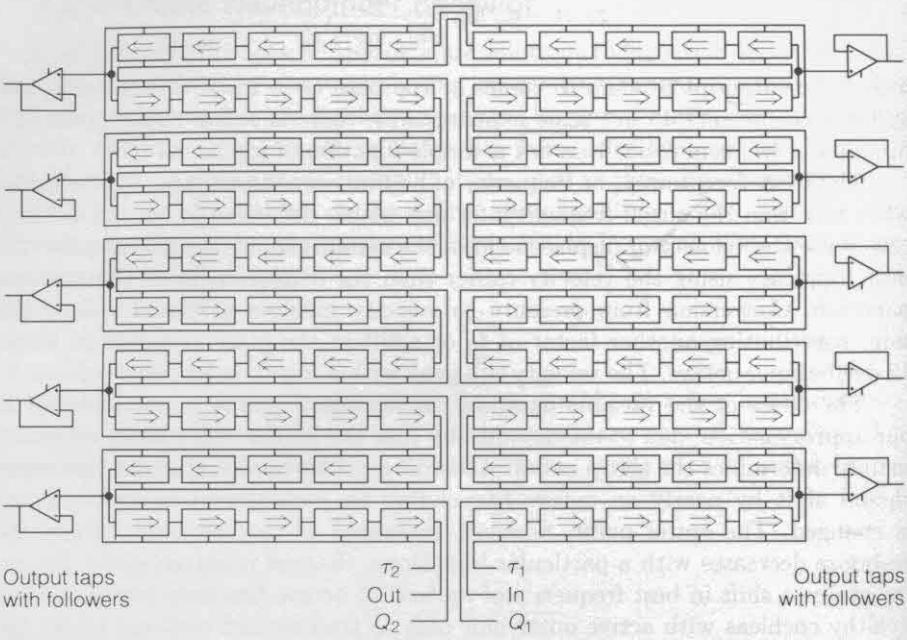

FIGURE 16.3 Floorplan of 100-stage cochlea chip, in serpentine arrangement. The wires that are shown connecting the τ and Q control terminals of the filter stages are built using a resistive polysilicon line, which acts as a voltage divider that adjusts the bias currents in the cascade as an exponential function of distance from the input. The second-order section used in this composition is shown in Figure 11.1 (p. 180).

there are many sections, each of which is longer than it is high. The chip has a reasonable aspect ratio with this floorplan. Circuit yields are good enough that we regularly are able to propagate a signal through the entire 480-stage delay line on chips from several fabrication runs.

The τ and Q bias inputs on the second-order sections are connected to polysilicon lines that run along one edge of the sections. We connect both ends of each of these resistive polysilicon lines to pads, so that we can set the voltages from offchip. Due to the subthreshold characteristics of the bias transistors in the amplifiers, the time constants of the sections are exponentially related to the voltages on the τ control line. If we put a different voltage on the two ends of the τ line, we get a gradient in voltage along the length of the polysilicon line. The subthreshold bias transistors in the transconductance amplifiers will turn this linear gradient in voltage into an exponential gradient in the delay per section. We can thereby easily make a transmission line where the velocity of propagation and cutoff frequency are exponential functions of the distance x along the line. We adjust all the sections to have the same Q value by putting a similar gradient on the Q control line, with a voltage offset that determines the ratio of feedback gain to forward gain in each section.

For a cochlea operating in the range of human hearing, time constants of about 10^{-5} to 10^{-2} second are needed. A convenient capacitor size is a fraction of a picofarad (10^{-12} farad), so the range of transconductance values needed is between 10^{-7} and 10^{-11} mho. If $kT/(q\kappa)$ is 40 millivolts, the range of bias currents will be between 10^{-8} and 10^{-12} amp. This range leaves several orders of magnitude of leeway from room-temperature thermal leakage currents at the low end, and from space-charge–limited behavior (transistors operating above threshold) at the high end, so the circuits are in many ways ideal. The total current supplied to a cascade of 480 stages is less than a microamp.

Second-Order Sections in Cascade

We can spatially discretize a nonuniform wave medium such as the cochlea by looking at the outputs of N short sections of length Δx; the section outputs are indexed by n, an integer place designator that corresponds to the x location $n\Delta x$. A cascade of second-order sections with transfer functions $H_1, H_2, \ldots, H_n, \ldots, H_N$ can be designed to approximate the response of the wave medium at the section outputs. In passing from output $n-1$ to output n, a propagating (complex) wave will be modified by a factor of $H_n(\omega)$, which should match the effect of the wave medium.

