SILICON MODELS OF EARLY AUDITION

Thesis by

John Lazzaro

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1990

Submitted August 23, 1989

Acknowledgments

Thanks to my advisor Carver Mead and committee members Richard Lyon and Mark Konishi for most of the ideas in this dissertation; my role in the project has been the integration of their potentials — very large scale integration! Also thanks to my other collaborators in Chapter 3 of the dissertation, Sylvie Ryckebusch and M. Mahowald. Also thanks to my other committee members, Charles Seitz and Slobodan Cúk, and to my undergraduate advisor, Paul Mueller.

Many other people have contributed to the project in some way. In particular, I thank Tim Allen, Walker Aumann, Al Barr, Mark Bell, Ron Benson, Jerry Burch, Jim Campbell, Chong Chen, Richard DeCarlo, Tobi Delbruck, Helen Derevan, Arlene DesJardins, Steve DeWeerth, Lyn Dupré, Diane Dysland, David Feinstein, I. Fujita, Andy Fyfe, Dave Gillespie, Sherif Gobran, Glenn Gribble, John Harris, Scott Hemphill, Nancy Lee Henderson, John Hopfield, Calvin Jackson, Doug Kerns, John Klemic, Allen Knutson, Rick Koshi, Tor Sverre Lande, John Le Moncheck, Jin Luo, Mary Ann Maher, John Mariano, Vincent Martello, Lars Nielsen, John Platt, Larry Proctor, David Saia, Terrence Sejnowski, Massimo Sivilotte, Don Speck, Jackie Tanaka, John Tanner, Susan Volman, John Wawrzynek, and John Wyatt. Finally, thanks to my mom.

This work was sponsored by the System Development Foundation and the Office of Naval Research. I thank Hewlett-Packard for computing support, and DARPA and MOSIS for chip fabrication.

Abstract

This dissertation describes silicon integrated circuits that model known and proposed physiological structures in the early auditory system. Specifically, it describes silicon models of auditory-nerve response, of auditory localization in the barn owl, and of pitch perception. The integrated circuits model the structure as well as the function of the physiology; all subcircuits in the chips have anatomical correlates. The chips, two of which contain over 100,000 transistors, compute all outputs in real time, using analog, continuous-time processing. In most respects, chip responses approximate physiological or psychophysical response of the modeled biological systems. The dissertation also describes a novel nonlinear-inhibition circuit, which is a key component of two of the silicon models.

Table of Contents

List of Illustrations

Chapter 1
Introduction

As a branch of science matures, it often develops a synthetic component. Although many open issues remain, advances in auditory physiology, psychophysics, and theoretical modeling offer detailed information on the structure, function, and fundamental limitations of the early auditory system. Creating an artificial early auditory system, in a physical medium that shares many of the strengths and weaknesses of the biological substrate, requires a concrete specification of auditory neural computations, and provides a challenge to the field different from those of computer simulation and theoretical modeling. My co-workers and I are building analog VLSI (very large-scale integrated) silicon circuits that model the early auditory system, both to explore the general computational principles of auditory neuroscience, and to create potentially useful devices for sound understanding, for sound localization, and for prostheses for the deaf. This dissertation documents some of our research.

1.1 Motivation for the Scientist: Why Use Analog VLSI Technology?

We believe that any medium imposes a direction on modeling; it is thus advantageous to choose a modeling medium that shares characteristics with the system under study. Analog VLSI and neural systems are different in detail, but the frameworks for computation in the two technologies are remarkably similar. Both media pack a large number of imperfect computational elements into a small space. Systems in both technologies must confront these imperfections, not as a second-order effect, but as a prerequisite for a working design. Both media offer a rich palette of primitives in which to build a structure; nonlinearities are resources for improved system performance. Neural and silicon systems are both

ultimately limited not by the density of devices, but rather by the density of interconnect. Modeling neural systems directly in a physical medium subjects the researcher to many of the same pressures faced by the nervous system over the course of evolutionary time.

The design style of the projects described in this dissertation accentuates the similarities between analog VLSI and neural systems. The circuits use the full analog nature of transistors, exploiting the exponential functions inherent in the devices. The chips use voltage, current, and charge as analog data representations of system quantities; the systems also use a pulse representation that models biological action potentials. All communication in the systems is asynchronous; there are no clock signals. The speed of processing is matched to the time scale of auditory signals. Each subcircuit is dedicated to modeling a known or proposed structure in the early auditory system; circuits are not time-multiplexed between multiple functions.

Analog VLSI systems, using this design style, have another distinct advantage: real-time performance. All the systems described in this dissertation work in real time. Real-time performance allows large, complex models to be tested with a large library of input stimuli. Real-time performance also encourages the construction of electromechanical systems to test sensorimotor models.

Analog VLSI technology is an excellent medium for auditory modeling. Processing time-varying information fully exploits the real-time performance of analog VLSI systems; modeling human audition requires the processing of time-varying waveforms at frequencies up to 20 kHz. In addition, many representations of early audition do not require wiring in three dimensions, and map well to a two-dimensional silicon die.

1.2 Motivation for the Engineer: Why Model Audition?

Why should an engineer study and model early biological auditory processing, and incorporate these models into practical sonar and speech-recognition systems? The early representations of a sensory system are crucial. Early representations serve to reduce the amount of sensory data, while preserving information essential for the task of the system. Moreover, early representations express this information in a way that simplifies the task of later modules in the system. Because early biological auditory representations are the first stage of the best speech-recognition system known, it makes sense to incorporate these representations into human-made speech recognizers. Likewise, these representations are the sonic substrate for animals specialized for passive and active sonar. The sonar performance of some of these animals exceeds the performance of human-made sonar systems; incorporating biologically inspired early auditory representations into human-made sonar systems may decrease this performance gap.

Practical speech-recognition and sonar systems often require real-time performance. Modeling early biological auditory representations using conventional digital computers, in real time, is not possible with existing technology. Analog VLSI technology is a medium that offers real-time computation of these representations. Specialized digital architectures are another way to compute these representations in real time. It is beyond the scope of this dissertation to judge which method offers the best engineering solution to the problem. As detailed in the Section 1.1, analog VLSI technology, by its similarity to the biological substrate, aids the scientist in understanding existing representations and inventing new models to explain neural data. In the same way, analog VLSI technology may aid the creative engineer in designing innovative solutions to engineering problems.

1.3 Overview of the Dissertation

This dissertation describes silicon models of three aspects of early audition. Chapter 2 describes an integrated circuit that models the evoked responses of the auditory nerve. The auditory nerve is the first neural representation of audition; its carries information from the cochlea, the sense organ of hearing, to the brain. Comparisons of chip and physiological responses to classical stimuli show a qualitative similarity, with quantitative differences.

Chapter 4 describes an integrated circuit that computes the spatial position of a sound in the azimuthal plane. The chip models a neural representation of auditory space found in the barn owl, an animal specialized for passive auditory localization. The 220,000-transistor chip models the processing of two cochleas and several brainstem nuclei, computing an output map of azimuthal space by processing interaural time differences.

Chapter 5 describes an integrated circuit that models human pitch perception. Chip output approximates human performance on a variety of classical pitch-perception stimuli. Some of the circuits on the chip correspond to known physiological computations; other parts of the chip model proposed neural structures.

Inhibition is an essential element of the auditory-localization and pitch-perception chips; Chapter 3 describes the nonlinear-inhibition circuit used in both chips. This novel circuit needs only two transistors for each neuron, and requires a length of interconnect that increases linearly with the number of neurons. Chapter 6 offers suggestions for future research.

Chapter 2
A Silicon Model of Auditory-Nerve Response

Nonlinear signal processing is an integral part of sensory transduction in the nervous system. Sensory inputs are analog, continuous-time signals with a large dynamic range, whereas central neurons encode information with limited dynamic range and temporal specificity, using fixed-width, fixed-height pulses. Sensory transduction uses nonlinear signal processing to reduce real-world input to a neural representation, with a minimal loss of information.

An excellent example of nonlinear processing in sensory transduction occurs in the cochlea, the organ that converts the sound energy present at the eardrum into the first neural representation of the auditory system, the auditory nerve. Humans can process sound input over a 120-dB dynamic range, yet the firing rate of an auditory-nerve fiber can encode only about 25 dB of sound intensity. Humans can sense binaural time differences of the order of 10 μs, yet an auditory-nerve fiber can fire at most once per millisecond. Using limited neural resources, the cochlea creates a representation that preserves the information essential for sound localization and understanding. Moreover, this neural code expresses auditory information in a way that facilitates feature extraction by higher neural structures.

This chapter describes the architecture and operation of an integrated circuit that models, to a limited degree, the evoked responses of the auditory nerve. The chip receives as input a time-varying voltage corresponding to sound input, and computes outputs that correspond to the responses of individual auditory-nerve fibers. The chip models the structure as well as the function of the cochlea; all subcircuits in the chip have anatomical correlates. The chip computes all outputs in real time, using analog continuous-time processing. The original research in

this chapter was done in collaboration with Carver Mead; parts of this chapter were originally published in (Lazzaro and Mead, 1989b).

2.1 Neural Architecture of the Cochlea

Both mechanical and electrical processing occur in biological cochleas. The sound energy present at the eardrum is coupled into a mechanical traveling-wave structure, the basilar membrane, which converts time-domain information into spatially-encoded information by spreading out signals in space according to their time scale (or frequency). Over much of its length, the velocity of propagation along the basilar membrane decreases exponentially with distance. The structure also contains active electromechanical elements; outer hair cells have motile properties, acting to reduce the damping of the passive basilar membrane and thus allowing weaker signals to be heard. Axons from higher brain centers innervate the outer hair cells; these centers may dynamically vary the local damping of the cochlea, providing frequency-specific automatic gain control (Kim, 1984).

Inner hair cells occur at regular intervals along the basilar membrane. Each inner hair cell acts as an electromechanical transducer, converting basilar-membrane vibration into a graded electrical signal. Several signal-processing operations occur during transduction. Inner hair cells half-wave rectify the mechanical signal, responding to motion in only one direction. Inner hair cells primarily respond to the velocity of basilar-membrane motion, implicitly computing the time derivative of basilar-membrane displacement (Dallos, 1985). Inner hair cells also compress the mechanical signal nonlinearly, reducing a large range of input sound intensities to a manageable excursion of signal level.

Spiral-ganglion neurons connect to each inner hair cell, and produce fixed-width, fixed-height pulses in response to inner-hair-cell electrical activity. The

synaptic connection between the inner hair cell and the spiral-ganglion neuron might implement a stage of automatic gain control, exploiting the dynamics of synaptic-transmitter release (Geisler and Greenberg, 1986). Auditory-nerve fibers are axons from spiral-ganglion neurons; these fibers present a neural representation of audition to the brain.

When pure tones are presented as stimuli, an auditory-nerve fiber is most sensitive to tones of a specific frequency. This characteristic frequency corresponds to maximum basilar-membrane velocity at the location of the inner hair cell associated with the nerve fiber. The spiral trunk of the auditory nerve preserves this ordering; the nerve fibers are mapped cochleotopically and tonotopically. The mean firing rate of an auditory fiber encodes sound intensity, over about 25 dB of dynamic range. The temporal pattern of nerve firings reflects the shape of the filtered and rectified sound waveform; this phase locking does not diminish at high intensity levels (Evans, 1982).

2.2 Silicon Models of the Cochlea

Both mechanical and electrical processing occur in the cochlea. In the chip, however, we model both types of computation using electronic processing. Lyon and Mead have designed a silicon model of the mechanical processing of the cochlea (Lyon and Mead, 1988a; Mead, 1989). Their circuit is a one-dimensional physical model of the traveling-wave structure formed by the basilar membrane. In this model of cochlear function, the exponentially tapered stiffness of the basilar membrane and the motility of the outer hair cells combine to produce a pseudoresonant structure.

A circuit model implements this view of cochlear hydrodynamics, using a cascade of second-order sections, with exponentially scaled time constants. This analog, continuous-time circuit model computes the pressure of selected discrete

points along the basilar membrane in real time. The cascade structure enforces unidirectionality, so a discretization in space does not introduce reflections that could cause instability in an active model. An amplifier in each second-order section provides active positive feedback in the circuit, modeling the active mechanical feedback provided by the outer hair cells in physiological cochleas. An automatic-gain-control system can increase sensitivity to weak sounds by locally varying the amount of positive feedback in the second-order sections. Appendix 2A provides a brief description of the circuits in the model.

This model of cochlear mechanics is the foundation of our model of auditory nerve response. We have not implemented an automatic-gain-control system to vary the damping of the structure locally; however, the chip has inputs that provide local control of membrane damping, allowing off-chip experiments with automatic gain control. To complete the circuit model of the auditory periphery, we have added circuits that model the functions of the inner hair cells and spiral-ganglion neurons; Figure 2.1 shows the complete architecture of the chip.

Our inner-hair-cell circuit models the signal-processing operations that occur during transduction: velocity-sensing, nonlinear compression, and half-wave rectification. Our spiral-ganglion-neuron circuit converts the analog output of the inner-hair-cell circuit into fixed-width, fixed-height pulses; the mean firing rate of the circuit encodes the intensity of inner-hair-cell circuit response, whereas the temporal pattern of pulses reflects the waveform shape of inner-hair-cell circuit response. Appendix 2B provides a description of these two circuits.

The synaptic connection between the inner hair cell and the spiral-ganglion neuron might implement a stage of automatic gain control, exploiting the dynamics of synaptic-transmitter release (Geisler and Greenberg, 1986). The nonlinear compression of our inner-hair-cell circuit approximates the static effect of the automatic gain control of the inner-hair-cell synapse. A direct

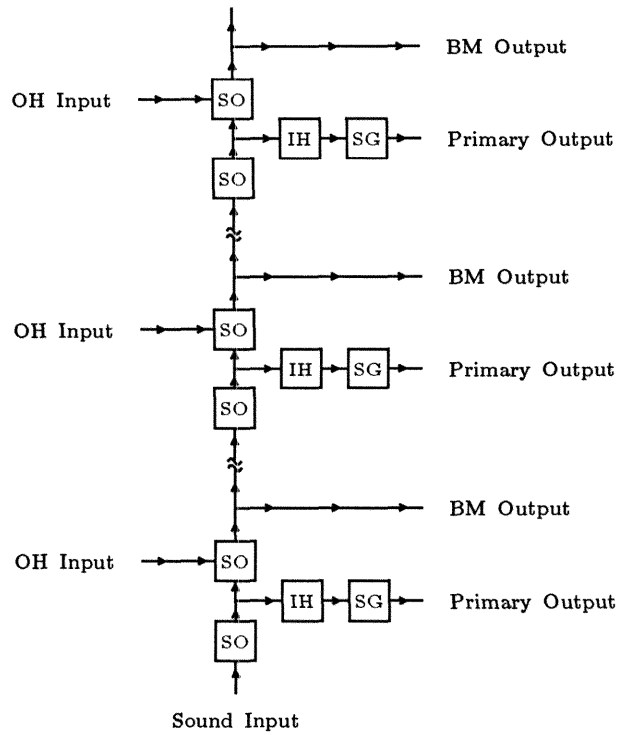


Figure 2.1. Block diagram of the auditory-nerve chip. Sound input travels down the basilar-membrane model, a cascade of second-order (SO) sections with exponentially increasing time constants. Basilar-membrane (BM) circuit outputs show pressure along the membrane, whereas inputs modeling efferent innervation of outer hair (OH) cells control local damping of the membrane circuit. Taps along the basilar membrane connect to a circuit model of inner hair (IH) cells; outputs from inner hair cells connect to circuits that model spiral-ganglion (SG) neurons. These neurons form the primary output of the chip, thus modeling auditory-fiber response.

implementation of dynamic automatic gain control would accurately model time adaptation and two-tone synchrony suppression; this enhancement is currently under development.

The outputs from spiral-ganglion-neuron circuits are the primary outputs of the chip. Additional outputs of the circuit display basilar-membrane pressure, for diagnostic purposes. The circuit we report in this chapter has eight auditory-fiber outputs; the projects described in Chapters 4 and 5 use cochlea circuits with 62 auditory-fiber outputs.

2.3 Silicon Basilar-Membrane Response

To test the tuning properties of the silicon auditory-nerve fibers, we duplicated a variety of classical auditory-nerve measurements. In these experiments, we tuned the basilar-membrane circuit to span about seven octaves, from 50 Hz to 10,000 Hz. We set the maximum firing rates of the auditory-fiber outputs at 150 to 300 spikes/s, with spike widths of 5 to 20 μs.

In this configuration, without an input signal, the auditory-fiber outputs fire at less than 0.1 spike/s. At the characteristic frequency of a fiber, pure tones of a few millivolts peak amplitude produce responses significantly above this spontaneous rate. The chip can process tones up to about 1 V of peak amplitude, yielding approximately 60 dB of usable dynamic range. Adding a preprocessor to the basilar-membrane circuit to limit intense input signals would extend the upper limit of the dynamic range. A biological cochlea has a mechanical limiter as a preprocessor — the stapedial reflex. Designing more sensitive inner-hair-cell circuits would extend the lower limit of dynamic range. Both dynamic-range enhancements are currently under development.

Figure 2.2(a) shows the amplitude of the silicon basilar-membrane response to pure tones, at a position with a best frequency of about 1900 Hz. We

set 0 dB at 3 mV peak, an input amplitude sufficient to produce responses above the spontaneous rate in auditory-fiber outputs near this basilar-membrane position. The 0-dB frequency-response curve shows a flat response for frequencies significantly below the best frequency, a 12-dB response peak at the best frequency, and a sharp dropoff to the noise floor for frequencies significantly above the best frequency. The 0-dB chip frequency-response curve is qualitatively similar to the 70-dB frequency-response curve taken from the basilar membrane of the squirrel monkey using the Mossbauer effect, shown in Figure 2.2(b) (Rhode 1971). Near the best frequency, basilar-membrane pressure, computed by the chip, is approximately equal to the basilar-membrane displacement, as measured by Rhode. Quantitatively, the bandwidth of the response peak of the chip is wider than that of the physiological data; a cascade of second-order sections does not yield an optimal model of cochlear hydrodynamics (Lyon and Mead, 1988b).

The 10-dB and 20-dB chip frequency-response curves show a decrease in the magnitude of the resonance peak, and a shift downward in best frequency. This qualitative behavior matches the nonlinear behavior of the squirrel-monkey basilar membrane. In the chip, the saturation of amplifiers that model the motility of the outer hair cells causes nonlinear behavior. A similar phenomenon may occur in the physiological system; for large sound intensities, outer hair cells may not be capable of a linear response to basilar-membrane motion.

Another explanation of nonlinear basilar-membrane response also is plausible. The efferent fibers that innervate the outer hair cells might be controlling frequency-selective automatic gain control. In this model, efferent fibers allow the outer hair cells to reduce the damping of the basilar membrane for soft sounds, whereas for louder sounds, efferent fibers inhibit the motility of outer hair cells, decreasing the size of the pseudoresonance peak (Kim, 1984). Our chip

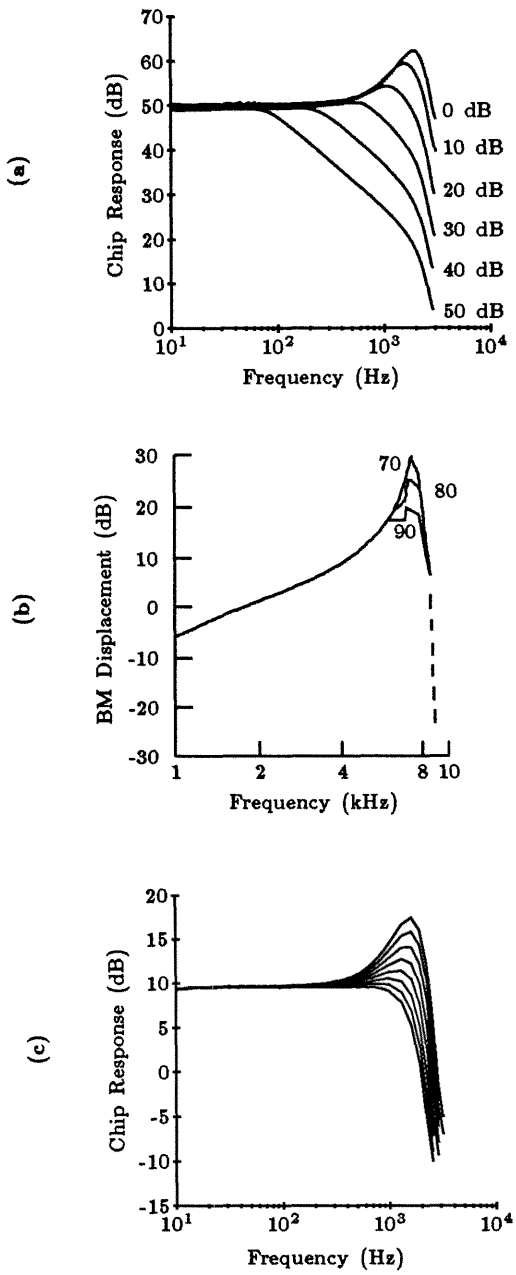


Figure 2.2. **a.** Plots showing the response of the basilar-membrane circuit at a single point, to pure tones at a fixed input amplitude (0 dB = 3 mV peak). The 0-dB to 40-dB curves are vertically shifted to match the low-frequency response of the 50-dB plot. **b.** Transfer functions of a single position on the basilar membrane of the squirrel monkey (Rhode, 1971). The curves show amplitude of vibration for constant malleus displacement. Two curves are vertically shifted to match the low-frequency response of the third curve. **c.** Plots showing the response of the basilar-membrane circuit to 10-dB pure tones, for different amounts of basilar-membrane damping at the measurement point.

does not model this automatic-gain-control system. There are, however, inputs into the chip that model the outer-hair-cell efferent fibers, allowing local control of damping at several points along the basilar-membrane model.

Figure 2.2(c) shows the effect of locally varying the strength of outer-hair-cell damping, near the position of recording, on the frequency response of the basilar-membrane model. The magnitude of the resonance peak decreases and the best frequency of the response shifts downward as the damping is increased. In the automatic-gain-control model, damping is increased for increasing signal levels; in this way, the model qualitatively matches the physiological responses of Figure 2.2(b). An integrated automatic-gain-control system for this chip is currently under development.

The frequency-response plots of the basilar-membrane model for 30-dB to 50-dB inputs, as shown in Figure 2.2(a), do not match physiological behavior. The saturation of the amplifiers that model the stiffness of the basilar membrane causes these undesired nonlinear effects. In practice, in many of the classical auditory-nerve experiments, the saturating nonlinearity of the inner-hair-cell circuit masks these nonphysiological effects. As mentioned earlier, future designs will incorporate more sensitive inner hair cells and a model of the stapedial reflex, to provide a large dynamic range without operating the basilar-membrane circuit in this regime; to manage this increased operating range, these designs will incorporate inner hair cells with dynamic automatic gain control.

2.4 Tuning Properties of Silicon Auditory-Nerve Fibers

We have characterized the tuning properties of the auditory-nerve-fiber circuit model, using both pure tones and clicks. In response to a click of medium intensity, a silicon auditory-nerve fiber produces one or several spikes. To extract the click response from these spikes, we present the click stimulus to the chip

many times, and record the responses of a silicon auditory-nerve fiber. These data are reduced to a poststimulus-time (PST) histogram, in which the height of each bin of the histogram indicates the number of spikes occurring within a particular time interval after the presentation of the click.

A PST histogram of the response of a silicon auditory-nerve fiber to a repetitive rarefaction click stimulus shows a half-wave-rectified version of a damped sinusoidal oscillation (Figure 2.3(a)). The frequency of this oscillation, 1724 Hz, is approximately the best frequency of the basilar-membrane position associated with this silicon nerve fiber. The half-wave rectification of the inner-hair-cell circuit removes the negative polarity of oscillatory waveform from the PST histogram of the click response. Repeating this experiment using a condensation click recovers the negative polarity of oscillation; a compound PST histogram, shown in Figure 2.3(b), combines data from both experiments to recreate the ringing waveform produced by the basilar-membrane circuit. Figure 2.3(c) shows a compound PST histogram of the click response of an auditory fiber in the cat (Kiang *et al.*, 1965). Qualitatively, the circuit response matches the physiological response.

Figures 2.3(a) and 2.3(b) are chip responses to a 60-mV click stimulus (26 dB, 0 dB = 3 mV peak). Higher-intensity clicks produce oscillatory responses with increased damping; a compound PST histogram of chip auditory-nerve response to a 36-dB click shows reduced ringing (Figure 2.3(d)). This effect is a direct result of the nonlinear response of the basilar-membrane model; physiological basilar-membrane click responses also show reduced ringing at high click-intensity levels (Robles *et al.*, 1976).

