

## A CMOS VLSI Cochlea

Richard F. Lyon

Schlumberger Palo Alto Research  
3340 Hillview Avenue  
Palo Alto, CA 94304

and

Carver A. Mead

California Institute of Technology  
Pasadena, CA 91125

## Abstract

An engineered system that hears, such as a speech recognizer, can be designed by modeling the cochlea, or inner ear, and higher levels of the auditory nervous system. To be useful in such a system, a model of the cochlea should incorporate a variety of known effects, such as an asymmetric lowpass/bandpass response at each output channel, a short ringing time, and active adaptation to a wide range of input signal levels. An analog electronic cochlea has been built in CMOS VLSI technology using micropower techniques to achieve this goal of usefulness via realism. The key point of the model and circuit is that a cascade of simple, nearly linear, second-order filter stages with controllable  $Q$  parameters suffices to capture the physics of the fluid-dynamic traveling-wave system in the cochlea, including the effects of adaptation and active gain involving the outer hair cells. Measurements on the test chip suggest that the circuit matches both the theory and observations from real cochleas.

## 1 Introduction

When we understand how hearing works, we will be able to build amazing machines with brain-like abilities to interpret the world through sounds (i. e., 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 systems that hear in simple ways. The relative success of these 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 approach used to model the nonuniform fluid-dynamic wave medium of the cochlea as a cascade of filters is based on the observation that the properties of the medium change only slowly and that wave energy is therefore not reflected to any significant degree [de Boer 80]. The effect of active outer hair cells is included as a variable negative damping term; the variable damping mechanism is shown to be effective as a wide-range automatic gain control (AGC) associated with a moderate change in sharpness of tuning with signal level. The cascade structure enforces unidirectionality, so a coarse discretization in space does not introduce reflections that could otherwise cause instability in an active model. The system is not highly tuned, but rather achieves a high-gain *pseudoresonance* by combining the modest gains of many stages; nevertheless, sharp iso-output tuning curves result from the interaction of the adaptive gains and the filters, as has been observed experimentally in both neural frequency threshold curves and basilar membrane iso-velocity curves.

The cascade filter model is implemented using continuous-time analog CMOS (complementary metal oxide semiconductor) VLSI (very large scale integrated) circuits. The building block of the filters is a transconductance amplifier operated in a subthreshold micropower regime, in which the long time constants needed for audio processing can be controlled easily without the use of clock signals. Variable- $Q$  second-order filter stages in cascade are adequate to capture most of the important qualitative effects of real cochleas.

## 2 Cochlear Model Background

The cochlea has been modeled by many researchers, from a wide variety of viewpoints. For our purposes, a good model is one that can take real sound as input, in real time, and produce output that resembles the signals on the cochlear nerve. The circuit models discussed in this paper are inherently real-time, and share important concepts of structure and function with a previous generation of computational models [Lyon 82, Lyon 85].

The two most important concepts kept from the computational models are that (1) the filtering is structured as a cascade of low-order filter stages, and (2) adaptation is accomplished by some form of *coupled* AGC that is able to change the gains to different output taps separately, but not independently.

Three significant differences between the older computational models and our current circuit models are that (1) pseudoresonance concepts from 2D and 3D models replace the older resonant-membrane 1D (or long-wave) model, (2) adaptive active outer hair cells replace hypothetical mechanisms of gain control, and (3) the implementation is continuous-time analog rather than discrete-time digital.

## 3 Nonuniform Media and Filter Cascades

The traveling-wave dispersion relations that must be satisfied by values of frequency  $\omega$  and wavenumber (spatial frequency)  $k$  involve the physical parameters of the cochlear partition and cochlear ducts, which are changing with the  $x$  (*place*) dimension. The differential equations that describe the nonuniform physical system are not solvable 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  $\Delta x$  over which the properties do not change much behaves just as would a small section in a uniform medium: It contributes a phase shift  $k_r \Delta x$  and a log gain  $-k_i \Delta x$  (where the complex wavenumber  $k$  is equal to  $k_r + jk_i$ ). The amplitude  $A(x)$  also may need to be adjusted to conserve energy as energy-storage parameters such as spring constants (membrane stiffness) change, even in a lossless medium (these observations are equivalent to the *WKB approximation* often invoked to solve cochlear model differential equations).