The equivalent transfer function $H_n(\omega)$, a function of place (output number n) and frequency, is thus directly related to the complex wavenumber $k(\omega, x)$, a function of place and frequency. The relation between the cascade of second-order sections and the wave medium is

$$H_n(\omega) = e^{jk\Delta x} \qquad \text{with } k \text{ evaluated at} \qquad x = n\Delta x \tag{16.2}$$

or

$$k(\omega, x) = \frac{\log H_n(\omega)}{j\Delta x} \qquad \text{for} \qquad x = n\Delta x$$

Because H and k both can be complex, we can separate the phase and loss terms using $\log H = \log|H| + j \arg H$:

$$\log H = jk\Delta x = jk_r\Delta x - k_i\Delta x$$

$$\log \text{ gain} = \log|H| = -k_i\Delta x$$

$$\text{phase lag} = \arg H = k_r\Delta x$$

Therefore, if we want to model the action of the cochlea by a cascade of second-order sections, we should design each section to have a phase lag or delay that matches k_r and a gain or loss that matches k_i, all as a function of frequency:

$$\text{gain} = e^{-k_i\Delta x}$$

$$\text{group delay} = \frac{d\text{phase}}{d\omega} = \frac{dk_r}{d\omega}\Delta x = \frac{\Delta x}{U} \tag{16.3}$$

where U is the group velocity; Equation 16.3 implies that the previous definition of group velocity, $d\omega/dk$, is correctly generalized to $d\omega/dk_r$. The overall transfer function of the cascade of second-order sections, from input to output m, which we call H^m, is

$$\begin{aligned}\mathrm{H}^m(\omega) &= \prod_{n=0}^{m} H_n(\omega) \\ &= \exp \sum_{n=0}^{m} \log H_n(\omega) \\ &= \exp j \sum_{n=0}^{m} k(\omega, n\Delta x)\Delta x \end{aligned} \tag{16.4}$$

In the cochlea, k is nearly real for frequencies significantly below cutoff; that is, the second-order sections are simply lossless delay stages at low frequencies. The gains may be slightly greater than unity at middle frequencies, when the input signal is small and the cochlea is actively undamped; at high frequencies, however, the gains always approach zero. Near cutoff, a small change in the value of k_i corresponds to a small change in the gain of any one section, but to a potentially large change in the overall gain of the cascade.

These formulae (from Equation 16.2 through Equation 16.4) provide a way to translate between a distributed-parameter wave view and a discrete delay-section view of the cochlea. The discrete-section model will be realistic to the extent that waves do not reflect back toward the base and that the sections are small enough that the value of k does not change much within a section. In our experimental circuits, k may change appreciably between the rather widely spaced output taps, so several delay sections are used per tap.

Figure 16.4(a) shows the response of a single second-order section from the transmission line. The curves were computed from Equation 16.4 for Q values of 0.7, 0.8, and 0.9; scaled versions of the $Q = 0.9$ transfer function, from earlier stages in a cascade, also are shown. For this application, we use Q values of less than 1.0, which means that the peak of the single-section response is very broad and has a maximum value just slightly greater than unity for Q greater than 0.707.

Because the system scales, each second-order section should have a similar response curve. The time-constant of each section is larger than that of its predecessor by a constant factor $e^{\Delta x/d_\omega}$, so each curve will be shifted along the log-frequency scale by a constant amount $\Delta x/d_\omega$. The overall response is the product of all the individual curves; the log response is the sum of all of the logs, as shown in Figure 16.4(b). In terms of the normalized log–log response G, the overall response is simply

$$\log \mathrm{H}^m(\omega) = \sum_{n=0}^{m} \log G\left(\log \frac{\omega}{\omega_0} + \frac{n\Delta x}{d_\omega}\right)$$