In response to a pure tone of sufficient intensity and appropriate frequency, the silicon auditory fiber produces spikes at a constant mean rate, as shown in Figure 2.4. The mean spike rate of a silicon fiber, in response to a constant tone,

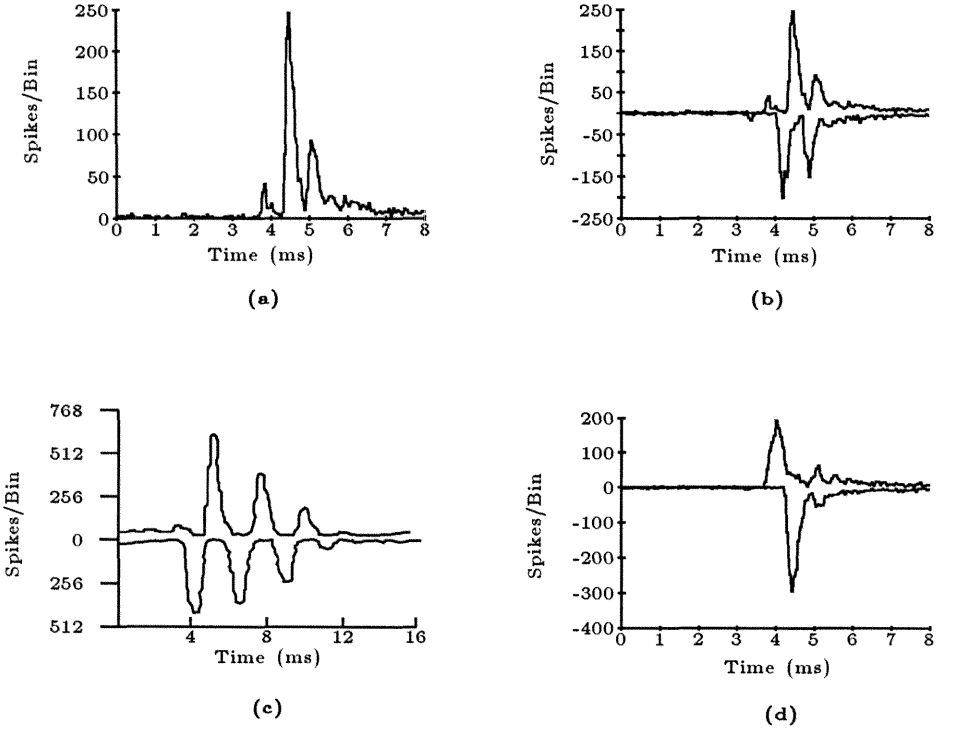


Figure 2.3. **a.** Post-stimulus time (PST) histogram of the rarefaction click response of a silicon auditory-nerve fiber. Click amplitude is 60 mV (26 dB peak); click width is 100 μs. Histogram is for 2000 click presentations; the width of each bin is 58 μs. **b.** Compound PST histogram of the click response of a silicon auditory-nerve fiber. Rarefaction click response is plotted as positive values; condensation click response is plotted as negative values. Conditions are identical to those of part a. **c.** Compound PST histogram of the click response of an auditory fiber in the cat (Kiang et al., 1965). Click level is 30 dB relative to threshold response level; click width is 100 μs. Rarefaction click response is plotted as positive values; condensation click response is plotted as negative values. **d.** Compound PST histogram of the click response of a silicon auditory-nerve fiber, for a 200-mV click (36-dB click). All other conditions are identical to those of part a.

does not decrease over time, unlike that of a physiological auditory fiber; this lack of adaptation indicates the absence of dynamic automatic gain control in our model.

Figure 2.5(a) shows the mean spike rate of a silicon auditory fiber as a function of pure tone frequency. For low-amplitude tones, the fiber responds to a narrow range of frequencies; for higher-intensity tones, the fiber responds to a wider range of frequencies. The saturating nonlinearities of the basilar-membrane circuit and of the inner-hair-cell circuit cause the bandwidth of the fiber to increase with sound intensity. Qualitatively, this behavior matches the iso-intensity plots from an auditory-nerve fiber in the squirrel monkey (Rose *et al.*, 1971), shown in Figure 2.5(b).

Figure 2.6(a) shows the mean spike rate of a silicon auditory fiber as a function of pure tone amplitude, at frequencies below, at, and above the best frequency of the fiber. In response to its characteristic frequency, 2100 Hz, the fiber encodes about 25 dB of tone amplitude before saturation. Figure 2.6(b) shows rate-intensity curves from an auditory fiber in the cat (Sachs and Abbas, 1974). At its characteristic frequency, the physiological fiber also encodes about 25 dB of tone amplitude before saturation. The shape of the biological and silicon curves at the characteristic frequency is remarkably similar, giving us some confidence in the validity of this modeling paradigm. In response to frequencies below and above the characteristic frequency, the functional forms of the silicon fiber responses are different from those of the physiological data. Most notably, the saturation rate of a silicon fiber for frequencies below the fiber's characteristic frequency exceeds the saturation rate of the silicon fiber at the fiber's characteristic frequency. This behavior is a direct result of the undesired saturation, at high input intensities, of amplifiers that model the stiffness of the

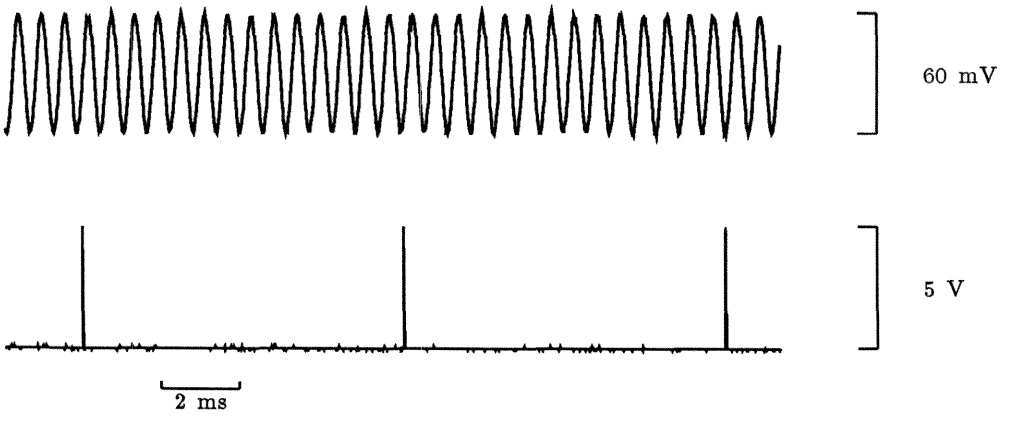


Figure 2.4. Output of a silicon auditory fiber (bottom trace) in response to a sinusoidal input (top trace). The frequency of the input is the characteristic frequency of the fiber.

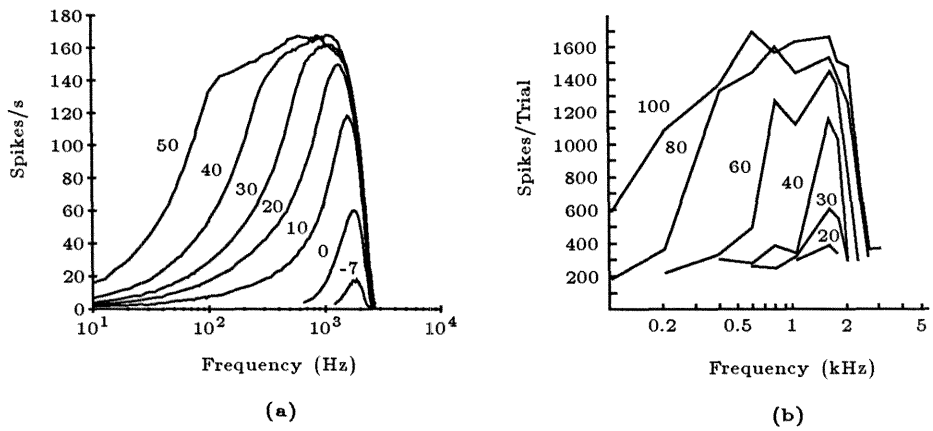


Figure 2.5. **a.** Plots showing the mean spike rate of a silicon auditory fiber as a function of pure tone frequency. Legend numbers indicate tone amplitude, in dB. **b.** Plots showing the number of discharges of an auditory fiber in the squirrel monkey, in response to a 10-s pure tone (Rose *et al.*, 1971). Legend numbers indicate tone amplitude, in dB.

basilar membrane. Above the best frequency of the silicon fiber, the response of the model decreases in a manner that is reminiscent of its biological counterpart.

Figure 2.7(a) shows iso-response curves for four silicon auditory-nerve fibers. These plots represent an iso-rate section through the iso-intensity curves of Figure 2.5(a), at a spike rate for each fiber that was comfortably above the spontaneous rate. The chip response accurately models the steep high-frequency tail of tuning curves from cat auditory fibers (Kiang, 1980), shown in Figure 2.7(b); the shapes of physiological and chip tuning curves are qualitatively similar. The bandwidth of the chip fibers for low sound intensities, however, is significantly wider than that of the physiological response. This problem stems from the wider bandwidth of the basilar-membrane circuit model, relative to that of the physiological data, as well as from the lack of a dynamic automatic-gain-control system for modulating the damping of the basilar-membrane circuit. The high-frequency cutoff of the iso-response curves, shown in Figure 2.7(a), is much steeper than is the cutoff of the iso-input curves shown in Figure 2.5(a). In a linear system, these two measurements would give identical results. The difference reflects the presence of a saturating nonlinearity in the system; the inner-hair-cell circuit and the basilar-membrane circuit provide this saturation function.

2.5 Timing Properties of Silicon Auditory-Nerve Fibers

As the click response of Figure 2.3 shows, the temporal firing patterns of the silicon auditory-nerve fibers encode information. Figure 2.8(a) shows period histograms of a chip fiber, in response to −5-dB to 50-dB pure tones at the fiber's characteristic frequency; these histograms show the probability of a spike output occurring within a particular time interval during a single cycle of the input sinusoid. The fiber preserves the shape of the input sinusoid throughout

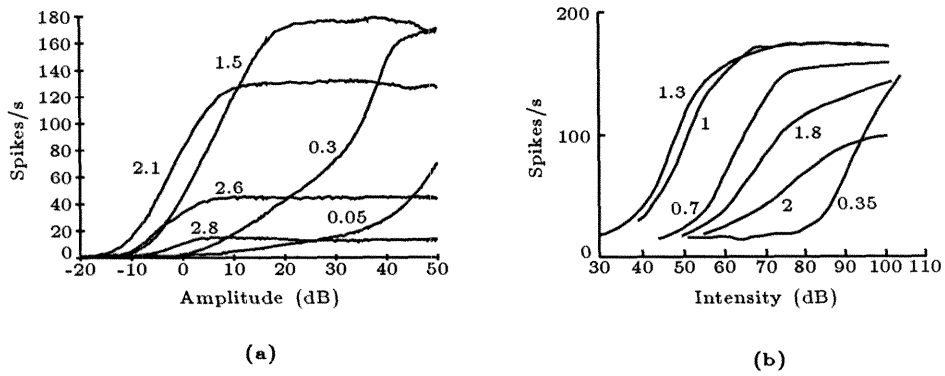
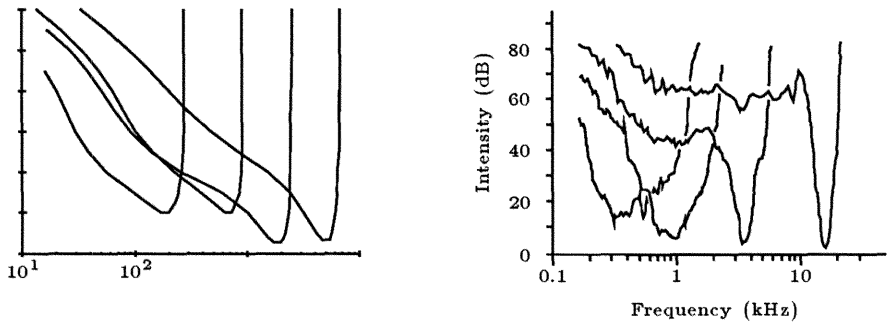


Figure 2.6. **a.** Plots showing the mean spike rate of a silicon auditory fiber as a function of pure tone amplitude. Label numbers indicate tone frequency, in Hz. **b.** Plots showing the mean spike rate of an auditory fiber in the cat, as a function of pure tone amplitude (Sachs and Abbas, 1974). Label numbers indicate tone frequency, in Hz.


10^1
10^2
Intensity (dB)
80
60
40
20
0
0.1
1
10
Frequency (kHz)

this intensity range; this behavior matches data from an auditory fiber in the cat (Rose *et al.*, 1971), shown in Figure 2.8(b). Unlike the cat fiber, however, the silicon fiber does not preserve absolute phase at higher intensities; this deficiency results from the saturation of the second-order section amplifiers that model basilar-membrane stiffness.

The temporal firing patterns of the silicon auditory-nerve fiber are, however, a good representation of signal periodicity; the synchronization ratios (normalized magnitude of the first Fourier coefficient) of the period histograms in Figure 2.8(a) are 0.5 to 0.6, comparable to those of physiological data at the same frequency. The firing patterns of silicon nerve fibers maintain partial synchrony in the presence of masking noise. Figure 2.9(a) shows the degree of synchronization of a silicon fiber to a pure tone at the best frequency of the fiber, in the presence of white noise. Figure 2.9(b) shows similar data from the auditory nerve of the squirrel monkey (Rhode *et al.*, 1978). Qualitatively, the data from the chip fiber and those from the cat fiber are similar.

The temporal firing patterns of the silicon auditory-nerve fiber preserve the shape of complex waveforms; Figure 2.10(a) shows the compound period histograms of a silicon fiber, in response to two harmonically related tones, combined with equal amplitudes. A Fourier analysis of the histogram shows strong peaks at 868 Hz and at 1156 Hz, along with distortion products; the fiber preserves periodicity information. Figure 2.10(b) shows an auditory fiber in the cat, which preserves periodicity information of a tone pair in the same manner (Goblick and Pfeiffer, 1969).

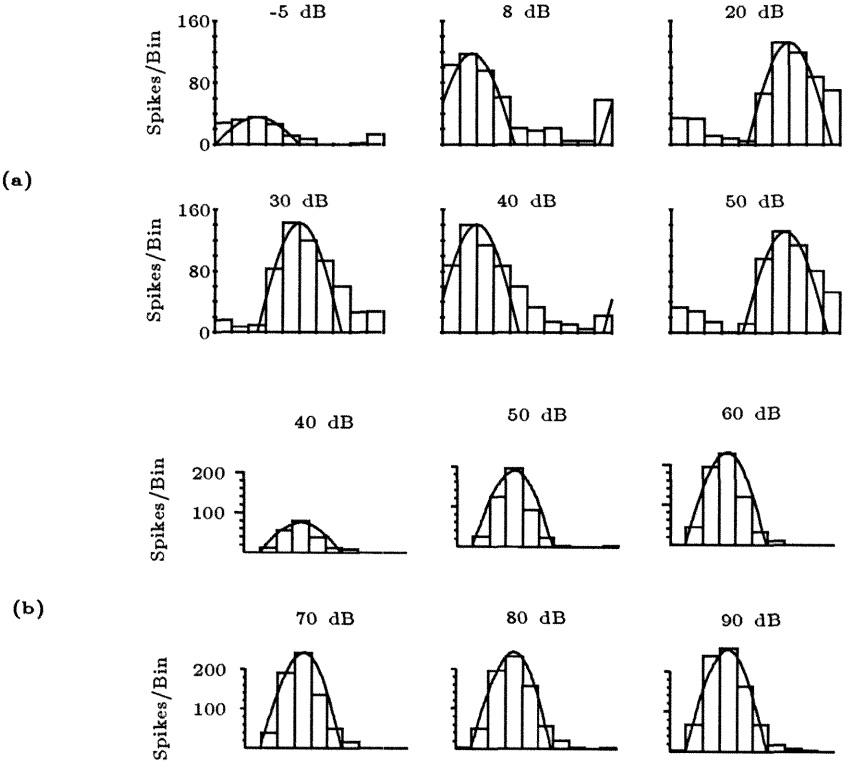


Figure 2.8. **a.** Period histograms of the silicon auditory-fiber response to a pure tone of 1840 Hz, near the fiber's best frequency. Amplitude of tone is shown above each plot. Histogram width is 54 μs. Each histogram begins at a constant position, relative to the input sinusoid; each is fitted to a sinusoid of best amplitude and phase. **b.** Period histograms of the response of an auditory fiber in the cat, to a low-frequency tone (Rose *et al.*, 1971). Amplitude of pure tone is shown above each plot. Each histogram is fitted to a sinusoid of best amplitude but fixed phase.

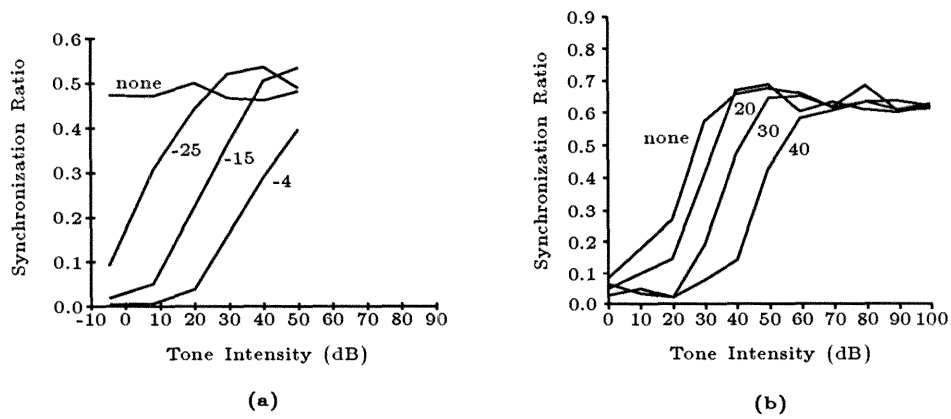


Figure 2.9. a. Plots of the synchronization ratio of PST histograms of a silicon auditory-fiber response to pure tones. Pure tone frequency is 1840 Hz, near the fiber's best frequency. Synchronization ratio is plotted as a function of tone amplitude. White noise is mixed with the tone; the legend indicates the amplitude of the noise signal, in $dB/\sqrt{Hz}$. **b.** Plots of the synchronization ratio of an auditory fiber in the squirrel monkey (Rhode *et al.*, 1978). Pure tone frequency is 900 Hz, the fiber's best frequency. Synchronization ratio is plotted as a function of tone amplitude. Band-limited (100- to 2000-Hz) noise is mixed with the tone; the legend indicates the spectrum density of the noise signal, in dB.

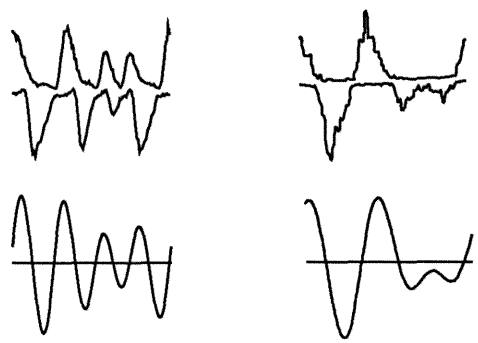


Figure 2.10. **a.** Top plot shows the compound period histogram of silicon auditory nerve response to two tones, at 868 Hz and 1156 Hz (f_3 : f_4), mixed at equal amplitudes (24 dB). Bottom plot shows the best-fit combination of 868 Hz and 1156 Hz to the histogram (arbitrary amplitude and phase). The best frequency of the fiber is about 1900 Hz. **b.** Top plot shows the compound period histogram of a cat auditory-nerve-fiber response to two tones (f_2 : f_3), at 538 Hz and 807 Hz (Goblick and Pfeiffer, 1969). Bottom plot shows the best-fit combination of 538 Hz and 807 Hz to the histogram (arbitrary amplitude, but the same relative phase as the stimulus).

2.6 Discussion for the Scientist

The integrated circuit captures many essential features of data representation in the auditory nerve; moreover, it computes the representation in real time. The performance of the present circuit, however, falls short of the performance of the auditory nerve in several significant respects. Each shortcoming in performance can be traced to the incomplete modeling of an aspect of cochlear physiology.

The bandwidth of the resonant peak in the basilar-membrane circuit is wider than that of physiological basilar-membrane response; a cascade of second-order sections does not yield an optimal model of cochlear hydrodynamics (Lyon and Mead, 1988b). The bandwidth of the silicon auditory-nerve-fiber tuning curves is also wider than that of physiological fiber response. The insufficient basilar-membrane circuit bandwidth obviously is a factor in this shortcoming; the lack of a dynamic automatic-gain-control system for modulating the damping of the basilar-membrane circuit is another important factor.

This absence of this automatic-gain-control system, as well as the absence of circuits that model the dynamics of the synaptic connection between the inner hair cell and the spiral-ganglion neuron, result in other performance deficiencies. Specifically, a silicon auditory nerve fiber does not exhibit time adaptation, and does not exhibit physiological two-tone rate supression and two-tone synchrony supression.

Finally, a silicon auditory fiber lacks the wide dynamic range of physiological cochleas; the chip can process sounds that range over only 60 dB. Improved inner-hair-cell circuits that are sensitive to smaller voltage excursions would improve circuit dynamic range, as would a circuit model of the stapedial reflex. With a stapedial-reflex circuit as a preprocessor, the undesired saturation of amplifiers that model the stiffness of the basilar membrane would not occur; this saturation causes the unwanted phase shift of the period histograms of

Figure 2.8(b), which show the temporal firing patterns of a silicon auditory nerve in response to sinusoids at varying intensities. This saturation also causes the saturation rate of a silicon fiber in response to frequencies below the fiber's characteristic frequency to exceed the saturation rate of a silicon fiber in response to the fiber's characteristic frequency, as shown in Figure 2.6(b).

Future research on this project involves implementing these enhancements to our model of auditory-nerve response. Modeling neural systems, in a physical medium that shares many of the strengths and weaknesses of the biological substrate, offers a unique perspective on the relationship of neural structure and function. For example, during the design of our silicon auditory nerve model, many iterations of the inner-hair-cell circuit design were required to achieve acceptable circuit performance. The major difficulty in circuit design was modeling the long-term adaptation of the cell to static cilia deflection. Duplicating this property of the cell was crucial to the successful operation of the chip. Without this auto-zeroing property, the inner-hair-cell circuit cannot adjust for variances in circuit elements due to fabrication tolerances; as a result, most auditory-fiber circuit outputs either refuse to fire or fire at a maximum rate, regardless of sound input.

Sensing basilar-membrane velocity, rather than pressure, provides this auto-zeroing function; our design goal evolved to be the synthesis of a sensitive time-differentiator circuit. Our inner-hair-cell circuit includes a time-differentiation circuit that works by comparing the instantaneous value of a voltage that represents basilar-membrane pressure with a time-averaged value of this voltage. Thus, the characteristics of the silicon medium dictated the functional form of our model. We believe that any medium imposes a direction on modeling; it is thus advantageous to choose a modeling medium — analog VLSI technology — that shares characteristics with the biological system under study.

2.7 Discussion for the Engineer

Our integrated-circuit model captures many essential features of data representation in the auditory nerve; moreover, it computes the representation in real time. There are many traditional engineering representations of audition, however, that are also amenable to analog implementation. What advantages does a silicon auditory-nerve representation offer to a designer of artificial sensory systems?

As shown in Figures 2.3, 2.8, and 2.10, an auditory-nerve fiber encodes a filtered, half-wave–rectified version of the input waveform, over a wide dynamic range, using the temporal patterning of fixed-width, fixed-height pulses. This representation supports the efficient, massively parallel computation of signal properties, using autocorrelations in time and cross-correlations between auditory fibers. In this representation, a correlation is simply a logical AND operation, performed by a few synapses in neural systems, or by a few transistors in silicon systems. Axonal delays in neural systems provide the time parameter for computing autocorrelations; in silicon systems, we model this delay with compact monostable circuits (Mead, 1989). We use these techniques in the projects in Chapter 4 and Chapter 5.

The nonlinear filtering properties of the auditory-nerve fibers, shown in Figures 2.5 and 2.7, enhance these correlations. In a quiet environment, auditory fibers have narrow bandwidths; each fiber carries independent information, yielding rich correlations. In noisier environments, the tuning of auditory fibers widens, increasing the number of fibers that carry information about the signal. This detuning ensures that some fibers still encode signal properties reliably (Greenberg, 1988).

As shown in Figure 2.6, auditory fibers encode about 25 dB of signal intensity. Dynamic automatic gain control, present in a physiological cochlea,

enhances this range; in addition, different populations of auditory fibers have different thresholds, further enhancing the encoding of signal intensity. Although not sufficient as a primary representation of sound, rate encoding of signal intensity is a valuable secondary cue, particularly for the detection of rapid spectral changes and the encoding of aperiodic sounds. Future versions of our chip will include these enhancements for rate encoding of signal intensity.

In conclusion, we have designed and tested an integrated circuit that computes, in real time, the evoked responses of auditory nerve, using analog, continuous-time processing. The chip offers a robust representation of audition, which can serve as a solid foundation for analog silicon systems that model higher auditory function.

Appendix 2A
Circuit Description of the Basilar-Membrane Model

The circuit model of cochlear mechanics developed by Lyon and Mead (Lyon and Mead, 1988a; Mead, 1989) is the foundation of our silicon model of auditory-fiber response. This appendix provides a brief description of the implementation of this basilar-membrane model.

The circuit is a one-dimensional physical model of the traveling-wave structure formed by the basilar membrane. In this viewpoint of cochlear function, the exponentially tapered stiffness of the basilar membrane and the motility of the outer hair cells combine to produce a pseudoresonant structure. The basilar-membrane circuit model implements this view of cochlear hydrodynamics using a cascade of second-order sections with exponentially scaled time constants.