The wave amplitudes  $A(x)$  for pressure or velocity potential in a passive lossless cochlea are roughly constant (in the case of a 2D short-wave approximation, their amplitudes decrease as  $k^{-0.25}$  as the wavenumber  $k$  increases). Conversion to basilar-membrane displacement or velocity involves a spatial differentiation, so amplitudes will increase as  $k^{0.75}$  as the waves travel. Loss terms tend to be high order (i. e., roughly proportional to a high power of frequency), so losses reduce the wave energy quickly near cutoff, more than canceling the slowly increasing  $A(x)$ . The resulting rise and fall has been termed a *pseudoresonance* [Holmes 83]; it does not involve a resonance between membrane mass and stiffness, as some models do, and it is not sharply tuned. Low-order passive-loss or active-gain terms mainly affect the height of the pseudoresonance, and have relatively less effect on its peak position and sharpness.

A wave-propagation medium can be approximated (for waves traveling in one direction) by a cascade of linear filters. A filter is typically characterized by its transfer function  $H(\omega)$ . A nonuniform wave medium such as the cochlea can be spatially discretized by looking at the outputs of  $N$  short sections of length  $\Delta x$ ; the section outputs, or *taps*, can be indexed by  $n$ , an integer place designator that corresponds to the  $x$  location  $n\Delta x$ . A cascade of filters  $H_1, H_2, \dots, H_n, \dots, H_N$  can be designed to approximate the response of the wave medium at the output taps. In passing from tap  $n-1$  to tap  $n$ , a propagating (complex) wave will be modified by a factor of  $H_n(\omega)$ , which should match the effect of the wave medium.

Therefore, if we want to model the action of the cochlea by a cascade of simple filters, each filter should be designed to have a phase shift or delay that matches  $k_r$  and a gain or loss that matches  $k_i$ , all as a function of frequency

$$\text{gain} = e^{-k_i \Delta x} \quad (1)$$

$$\text{group delay} = \frac{d\text{phase}}{d\omega} = \frac{dk_r}{d\omega} \Delta x \quad (2)$$

The overall transfer function of the cascade of filters, from input to tap  $m$ , which we call  $H^m$ , is

$$\begin{aligned} H^m(\omega) &= \prod_{n=0}^m H_n(\omega) \\ &= \exp \sum_{n=0}^m \log H_n(\omega) \\ &= \exp j \sum_{n=0}^m k(\omega, n\Delta x) \Delta x \end{aligned} \quad (3)$$

(The integral form of the sum in the last line in Equation 3 is the form usually used as the WKB approximation:  $\int k dx$ .)

In the cochlea,  $k$  is nearly real for frequencies significantly below cutoff; that is, the filters 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 frequencies significantly above cutoff, 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 a section and a potentially large change in the overall gain of the cascade.

#### 4 Active and Adaptation Undamping

The lowest-order loss mechanism in the cochlea probably is a 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. Apparently, at low sound levels, the outer hair cells can supply more than enough energy to make up for the energy lost to viscous drag. Looking at the effect of the outer hair cells as negative resistances (being added to the positive resistances of viscous loss), we find that the low-order loss term becomes a low-order gain term; that is, the wavenumber has a negative imaginary part until, at high enough frequency, a higher-order loss mechanism dominates.

A system with active gain can be modeled easily as a time-invariant linear system; it will have the same gain no matter how loud the sound input is. 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 [Robles 85]. This nonlinear (but short-term nearly linear) behavior makes sense for several reasons: first, the mechanism that adds energy must of necessity be energy-supply limited; and second, even at relatively low levels the variation of the gain can be useful in compressing inputs into a usable dynamic range, without causing excessive distortion.

#### 5 Fluid Mechanical Results

The analysis of the hydrodynamic system of the cochlea yields a relation among frequency, place, and complex wavenumber.

The details of the hydrodynamic analysis and reasoning about physical approximations are too lengthy to include in this paper, but we can summarize the results by the short-wave dispersion relation (with complex  $k$ ), which is

$$\omega^2 \rho = \pm k[S - j\omega(\beta + k^2 \gamma)] \quad (4)$$

where  $\beta$  is a low-order (viscous) loss coefficient (which may be negative in the actively undamped case),  $\gamma$  is a high-order (bending) loss coefficient,  $S$  is the membrane stiffness, and  $\rho$  is the mass density of the fluid. Parameters  $S$ ,  $\beta$ , and  $\gamma$  change with place, but only  $\beta$  is considered to change with time as the system adapts.

### 6 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. Signal output taps are located at intervals along the line; we can think about the taps as crude inner hair cells that transduce the propagating wave into an observable form.