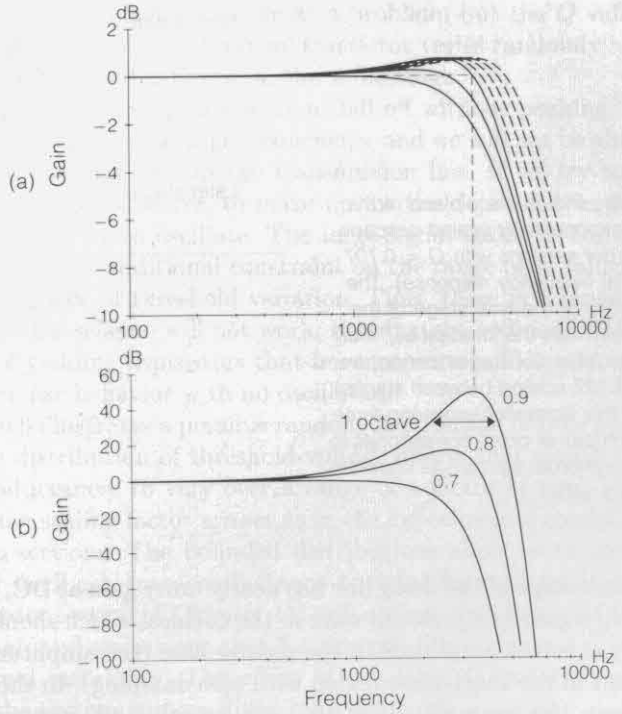

FIGURE 16.4 (a) Frequency response of a single second-order filter section, for Q values 0.7, 0.8, and 0.9, including scaled copies (dashed lines) of the $Q = 0.9$ response to represent earlier sections in a cascade. (b) Corresponding overall response of a cascade of 120 stages with a factor of 1.0139 scaling between adjacent stages. The dashed line between (a) and (b) indicates that the overall response peak occurs at the frequency for which the final section gain crosses unity. Note the different decibel scale factors on the ordinates.

Taking the section illustrated in Figure 16.4(a) as the last section before the output tap and working backward to sections of shorter time constants, we obtain the overall response in Figure 16.4(b). Each response curve has a maximum gain slightly larger than unity. There are many sections, and each one is shifted over from the other by an amount that is small compared with the width of the peak. Although there is not much gain in each section, the cumulative gain causes a large peak in the response. This overall-gain peak is termed a **pseudoresonance**, and is much broader (less sharply tuned) than is a single resonance of the same gain.

Figure 16.5 shows the s-plane plot of the poles of an exponentially scaling cascade of second-order sections with $Q = 0.707$. The time constant τ of the final stage of the cascade determines the smallest $\omega_N = 1/\tau$, indicated by the circle in the figure. For clarity, only six stages per octave (a factor of 1.122), covering only three octaves, are used for this illustration.

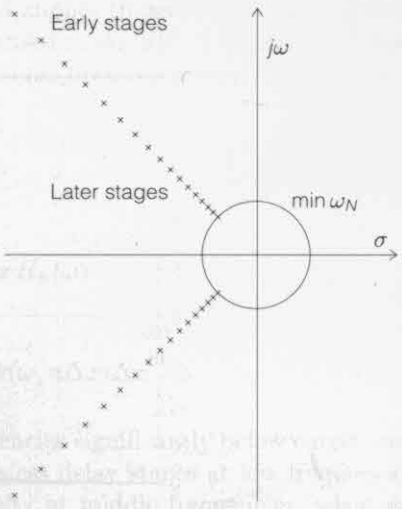

FIGURE 16.5 Plot in the s-plane of the poles of an exponentially scaled cascade of second-order sections with $Q = 0.707$ (maximally flat frequency response). The time constant τ of the final stage of the cascade determines the smallest ω_N, indicated by the circle. Six stages per octave (a factor of 1.122 scaling between stages) are used for this illustration, covering three octaves. Each pair of poles corresponds to a single second-order section, as shown in Figure 11.1 (p. 180).

Because each stage of the delay line has nearly unity gain at DC, we interpret the propagating signal as a pressure wave in the cochlea, which should propagate with a roughly constant amplitude in the passive case (the amplitude should be exactly constant in the short-wave region with zero damping). In the active case, a significant gain (for example, 10 to 100) can be achieved, measured in terms of the pressure wave.

We can design output-tap circuits to convert the propagating pressure wave into a signal analogous to a membrane deflection-velocity wave. Ideally, we would do this by a spatial differentiation to convert pressure to acceleration, followed by a time integration to convert to velocity. The combination would be exactly equivalent to a single time-domain filter, which can be approximated in the short-wave region by the approximate differentiator $\tau s/(\tau s+1)$, with τ adjusted to correspond roughly to the τ of each section; this filter tilts the low side of the response to 6 decibels per octave, without having much effect on the shape or sharp cutoff of the pseudoresonance. The most effective such circuit we have built and tested is the hysteretic differentiator described in Chapter 10. The results reported in this chapter, however, are based on the second-order delay line alone.

Transistor Parameter Variation

Our electronic cochlea would be ideal if Figure 16.4 showed the real picture. As we have noted in Chapter 5, however, MOS transistors are not inherently well matched. For a given gate voltage, the current in the subthreshold region where our circuits operate can vary randomly over a range of a factor of about two. In the response curves of the real second-order sections, there is a dispersion in the center frequencies of the sections because of this random variation in currents.

That dispersion would not be too great a problem, but the Q values also vary, because the threshold of the Q control transistor varies randomly and is not well correlated with the τ adjustment on the same stage.