Figure 2A.1 shows the CMOS circuit implementation of a second-order section. Input and output signals for the circuit are time-varying voltages. The gain blocks are transconductance amplifiers, operated in the subthreshold regime. Capacitors are formed using the gate capacitance of n-channel and p-channel MOS transistors in parallel. Because of subthreshold amplifier operation, the time constant of the second-order section is an exponential function of the voltage applied to the transconductance control inputs of $A1$ and $A2$, labeled τ in Figure 2A.1. Thus, a cascade of second-order circuits, with a linear gradient applied to the τ control inputs, has exponentially scaled time constants. To implement this gradient, we used a polysilicon wire that travels along the length of circuit, and that connects to the τ control input of each second-order section. A voltage difference across this wire, applied from off the chip, produces exponentially scaled time constants. The amplifier $A3$ provides active positive

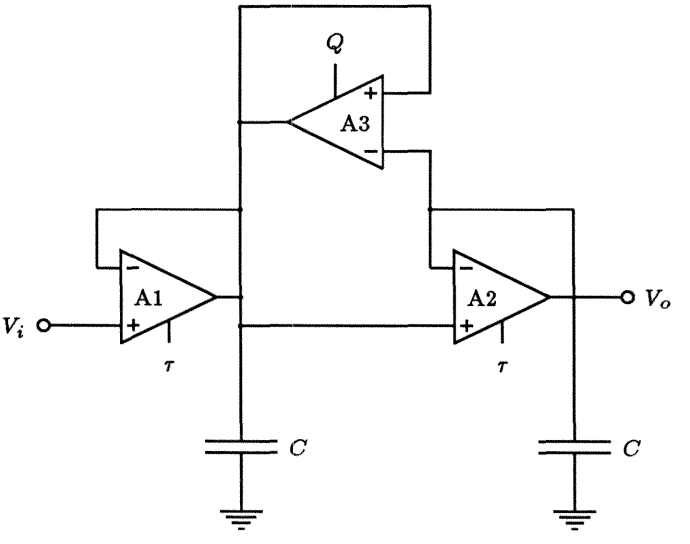


Figure 2A.1. Circuit implementation of a second-order section. Input V_i and output V_o are time-varying voltages. The τ and Q control inputs set bias currents on transconductance amplifiers $A1$, $A2$, and $A3$, to control both the characteristic frequency and the peak height of the lowpass-filter response.

feedback to the membrane, modeling the active mechanical feedback provided by the outer hair cells in biological cochleas. A second polysilicon wire is connected to the transconductance inputs of the $A3$ amplifiers in each second-order section (labeled Q in Figure 2A.1); a voltage gradient across this wire similar to that on the τ control inputs sets all the second-order sections to the same response shape.

A way to model the adjustment of basilar-membrane damping by higher brain centers is to use an automatic-gain-control system that varies the damping of the second-order sections locally. We have not implemented this automatic-gain-control system; however, we have brought off-chip several taps from the polysilicon wire that connects to the Q control of the second-order sections, allowing off-chip experiments with automatic gain control.

Appendix 2B

Circuit Description of the Auditory-Transduction Model

Our inner-hair-cell circuit models the signal-processing operations that occur during transduction: velocity sensing, nonlinear compression, and half-wave rectification. Our spiral-ganglion-neuron circuit converts the analog output of the inner-hair-cell circuit into fixed-width, fixed-height pulses; the mean firing rate of the circuit encodes the intensity of inner-hair-cell circuit response, whereas the temporal pattern of pulses reflects the waveform shape of inner-hair-cell circuit response. This appendix provides a description of the implementations of the inner-hair-cell circuit and of the spiral-ganglion-neuron circuit.

Figure 2B.1 shows our inner-hair-cell circuit model. A hysteretic-differentiator circuit (Mead, 1989) processes the input-voltage waveform from the basilar-membrane circuit, performing time differentiation and logarithmic compression. The circuit enhances the zero-crossings of the input waveform, accentuating phase information in the signal. The output voltage of the hysteretic differentiator connects to a novel implementation of a half-wave current rectifier.

Figure 2B.2 shows our half-wave current-rectifier circuit. To understand its operation, we consider the state of this circuit when the input voltage V_h is constant. If V_h is constant, $I_h = 0$, and V_a adapts such that $I_p = I_n$. For $I_h = 0$, we define the quiescent conditions $I_q \equiv I_p = I_n$ and $V_q \equiv V_a$. The value of I_q depends on the circuit bias voltage, V_s. A current mirror reflects this quiescent current to the circuit output. Thus, the output of the half-wave current-rectifier circuit in response to a constant voltage input is an adjustable bias current.

Now consider the circuit state when the input voltage V_h is a time-varying waveform. During the positive-going phase of the waveform, the current I_h is

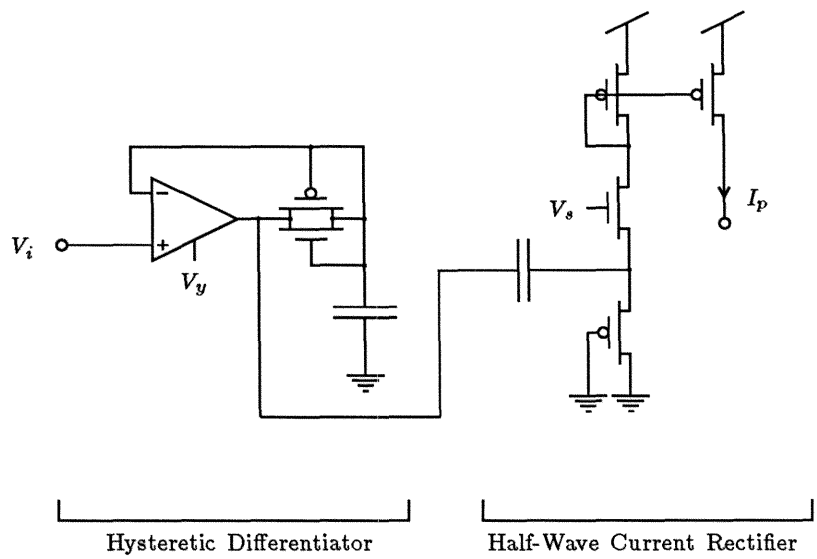


Figure 2B.1. The inner-hair-cell circuit model. Input V_i, from the basilar-membrane circuit, is a time-varying voltage. The hysteretic-differentiator circuit, biased by voltage V_y, performs time differentiation and logarithmic compression. The output of the hysteretic differentiator, a time-varying voltage, connects to the half-wave current-rectifier circuit, which is shown in more detail in Figure 2B.2.

positive, and $I_n = I_h + I_p$. As I_n increases, V_a must also increase; the amount of increase depends on the circuit bias voltage, V_s, as shown in the bottom graph in Figure 2B.2. If V_a increases, however, then I_p must decrease. So, during the positive-going phase of the waveform, the output current I_p decreases from the quiescent current I_q.

During the negative-going phase of the waveform, the current I_h is negative, $I_p = |I_h| + I_n$, and the output current of the circuit increases from the quiescent current I_q. Thus, the circuit converts the input time-varying voltage waveform V_h into a unidirectional current waveform I_p. For large $|I_h|$ relative to I_q, the current waveform I_p is not symmetrical about I_q, and the average value of I_p is greater than that of I_q; thus, the circuit performs the rectification function, as shown in the top graph in Figure 2B.2.

The current I_p is the output of the inner-hair-cell circuit. The spiral-ganglion neuron circuit model, shown in Figure 2B.3, converts this current into fixed-width, fixed-height pulses. The circuit — a slightly modified version of the neuron circuit described in (Mead, 1989) — creates a pulse rate that is linear in input current, for sufficiently low pulse rates. Thus, the average pulse rate of the circuit reflects the average value of I_p, whereas the temporal placement of each pulse reflects the shape of the current waveform I_p.

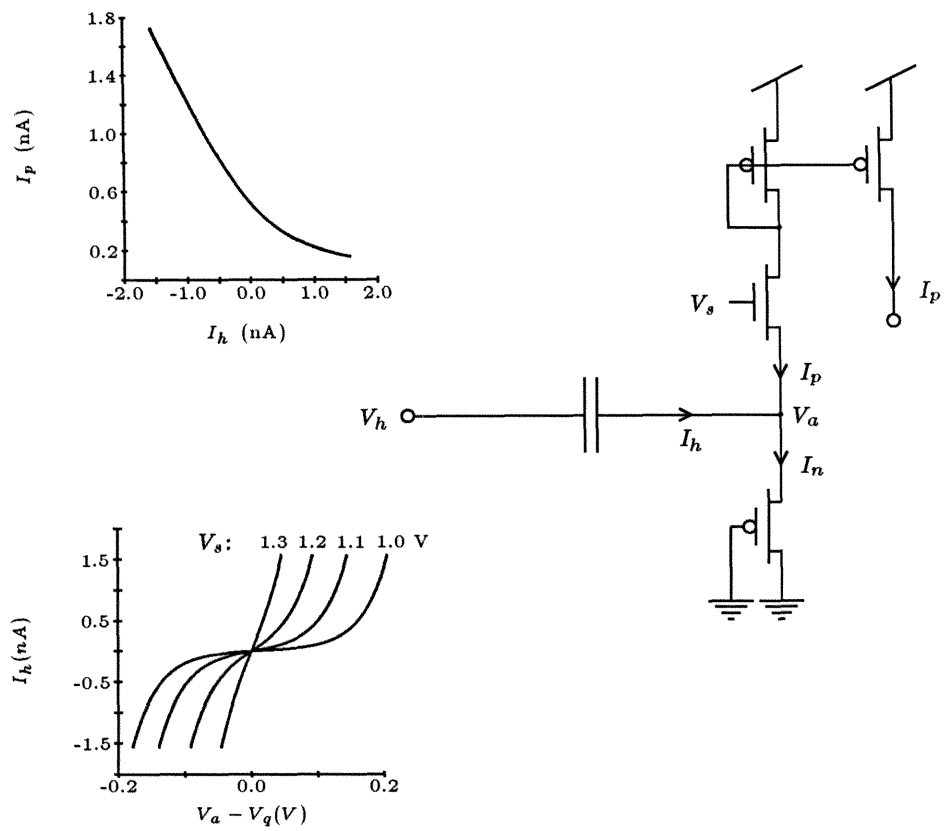


Figure 2B.2. The half-wave current-rectifier circuit. Input V_h, from the hysteretic-differentiator circuit, is a time-varying voltage. A floating capacitor couples V_h into the node associated with V_a, as the bidirectional time-varying current I_h. The bottom graph shows the change in V_a required to sink or source I_h, for several values of bias voltage V_s; the voltage V_q is the value of V_a when $I_h = 0$. When $V_a = V_q$ and $I_h = 0$, the circuit output, the unidirectional current I_p, is at a quiescent value, I_q, set by V_s. Nonzero values of I_h modulate the output current I_p about I_q; for large $|I_h|$ relative to I_q, the circuit output I_p is a half-wave–rectified version of I_h, as shown in the top graph. Graphs show theoretical responses.

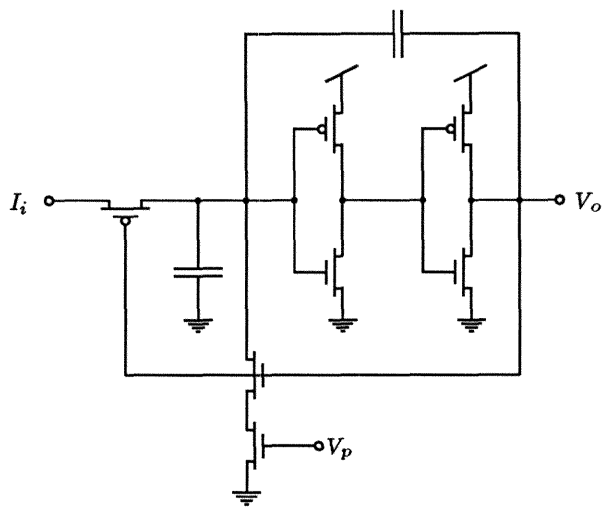


Figure 2B.3. The spiral-ganglion-neuron circuit. Circuit input, from the half-wave–rectification circuit, is the unidirectional current I_i. The circuit converts this current into fixed-width, fixed-height voltage pulses, at output V_o. The bias voltage V_p sets pulse width; the output voltage V_o pulses between V_{dd} and ground.

Chapter 3
Circuit Models of Nonlinear Inhibition

The silicon models of auditory localization and pitch perception, presented in Chapters 4 and 5, use inhibitory processing to improve the selectivity of their output representations. This chapter describes the inhibitory processing used in these projects.

Two general types of inhibition mediate activity in neural systems: subtractive inhibition, which sets a zero level for the computation, and multiplicative (nonlinear) inhibition, which regulates the gain of the computation. The circuit described in the chapter implements general nonlinear inhibition in its extreme form, known as *winner-take-all.* In addition to the projects in Chapters 4 and 5, the winner-take-all circuit has been used successfully in a model of visual stereopsis (Mahowald and Delbruck, 1989).

The chapter also describes a modification to this global winner-take-all circuit, which computes local nonlinear inhibition. The circuit allows multiple winners in the network, and is well suited for use in systems that represent a feature space topographically and that process several features in parallel. The research in this chapter was done in collaboration with Sylvie Ryckebusch, M. A. Mahowald, and Carver Mead; parts of this chapter were originally published in (Lazzaro *et al.*, 1988).

3.1 The Winner-Take-All Circuit

Figure 3.1 is a schematic diagram of the winner-take-all circuit. Each neuron receives a unidirectional current input I_k; the output voltages $V_1 \ldots V_n$ represent the results of the winner-take-all computation. A single wire, associated with the potential V_c, computes the inhibition for the entire circuit; for an n-neuron

circuit, this wire is $O(n)$ long. To compute the global inhibition, each neuron k contributes a current onto this common wire, using transistor T_{2_k}. To apply this global inhibition locally, each neuron responds to the common wire voltage V_c, using transistor T_{1_k}. This computation is continuous in time; no clocks are used. The circuit exhibits no hysteresis, and operates with a time constant related to the size of the largest input. The output representation of the circuit is not binary; the winning output encodes the logarithm of its associated input.

A static and dynamic analysis of the two-neuron circuit illustrates these properties. Figure 3.2 shows a schematic diagram of a two-neuron winner-take-all circuit. To understand the behavior of the circuit, we first consider the input condition $I_1 = I_2 \equiv I_m$. Transistors T_{1_1} and T_{1_2} have identical potentials at gate and source, and are both sinking I_m; thus, the drain potentials V_1 and V_2 must be equal. Transistors T_{2_1} and T_{2_2} have identical source, drain, and gate potentials, and therefore must sink the identical current $I_{c_1} = I_{c_2} = I_c/2$. In the subthreshold region of operation, the equation $I_m = I_o \exp(V_c/V_o)$ describes transistors T_{1_1} and T_{1_2}, where I_o is a fabrication parameter, and $V_o = kT/q\kappa$. Likewise, the equation $I_c/2 = I_o \exp((V_m - V_c)/V_o)$, where $V_m \equiv V_1 = V_2$, describes transistors T_{2_1} and T_{2_2}. Solving for $V_m(I_m, I_c)$ yields

$$V_m = V_o \ln(\frac{I_m}{I_o}) + V_o \ln(\frac{I_c}{2I_o}). \tag{3.1}$$

Thus, for equal input currents, the circuit produces equal output voltages; this behavior is desirable for a winner-take-all circuit. In addition, the output voltage V_m logarithmically encodes the magnitude of the input current I_m.

The input condition $I_1 = I_m + \delta_i$, $I_2 = I_m$ illustrates the inhibitory action of the circuit. Transistor T_{1_1} must sink δ_i more current than in the previous example; as a result, the gate voltage of T_{1_1} rises. Transistors T_{1_1} and

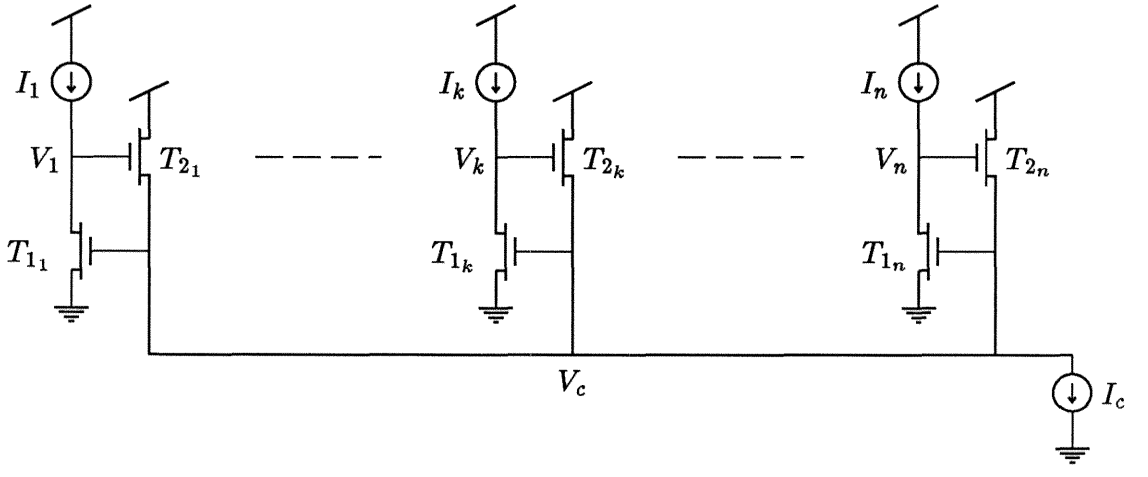


Figure 3.1. Schematic diagram of the winner-take-all circuit. Each neuron receives a unidirectional current input I_k; the output voltages $V_1 \ldots V_n$ represent the result of the winner-take-all computation. If $I_k = \max(I_1 \ldots I_n)$, then V_k is a logarithmic function of I_k; if $I_j \ll I_k$, then $V_j \approx 0$.

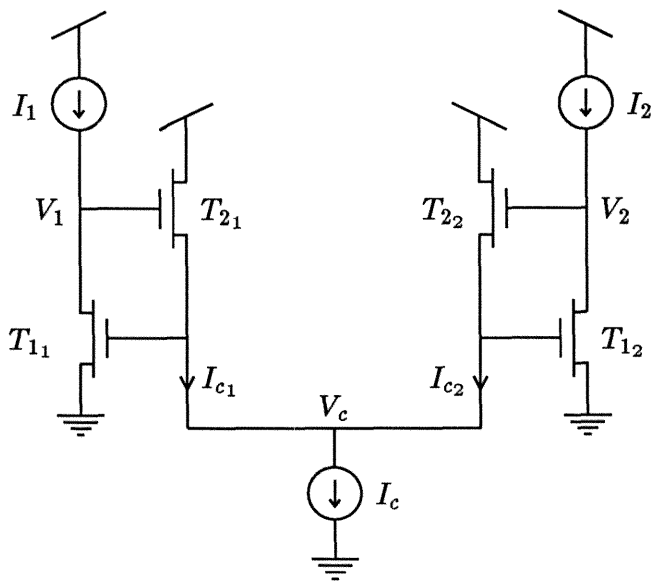


Figure 3.2. Schematic diagram of a two-neuron winner-take-all circuit.

T_{1_2} share a common gate, however; thus, T_{1_2} must also sink $I_m + \delta_i$. But only I_m is present at the drain of T_{1_2}. To compensate, the drain voltage of T_{1_2}, V_2, must decrease. For small δ_is, the Early effect decreases the current through T_{1_2}, decreasing V_2 linearly with δ_i. For large δ_is, T_{1_2} must leave saturation, driving V_2 to approximately 0 V. As desired, the output associated with the smaller input diminishes. For large δ_is, $I_{c_2} \approx 0$, and $I_{c_1} \approx I_c$. The equation $I_m + \delta_i = I_o \exp(V_c/V_o)$ describes transistor T_{1_1}, and the equation $I_c = I_o \exp((V_1 - V_c)/V_o)$ describes transistor T_{2_1}. Solving for V_1 yields

$$V_1 = V_o \ln(\frac{I_m + \delta_i}{I_o}) + V_o \ln(\frac{I_c}{I_o}). \tag{3.2}$$

The winning output encodes the logarithm of the associated input. The symmetrical circuit topology ensures similar behavior for increases in I_2 relative to I_1.

Equation 3.2 predicts the winning response of the circuit; a more complex expression, derived in Appendix 3A, predicts the losing and crossover response of the circuit. Figure 3.3 is a plot of this analysis, fit to experimental data. Figure 3.4 shows the wide dynamic range and logarithmic properties of the circuit; the experiment in Figure 3.3 is repeated for several values of I_2, ranging over four orders of magnitude.

The conductance of transistors T_{1_1} and T_{1_2} determines the losing response of the circuit. The Early voltage, V_e, is a measure of the conductance of a saturated MOS transistor. The expression

$$V_e = L\frac{\partial V_d}{\partial L} \tag{3.3}$$

defines the Early voltage, where V_d is the drain potential of a transistor, and L is the channel length of a transistor. Thus, the width of the losing response

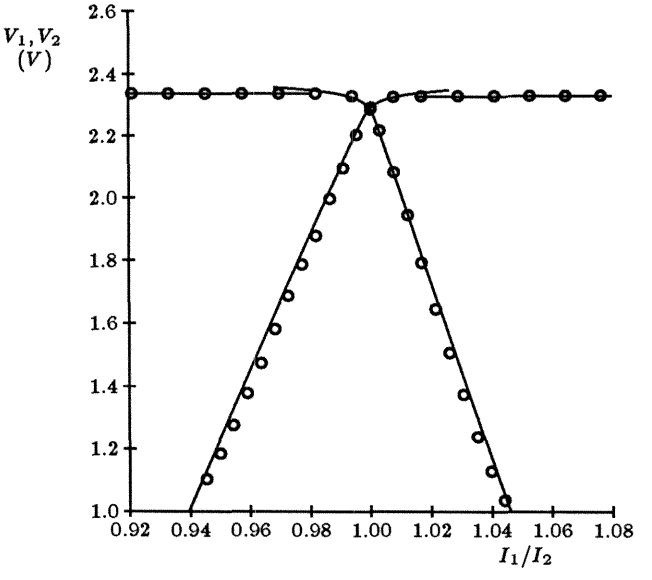


Figure 3.3. Experimental data (circles) and theoretical statements (solid lines) for a two-neuron winner-take-all circuit. I_1, the input current of the first neuron, is swept about the value of I_2, the input current of the second neuron; neuron voltage outputs V_1 and V_2 are plotted versus normalized input current.

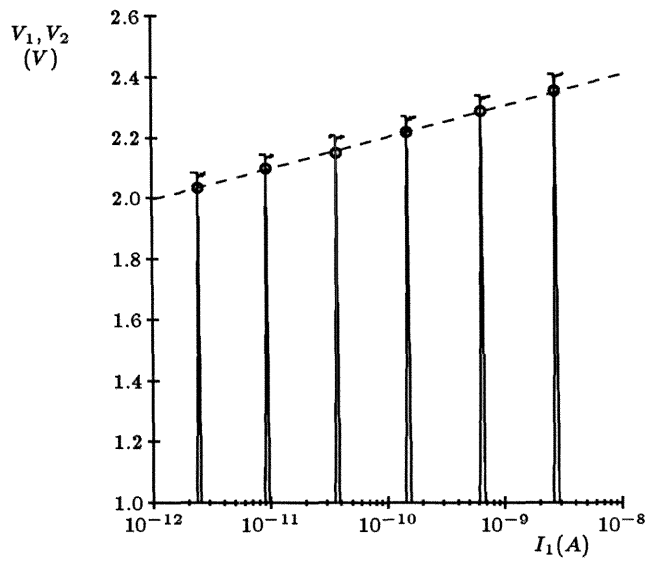


Figure 3.4. The experiment of Figure 3.3 is repeated for several values of I_2; experimental data of output voltage response are plotted versus absolute input current on a log scale. The output voltage $V_1 = V_2$ is highlighted with a circle for each experiment. The dashed line is a theoretical expression confirming logarithmic behavior over four orders of magnitude (Equation 3.1).

of the circuit depends on the channel length of transistors T_{1_1} and T_{1_2}. Figure 3.3 shows data for a circuit where the channel length of transistors T_{1_1} and T_{1_2} is 13.5 μm. Figure 3.5 shows data for a circuit with a wider losing response; in this circuit, the channel length for transistors T_{1_1} and T_{1_2} is 3 μm, the smallest allowable in the fabrication technology used.

Increasing the channel length of transistors T_{1_1} and T_{1_2} narrows the losing response of the circuit; alternatively, circuit modification also can narrow the losing response. The circuit shown in Figure 3.6 approximately halves the width of the original losing response, through source degeneration of transistors T_{1_1} and T_{1_2} by the added diode-connected transistors T_{3_1} and T_{3_2}. Figure 3.7 shows experimental data for this modified circuit.

3.2 Time Response of the Winner-Take-All Circuit

A good winner-take-all circuit should be stable, and should not exhibit damped oscillations (ringing) in response to input changes. This section explores these dynamic properties of our winner-take-all circuit, and predicts the temporal response of the circuit. Figure 3.8 shows the two-neuron winner-take-all circuit, with capacitances added to model dynamic behavior.

Appendix 3B shows a small-signal analysis of this circuit. This analysis shows that the circuit is stable and does not ring if $I_c > 4I(C_c/C)$, where $I_1 \approx I_2 \approx I$. Figure 3.9 compares this bound with experimental data.