#### 6.1 CMOS Circuit Building Blocks

The most useful fundamental building blocks used in all our analog integrated circuit work are transconductance amplifiers and capacitors, both of which are made up of  $p$ -type and  $n$ -type MOS transistors.

The MOS transistor is used in strong inversion (above threshold) as a capacitor, and in weak inversion (subthreshold) as an active device. In the subthreshold region, the transistor's drain current is exponential in the gate-source voltage. The relative advantages of operating transconductance amplifiers in this low-current exponential region are discussed by Vittoz [Vittoz 85].

We can think of the transconductance amplifier as an operational amplifier with well-controlled limitations; in particular, although we can think of open-loop voltage gain as being very high (greater than 1000), the gain from voltage difference input to current output (i. e., the transconductance) is limited and well controlled. A single bias-control voltage input on the transconductance amplifier sets the current through a differential-amplifier stage, thereby setting both the transconductance and the saturated output current.

The ideal transfer characteristic of the transconductance amplifier, for MOS transistors in the subthreshold region, is

$$I_{\text{out}} = I_B \tanh \frac{V_+ - V_-}{V_T} \quad (5)$$

where  $I_B$  is the amplifier's bias current and  $V_T$  is a characteristic voltage near 40 mV equal to  $kT/q\kappa$ , in which  $k$  is Boltzmann's constant,  $T$  is absolute temperature,  $q$  is the charge of an electron, and  $\kappa$  is a capacitive divider ratio, typically around 0.6, that characterizes how much of the potential change on the gate of a transistor causes a change in surface potential in the channel region of the transistor.

For input voltage differences  $V_+ - V_-$  less than about 40 mV, we can approximate the tanh function as an identity, so that the amplifier is roughly linear with a transconductance of  $I_B/V_T$ . For large input differences, the output current saturates smoothly at  $\pm I_B$ .

For the cochlea chip we use the wide-range transconductance amplifier circuit, which uses nine transistors [Vittoz 85]; the final output transistors are made extra-long, to maximize output impedance and hence voltage gain.

The next-level building block used in many analog systems is a first-order lowpass filter, consisting of a transconductance amplifier with a capacitor on its output, which we refer to as a follower-integrator [Mead 88]. For low frequencies, it acts as a unity-gain follower; for high frequencies, it acts as an integrator. The amplifier's bias control terminal, often referred to as the " $\tau$  knob," sets the transconductance  $G$ , and hence the time constant  $\tau = C/G$  in the first-order transfer function

$$\frac{V_2}{V_1} = \frac{1}{\tau s + 1} \quad (6)$$

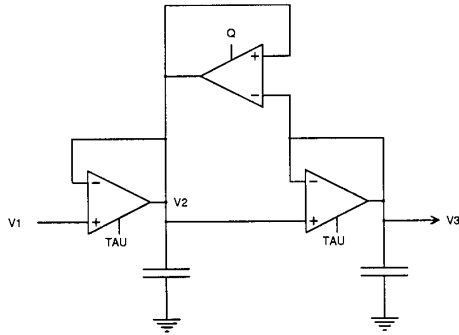


Figure 1: Second-order filter-section circuit.

### 6.2 Second-Order Section

The second-order circuit is shown in Figure 1; it contains two cascaded follower-integrators and an extra amplifier. The capacitance  $C$  is the same for both stages ( $C_1 = C_2 = C$ ), and the transconductances of the two feed-forward amplifiers, A1 and A2, are the same:  $G_1 = G_2 = G$ . An underdamped response is obtained by adding the *feedback amplifier* A3, with transconductance  $G_3$ ; its output current is proportional to the difference between  $V_2$  and  $V_3$ , but the sign of the feedback is *positive*: For small signals,  $I_3 = G_3(V_2 - V_3)$ .

The transfer function of the circuit is

$$H(s) = \frac{V_3}{V_1} = \frac{1}{\tau^2 s^2 + 2\tau s(1 - \alpha) + 1} \quad (7)$$

where  $\tau = C/G$  and  $\alpha = G_3/(G_1 + G_2)$ .

The second-order transfer function can be expressed in terms of the  $Q$  parameter that describes the degree of resonance (or *quality*) of a second-order system:

$$H(s) = \frac{1}{\tau^2 s^2 + \frac{1}{Q}\tau s + 1} \quad (8)$$

By comparison with Equation 7, we find that

$$Q = \frac{1}{2(1 - \alpha)} \quad (9)$$

Note that  $Q$  starts from 0.5 with no feedback ( $\alpha = 0$ ) and grows without bound as the feedback gain approaches the total forward gain ( $\alpha = 1$ , or  $G_3 = G_1 + G_2$ ).