If the responses of too many sections fall off without peaking, the collective response will be depressed at high frequencies, and we will not be able to maintain a good bandwidth and gain in the transmission line. If we try to increase the value of Q too much, however, to make up for the depressed response, some of the sections will start to oscillate. The large-signal stability limit described in Chapter 11 places an additional constraint on the range of Q values than can be tolerated in the face of threshold variation. Thus, there is a range of variations beyond which this scheme will not work. Fortunately, ordinary CMOS processes are capable of yielding transistors that have currents sufficiently well controlled to obtain cochlear behavior with no oscillation.

Figure 16.6 illustrates a possible random distribution of pole positions, based on a uniform distribution of threshold-voltage offsets that would cause currents and transconductances to vary over a range of a factor of two. The nominal Q value and the τ scaling factor are set as in the experimental conditions discussed the next two sections. The bounded distributions allow us to compute that α (the ratio of feedback transconductance to total forward transconductance in the second-order section of Chapter 11) will change by a factor of two above and below nominal, and we can see that for this condition it is not difficult to avoid the large-signal instability. The effect of the large Q variation on the overall response of the system is more difficult to estimate.

Frequency Response

The most straightforward behavior of the silicon cochlea that can be compared with theory is the magnitude of the frequency response. The circuit is set up with a gradient in the τ such that each section is slower than its predecessor by a factor of 1.0139. With this value, the auditory range is covered in the 480 sections. The Q control voltage is set so that the peak response is about five times the DC response, as seen at several different output taps. Experimental data were taken with a sine wave of 14 millivolts peak-to-peak amplitude applied to the input. Results for two taps 120 sections apart are shown in Figure 16.7. The solid points are measured values, and the smooth curves are theoretical predictions. Each curve is constructed as a product of individual section response curves as given by Equation 11.4 (p. 181). The value of the DC gain of the amplifiers is determined from the ratio of the response peaks. The value of Q used in the theory is adjusted until the predicted peak heights agreed with observation. The resulting Q value is 0.79. The lower-frequency peak corresponds to a tap farther from the input by 120 sections than is the higher-frequency peak; the signal level at the second tap has thus suffered a degradation due to the DC gain of 120 followers in cascade. The open-circuit amplifier gain inferred from this observation is 1800, in good agreement with measurements on other amplifiers of similar design.