If $I_c > 4I(C_c/C)$, the circuit exhibits first-order behavior. The time constant CV_o/I sets the dynamics of the winning neuron, where $V_o = kT/q\kappa \approx 40\,\mathrm{m}V$. The time constant CV_e/I sets the dynamics of the losing neuron, where $V_e \approx 50\,V$. Figure 3.10 compares these predictions with experimental data, for several variants of the winner-take-all circuit.

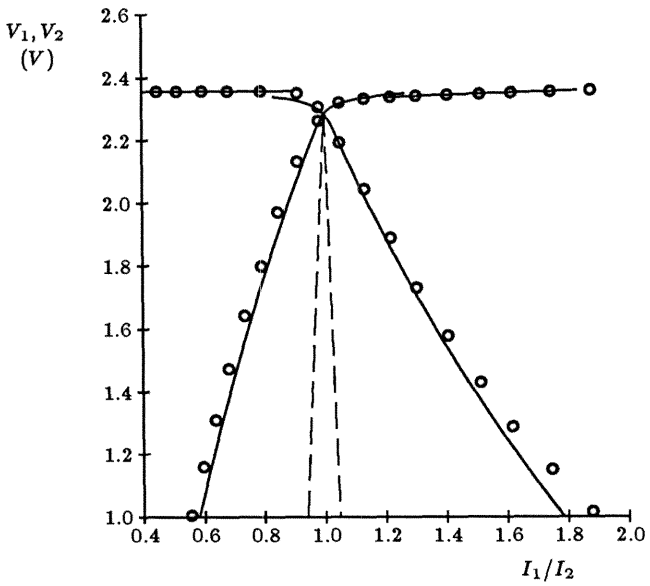


Figure 3.5. Experimental data (circles) and theoretical statements (solid lines) for a two-neuron winner-take-all circuit with a channel length for transistors T_{1_1} and T_{1_2} of 3 μm. The dotted lines show the losing response for the circuit used in Figure 3.3, which has a channel length for transistors T_{1_1} and T_{1_2} of 13.5 μm.

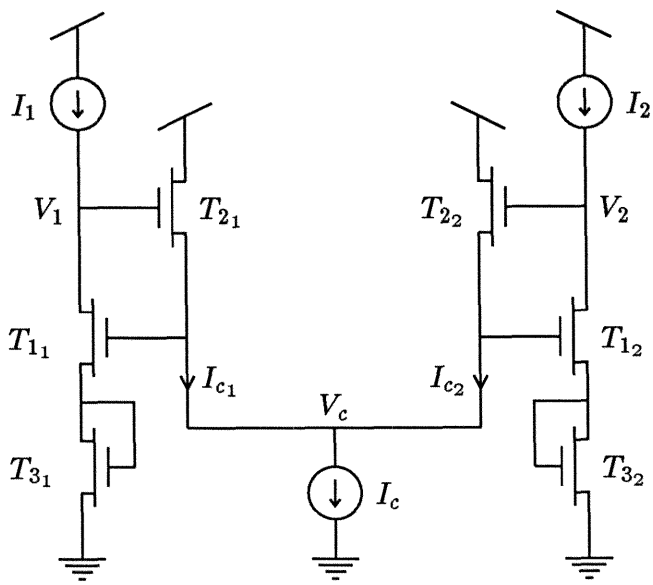


Figure 3.6. Schematic diagram of a two-neuron winner-take-all circuit, modified to produce a narrower losing response.

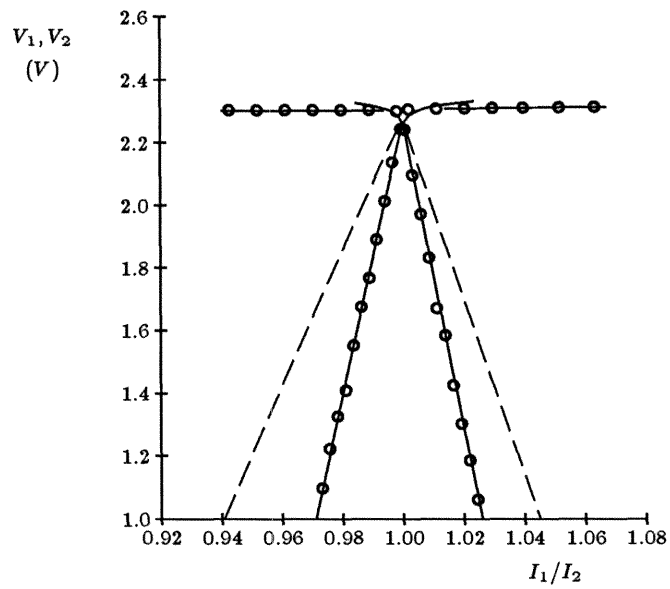


Figure 3.7. Experimental data (circles) and theoretical statements (solid lines) for a two-neuron winner-take-all circuit, modified to produce a narrower losing response. The dotted lines show losing response for the circuit used in Figure 3.4.

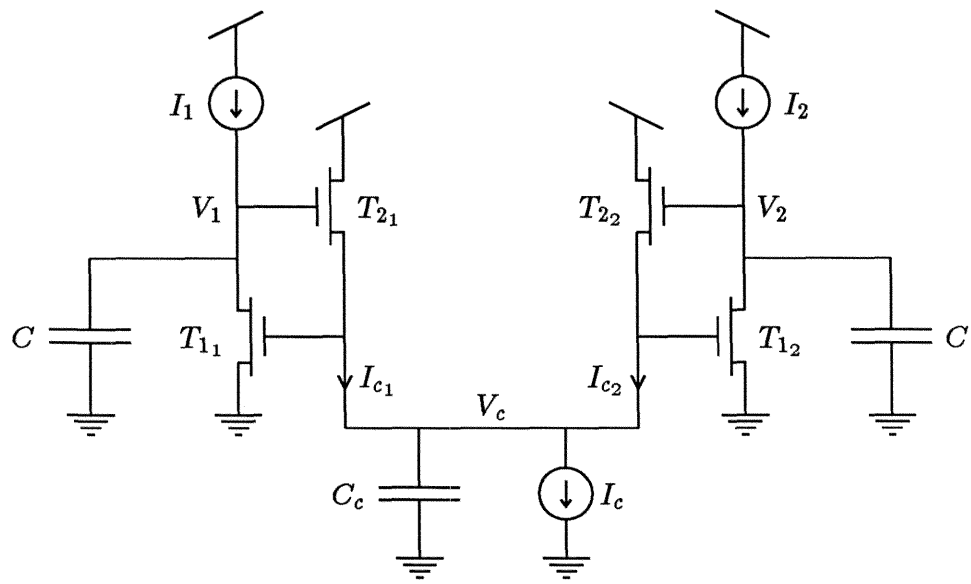


Figure 3.8. Schematic diagram of a two-neuron winner-take-all circuit, with capacitances added for dynamic analysis. C is a large MOS capacitor added to each neuron for smoothing; C_c models the parasitic capacitance contributed by the gates of T_{11} and T_{12}, the drains of T_{21} and T_{22}, and the interconnect.

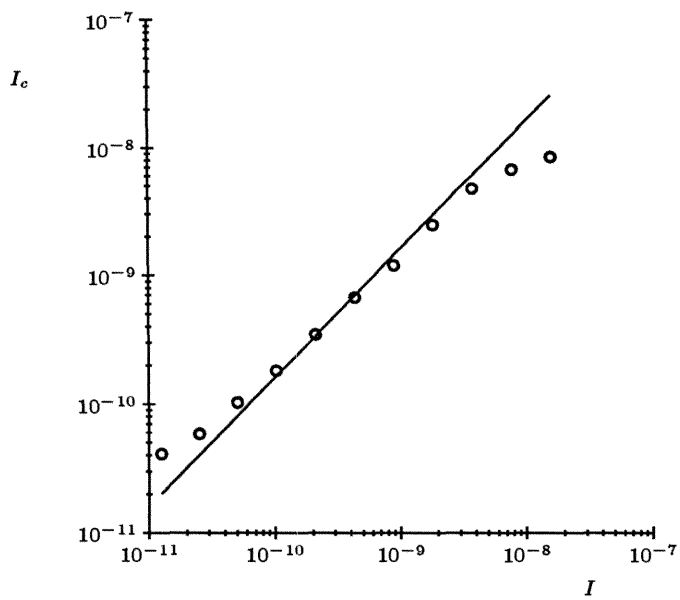


Figure 3.9. Experimental data (circles) and theoretical statements (solid line) for a two-neuron winner-take-all circuit, showing the smallest I_c, for a given I, necessary for a first-order response to a small-signal step input.

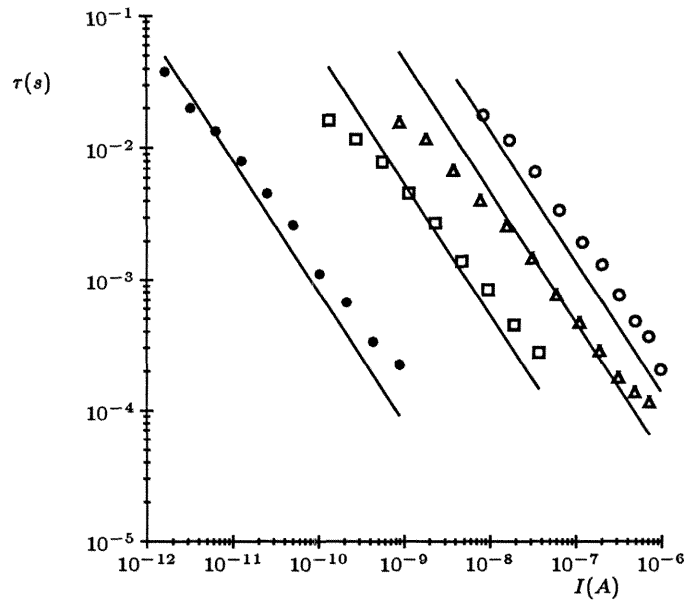


Figure 3.10. Experimental data (symbols) and theoretical statements (solid lines) for a two-neuron winner-take-all circuit, showing the time constant of the first-order response to a small-signal step input. The winning response (filled circles) and losing response (triangles) of a winner-take-all circuit with the static response of Figure 3.3 are shown; the time constants differ by several orders of magnitude. Losing responses for winner-take-all circuits with the static responses shown in Figure 3.5 (squares) and in Figure 3.7 (open circles) are also shown, demonstrating the effect of the width of static response on dynamic behavior.

3.3 The Local Nonlinear Inhibition Circuit

The winner-take-all circuit in Figure 3.1, as previously explained, locates the largest input to the circuit. Figure 3.11 shows this behavior. Figure 3.11(a) is the spatial input to a winner-take-all circuit with 16 neurons, with input 8 much higher than all other inputs. Figure 3.11(b) shows the circuit response to this input; only neuron 8 has significant response.

Certain applications require a gentler form of nonlinear inhibition. Sometimes, a circuit that can represent multiple intensity scales is necessary. Without circuit modification, the winner-take-all circuit in Figure 3.1 can perform this task. Appendix 3C explains this mode of operation.

Other applications require a local winner-take-all computation, with each winner having influence over only a limited spatial area. Figure 3.11(c) shows the desired computation. As in Figure 3.11(b), neuron 8 has the largest response in the circuit. However, neuron 8 suppresses the output of only nearby neurons; neurons far from neuron 8 have significant responses, encoding their input signals.

Figure 3.12 shows a circuit that computes the local winner-take-all function. The circuit is identical to the original winner-take-all circuit, except that each neuron connects to its nearest neighbors with a nonlinear resistor circuit (Mead, 1989). Each resistor conducts a current I_r in response to a voltage ΔV across it, where

$$I_r = I_s \tanh(\Delta V/(2V_o)). \tag{3.4}$$

I_s, the saturating current of the resistor, is a controllable parameter. The current source I_c, present in the original winner-take-all circuit, is distributed between the resistors in the local winner-take-all circuit.

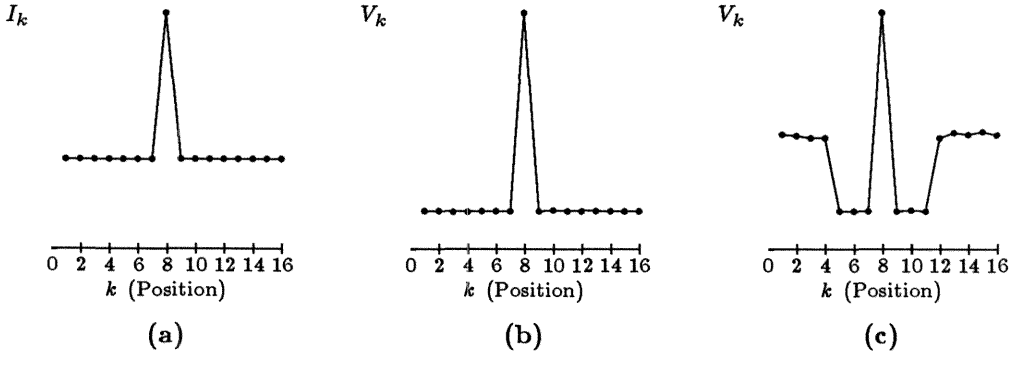


Figure 3.11. Comparison of idealized winner-take-all spatial response and the desired local winner-take-all response. The horizontal axis of each plot represents spatial position in a 16-neuron network. **a.** The plot shows a spatial impulse function, used as input to compare the two concepts. The vertical axis shows the input current to each neuron, with $I_8 \gg I_{k\neq 8}$. **b.** The plot shows the winner-take-all response. **c.** The plot shows the local winner-take-all response.

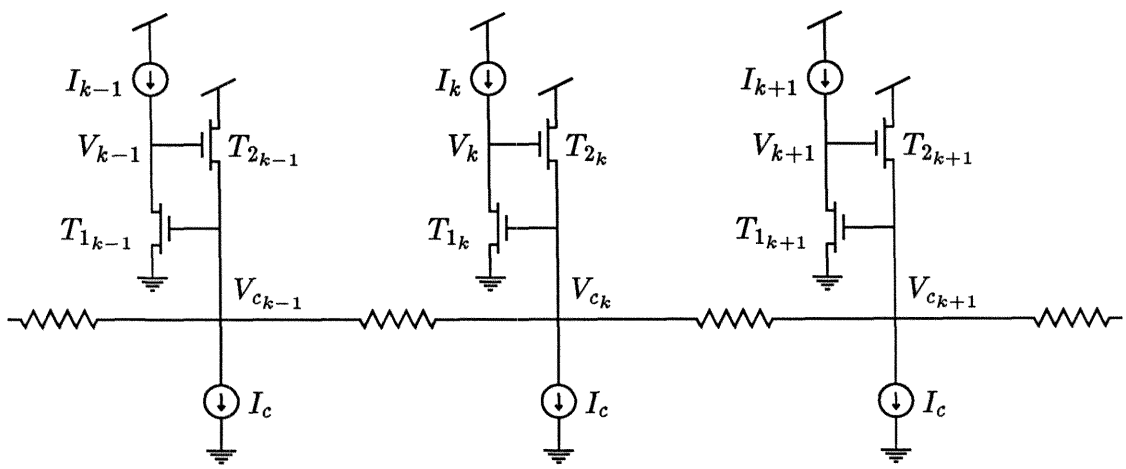


Figure 3.12. Schematic diagram of a section of the local winner-take-all circuit. Each neuron i receives a unidirectional current input I_i; the output voltages V_i represent the result of the local winner-take-all computation.

To understand the operation of the local winner-take-all circuit, we consider the circuit response to a spatial impulse, defined as $I_k \gg I$, where $I \equiv I_{i \neq k}$. $I_k \gg I_{k-1}$ and $I_k \gg I_{k+1}$, so V_{c_k} is much larger than $V_{c_{k-1}}$ and $V_{c_{k+1}}$, and the resistor circuits connecting neuron k with neuron $k-1$ and neuron $k+1$ saturate. Each resistor sinks I_s current when saturated; transistor T_{2_k} thus conducts $2I_s + I_c$ current. In the subthreshold region of operation, the equation $I_k = I_o \exp(V_{c_k}/V_o)$ describes transistor T_{1_k}, and the equation $2I_s + I_c = I_o \exp((V_k - V_{c_k})/V_o)$ describes transistor T_{2_k}. Solving for V_k yields

$$V_k = V_o \ln((2I_s + I_c)/I_o) + V_o \ln(I_k/I_o). \tag{3.5}$$

As in the original winner-take-all circuit, the output of a winning neuron encodes the logarithm of that neuron's associated input.

As mentioned, the resistor circuit connecting neuron k with neuron $k-1$ sinks I_s current. The current sources I_c associated with neurons $k-1$, $k-2$, ... must supply this current. If the current source I_c for neuron $k-1$ supplies part of this current, then the transistor $T_{2_{k-1}}$ carries no current, and the neuron output V_{k-1} approaches zero. Similar reasoning applies to neurons $k+1$, $k+2$, ... In this way, a winning neuron inhibits its neighboring neurons.

This inhibitory action does not extend throughout the network. Neuron k needs only I_s current from neurons $k-1$, $k-2$, ... Thus, neurons sufficiently distant from neuron k maintain the service of their current source I_c, and the outputs of these distant neurons can be active. Since, for a spatial impulse, all neurons $k-1$, $k-2$, ... have an equal input current I, all distant neurons have the equal output

$$V_{i \ll k} = V_o \ln(I_c/I_o) + V_o \ln(I/I_o). \tag{3.6}$$

Similar reasoning applies for neurons $k+1$, $k+2$, ...

The relative values of I_s and I_c determine the spatial extent of the inhibitory action. Figure 3.13 shows the spatial impulse response of the local winner-take-all circuit, for different settings of I_s/I_c.

3.4 Discussion

The circuits described in this chapter use the full analog nature of MOS devices to realize an interesting class of neural computations efficiently. The circuits exploit the physics of the medium in many ways. The winner-take-all circuit uses a single wire to compute and communicate inhibition for the entire circuit. Transistor T_{1_k} in the winner-take-all circuit uses two physical phenomena in its computation: its exponential current function encodes the logarithm of the input, and the finite conductance of the transistor defines the losing output response. As evolution exploits all the physical properties of neural devices to optimize system performance, designers of synthetic neural systems should strive to harness the full potential of the physics of their media.

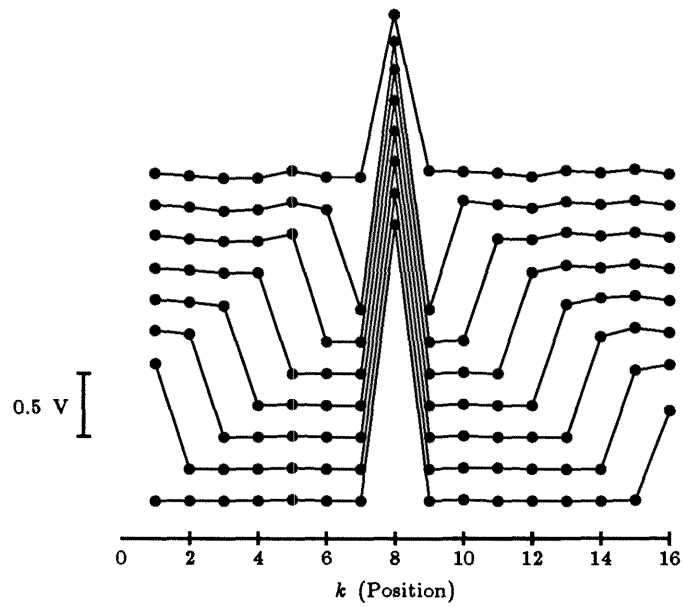


Figure 3.13. Experimental data showing the spatial impulse response of the local winner-take-all circuit, for values of I_s/I_c ranging over a factor of 12.7. Wider inhibitory responses correspond to larger ratios. For clarity, the plots are vertically displaced in 0.25 V increments.

Appendix 3A

Static Response of the Winner-Take-All Circuit

Figure 3.3 compares data from the two-neuron winner-take-all circuit with a closed-form theoretical statement describing the losing and crossover response of the circuit. This appendix derives that theoretical statement.

Figure 3A.1 shows a small-signal circuit model of the two-neuron winner-take-all circuit (Figure 3.2). For a particular operating point $[I_1, I_2, I_{c_1}, I_{c_2}]$, the model shows the effect of a small change in I_1, denoted i_1, on the circuit voltages V_1, V_2, and V_c, indicated by the small-signal voltages v_1, v_2, and v_c. In this model, a linear resistor r_{i_j}, in parallel with a linear dependent current source, with a conductance g_{i_j}, replaces each transistor T_{i_j} from Figure 3.2. For a particular operating point in subthreshold, the small-signal parameters are

$$\begin{aligned} g_{1_1} &= I_1/V_o, & g_{2_1} &= I_{c_1}/V_o, & r_{1_1} &= V_e/I_1, & r_{2_1} &= V_e/I_{c_1}, \\ g_{1_2} &= I_2/V_o, & g_{2_2} &= I_{c_2}/V_o, & r_{1_2} &= V_e/I_2, & r_{2_2} &= V_e/I_{c_2}, \end{aligned} \tag{3A.1}$$

where V_e, the Early voltage, is a measure of transistor resistance, and $V_o = kT/q\kappa$. This small-signal model is a linear system, which we can solve analytically using conventional techniques; applying the approximation $V_e + V_o \approx V_e$ to the solution yields the simplified equations

$$\begin{aligned} \frac{v_1}{i_1} &= (1/I_1)(V_o + V_e(I_{c_2}/I_c)), \\ \frac{v_2}{i_1} &= -V_e(1/I_1)(I_{c_1}/I_c). \end{aligned} \tag{3A.2}$$

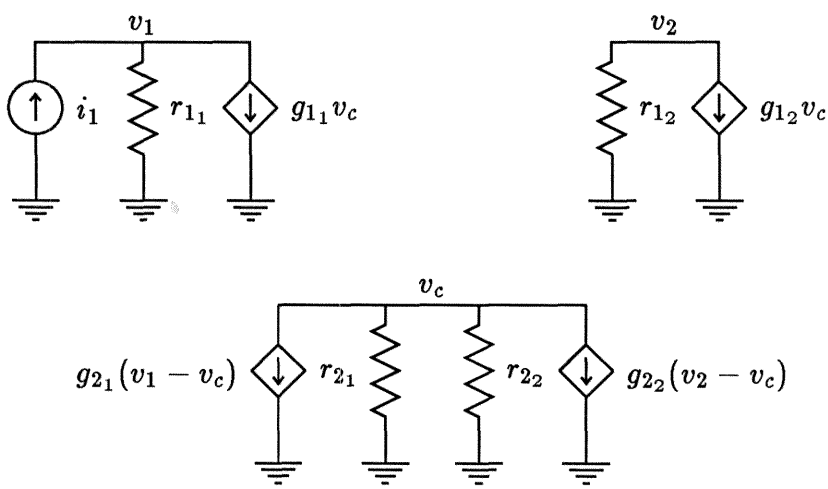


Figure 3A.1. Small-signal model of the two-neuron winner-take-all circuit.