Examination of this well-known transfer function reveals that the gain is always less than unity whenever  $Q$  is less than 0.707. We will use this region in modeling the cochlea without active gain. When active gain is included, a gain slightly greater than unity is needed at middle frequencies, and this can be achieved by using  $Q$  values up to about 1.0.

We have not yet defined damping quantitatively. In second-order systems, damping often is defined as  $\xi = \frac{1}{2Q}$ , which goes to 0 as  $Q$  goes to infinity, and becomes negative for poles in the right half of the  $s$  plane. Second-order systems with any amount of negative damping are unstable—any small disturbance will grow exponentially, at a controlled rate but without bound. We have used damping differently, as applied to wave propagation: When wave damping is negative, a disturbance will grow exponentially in space, but the growth will be limited due to the limited range of places for which the damping remains negative. Therefore, in second-order stages cascaded to model wave propagation, negative damping corresponds to a stage gain greater than unity (i. e.,  $Q > 0.707$ ).

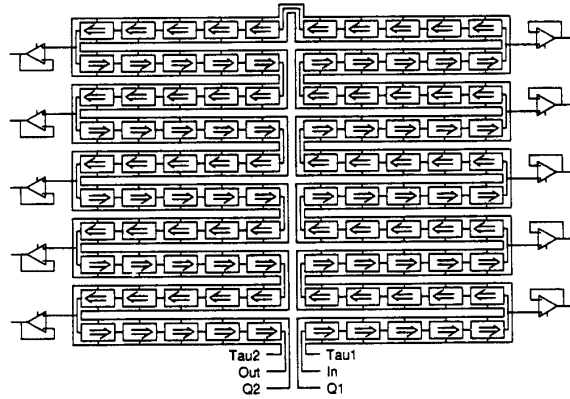


Figure 2: Floorplan of 100-stage cochlea chip.

### 6.3 Basilar Membrane Delay Line

Our fabricated basilar membrane model has 480 sections in a boustrophedonic arrangement, as shown in a 100-section version in Figure 2. 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  $\tau$  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 off chip. 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  $\tau$  control line. If we put a different voltage on the two ends of the  $\tau$  line, we get a gradient in voltage along the length of the polysilicon line. A linear gradient in voltage will turn into an exponential gradient in the delay per section. We can thereby easily make a transmission line with exponential velocity of propagation and cutoff frequency. Adjusting all the sections to have the same  $Q$  value then simply requires 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.

### 6.4 Second-Order Sections in Cascade

Figure 3(a) shows the response of a single second-order section from the transmission line, for  $Q$  values 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  $Q > 0.707$ ).

Because the system scales, each second-order section should have a similar response curve. That is, 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  $\Delta x/d_\omega$  (where  $d_\omega$  is the characteristic scaling distance, in which the best frequency changes by a factor of  $e$ ). 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 3(b).

### 6.5 Frequency Response of the Silicon Cochlea

The most straightforward behavior of the silicon cochlea that can be compared with theory is the magnitude of the frequency response. The circuit was set up with a gradient in the  $\tau$  such that each section was slower than its predecessor by a factor of 1.0139. With this value, the auditory range was covered in the 480 sections. The  $Q$ -control voltage was set so that the peak response was about 5 times the DC response, as seen at several different output taps.

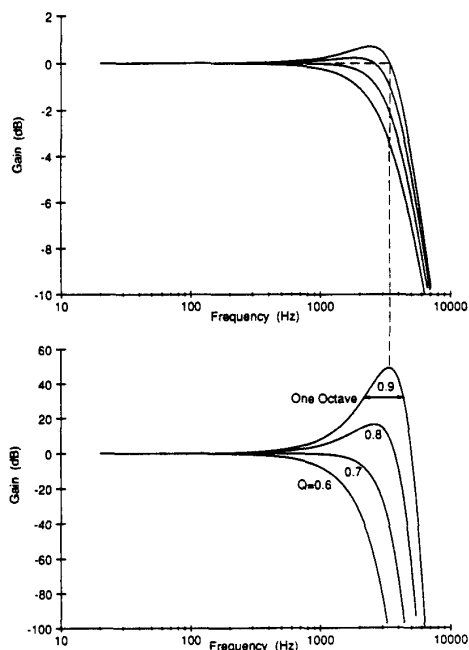


Figure 3: (a) Log-log frequency response of a single second-order filter section, for  $Q$  values 0.6, 0.7, 0.8, and 0.9, including scaled copies of the  $Q = 0.9$  response. (b) Corresponding overall response of a cascade of 120 stages with a factor 1.0139 scaling between adjacent stages.