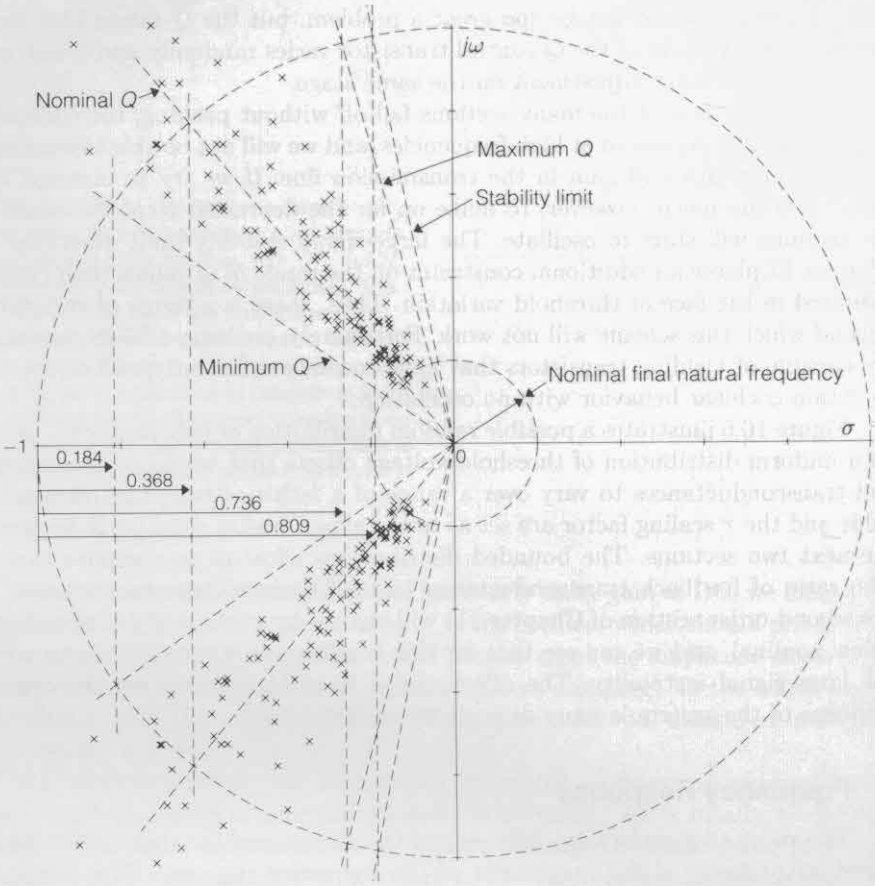

FIGURE 16.6 Plot in the *s*-plane of the poles of an exponential cascade of second-order filter sections with random threshold offsets (Q = 0.79 nominal, stage ratio = 1.0139, 120 stages covering about a factor of five in frequency). The threshold offset was modeled as a uniform distribution of bias current over a factor-of-two range, resulting in α parameter variation of up to a factor of two above and below nominal (factor-of-four range).

The remarkable agreement between theory and experiment is surprising in view of known random variations in transistor input offset voltages. We would expect a variation in Q and τ values for each section that is much larger than is the systematic progression between adjacent amplifiers. This variation need not, however, have a drastic effect on the result. The total response is the product of the responses of a large number of amplifiers. The product is an associative operation—it does not depend on the order of the terms. The fact that amplifiers in a particular physical location do not have precisely the τ value that we desire does not matter. Some amplifier somewhere will have that τ value, and it will make its contribution as required. It is more surprising that the random variation

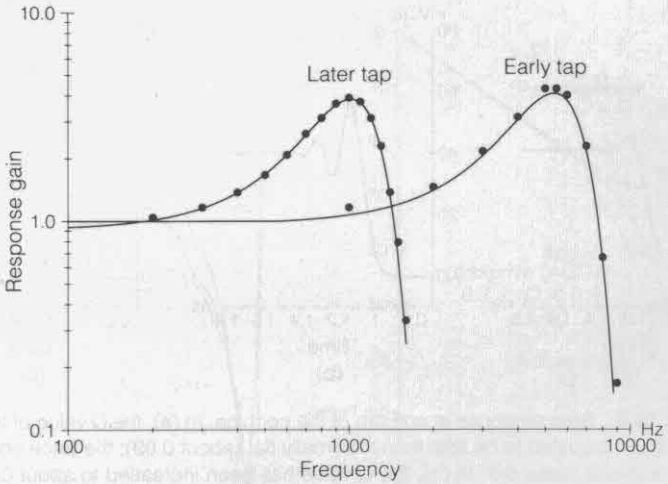

FIGURE 16.7 Log–log frequency response measured at two output taps of an experimental silicon cochlea 120 sections apart. The experimentally measured points (dots) agree quite well with the theoretical curves (using empirically fitted values of $Q = 0.79$, stage ratio = 1.0139, and DC gain = 1800). The DC-gain parameter (or open-loop gain of the transconductance amplifier) provides a correction that shifts the response at the later tap downward by $(1800/1801)^{120} = 0.936$ relative to the earlier tap.

of Q values does not affect the result in a more violent way. As of now, we do not have a satisfactory theory explaining the composite response curve resulting from many curves of different individual Q values.

Transient Response

The response at one tap of the cochlea to a step input is shown in Figure 16.8. In part (a), the Q value of the delay line has been adjusted to be just slightly less than 0.707; the trace shows only a slight resonant overshoot. In part (b), the Q value has been increased, and more overshoot is evident. In part (c), the Q value is considerably higher, and the delay line rings for several cycles. If the Q value were automatically adjusted, as it is in living systems, part (a) would correspond to the response at high background sound levels, and part (c) would correspond to a very quiet environment. With the Q value adjustment corresponding to part (c), we can observe the response at the two taps along the delay line where the frequency response of Figure 16.7 was measured. The result is shown in Figure 16.9. These results illustrate the scale-invariance property that is unique and valuable about this structure: When we adjust the time scale on the oscilloscope to correspond to the τ value of the particular section being observed, we obtain a similar output waveform at every tap. Living systems use this principle so that the detection machinery does not depend on the position along the basilar membrane. We will use it for the same purpose.