Note that both small-signal and large-signal quantities appear in Equation 3A.2. We can view the small-signal quantities as differential elements of large-signal quantities; as a result, we can rewrite Equation 3A.2 as the pair of nonlinear differential equations

$$\begin{aligned} \frac{dV_1}{dI_1} &= (1/I_1)(V_o + V_e(I_{c_2}/I_c)), \\ \frac{dV_2}{dI_1} &= -V_e(1/I_1)(I_{c_1}/I_c). \end{aligned} \tag{3A.3}$$

Solving this pair of nonlinear differential equations yields a complete description of circuit response. We begin by eliminating I_{c_1} and I_{c_2} from the equations. Referring to Figure 3.2, the equations

$$\begin{aligned} I_{c_1} &= I_o \exp((V_1 - V_c)/V_o), \\ I_{c_2} &= I_o \exp((V_2 - V_c)/V_o) \end{aligned} \tag{3A.4}$$

describe transistors T_{2_1} and T_{2_2}. From Kirchoff's current law, we know that $I_{c_1} + I_{c_2} = I_c$; substitution of Equation 3A.4 into this equation yields the expression

$$I_c = I_o \exp((V_1 - V_c)/V_o) + I_o \exp((V_2 - V_c)/V_o). \tag{3A.5}$$

Dividing Equation 3A.4 by Equation 3A.5 eliminates V_c, leaving, after rearrangement,

$$\begin{aligned} I_{c_1}/I_c &= \frac{1}{1 + \exp((V_2 - V_1)/V_o)}, \\ I_{c_2}/I_c &= \frac{1}{1 + \exp((V_1 - V_2)/V_o)}. \end{aligned} \tag{3A.6}$$

These expressions fit nicely into Equation 3A.3, eliminating I_{c_1} and I_{c_2}, and leaving a set of differential equations involving only V_1, V_2, and I_1:

$$\frac{dV_1}{dI_1} = (1/I_1)(V_o + V_e(\frac{1}{1+\exp((V_1 - V_2)/V_o)})), \tag{A7a}$$

$$\frac{dV_2}{dI_1} = -V_e(1/I_1)(\frac{1}{1+\exp((V_2 - V_1)/V_o)}). \tag{A7b}$$

Equation 3A.7a contains V_2 only in the subexpression

$$\frac{1}{1+\exp((V_1 - V_2)/V_o)}; \tag{3A.8a}$$

Equation 3A.7b contains V_1 only in the subexpression

$$\frac{1}{1+\exp((V_2 - V_1)/V_o)}. \tag{3A.8b}$$

These subexpressions are both *Fermi functions* of the difference $V_1 - V_2$. For $V_1 - V_2 \gg V_o$, the value of the subexpression 3A.8a is approximately 0, whereas the value of the subexpression 3A.8b is approximately 1; for $V_2 - V_1 \gg V_o$, the value of the subexpression 3A.8a is approximately 1, whereas the value of the subexpression 3A.8b is approximately 0. In the region $V_1 \approx V_2$, we can assume that V_1 and V_2 are both changing with the same magnitude of slope relative to I_1. We can write this approximation as $V_1 - V_2 \approx 2(V_1 - V_m)$ and $V_2 - V_1 \approx 2(V_2 - V_m)$, where, from the qualitative analysis in the chapter, $V_m \equiv V_1 = V_2$ when $I_1 = I_2 \equiv I_m$. We can use this approximation to decouple Equations 3A.7a and 3A.7b, producing

$$\begin{aligned} \frac{dV_1}{dI_1} &= (1/I_1)(V_o + V_e(\frac{1}{1+\exp(2(V_1 - V_m)/V_o)})), \\ \frac{dV_2}{dI_1} &= -V_e(1/I_1)(\frac{1}{1+\exp(2(V_2 - V_m)/V_o)}). \end{aligned} \tag{3A.9}$$

We can solve these equations by straightforward integration, yielding, after application of the approximation $V_e + V_o \approx V_e$,

$$\ln(I_1/I_m) = \frac{(V_1 - V_m)}{V_e} + \frac{1}{2}\ln(1 + (V_o/V_e)\exp(2(V_1 - V_m)/V_o)), \quad (3A10a)$$

$$\ln(I_1/I_m) = \frac{(V_m - V_2)}{V_e} + \frac{1}{2}(V_o/V_e)(1 - \exp(2(V_2 - V_m)/V_o)). \quad (3A10b)$$

Equation 3A.10a predicts the value of I_1 for a given value of V_1, whereas Equation 3A.10b predicts the value of I_1 for a given value of V_2; in this way, these equations are a closed-form approximation of circuit response. To understand the behavior of the circuit, and to evaluate the effect of the approximations $V_1 - V_2 \approx 2(V_1 - V_m)$ and $V_2 - V_1 \approx 2(V_2 - V_m)$, we can simplify Equations 3A.10a and 3.A10b for three regions of interest: $V_1 \approx V_2 \approx V_m$, $V_1 \gg V_m$ while $V_2 \ll V_m$, and $V_1 \ll V_m$ while $V_2 \gg V_m$.

First consider the condition $V_1 \approx V_2 \approx V_m$. In this case, $|V_1 - V_2| \rightarrow 0$, $I_1/I_m \rightarrow 1$, and we can linearize the transcendental functions in Equations 3A.10a and 3A.10b, yielding the simpler relations

$$\begin{aligned} V_1 &= (V_e/2)((I_1/I_m) - 1)) + V_m, \\ V_2 &= (V_e/2)(1 - (I_1/I_m)) + V_m. \end{aligned} \quad (3A.11)$$

In this region, V_1 and V_2 are a linear function of I_1, with a slope of $\pm V_e/(2I_m)$.

Next, consider the condition $V_1 \gg V_m$ while $V_2 \ll V_m$, valid when $I_1 > I_m$. In Equation 3A.10b, $V_2 \ll V_m$ implies $\exp(2(V_2 - V_m)/V_o) \rightarrow 0$. This simplification yields, after rearrangement,

$$V_2 = V_o/2 + V_m - V_e \ln(I_1/I_m). \quad (3A.12)$$

If we use the notation $I_1 = I_m + \delta_i$, as in the earlier qualitative analysis, we can rewrite the subexpression $\ln(I_1/I_m)$ as $\ln(1 + (\delta_i/I_m))$, which we can approximate as δ_i/I_m for small δ_i/I_m, yielding the simplified result

$$V_2 = V_o/2 + V_m - (V_e/I_m)\delta_i. \tag{3A.13}$$

Thus, in this region, V_2 decreases linearly with δ_i, with a slope of V_e/I_m, which is twice as large as in the previous condition.

We can similarly derive a simplified expression for V_1, for the same condition $V_1 \gg V_m$ while $V_2 \ll V_m$. In Equation 3A.10a, $V_1 \gg V_m$ implies $(V_o/V_e)\exp(2(V_1 - V_m)/V_o) \gg 1$. This approximation yields, after rearrangement,

$$V_1 = V_o \ln(I_1/I_m) + (V_o/2)\ln(V_e/V_o) + V_m. \tag{3A.14}$$

For this condition, as predicted by Equation 3.2 in the chapter, V_1 is a logarithmic function of I_1. However, when does the approximation $(V_o/V_e)\exp(2(V_1 - V_m)/V_o) \gg 1$ hold? This inequality, when rearranged, yields the constraint

$$(V_1 - V_m) \gg (V_o/2)\ln(V_e/V_o). \tag{3A.15}$$

Therefore, for a typical fabrication process, $V_1 - V_m$ must be much greater than 0.15 V for Equation 3A.14 to hold! This error stems from the central approximation $V_1 - V_2 \approx 2(V_1 - V_m)$, which is valid for only $V_1 - V_2 \leq V_o$. Thus, for this region of operation, Equation 3.2 best predicts circuit behavior.

Finally, we consider the condition $V_1 \ll V_m$ while $V_2 \gg V_m$, valid when $I_1 < I_m$. In Equation 3A.10a, $V_1 \ll V_m$ implies $(V_o/V_e)\exp(2(V_1 - V_m)/V_o) \to 0$. This simplification yields, after rearrangement,

$$V_1 = V_m + V_e \ln(I_1/I_m). \tag{3A.16}$$

If we use the notation $I_1 = I_m - \delta_i$, as in the earlier analysis, we can rewrite the subexpression $\ln(I_1/I_m)$ as $\ln(1 - (\delta_i/I_m))$, which we can approximate as $-\delta_i/I_m$ for small $|-\delta_i/I_m|$, yielding the simplified result

$$V_1 = V_m - (V_e/I_m)\delta_i. \tag{3A.17}$$

Thus, in this region, V_1 decreases linearly with δ_i, with a slope of V_e/I_m. The losing responses for V_1 and V_2 are thus identical.

We can similarly derive a simplified expression for V_2, for the same condition $V_1 \ll V_m$ while $V_2 \gg V_m$. For Equation 3A.10b, $V_2 \gg V_m$ implies $\exp(2(V_2 - V_m)/V_o) \gg 1$. This approximation yields, after rearrangement,

$$\ln(I_1/I_m) = (V_m - V_2)/V_e - (1/2)(V_o/V_e)\exp(2(V_2 - V_m)/V_o). \tag{3A.18}$$

As $V_2 - V_m$ increases, the right side of this equation grows exponentially large and negative, forcing I_1 to grow closer and closer to zero; thus, V_2 is constant with I_1. However, the poor approximation $V_2 - V_1 \approx 2(V_2 - V_m)$ for $V_2 - V_1 \geq V_o$ stunts this exponential growth. The qualitative analysis in the chapter predicts this constant value accurately, as

$$V_2 = V_o \ln(\frac{I_m}{I_o}) + V_o \ln(\frac{I_c}{I_o}). \tag{3A.19}$$

In summary, Equations 3A.10a and 3A.10b predict the losing and crossover response of the circuit, whereas Equations 3.2 and 3A.19 predict the winning response of the circuit. Figure 3.4 is a plot of this analysis, fitted to experimental data. Figure 3A.2 expands the crossover region of Figure 3.4, showing the crossover region between losing and winning analyses. The theoretical predictions in Figure 3.5 and Figure 3.7 also use this analysis, with altered values of V_e.

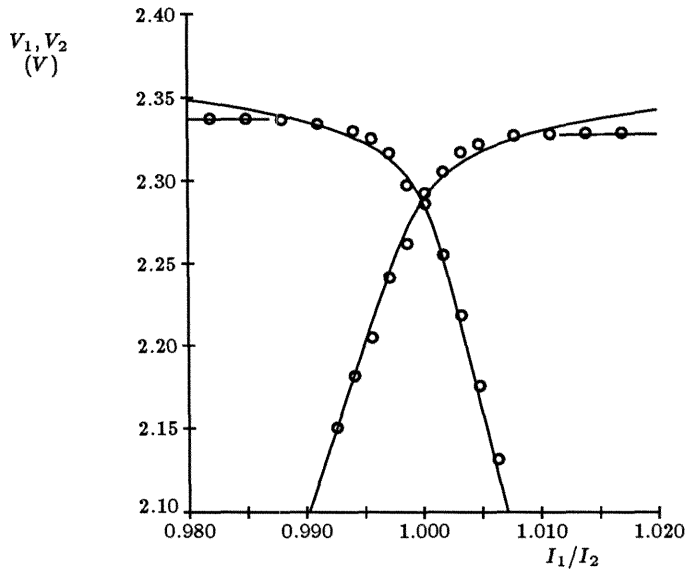


Figure 3A.2. Experimental data (circles) and theoretical statements (solid lines) for a two-neuron winner-take-all circuit in the crossover region.

Appendix 3B
Dynamic Response of the Winner-Take-All Circuit

In the chapter, we presented theoretical predictions of the time response of the winner-take-all circuit; in Figure 3.9 and 3.10, we compared these predictions with experimental data. In this appendix, we derive these theoretical predictions.

Figure 3.8 in the chapter shows a schematic diagram for a two-neuron winner-take-all circuit, with capacitances added to model dynamic behavior. Figure 3B.1 shows a small-signal circuit model for this circuit. For a particular operating point $[I_1, I_2, I_{c_1}, I_{c_2}]$, the model shows the effect of a small change in I_1, denoted by i_1, on the circuit voltages V_1, V_2, and V_c, indicated by the small-signal voltages v_1, v_2, and v_c. In this model, a linear resistor r_{i_j}, in parallel with a linear dependent current source, with a conductance g_{i_j}, replaces each transistor T_{i_j} from Figure 3.2. For a particular operating point in subthreshold, the small-signal parameters are

$$\begin{aligned} g_{1_1} &= I_1/V_o, \qquad g_{2_1} = I_{c_1}/V_o, \qquad r_{1_1} = V_e/I_1, \qquad r_{2_1} = V_e/I_{c_1}, \\ g_{1_2} &= I_2/V_o, \qquad g_{2_2} = I_{c_2}/V_o, \qquad r_{1_2} = V_e/I_2, \qquad r_{2_2} = V_e/I_{c_2}, \end{aligned} \tag{3B.1}$$

where V_e, the Early voltage, is a measure of transistor resistance, and $V_o = kT/q\kappa$. This small-signal circuit model is a linear system, which we can solve analytically using conventional techniques. The resulting solution, unfortunately, is a function of the unsolved large signal I_{c_1} and I_{c_2}. However, for the input conditions $I_2 = I_m$ and $I_1 = I_m + \delta_i$, we can reasonably make the approximations $I_{c_1} \approx I_c$ and $I_{c_2} \approx 0$ for relatively small δ_i, due to the exponential dependence of T_{21} and T_{22} on V_1 and V_2. Using these approximations, we can express the small-signal voltages v_1 and v_2 as linear functions of the small-signal input current i_1, as

$$\frac{v_1}{i_1} = (\frac{V_o}{I_1})\frac{((C_cV_o/I_c)s+1)}{(s/(a+b)+1)(s/(a-b)+1)} \tag{3B.2}$$

and

$$\frac{v_2}{i_1} = -(\frac{V_e}{I_1})\frac{1}{((CV_e/I_2)s+1)(s/(a+b)+1)(s/(a-b)+1)}, \tag{3B.3}$$

where

$$a = \frac{I_1}{2CV_e} + \frac{I_c}{2C_cV_o} \tag{3B.4}$$

and

$$b = \sqrt{(\frac{I_1}{2CV_e})^2 + (\frac{I_c}{2C_cV_o})^2 - (\frac{I_cI_1}{CC_cV_o^2})}. \tag{3B.5}$$

If b is an imaginary number, the circuit has complex poles, and exhibits undesirable ringing behavior. If $I_c > 4I_1(C_c/C)$, then b is real, and ringing does not occur. Figure 3.9 in the chapter compares experimental data with this inequality.

When b is real, the circuit exhibits first-order behavior. We can simplify Equations 3B.2 and 3B.3, and show that the first-order time constant for V_1 is CV_o/I, and the first-order time constant for V_2 is CV_e/I, where $I_1 \approx I_2 \equiv I$. Figure 3.10 in the chapter compares experimental data with these time constants.

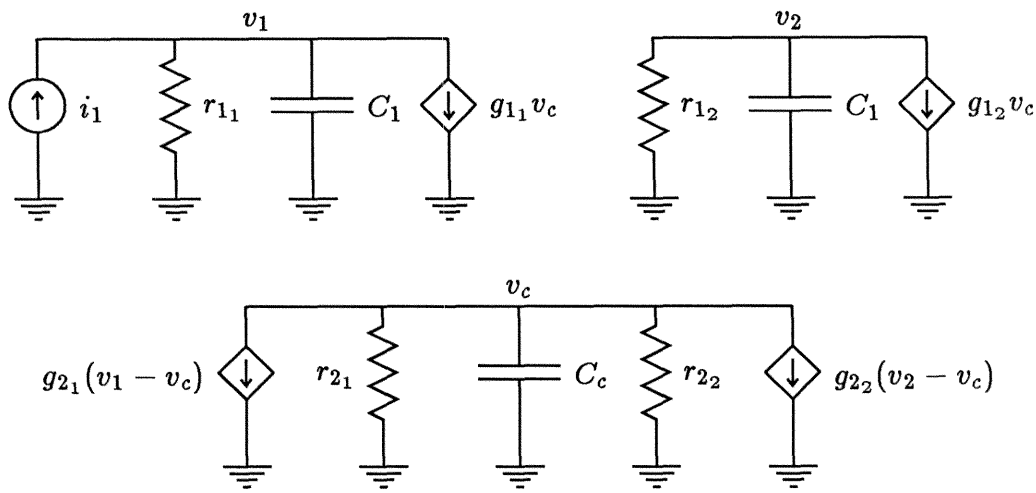


Figure 3B.1. Small-signal model of the two-neuron winner-take-all circuit, with capacitances added to model dynamic behavior.

Appendix 3C

Representation of Multiple Intensity Scales

This appendix explains a regime of operation of the winner-take-all circuit that represents multiple input intensity scales in the output, while still functioning as an inhibitory circuit.

Consider an n-neuron winner-take-all circuit, with input currents $I_1 \gg I_2 \gg \ldots \gg I_n$. As shown in Equation 3.1 in the chapter, the output voltage V_1 is

$$V_1 = V_o \ln(\frac{I_1}{I_o}) + V_o \ln(\frac{I_c}{I_o}), \tag{3C.1}$$

while $V_2 \ldots V_n$ are approximately zero. The output does not represent the input ordering $I_2 \gg I_3 \gg \ldots \gg I_n$; the largest input wins, and all other inputs lose. We can operate the circuit in another regime, however, which allows inputs $I_1 \ldots I_k$ to win, and inputs $I_{k+1} \ldots I_n$ to lose, where the magnitude of I_k is under external control. Voltage outputs $V_1 \ldots V_{k-1}$ are now binary representations, whereas V_k maintains a logarithmic encoding of the input current I_k.

In our previous analysis in the chapter, we used ideal current sources to represent $I_1 \ldots I_n$. In Figure 3C.1, we replace these ideal sources with transistor realizations. Transistors $T_{3_1} \ldots T_{3_n}$, when operating in the subthreshold region, realize ideal current sources if $V_{dd} - V_k > 2V_o$. Recall our input $I_1 \gg I_2 \gg \ldots \gg I_n$, and consider the effect of increasing the value of current source I_c. As shown in Equation 3C.1, the neuron output V_1 increases with I_c. For large I_c, transistor T_{2_1} is no longer operating in the subthreshold region. In this case, the equation $I_{c_1} = k'(W/L)(V_1 - V_c - V_T)^2$ describes T_{2_1}, where W and L are the width and

length of T_{2_1}, and k' and V_T (the threshold voltage) are fabrication constants. We can solve for V_1 for this situation, as

$$V_1 = V_o \ln(\frac{I_1}{I_o}) + \sqrt{\frac{I_c}{k'(W/L)}} + V_T. \tag{3C.2}$$

If we increase I_c further, V_1 continues to increase. For a sufficiently large I_c, V_1 can approach V_{dd}. In this situation, T_{3_1} begins to turn off, and no longer acts as an ideal current source supplying I_1. In this case, we can model T_{2_1} as an independent current source, supplying the current $I_s \equiv k'(W/L)(V_{dd} - V_c)^2$, as shown in Figure 3C.2. To a first approximation, Figure 3C.2 shows a winner-take-all circuit with $(n-1)$ neurons, with an effective control current of $I_c - I_s$.

We can apply this technique to represent multiple input intensity scales. Recall the input condition $I_1 \gg I_2 \gg \ldots \gg I_n$, and the desired behavior of outputs: $V_1 \ldots V_{k-1}$ to be approximately V_{dd}, V_k to maintain a logarithmic encoding of the input current I_k, and all other output voltages to be approximately 0. To produce this behavior, we simply increase I_s, until $V_1 \ldots V_{k-1}$ are approximately V_{dd}, but $V_k < V_{dd}$.

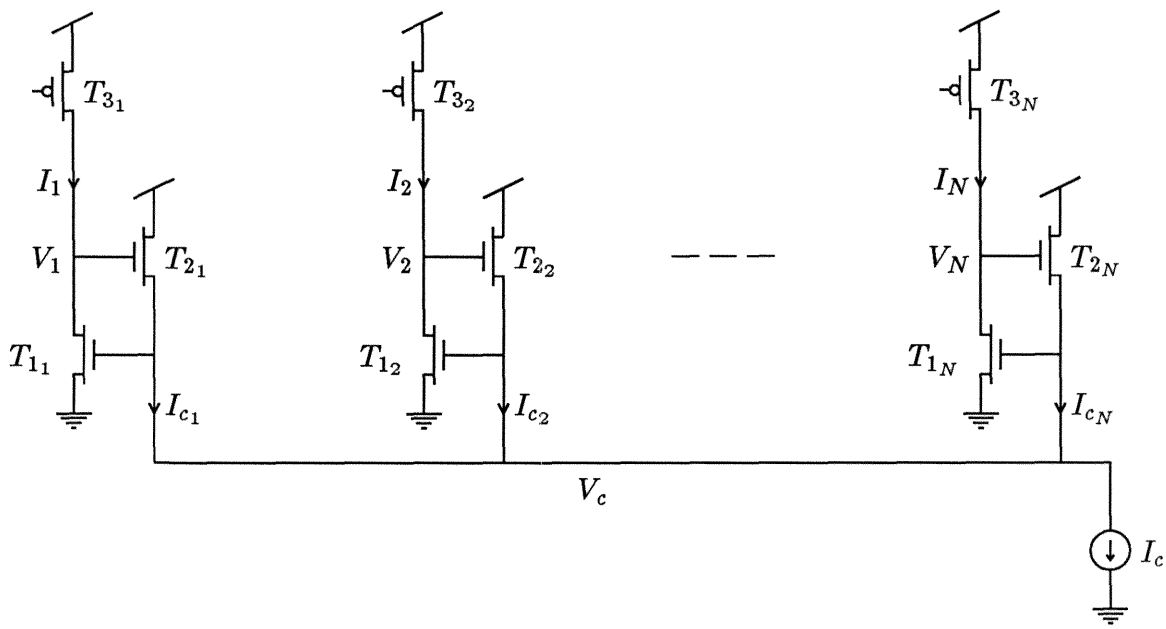


Figure 3C.1. Winner-take-all circuit, with transistor realizations replacing ideal input current sources.

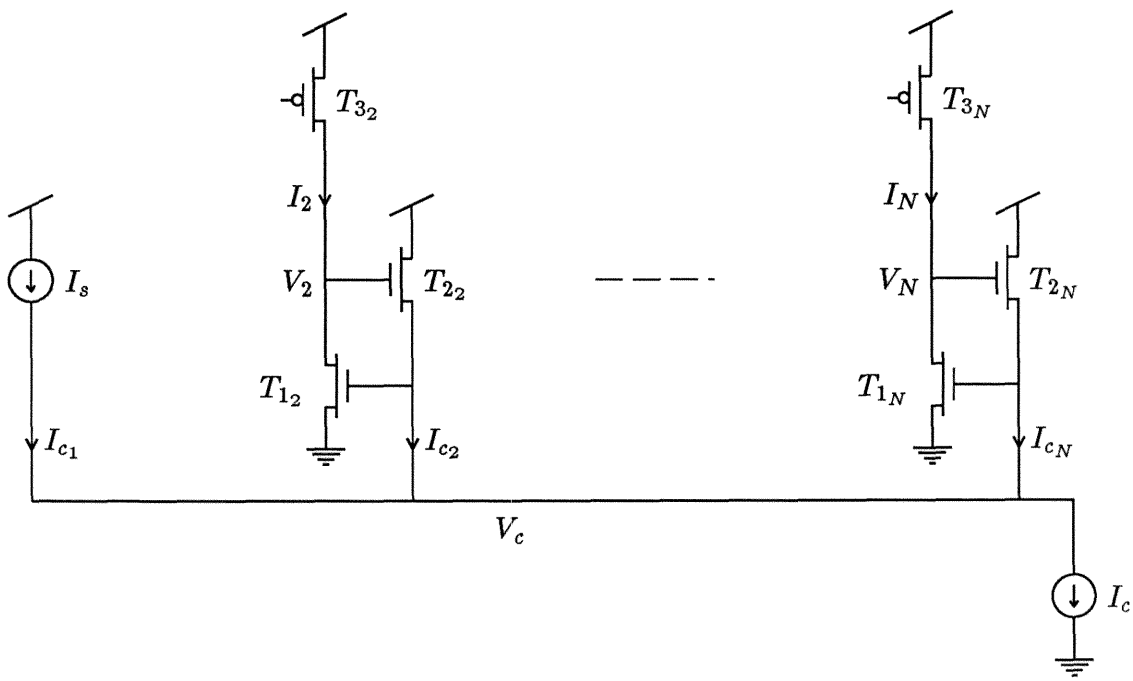


Figure 3C.2. Winner-take-all circuit, after modeling a saturated neuron with the independent current source I_s.

Chapter 4
A Silicon Model of Auditory Localization

The principles of organization of neural systems arose from the combination of the performance requirements for survival and the physics of neural elements. From this perspective, the extraction of time-domain information from auditory data is a challenging computation; the system must detect changes in the data that occur in tens of microseconds, using neurons that can fire only once per several milliseconds. Neural approaches to this problem succeed by closely coupling algorithms and implementation, unlike standard engineering practice, which aims to define algorithms that are easily abstracted from hardware implementation.

The echolocation system of the bat and the passive localization system of the barn owl are two neural systems that use extensive time-domain processing to create an accurate representation of auditory space. While the bat has developed specialized algorithms to implement an active sonar system, the barn owl and mammals use similar techniques to perform auditory localization. This chapter describes an integrated circuit that models, to a limited degree, the passive localization system of the barn owl.

The barn owl (*Tyto alba*) uses hearing to locate and catch small rodents in total darkness. The owl localizes the rustles of the prey to within 1 to 2 degrees in azimuth and elevation (Knudsen *et al.*, 1979). The owl cannot localize sounds when one ear is completely occluded; the auditory cues are differences in sound received by the two ears. The owl uses different binaural cues to determine azimuth and elevation. The elevational cue for the owl is binaural intensity difference. This cue is a result of a vertical asymmetry in the placement of the barn owl's ear openings, as well as a slight asymmetry in the left and right

halves of the barn owl's facial ruff, which acts as a sound-collecting device (Knudsen and Konishi, 1979). The azimuthal cue is binaural time difference. The binaural time differences are in the microsecond range, as expected from the short interaural distance of the owl, and vary as a function of azimuthal angle of the sound source (Moiseff and Konishi, 1981).

Unlike those in the visual system, the primary sensory neurons of the auditory system do not directly record the spatial location of a stimulus. Neuronal circuits must compute the position of sounds in space from acoustic cues in the stimulus. In the barn owl, the external nucleus of the inferior colliculus (ICx) contains the neural substrate of this computation, a spatial map of auditory space (Knudsen and Konishi, 1978). Neurons in the ICx respond maximally to stimuli located in a small area in space, corresponding to a specific combination of binaural intensity difference and binaural time difference (Figure 4.1). Recordings of the ICx neurons made while binaural intensity difference (Knudsen and Konishi, 1980) and binaural time difference (Moiseff and Konishi, 1981) are varied separately confirm the role of these acoustic cues.

There are several stages of neural processing between the primary sensory neurons at each cochlea and the computed map of space in the ICx. Each primary auditory fiber initially divides into two distinct pathways (Figure 4.2). One pathway processes intensity information, and thus encodes elevation cues, whereas the other pathway processes timing information, and thus encodes azimuthal cues. The two pathways recombine in the ICx to produce a complete map of space (Takahashi and Konishi, 1988b).