Experimental data were taken with a sine wave of 14 mV peak-to-peak amplitude applied to the input. Results for two taps 120 sections apart are shown in Figure 4. 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 7. The value of the DC gain of the amplifiers was determined from the ratio of the response peaks. The value of  $Q$  used in the theory was adjusted until the predicted peak heights agreed with the observed ones. The resulting  $Q$  value was 0.79. The lower frequency peak corresponds to a tap further from the input by 120 sections than 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.

The remarkable agreement between theory and experiment is surprising in view of random variations in transistor input offset voltages (thresholds). We would expect a variation in  $Q$  and  $\tau$  values for each section that is much larger than 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  $\tau$  value that we desire does not matter. Some amplifier will have about that  $\tau$  value, and it will make its contribution as required. It is more surprising that the random variation of the value of  $Q$  does not affect the result in a more violent way. As of now we do not have a satisfactory explanation for the composite response curve as a result of many curves of different individual  $Q$  values.

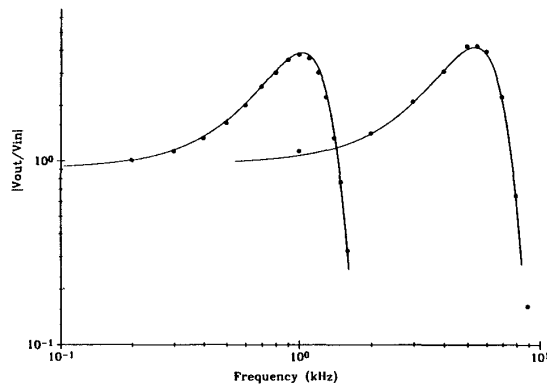


Figure 4: Log-log frequency responses to two output taps 120 sections apart: experimental points and theoretical curves.

## 7 Summary and Conclusions

The problems we must solve to build perception machines are mostly similar to those that nature had to solve biologically in the evolution of intelligent animal behavior. The key problems in sound perception are to cope with a very wide dynamic range of loudness and to separate sounds on the basis of their properties, such as frequency content and time structure. We believe that, by developing circuits that solve the same problems using imprecise analog components, we will increase our understanding of how animals hear.

In order to make further progress in hearing research it is also important for researchers to adopt a coherent set of working hypotheses, such as those presented by Kim [Kim 84], that emphasize the nonlinear, active, and adaptive aspects of hearing. The circuit model presented in this paper incorporates some of Kim's ideas as well as other observations on hearing and on circuits. Several more generations of experimental circuits will be built to follow up on these ideas.

## References

- [de Boer 80] E. de Boer and R. MacKay. "Reflections on reflections." *J. Acoust. Soc. Am.*, 67:882-890, 1980.
- [Lyon 82] R. F. Lyon. "A Computational Model of Filtering, Detection, and Compression in the Cochlea." *Proc. IEEE Intl. Conf. on Acoust. Speech and Signal Proc.*, Paris, May 1982.
- [Lyon 85] R. F. Lyon and N. Lauritzen. "Processing Speech with the Multi-Serial Signal Processor." *Proc. IEEE Intl. Conf. on Acoust. Speech and Signal Proc.*, Tampa, March 1985.
- [Holmes 83] M. Holmes and J. D. Cole. "Pseudo-resonance in the cochlea." In E. de Boer and M. A. Viergever (eds). *Mechanics of Hearing*. the Hague: Martinus Nijhoff Publishers, 1983.
- [Robles 85] L. Robles, M. A. Ruggero, and N. C. Rich. "Mössbauer measurements of the mechanical response to single-tone and two-tone stimuli at the base of the chinchilla cochlea." In J. B. Allen et al. (eds) *Peripheral Auditory Mechanisms*. New York: Springer-Verlag, 1985.
- [Vittoz 85] E. A. Vittoz. "Micropower Techniques." In Y. Tzividis and P. Antognetti (eds). *Design of MOS VLSI Circuits for Telecommunications*. Ch. 4. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [Mead 88] C. A. Mead. *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, (in press).
- [Kim 84] D. O. Kim. "Functional Roles of the Inner- and Outer-Hair-Cell Subsystems in the Cochlea and Brainstem." In C. Berlin (ed). *Hearing Science*. San Diego: College-Hill Press, 1984.