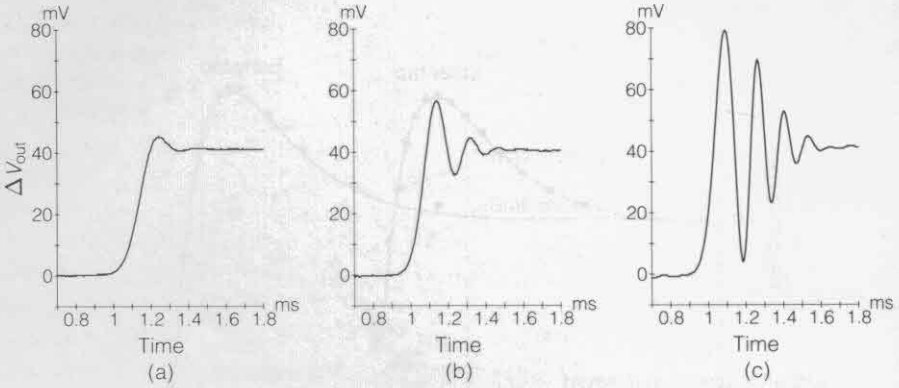

FIGURE 16.8 Step response at one tap of the cochlea. In (a), the Q value of the delay line has been adjusted to be less than maximally flat (about 0.69); the trace shows only a slight resonant overshoot. In (b), the Q value has been increased to about 0.74, and more overshoot is evident. In (c), the Q value is considerably higher (about 0.79), and the delay line "rings" for several cycles.

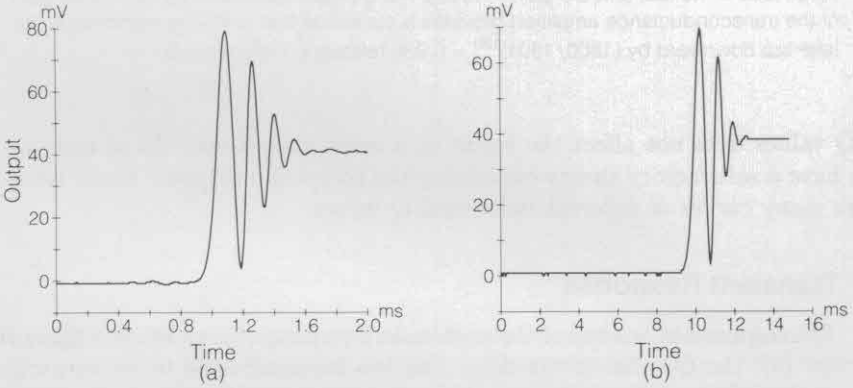

FIGURE 16.9 Step responses at two taps separated by 120 delay stages. Note that the second response (b) has about 10 times as much delay as the first one has (a), and that it has a slightly faster rise time relative to the delay. Aside from the relation of rise time to delay, the response at the two taps is similar, with the time scaled by about a factor of 10. This behavior gives rise to the scale-invariance property of the cochlea.

GAIN CONTROL

The function of the outer-hair-cell arrangement is to provide not just gain, but also control of the gain, which it does by a factor of about 100 in amplitude (10,000 in energy). When the signal is small, the outer hair cells are not inhibited and they feed back energy. This AGC system works for sound power levels within a few decades of the bottom end of our hearing range by making the