Recent anatomical and physiological studies provide a model of the structure and function of the time-coding pathway of the barn owl (Carr and Konishi, 1988; Takahashi and Konishi, 1988a; Fujita and Konishi, unpublished). This pathway has clear anatomical and physiological homologues in other avian

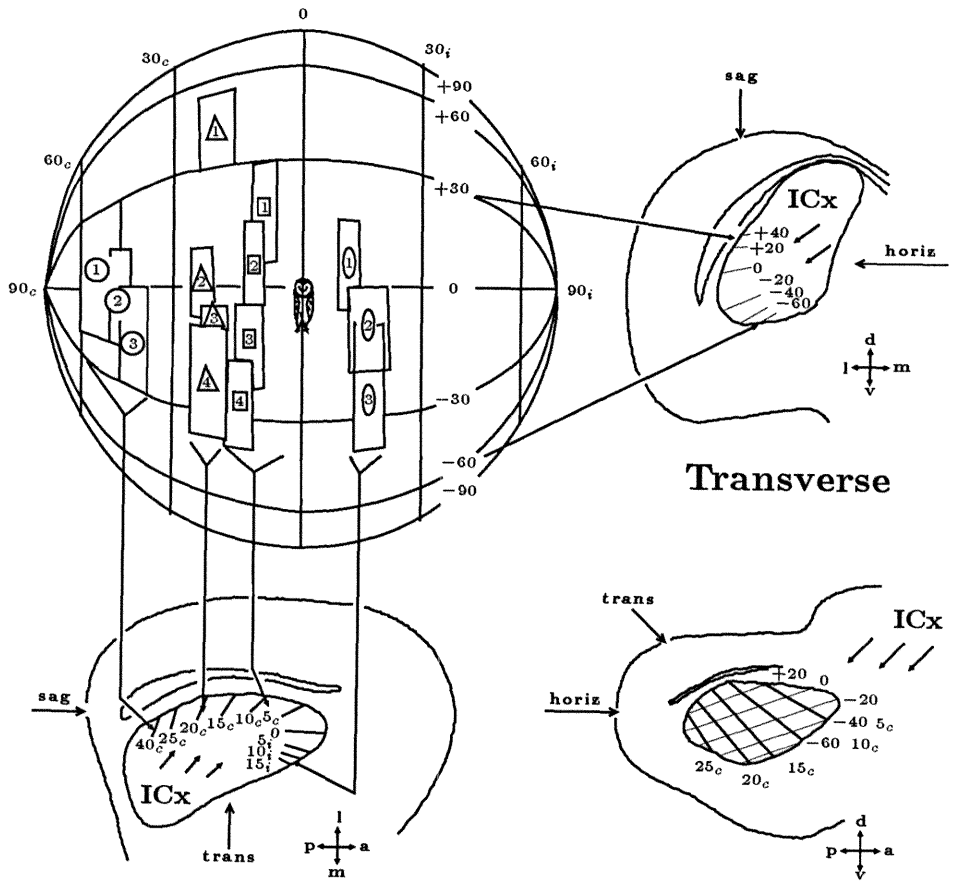

0
30c
30i
+90
+60
60c
60i
+30
90c
0
90i
−30
−60
−90
sag
ICx
+40
+20
0
−20
−40
−60
horiz
d
l
m
v
Transverse
trans
ICx
+20
0
−20
−40 5c
−60 10c
25c
20c
15c
horiz
d
p
a
v
sag
40c 25c 20c 15c 10c 5c
0
5i
10i
15i
ICx
trans
l
p
a
m
Horizontal
Sagittal

Figure 4.1. Neural map of auditory space in the barn owl. Auditory space is depicted as an imaginary globe surrounding the owl. Projected onto the globe are the receptive field best areas of 14 neurons. The best area, a zone of maximal response within a receptive field, remains unaffected by variations in sound intensity and quality. The numbers surrounding the same symbols (circles, rectangles, triangles, ellipses) represent neurons from the same electrode penetration; the numbers themselves denote the order in which the neurons were encountered. Below and to the right of the globe are illustrations of three histological sections through the inferior colliculus; arrows point to the external nucleus, the ICx. Iso-azimuth contours are shown as thick lines in the horizontal and sagittal sections; iso-elevation contours are represented by thin lines in the transverse and sagittal sections. Abbreviations: a, anterior; c, contralateral; d, dorsal; i, ipsilateral; l, lateral; m, medial; p, posterior; v, ventral. From (Knudsen and Konishi, 1978) and (Konishi, 1986);

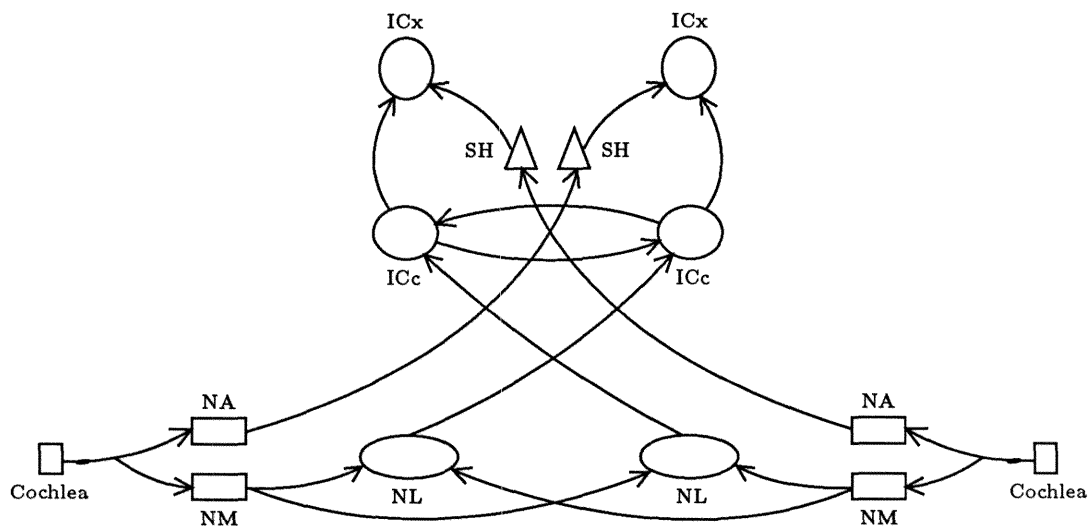


Figure 4.2. Schematic drawing of the auditory pathway of the barn owl. The pathway begins at the cochlea, the sound transducer. The cochlea projects to the nucleus angularis (NA), to start the amplitude-coding pathway, and to the nucleus magnocellularis (NM), to start the time-coding pathway. The NA projects contralaterally to the shell of central nucleus of the inferior colliculus (SH), which projects to the external nucleus of the inferior colliculus (ICx), the location of the complete map of auditory space. The NM projects bilaterally to nucleus laminaris (NL); the NL projects contralaterally to the central nucleus of the inferior colliculus (ICc). Fibers connect the ipsilateral and contralateral sides of the ICc, returning information to the original side. The ICc then projects to the ICx, the location of the complete map of auditory space. Adapted from (Fujita and Konishi, unpublished) and (Takahashi and Konishi, 1988b).

species and in mammals, although the most completely studied example of time-coding processing is that done by the barn owl.

The chapter describes a 220,000-transistor integrated circuit that models the time-coding pathway of the barn owl, using analog, continuous-time processing. The chip receives two inputs, corresponding to the sound pressure at each ear of the owl. The chip computes as its output a map of the interaural time difference between the input signals, encoding the azimuthal localization cue. The architecture of the chip directly reflects the anatomy and physiology of the neural pathway; intermediate outputs of the chip correspond to different neurons in this pathway. The original research in this chapter was done in collaboration with Carver Mead; parts of this chapter were originally published in (Lazzaro and Mead, 1989a).

4.1 The Time-Coding Pathway of the Owl

By installing small microphones near the eardrums of the owl, Moiseff and Konishi have demonstrated the dependence of binaural time difference on the azimuthal angle of a sound source (Figure 4.3). The time-coding pathway of the owl computes azimuth, using this binaural time difference as a cue.

The time-coding pathway receives as input a collateral of the primary auditory fibers. The primary auditory fibers originate at the cochlea; Chapter 2 described the structure and function of the cochlea. The mean firing rate of a primary auditory fiber encodes sound intensity. The temporal pattern of auditory-nerve firings reflects the shape of the filtered and rectified sound waveform; this phase locking does not diminish at high intensity levels (Evans, 1982). This temporal patterning preserves the timing information needed to encode interaural time differences.

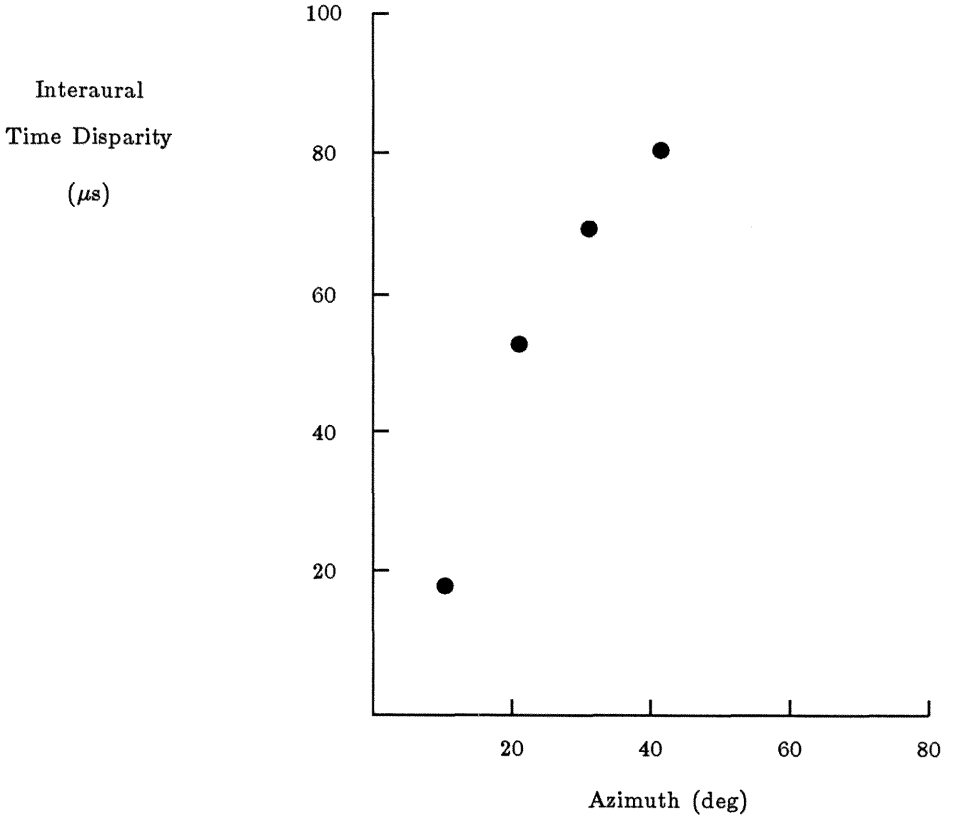


Figure 4.3. Binaural time difference in the owl, as a function of azimuthal angle. Small microphones installed near the eardrums were used to measure the interaural time differences in the microsecond range. From (Moiseff and Konishi, 1981).

In the owl, the primary auditory fibers project to the nucleus magnocellularis (NM), the first nucleus of the time-coding pathway (Figure 4.2). Neurons in the NM preserve the temporal patterning of nerve firings conveyed by the primary auditory fibers (Sullivan and Konishi, 1984). The NM projects bilaterally to the nucleus laminaris (NL), the first nucleus in the time-coding pathway that receives input from both ears. The mammalian homologue of the NL is the medial superior olive (MSO), which also receives input from both ears.

Neurons in both the NL (Moiseff and Konishi, 1983) and the MSO (Goldberg and Brown, 1969) are most sensitive to binaural sounds that have a specific interaural time delay (Figure 4.4). In 1948, Jeffress proposed a model to explain the encoding of interaural time differences in neural circuits (Jeffress, 1948). In the model, as applied to the anatomy of the owl, axons from the ipsilateral (same side) NM and the contralateral (opposite side) NM, with similar characteristic frequencies, enter the NL from opposite surfaces (Figure 4.5). The axons travel antiparallel, and action potentials counterpropagate across the NL; the axons act as neural delay lines. NL neurons are adjacent to both axons. Each NL neuron receives synaptic connections from both axons, and fires maximally when action potentials present in both axons reach that particular neuron at the same time. In this way, interaural time differences map into a neural place coding; the interaural time difference that maximally excites an NL neuron depends on the position of the neuron in the NL (Konishi, 1986). In engineering terms, each NL neuron computes the cross-correlation function of a filtered version of the sound inputs; the value of this function is the relative probability of the sounds being offset by a particular interaural time difference (Yin and Kuwada, 1984).

There is considerable evidence to support the application of this theory to the barn owl. Anatomical surveys of the NL of the owl show antiparallel axons, which enter from opposite sides, and which originate from the ipsilateral

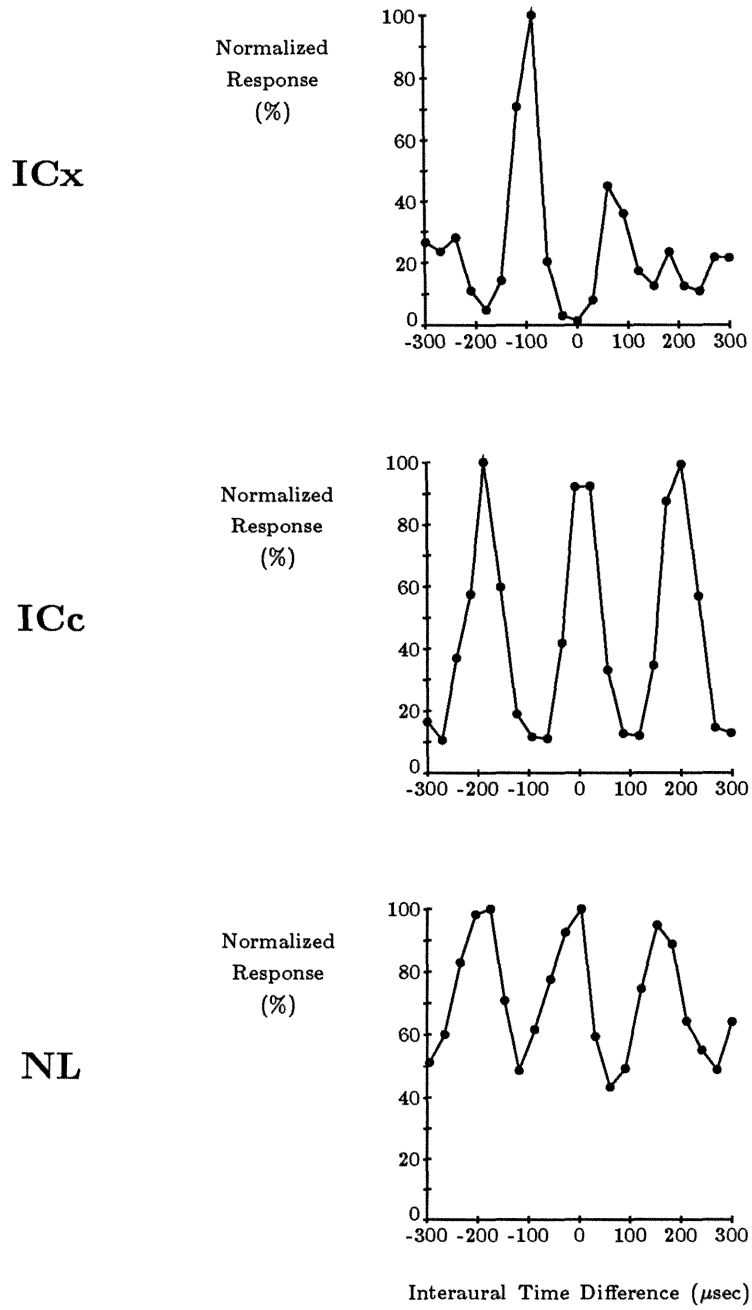


Figure 4.4. The delay curves obtained under physiological conditions from single units in the NL, the ICc, and the ICx. Broadband noise was used as a stimulus. Neurons with similar best frequencies (5.4 to 5.7 kHz) are shown. Neuronal selectivity improves markedly between the NL and the ICc. Responses are phase ambiguous in the NL and the ICc, but not in the ICx, when stimulated with noise. From (Fujita and Konishi, unpublished).

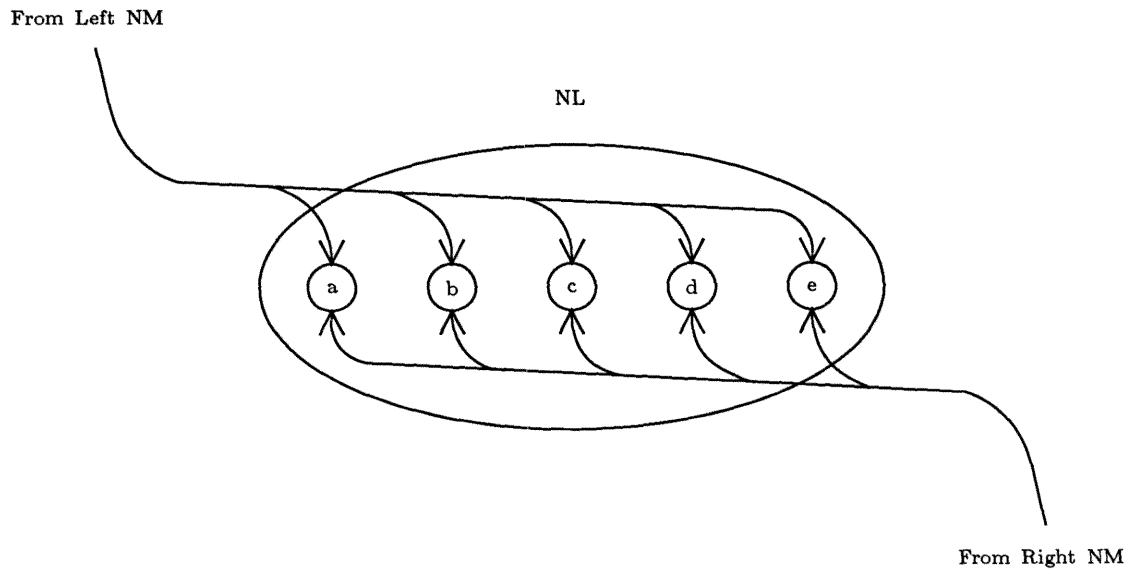


Figure 4.5. The Jeffress model for measuring and encoding interaural time differences, as applied to the barn owl. Fibers from the left and right NM converge on the NL; the NM fibers encode timing information in the temporal patterning of their action potentials. Each fiber has a uniform time delay. The NL neurons (circles labeled a through e) act as coincidence detectors, firing maximally when signals from two sources arrive simultaneously. The pattern of innervation of the NM fibers in the NL create left–right asymmetries in transmission delays; when the binaural disparities in acoustic signals exactly compensate for the asymmetry of a particular neuron, this neuron fires maximally. The position in the array thus encodes the interaural time difference. Adapted from (Konishi, 1986).

and contralateral NMs (Figure 4.6). Physiological data from NL neurons show a linear relationship between axonal time delay and axonal position in the NL (Carr and Konishi, 1988). Note that in Figure 4.4, NL neurons respond maximally to several different interaural time delays. As predicted by the Jeffress model, NL neurons respond to all interaural time differences that result in the same interaural phase difference of the characteristic frequency of the input fibers. In mammals, neuronal recordings in the MSO are also consistent with the Jeffress model (Goldberg and Brown, 1969).

NL outputs project contralaterally to a subdivision of the central nucleus of the inferior colliculus (ICc). ICc neurons are more selective for interaural time differences than are NL neurons (Figure 4.4); application of bicuculline methiodide (BMI, a selective gamma aminobtyric acid (GABA) antagonist) to the ICc reduces this improved selectivity. BMI acts as a blocker of inhibition; the action of BMI suggests that the increased selectivity of ICc neurons is a result of inhibitory circuits between neurons that are tuned to different interaural time differences (Fujita and Konishi, unpublished). Inhibitory circuits between neurons tuned to the same interaural time differences might also be present. In addition, evidence suggests that ICc neurons temporally integrate information to improve selectivity for interaural time differences. The sensitivity of ICc neurons to interaural time differences increases as a function of the duration of the sound stimulus, for sounds lasting less than 5 ms (Wagner and Konishi, unpublished).

The ICc projects to the external nucleus of the inferior colliculus (ICx), after crossing sides in the ICc. The ICx integrates information from the time-coding pathway and from the amplitude-coding pathway to produce a complete map of auditory space. Neurons in the ICx, unlike those in earlier stations of the time-coding pathway, respond to input sounds over a wide range of frequencies. ICx neurons are more selective for interaural time differences than are ICc neurons,

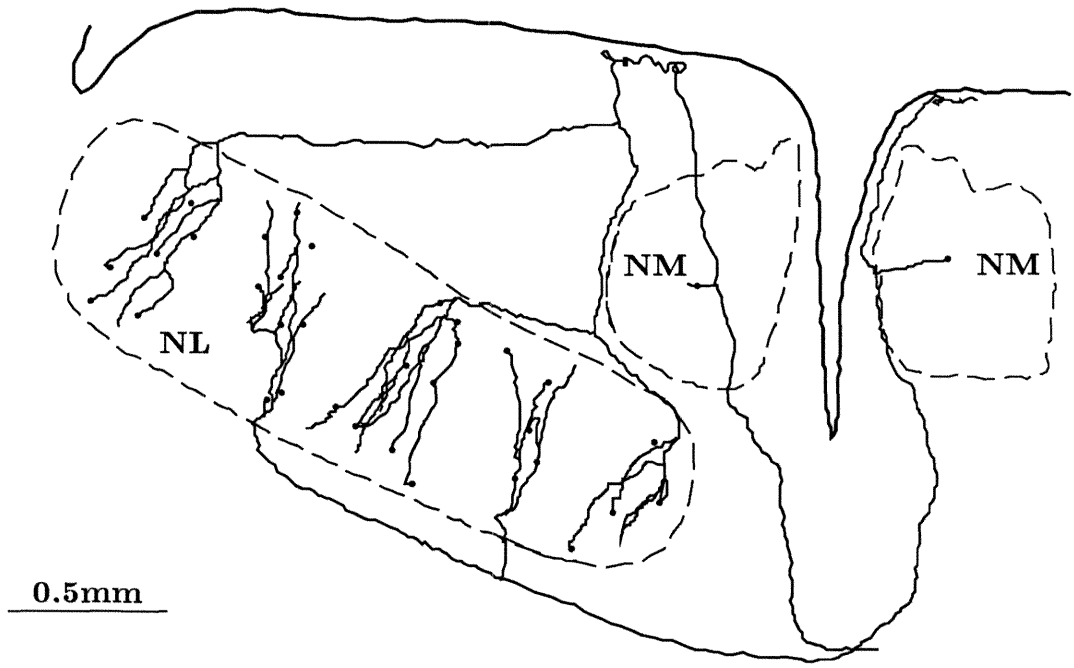


Figure 4.6. Innervation of nucleus laminaris. Axons from the ipsilateral and contralateral NM enter, respectively, the dorsal and ventral surfaces of the NL. Each axon runs along the respective surface until it comes to the iso-frequency band that it is destined to innervate. Within this band, the axon sends two or three collaterals into the NL, each of which divides into several branches. These fibers, which might be as long as 1 mm, run along a relatively straight course toward the opposite surface. The fibers appear to establish contacts with the NL cell bodies either by *en passant* endings or by small branchlets. From (Carr and Konishi, 1988) and (Konishi, 1986).

for low frequency sounds. In addition, neurons in earlier stations of the time-coding pathway respond to all interaural time differences that result in the same interaural phase difference of the neuron's characteristic frequency; neurons in the ICx respond maximally to one interaural time difference (Figure 4.4). This behavior suggests that ICx neurons combine information over many frequency channels in the ICc, to disambiguate interaural time differences from interaural phase differences. Application of BMI to the ICx results in phase ambiguity and decreased selectivity, suggesting an inhibitory mechanism for this computation (Fujita and Konishi, unpublished).

4.2 A Silicon Model of the Time-Coding Pathway

Figure 4.7 shows the floorplan of the silicon model of the time-coding pathway. The chip receives two inputs, corresponding to the sound pressure at each ear of the owl. Each input connects to a silicon model of the mechanical processing of the cochlea (Lyon and Mead, 1988a); Chapter 2 describes the cochlea model. Silicon models of inner hair cells connect to the cochlea circuit at constant intervals; 62 inner-hair-cell circuits connect to each silicon cochlea. The output of each inner-hair-cell model connects directly to a spiral-ganglion-neuron model, to complete a silicon model of auditory-nerve response. Chapter 2 described the auditory-nerve model. In the owl, the spiral-ganglion neurons project to the NM; for simplicity, our integrated circuit does not model the NM; each spiral-ganglion-neuron circuit directly connects to the NL.