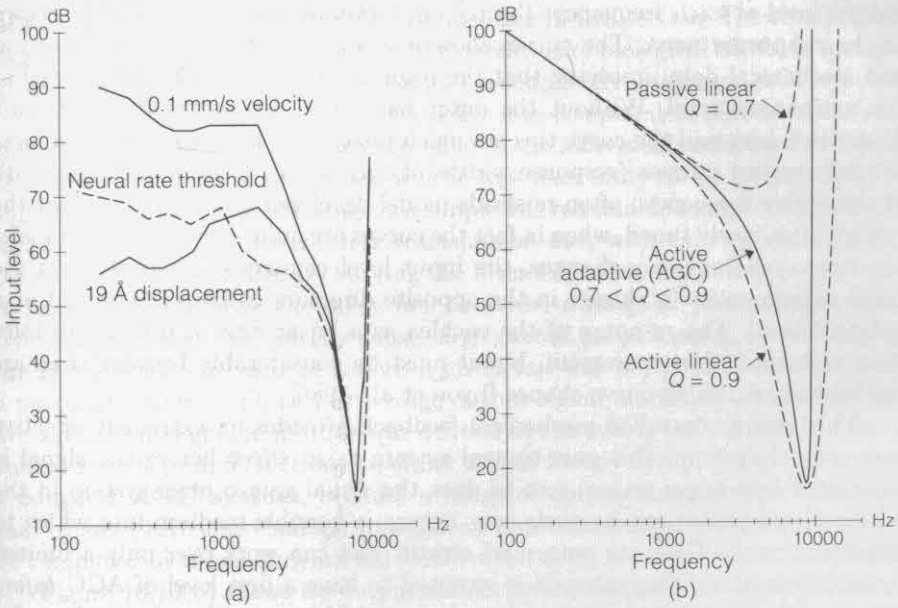

FIGURE 16.10 (a) Mechanical and neural iso-output tuning curves. Based on data from Robles and his associates [Robles et al., 1986]. The mechanical measurements (amount of input needed to get 1 millimeter per second basilar-membrane displacement velocity, or 19-angstrom basilar-membrane displacement amplitude) were made by measuring doppler-shifted gamma rays (Mössbauer effect) from a small radioactive source mounted on the cochlear partition. The neural tuning curve was measured by looking for a specified increase in firing rate of a single fiber in the cochlear nerve. (b) Iso-output tuning curves for the second-order model described in this chapter, under three operating conditions. The dashed curves are the response for fixed Q, independent of input level. The solid curve is the result for a particular nonlinear AGC scheme that reduces the Q of the cascaded filter stages as the output signal increases. The similarity in the response area shape and width between the active adaptive model and the biological system (a) is striking. For this simulation, all filter-stage Q values are equal, and are computed from a feedback gain α that is a maximum value (0.5) minus a constant, times the total output of 100 channels covering about an octave in each direction from the channel being measured. The relation of Q values to overall gains and overall transfer functions is discussed in the text.

structure slightly more resonant and thereby of much higher gain—by reducing the damping until it is negative in some regions.

Figure 16.10(a) shows data from a biological cochlea. The sound-pressure level required to produce a fixed membrane-displacement amplitude and velocity is plotted as a function of frequency. Also plotted is the input level required to produce a certain increase in the rate of firing of a single auditory nerve fiber. The data were obtained by Robles and his associates [Robles et al., 1985], using the Mössbauer effect in the chinchilla cochlea. Curves such as these are termed **iso-output** curves, because the input level is adjusted to produce the same

output level at each frequency; the region above an iso-output curve is known as the **response area**. The curves show reasonable agreement between neural and mechanical data, implying that the response area is already determined at the mechanical level. Without the outer hair cells, the sensitivity is at least 30 decibels less, and the curve tips are much broader [Kim, 1984]. The sharpness of such **tuning curves** (response widths of only about one-fourth to one-tenth of the center frequency) often misleads model developers into thinking that the system is narrowly tuned, when in fact the curves are quite different from transfer functions. As frequency changes, the input level changes enormously, and the AGC system causes a change in the opposite direction to keep the output at a constant level. The response of the cochlea as a linear filter is difficult to infer from this kind of measurement, but it must be considerably broader than are the iso-output tuning-curve shapes [Lyon et al., 1986].

This use of controlled mechanical feedback provides an extremely effective gain-control system. This gain-control system takes effect before the signal is translated into nerve pulses, just as does the visual gain-control system in the retina. Nerve pulses are, by their very nature, a horrible medium into which to translate sound. They are noisy and erratic, and can work over only a limited dynamic range of firing rates. It is essential to have a first level of AGC *before* the signal is turned into nerve pulses, because this approach reduces the noise associated with the quantization of the signal.

We can model the effect of the outer hair cells as a negative damping. When the low-order loss term β in Equation 16.1 is negative, the system exhibits gain until, at high enough frequency, the higher-order γ loss mechanism dominates. If we were to model such a traveling-wave system with active gain as a time-invariant linear system, the system would have a fixed gain, independent of the sound input level. The live cochlea, however, is known to be highly adaptive and compressive, such that the mechanical gain is much less for loud inputs than for soft inputs. This nonlinear (but short-term nearly linear) behavior is necessary for two reasons. First, the mechanism that adds energy must be energy-supply limited, and therefore the gain cannot extend to arbitrarily high signal levels. Second, even at relatively low sound levels, the variation of the gain is needed to compress inputs into a usable dynamic range, without causing excessive distortion.

Many researchers who have attempted to model active wave amplification in the cochlea have met with difficulties, especially when the place dimension was discretized for numerical solution. De Boer has shown that slight irregularities in the cochlea can reflect enough energy to make the system break into unstable oscillations (as in tinnitus) [de Boer, 1983]; models that use discrete sections and allow waves to propagate in both directions sometimes suffer from the same problem. By taking advantage of the known normal mode of cochlea operation in which signals propagate in only one direction, our circuit model avoids the stability problem, as long as each section is independently stable.

We have not, at the present time, integrated the control system for automatically adjusting the Q values of the second-order sections onto the same silicon as the basilar-membrane model. We have, however, built a computer simulation

of such a control system, which we describe later in this section. The results from this model shed considerable light on the operation of biological cochleas, and are guiding our design of a silicon cochlea with a completely integrated AGC system.