In both owls and mammals, two instances of each nucleus in the auditory pathway appear in the brainstem, symmetric about the midline. Obviously, for binaural processing, two cochleas and two NMs are required. The NL, however, receives bilateral projection from the NM; each NL primarily encodes one-half of the azimuthal plane. This partitioning continues at higher stations;

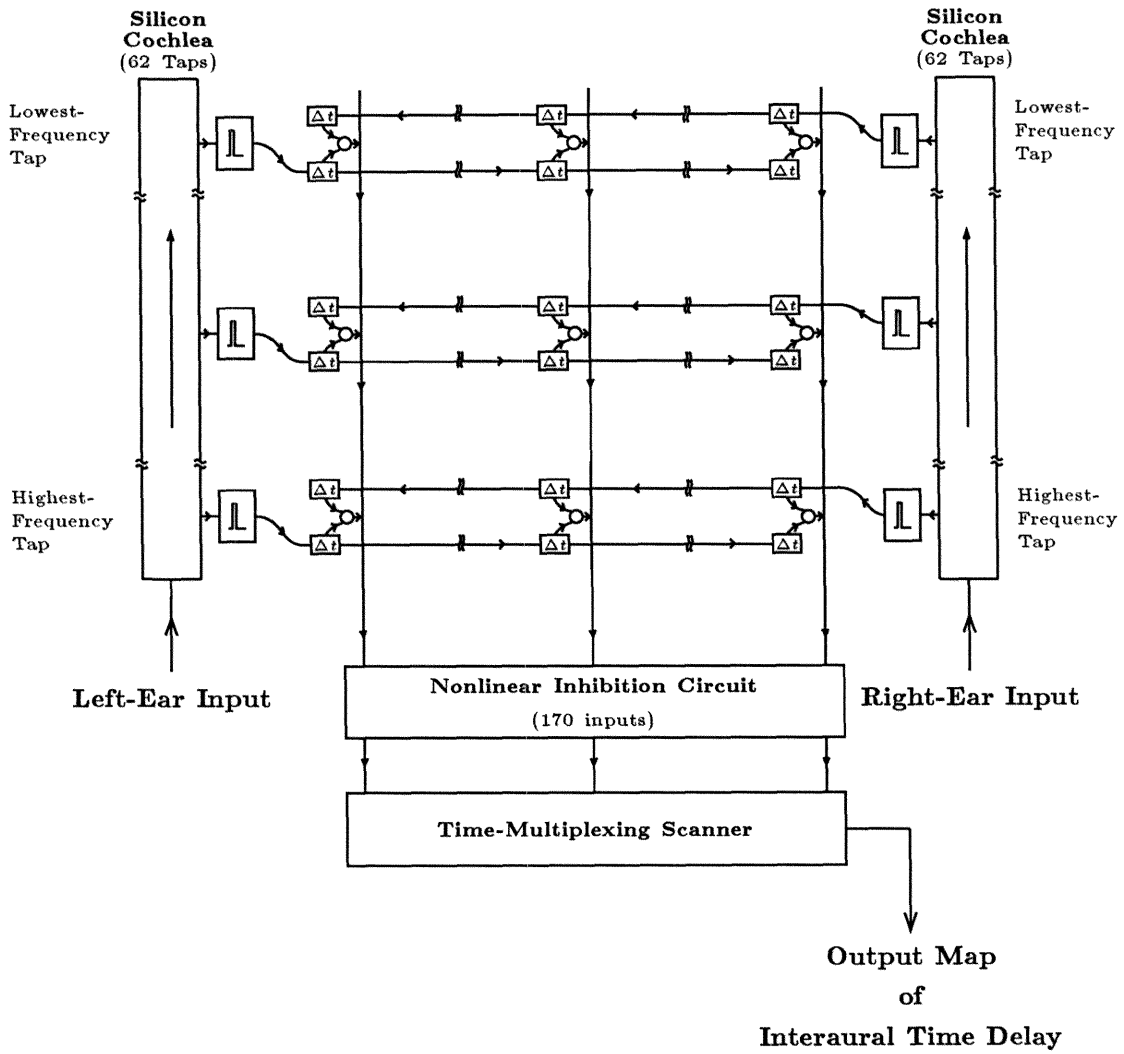

Silicon Cochlea (62 Taps)
Silicon Cochlea (62 Taps)
Lowest-Frequency Tap
Lowest-Frequency Tap
Highest-Frequency Tap
Highest-Frequency Tap
Left-Ear Input
Right-Ear Input
Nonlinear Inhibition Circuit (170 inputs)
Time-Multiplexing Scanner
Output Map of Interaural Time Delay

Figure 4.7. Floorplan of the silicon model of the time-coding pathway of the owl. Sounds for the left ear and right ear enter the respective silicon cochleas at the lower left and lower right of the figure. Silicon inner hair cells tap each silicon cochlea at 62 equally spaced locations; each silicon inner hair cell connects directly to a silicon spiral-ganglion cell. The square box marked with a pulse represents both the silicon inner hair cell and the silicon spiral-ganglion cell. Each silicon spiral-ganglion cell generates action potentials; these signals travel down silicon axons, which propagate from left to right for silicon spiral-ganglion cells from the left cochlea, and from right to left for silicon spiral-ganglion cells from the right cochlea. The rows of small rectangular boxes, marked with the symbol "Δt," represent the silicon axons. One hundred seventy silicon NL cells, represented by small circles, lie between each pair of antiparallel silicon axons. Each silicon NL cell connects directly to both axons, and responds maximally when action potentials present in both axons reach that particular neuron at the same time. In this way, interaural time differences map into a neural place code. Each vertical wire that spans the array combines the outputs of all silicon NL neurons with a specific interaural time difference. These 170 vertical wires form a temporally smoothed map of interaural time difference, which responds to a wide range of input sound frequencies. The nonlinear inhibition circuit near the bottom of the figure increases the selectivity of this map. The time-multiplexing scanner transforms this map into a signal suitable for display on an oscilloscope.

for simplicity, our integrated circuit has only one copy of the NL and of later stations, which encodes all azimuthal angles.

The chip models the anatomy of the NL directly. The two silicon cochleas lie at opposite ends of the chip; the sounds in each silicon cochlea travel in parallel up the chip. Spiral-ganglion-neuron circuits from each ear, which receive input from inner-hair-cell circuits at the same cochlear position, project to separate axon circuits, which travel antiparallel across the chip. When a sound is presented to both chip inputs, action potentials counterpropagate across the chip.

The axon circuit is a discrete neural delay line; for each action potential at the axon's input, a fixed-width, fixed-height pulse travels through the axon, section by section, at a controllable velocity (Mead, 1989). After excitation of the circuit with a single action potential, only one section of the axon is firing at any point in time. NL neuron circuits lie between each pair of antiparallel axons at every discrete section, and connect directly to both axons. Simultaneous action potentials at both inputs excite the NL neuron circuit; if only one input is active, the neuron generates no output. This simplified model differs from the owl in several ways. The silicon NL neurons are perfect coincidence detectors; in the owl, NL neurons also respond, with reduced intensity, to monaural input. In addition, only two silicon axons converge on each silicon NL neuron; in the owl, many axons from each side converge on an NL neuron.

In the owl, the NL projects to the ICc, which in turn projects to the ICx. The final output of our integrated circuit models the responses of ICx neurons to interaural time differences. In response to interaural time differences, ICx neurons act different than NL neurons. Experiments suggest three mechanisms for these differences: nonlinear inhibition, temporal integration, and combining

of information over frequency channels. Our integrated circuit implements these mechanisms to produce a neural map of interaural time difference.

In our chip, all NL neuron outputs corresponding to a particular interaural time delay are summed to produce a single output value. NL neuron outputs are current pulses; a single wire acts as a dendritic tree to perform the summation. In this way, a two-dimensional matrix of NL neurons reduces to a single vector; this vector is a map of interaural time difference, for all frequencies. In the owl, inhibitory circuits between neurons tuned to the same interaural time differences might be present in the ICc, before combination across frequency channels. Our model does not include these circuits.

The ICc and perhaps the ICx might use temporal integration to increase selectivity. Our chip temporally integrates the vector that represents interaural time differences; the time constant of integration is adjustable. Nonlinear inhibitory connections between neurons tuned to different interaural time differences in the ICx and the ICc also increase sensitivity to interaural time differences. In our chip, a nonlinear inhibition circuit, described in Chapter 3, processes the temporally integrated vector that represents interaural time differences. The circuit performs a winner-take-all function, producing a new map of interaural time differences in which only one neuron has significant energy. The chip time multiplexes this final map of interaural time differences on a single wire for display on an oscilloscope.

4.3 Comparison of Responses

We presented periodic click stimuli to the chip; one input received the sound directly, while the other input received a time-delayed replica of the sound (Figure 4.8). We recorded chip response at several intermediate outputs and at the final output map of interaural time differences, and compared chip response

to similar responses recorded from the time-coding pathway of the barn owl. We also separately fabricated and tested several parts of the complete model.

Figure 4.9 shows the basilar-membrane response of the chip to click stimuli, with no interaural time difference. The figure shows output response at several equally spaced basilar-membrane locations, for both left and right cochleas. The frequency content of the response progressively decreases at locations farther from the beginning of the cochlea; in addition, the response at each location has a slight resonant overshoot. The response also shows a time delay between each pair of adjacent locations; for more distant locations, this time delay exceeds the period of the input waveform. These observations match the behavior of the traveling wave structure of physiological cochleas; a comparison of the responses of the physiological and silicon cochleas was given in Chapter 2.

The left and right cochlea responses are nearly identical; there are small differences in the velocity of propagation, due to imperfections in the circuit elements. Because of these differences, the time delays in the left and right cochleas at a particular location are not equal; these differences, shown in Figure 4.9, are a potential source of localization error.

Chapter 2 compared the responses of the silicon auditory-nerve model with physiological data. As shown in Figures 2.3, 2.8, and 2.10, a silicon auditory-nerve fiber encodes a filtered, half-wave-rectified version of the cochlea input, over a wide dynamic range, using the temporal patterning of fixed-width, fixed-height pulses. This property is vital for the correct function of the silicon model of the time-coding pathway. Figure 2.8(b) shows that the absolute phase of the temporal firing patterns of a silicon auditory nerve fiber shifts with input intensity, for high signal levels; this behavior is unlike a physiological fiber. Because of this shortcoming, we did not apply binaural intensity differences to the chip.

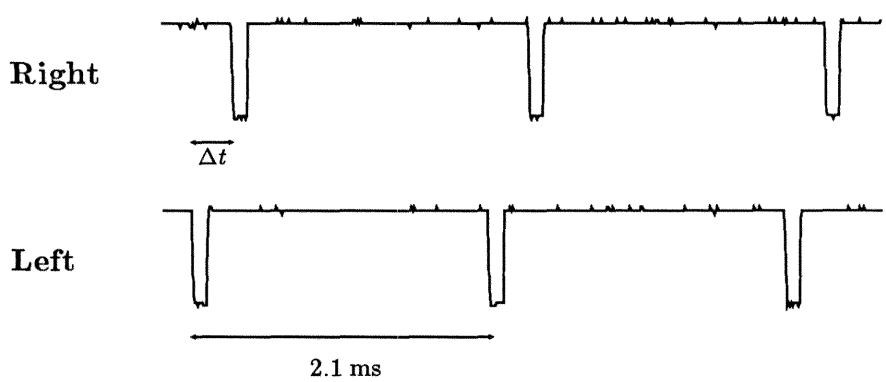


Figure 4.8. Input stimulus for the chip. Both left and right ears receive a periodic click waveform, at a frequency of 475 Hz. The time delay between the two signals, denoted by "Δt," is variable.

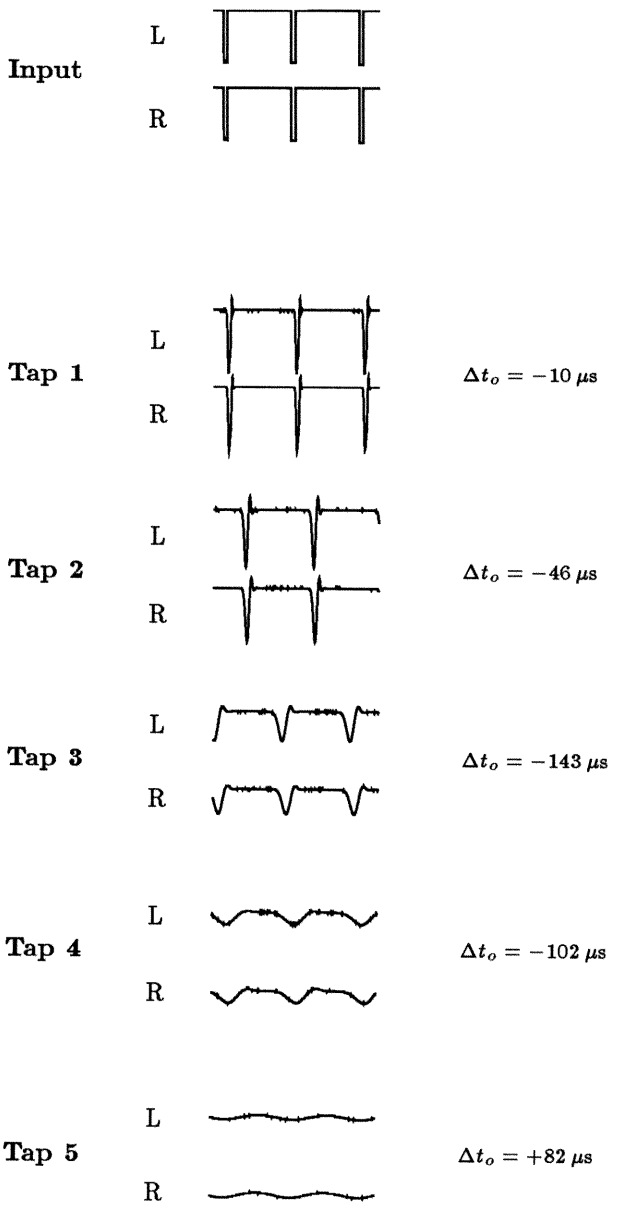

Input
L
R
Tap 1
L
R
$\Delta t_o = -10\,\mu s$
Tap 2
L
R
$\Delta t_o = -46\,\mu s$
Tap 3
L
R
$\Delta t_o = -143\,\mu s$
Tap 4
L
R
$\Delta t_o = -102\,\mu s$
Tap 5
L
R
$\Delta t_o = +82\,\mu s$

Figure 4.9. Response of both cochleas to the input stimulus of Figure 3.8; this stimulus is reproduced at the top of this figure. The horizontal axis is time; each vertical axis is cochlear response. The figure shows the output response at five cochlear positions, for both left and right cochleas. Tap 1 is closest to the beginning of the cochlea, and retains most of the high-frequency content of the signal. Later taps show less and less high-frequency content; tap 5 contains only the fundamental frequency of 475 Hz. Each tap accentuates a particular frequency, the characteristic frequency of the cochlear position; this frequency decreases with increasing tap numbers. The cochleas act also as exponentially tapered delay lines; notice the response of each tap occurs progressively further from the input waveform. For taps 4 and 5, the delay of the cochlea exceeds a single period of the input waveform. The cochleas are not exactly matched; a difference in time delay between the two cochleas is denoted by "Δt_o," on the right of each pair of outputs. For lower-numbered taps, the delay is small compared to the time delay of the silicon axons; for higher-numbered taps, it is significant.

Figure 4.10 shows the output of several sections of the axon circuit, in response to a single input pulse. Only one section is active at any point in time; the pulse width does not grow or shrink systematically, but rather stays roughly constant. Figure 4.11 shows the variation in axon pulse width over about 100 sections of axon, due to circuit element imperfections. In this circuit, a variation in axon pulse width indicates a variation in the velocity of axonal propagation; this variation is a potential source of localization error.

The final output of the chip is a map of interaural time difference. Three signal-processing operations, computed in the ICx and ICc, improve the original encoding of interaural time differences in the NL: temporal integration, integration of information over many frequency channels, and inhibition among neurons tuned to different interaural time delays. In our chip, we can disable the inhibition and temporal-integration operations, and observe the unprocessed map of interaural time differences.

Figure 4.12(a) shows these unprocessed maps of interaural time differences, created in response to click stimuli. Each map corresponds to an interaural time difference that is 100 μs greater than that for the previous map; the first map corresponds to simultaneous excitation of both ears. The vertical axis of each map corresponds to neural activity level, whereas the horizontal axis of each map corresponds to linear position within the map. Each map is an average of several maps recorded at 100-ms intervals; averaging is necessary to capture a representation of the quickly changing, temporally unsmoothed response. The encoding of interaural time differences is present in the maps, but false correlations add unwanted noise to the desired signal. By combining the outputs of 62 rows of NL neurons, each tuned to a separate frequency region, the maps in Figure 4.12(a) correctly encode interaural time differences, despite variation in axonal velocity, mismatches in cochlear delay, and noise.

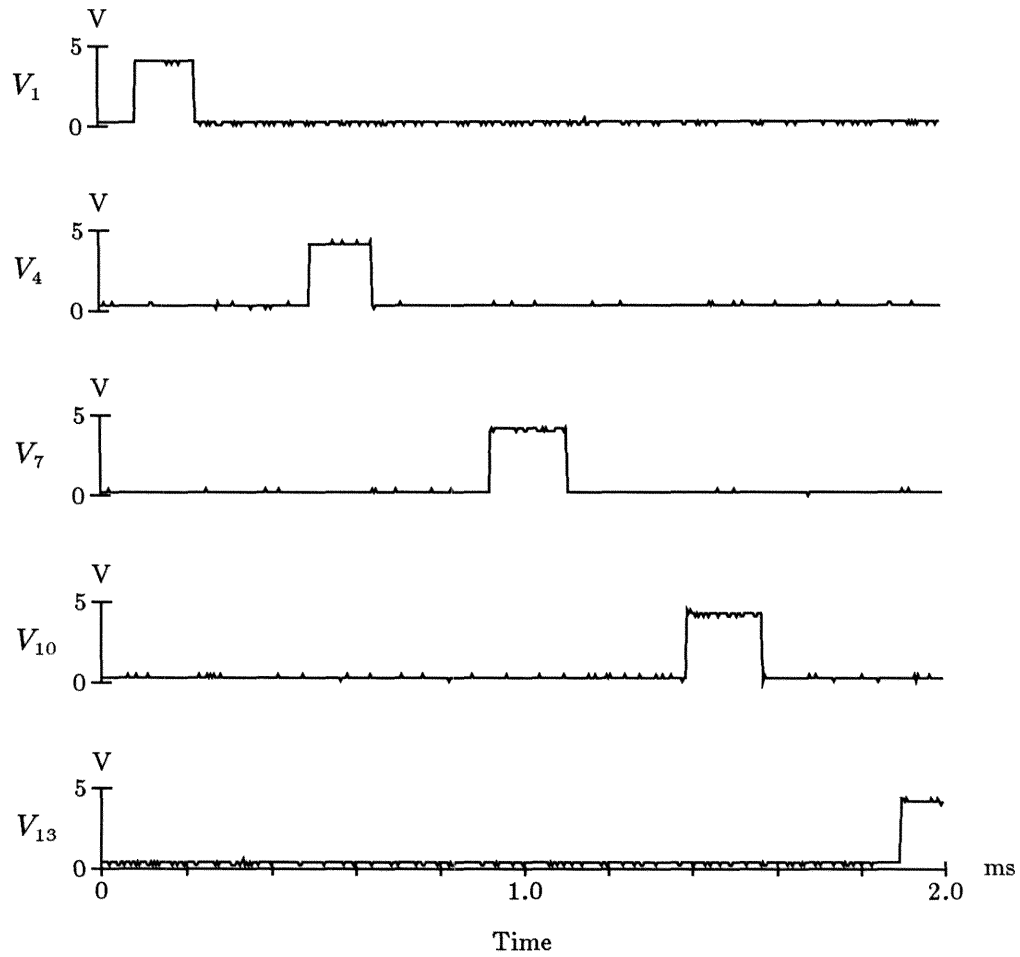


Figure 4.10. Waveform measured at several sections of a silicon axon. Outputs were located at every third section; thus, two outputs are not shown for every one that is. Note that pulse height and pulse width stay fairly constant, and that only one axon section is active at any point in time. From (Mead, 1989).

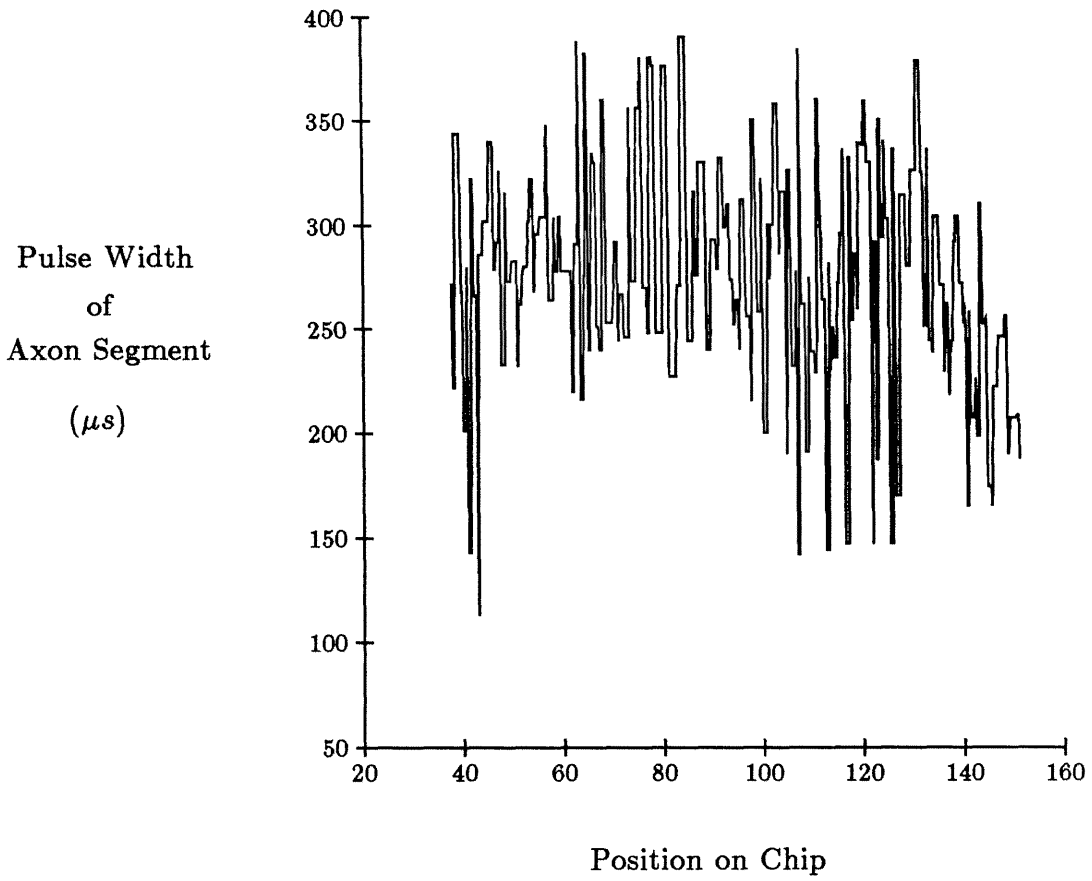


Figure 4.11. Variation in the pulse width of a silicon axon, over about 100 axonal sections. Axons were set to fire at a slower velocity than in the owl model, for more accurate measurement. The line represents the section-by-section pulse width.