We have seen that the gain of a delay line composed of second-order sections is a sensitive function of the Q value of the individual sections. For Q less than 0.707, the gain of each stage will be less than unity at all frequencies; for slightly higher Q values, the stage is a simple but reasonable model of an actively undamped section of the cochlear transmission line, with gain exceeding unity over a limited bandwidth. By varying the filter's Q value adaptively in response to sound, we can cause the delay line to model a range of positive and negative damping, and can thereby cause large overall gain changes. In our model, the Q value for 120 sections before a given tap was adjusted downward from a maximum value of 1.0 as the average output-signal amplitude increased. The average output-signal amplitude was defined as the average of the amplitudes of output-signals from 50 sections on either side of the given tap; the model is thus a **coupled AGC system**, because a range of output channels can affect the gain of any particular channel. In a silicon implementation, this average would be computed by a one-dimensional resistive network, as described in Chapter 7.

Figure 16.10(b) shows iso-output curves for the cochlear model described in the previous paragraph, under two linear conditions (a passive low-gain condition, as in a cochlea with dead or damaged outer hair cells, and an active high-gain condition, as in a hypothetical cochlea with active outer hairs of constant gain and unlimited energy), and under the condition where the AGC system acts to adapt the gain between the two linear conditions in response to the output level averaged over nearby channels. The curves show that a simple coupled gain-control loop can cause a broadly tuned filter to appear to have a much narrower response than does a similar filter without AGC, when observed with an iso-output criterion; this result is in excellent agreement with the experimental data on the biological system in Figure 16.10(a).[1] The model also predicts that higher signal levels will cause an increase in effective bandwidth and a reduction in phase shift or delay near the best frequency (but a slightly increased phase lag below the best frequency), in agreement with physiological observations [Pickles, 1985; Sellick et al., 1982].

SUMMARY

The cochlea is a traveling-wave structure that creates the first-level representation in the auditory system. It converts time-domain information into spatially encoded information by spreading out signals in space according to their time

[1] Our scale-invariant model produces tuning curves that are similar to those observed in biological systems in the basal region, but are too sharp in the apical region [Dallos, 1988]. In the real cochlea the frequency-place mapping becomes nearly linear in the apical region, so waves are amplified over a more limited region, resulting in less AGC effect and less iso-output sharpness.

scale. The velocity of propagation along the structure decreases exponentially with distance, so the spatial pattern generated by a certain time sequence is independent of the rate at which the sequence is presented. Faster sequences create output patterns closer to the input of the structure; slower sequences generate output patterns nearer the output of the structure.

The silicon model of this traveling-wave structure exhibits behaviors that bear an uncanny resemblance to those of the living system. The effect of variation in transistor parameters is insignificant in the operation of the system, which is determined by the collective action of many sections. This model has allowed us to sharpen our understanding of nature's solution to the hearing problem. It also is an effective solution to a difficult engineering problem.

REFERENCES

Dallos, P. Biophysics of the cochlea. In Carterette, E.C. and Friedman, M.P. (eds), *Handbook of Perception*, vol 4. New York: Academic Press, 1978, p. 125.

Dallos, P. Personal communication, 1988.

de Boer, E. Wave reflection in passive and active cochlea models. In de Boer, E. and Viergever, M.A. (eds), *Mechanics of Hearing. Proceedings of the IUTAM/ICA Symposium*, The Hague: The Netherlands, Martinus Nijhoff Publishers, 1983, p. 135.

Kim, D.O. Functional roles of the inner– and outer–hair-cell subsystems in the cochlea and brainstem. In Berlin, C.I. (ed), *Hearing Science: Recent Advances*. San Diego, CA: College-Hill Press, 1984, p. 241.

Lyon, R.F. and Dyer, L. Experiments with a computational model of the cochlea. *Proceedings of the IEEE International Conference on Acoustical Speech and Signal Processing*, vol 3. Tokyo, Japan: IEEE, 1986, p. 1975.

Pickles, J.O. Recent advances in cochlear physiology. *Progress in Neurobiology*, 24:1, 1985.

Robles, L., Ruggero, M.A., and Rich, N.C. Mössbauer measurements of the mechanical response to single-tone and two-tone stimuli at the base of the chinchilla cochlea. In Allen, J.B., Hall, J.L., Hubbard, A., Neely, S.T., and Tubis, A. (eds), *Peripheral Auditory Mechanisms*. Berlin, New York: Springer-Verlag, 1986.

Sellick, P.M., Patuzzi, R., and Johnstone, B.M. Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique. *Journal of the Acoustical Society of America*, 72:131, 1982.

von Békésy, G. *Sensory Inhibition*. Princeton, NJ: Princeton University Press, 1967.

Zurek, P.M. and Clark, W.W. Narrow-band acoustic signals emitted by chinchilla ears after noise exposure. *Journal of the Acoustical Society of America*, 70:446, 1981.