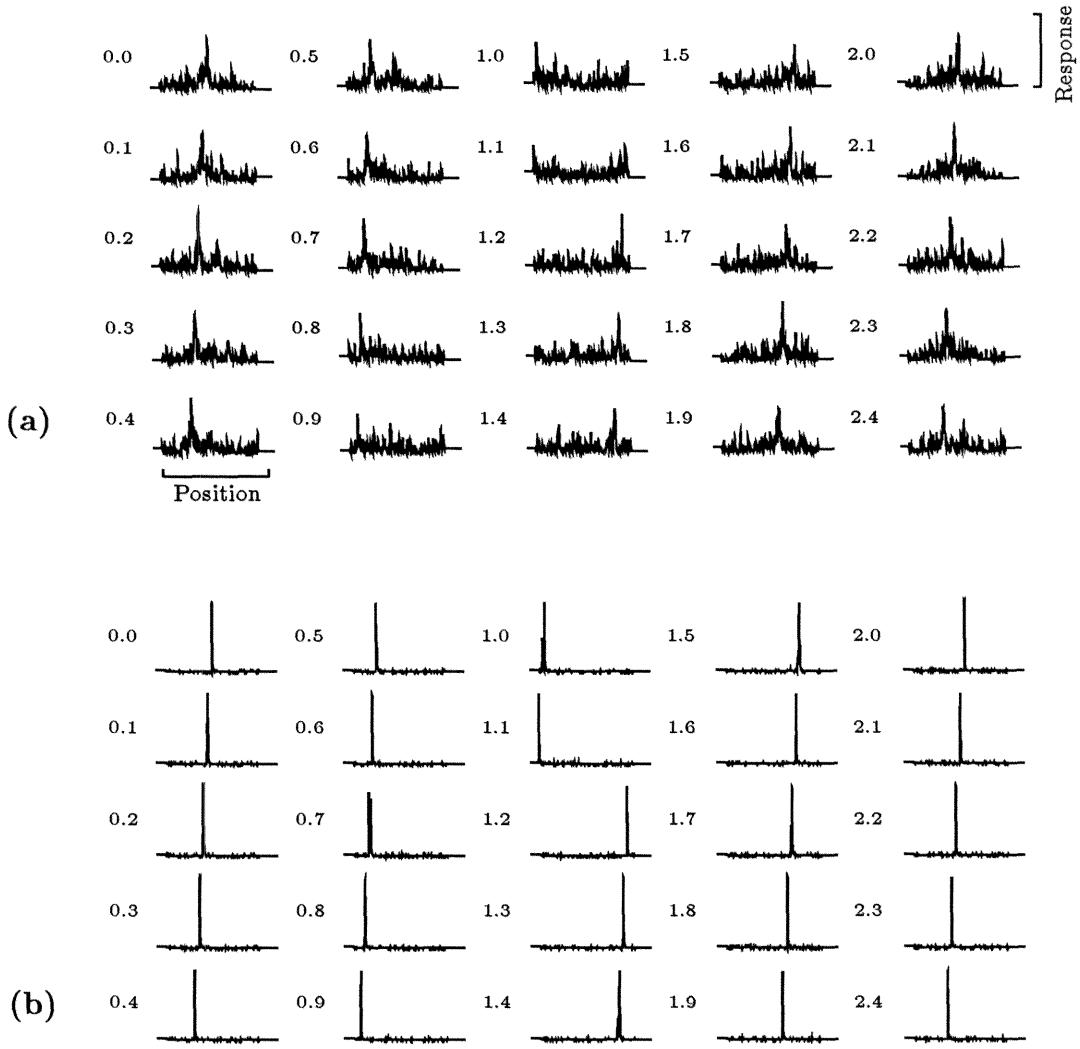

0.0
0.5
1.0
1.5
2.0
0.1
0.6
1.1
1.6
2.1
0.2
0.7
1.2
1.7
2.2
0.3
0.8
1.3
1.8
2.3
0.4
0.9
1.4
1.9
2.4
Response
Position
(a)
(b)

Figure 4.12. Maps of interaural time differences, taken from the chip. **a.** The nonlinear inhibition and temporal smoothing operations were turned off, showing the unprocessed map of interaural time delays. Each individual plot shows the neural response at every position across the map. The stimuli for each plot are the periodic click waveforms, offset by an interaural time difference, shown in Figure 4.8; the interaural time difference is shown in the upper-left corner of each plot, measured in milliseconds. These responses are the averaged result of four maps, at 100-ms intervals, using a digital oscilloscope, because the on-chip temporal smoothing is disabled. The encoding of interaural time differences is present in the maps, but false correlations add unwanted noise to the desired signal. Because we are using a periodic stimulus, large time delays are interpreted as negative delays, and the map response wraps from one side to the other at an interaural time difference of 1.2 ms. **b.** The nonlinear inhibition and temporal-smoothing operations were turned on, showing the final output map of interaural time delays. Format is identical to that in part a. Off-chip averaging was not used, because the chip temporally smooths the data. Each map shows a single peak, with little activity at other positions, due to nonlinear inhibition.

Figure 4.12(b) shows maps of interaural time differences taken with inhibition and temporal-integration operations enabled, using stimuli identical to those for Figure 4.12(a). For most interaural time differences, only one peak exists. Off-chip averaging of maps over time is not necessary, as the temporal-integration operation on the chip smooths the time response of the chip. The maps do not reflect the periodicity of the individual frequency components of the sound stimulus; experiments with a noise stimulus confirm the phase-disambiguation property of the chip.

Figure 4.13(a) is an alternative representation of the map of interaural time differences computed by the chip. We recorded the map position of the neuron with maximum signal energy, for different interaural time differences, using periodic click stimuli. Carr and Konishi (Carr and Konishi, 1988) performed a similar experiment in the NL of the barn owl (Figure 4.13b), mapping the time delay of an axon innervating the NL, as a function of position in the NL. The linear properties of our chip map are the same as those of the owl map.

4.4 Discussion

The experiments on the chip raise several interesting issues about the algorithms implemented by the auditory-localization system of the barn owl. These issues concern the computational errors that occur due to imperfect components, and how nature might attempt to correct for these errors.

Figure 4.9 shows the responses of two silicon cochleas to a click stimulus. Because of component imperfections, the delays at the same position on the two silicon cochleas are not identical. As a result, the row of NL neuron circuits associated with this position has a spatial offset; in response to an interaural time difference of zero, the NL neuron circuit with maximum energy is not located exactly in the center of the row. If each cochlear position has a different

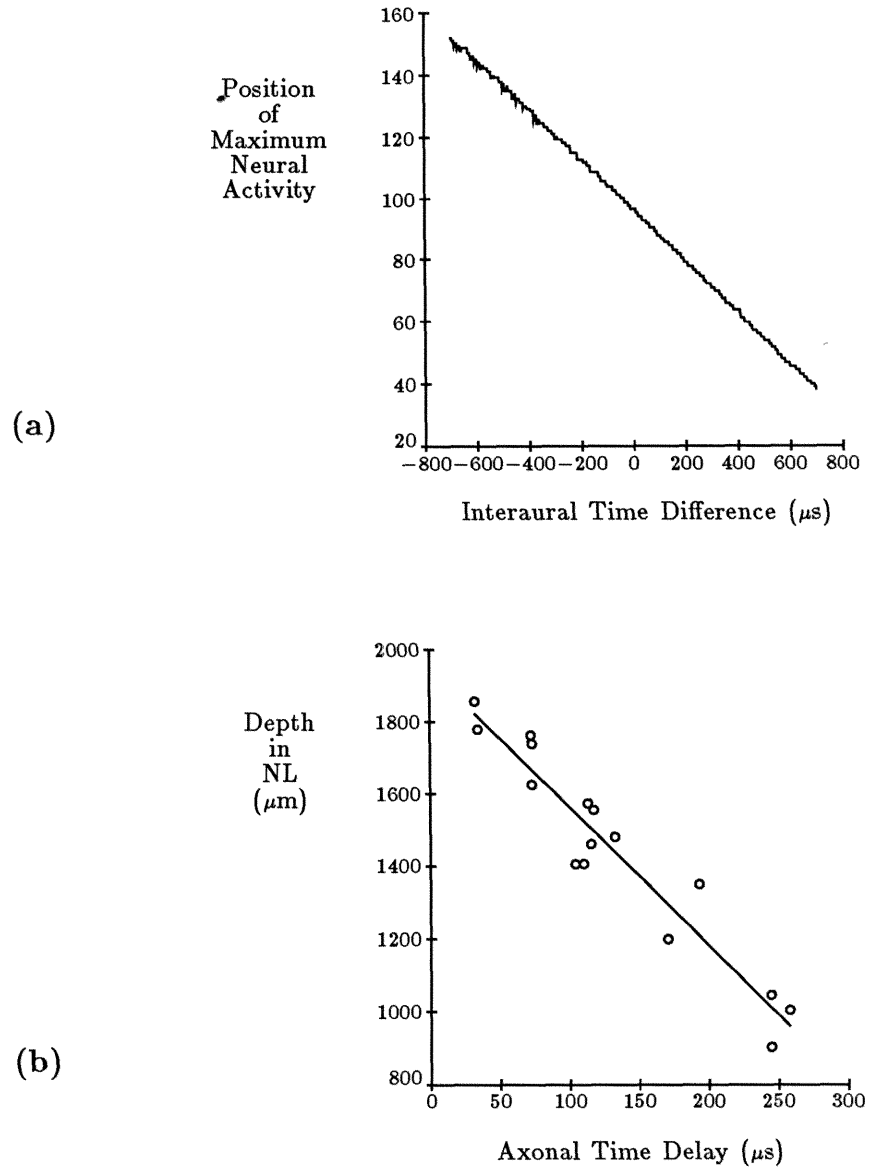


Figure 4.13. **a.** Chip data showing the linear relationship of silicon NL neuron position and interaural time difference. For each interaural time difference presented to the chip, the output map position with the maximal response is plotted (170 map positions). The linearity shows that silicon axons have a uniform mean time delay per section. **b.** Recordings of the NM axons innervating the NL in the barn owl. The figure shows the mean time delays of contralateral fibers recorded at different depths during one penetration through the 7-kHz region. From (Carr and Konishi, 1988).

temporal offset, each row of NL neurons has a different spatial offset; thus, the two-dimensional silicon NL map is not correctly registered in the place dimension.

This incorrect registration is a problem because the projection to the silicon model of ICx and ICc is hardwired; there is no mechanism to adapt to the offsets in the silicon NL map. The biological system might also have the same interaural matching problem; to solve the problem, the system might use adaptation. Adapting time delays in the system could result in a well-registered NL map; in addition, adaptive projections from the NL to higher centers could decrease registration errors in the NL map.

Figure 4.11 shows the variation in propagation velocity of a silicon axon. This variance is a potential source of localization error. The error in the output map of the chip from this error is minimized as a byproduct of the processing algorithm of the system; the output map averages the responses of 62 rows of NL neurons. In the biological system, ICx neurons combine information over many frequency channels in the ICc; this operation also reduces errors in individual ICc responses. In addition, the axons innervating NL might adapt to reduce the variance in their propagation velocity, eliminating the source of the error.

Future enhancements of the chip include modeling these methods of adaptation to improve system performance, modeling the intensity-coding localization pathway, and using the chip as part of a electromechanical system to model sensorimotor integration in the owl. In conclusion, the chip confirms the plausibility of the model of time-coding pathway of the owl formulated by Konishi and collaborators. The chip also demonstrates the utility of analog VLSI technology as a modeling tool in computational neuroscience.

Chapter 5
A Silicon Model of Pitch Perception

Many people can sing, in key, in unison with a melody. Perceiving the pitch of a sound is an essential part of this task. The diversity of sounds that evoke a distinct pitch indicates the complexity of human pitch perception. We perceive a pure sinusoid as having a pitch that depends directly on its frequency. A weighted sum of sinusoids, with harmonic (integer-related) frequencies $f,\ 2f,\ 3f,\ 4f, \ldots$, evokes a pitch identical to a sinusoid of frequency f, even if the sinusoid of frequency f in the sum has a weight of zero. We also perceive a distinct pitch, in a stereotypical fashion, in response to a sum of sinusoids with arithmetically related frequencies, and in response to a sum of time-delayed, correlated noise signals.

Explanations of the ability to perceive pitch initially used physiological models of auditory processing. An early explanation, suggested by Helmholtz, modeled the cochlea as a resonant frequency analyzer (Helmholtz, 1895). Models developed in the 1950s (Licklider, 1951; Licklider, 1959) advanced beyond the auditory periphery, specifying several stations of neural computation in explicit detail. However, two limitations impeded further development of physiological models of pitch perception. Auditory neurophysiology was in its infancy, and could offer researchers little evidence with which to judge proposed theories. In addition, computer simulation and circuit modeling of neural systems were both technologically limited; published computational verification of the Licklider model, by Lyon, did not occur until 1984 (Lyon, 1984).

As a result, models developed in the 1970s (Goldstein, 1973; Wightman, 1973) involved abstract models; the goal of the research was the description of algorithms, computed by an unspecified "central processor," that exactly

matched psychophysical pitch-perception data. These studies contributed essential insights into pitch perception; however, they did not address the implementation strengths and constraints of neural systems.

Advances in auditory physiology and computational neuroscience in the last decade encourage us to return to physiological models of pitch perception. Recent physiological studies have provided new insights into the structure and function of both the auditory periphery (Rhode, 1971; Kiang, 1980; Kim, 1984; Dallos, 1985) and the auditory brainstem nuclei (Carr and Konishi, 1988; Fujita and Konishi, unpublished). Analog VLSI technology is a appropriate medium for pitch-perception modeling; the projects described in Chapters 2, 3, and 4 provide useful tools for building these models.

This chapter describes a silicon integrated-circuit model of pitch perception. The chip receives as input a time-varying voltage corresponding to sound pressure at the ear, and produces as output a map of perceived pitch. The chip is a physiological model; subcircuits on the chip correspond to known and proposed structures in the peripheral auditory system and in the auditory brainstem nuclei. The algorithms of the chip share many details with the work of Licklider (Licklider, 1959); the chip is an analog integrated-circuit implementation of the work of Lyon, who proposed computational experiments with the Licklider model, and published computer simulations of the performance of the model (Lyon, 1984). The chip output approximates human performance on a variety of classical pitch-perception stimuli. The research in this chapter was done in collaboration with Carver Mead.

5.1 System Architecture

Figure 5.1 is a block diagram of the chip. The chip receives as input a time-varying signal, corresponding to the sound pressure at the ear. This input connects to a silicon model (Lyon and Mead, 1988a) of the mechanical processing of the cochlea; Chapter 2 described the cochlea circuit. Silicon models of inner hair cells connect to the cochlea circuit at constant intervals; 62 inner-hair-cell circuits connect to each silicon cochlea. The output of each inner-hair-cell circuit connects directly to a spiral-ganglion-neuron circuit, to complete a silicon model of auditory-nerve response. Chapter 2 describes the auditory-nerve model.

The portion of the chip explained thus far models the known structures of the auditory periphery. The remainder of the chip implements proposed neural structures in the brain. In the chip, each spiral-ganglion-neuron circuit connects to a discrete delay line; for each input pulse, a fixed-width, fixed-height pulse travels through the delay line, section by section, at a controllable velocity (Mead, 1989). The circuit is identical to the silicon axon circuit described in Chapter 4; Figures 4.10 and 4.11 illustrate the operation of the circuit.

A correlation-neuron circuit is associated with each delay-line section; this circuit receives a connection from the output of its delay-line section, and from the spiral-ganglion-neuron circuit that drives the delay line. Simultaneous pulses at both inputs excite the correlation-neuron circuit; if only one input is active, the circuit generates no output. Each row of correlation neurons associated with a spiral-ganglion neuron forms a place code of periodicity. A spiral-ganglion neuron fires in a repeating pattern, on average, in response to a periodic signal in the appropriate frequency range. Correlation neurons that fire maximally receive this repeating pattern simultaneously on both inputs; the time delay associated with this correlation neuron is an integer multiple of the period of the signal. In

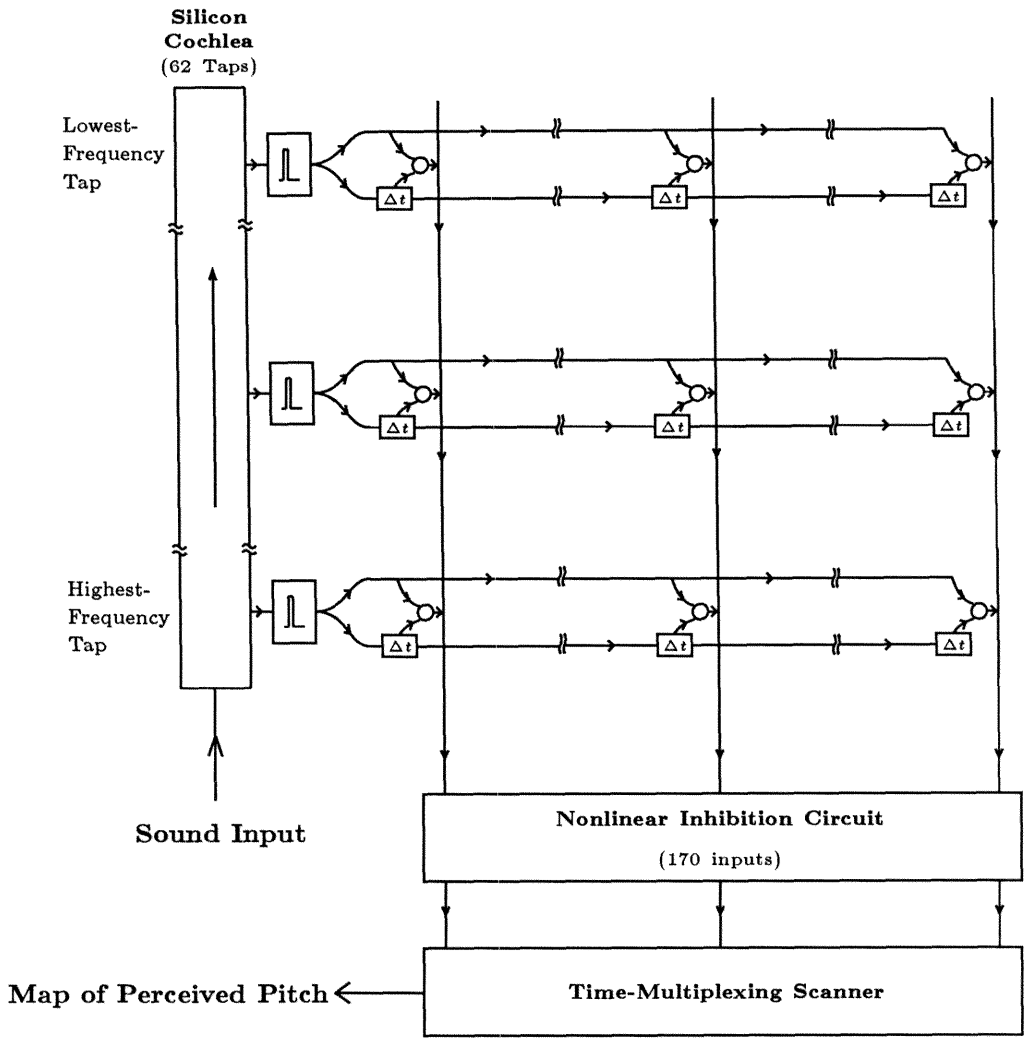

Silicon
Cochlea
(62 Taps)
Lowest-
Frequency
Tap
Highest-
Frequency
Tap
Δt
Sound Input
Nonlinear Inhibition Circuit
(170 inputs)
Time-Multiplexing Scanner
Map of Perceived Pitch

Figure 5.1. Block diagram of the pitch-perception chip. Sound enters the silicon cochlea at the lower left of the figure. Circuits that model inner hair cells and spiral-ganglion neurons tap the silicon cochlea at 62 equally spaced locations; a square box marked with a pulse represents these circuits. Spiral-ganglion-neuron circuits connect to discrete delay lines that span the width of the chip. A small rectangular box, marked with the symbol "Δt," represents a delay-line section; there are 170 sections in each delay line. A correlation-neuron circuit, represented by a small circle, is associated with each delay-line section. A correlation neuron receives connection from its delay-line section, and from the spiral-ganglion-neuron circuit that drives its delay line. Vertical wires, which span the array, sum the response of all correlation neurons that correspond to a specific time delay. These 170 vertical wires form a temporally smoothed map of perceived pitch. The nonlinear inhibition circuit near the bottom of the figure increases the selectivity of this map; the time-multiplexing scanner sends this map off the chip.

engineering terms, each correlation neuron computes the running autocorrelation function of a filtered version of the sound input, for a particular time delay.

In 1951, Licklider (Licklider, 1951) proposed this neural autocorrelation structure as a periodicity representation that could be implemented plausibly with synaptic delays in neural circuitry. Although no direct physiological evidence for these autocorrelation structures has been discovered, Carr and Konishi (Carr and Konishi, 1988) have found direct evidence for cross-correlation structures for auditory localization in the midbrain of the barn owl; these structures, shown in Figure 4.6 in Chapter 4, use axonal time delays to compute a place code of interaural time delay.

Figure 5.2 shows the utility of neural autocorrelation structures in the perception of the pitch of a weighted, harmonic sum of sinusoids, with frequencies (f, $2f$, $3f$, $4f$,...,). Due to the filtering action of the cochlea, different sinusoids are predominant in different autocorrelators throughout the chip. Cochlear processing is idealized in Figure 5.2; the figure shows an analog representation of the signals in the delay lines across the chip, assuming that all sinusoids are in phase. The peaks in all the delay lines coincide with the peaks of the sinusoid of frequency f. Thus, even if the sinusoid of frequency f has zero weight, the representation still encodes the frequency f, the perceived pitch of the sum. The outputs of the correlation neurons reflect this representation; in addition, they are invariant to the relative phase of the sinusoids.

To complete his model, Licklider proposed a self-organizing neural network, which received connections from the autocorrelation structures, and which learned to associate firing patterns with the perception of pitch. For our chip, we designed a simple recognition algorithm, suitable for the perception of a single pitch. The algorithm shares its structure with the silicon model of ICc and ICx processing in the time-coding pathway of the owl, presented in Chapter 4.

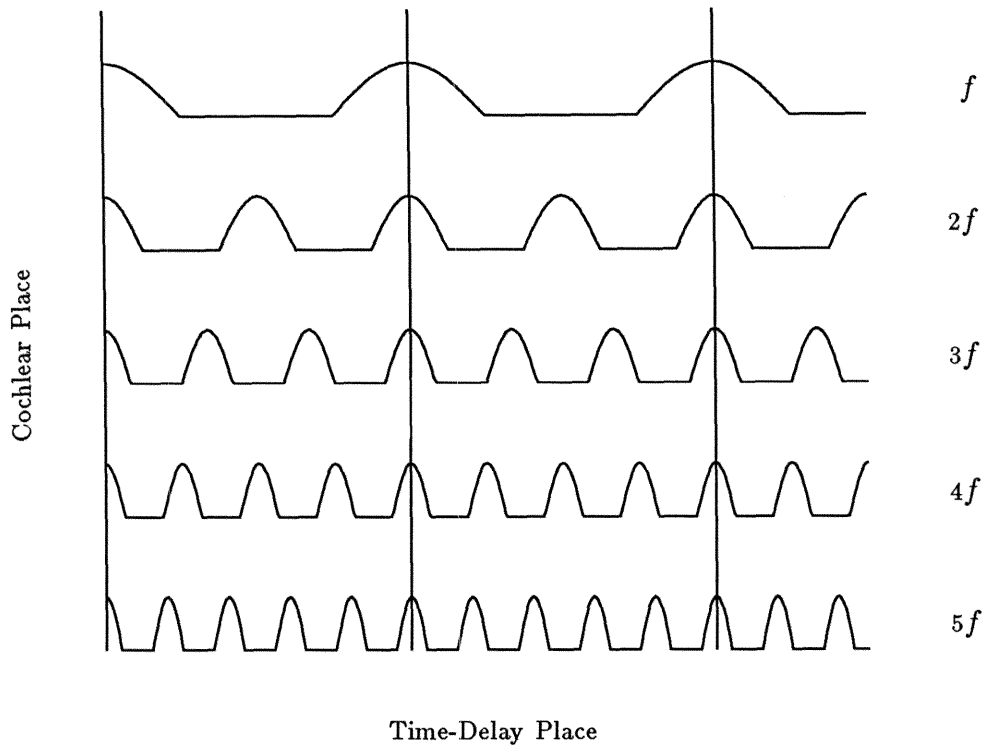


Figure 5.2. Analog representation of the signals in the delay lines across the chip, in response to a harmonic signal. Cochlear processing is idealized (fully resolved harmonic components, perfect half-wave rectification, no temporal smoothing). Peaks of activity in the horizontal direction coincide with peaks in f, shown by vertical lines.

The recognition algorithm sums all correlation-neuron outputs corresponding to a particular time delay, across frequency channels, to produce a single output value. Correlation-neuron outputs are current pulses; a single wire, running vertically through the chip, acts as a dendritic tree to perform the summation for each time delay. In this way, a two-dimensional representation of correlation neurons reduces to a single vector; this vector is the map of perceived pitch. The chip then integrates this vector temporally, with an adjustable time constant, providing a stable representation over many cycles of the input signal. Finally, a global nonlinear-inhibition circuit, described in Chapter 3 and shown in Figure 3.1, processes this temporally integrated vector. This nonlinear circuit performs a winner-take-all function, producing a more selective map of perceived pitch. The chip time multiplexes this output map on a single wire for display on an oscilloscope.

5.2 Chip Responses

To show the capabilities and limitations of the silicon model, we recorded chip responses to a variety of classical pitch-perception stimuli. In these experiments, we tuned the basilar-membrane circuit to span about five octaves; lowpass cutoff frequencies ranged from 300 Hz to 10,000 Hz. The delay lines were tuned to provide about 3.3 ms of total delay; with this tuning, the chip perceives pitches above 300 Hz. Temporal smoothing by the recognition algorithm acted with a time constant of tens of milliseconds.

Figure 5.3(a) shows maps of perceived pitch period, generated by the chip in response to sine, triangle, and square waves at various frequencies. As desired, chip response is invariant to the harmonic content of the signal. The chip response shows the first global peak of the autocorrelation representation; the spatial variation in the delay-line timing weakens the strength of subsequent

peaks. In Figure 5.3(b), we recorded the map position of the neuron with maximum signal energy, for square waves of different frequencies; the graph shows a linear relationship between the input period of the waveform and the pitch period predicted by the chip.

The stimuli in Figure 5.4 illustrate the classical "missing fundamental" aspect of pitch perception. Figure 5.4(a) shows a narrow-pulse waveform, whereas Figure 5.4(b) shows the sum of this narrow-pulse waveform and a synchronized sinusoid with appropriate frequency, amplitude, and phase to cancel exactly the fundamental frequency of the pulse waveform. Human subjects perceive the pitch of both waveforms to be identical (Schouten, 1940); Figure 5.4(c) shows identical maps from the chip in response to both waveforms, at various frequencies.

As in the biological system, the chip circuits that model the cochlear periphery are in some aspects nonlinear, and resynthesize the fundamental frequency of the signal in Figure 5.4(b). We have done several experiments to show that the effect of distortion products is negligible. Decreasing the intensity of the stimulus, within the operating range of the chip, does not alter the response map; at lower intensities, spectral analysis of cochlear-circuit outputs shows the strength of the fundamental component of the signal to be near the circuit's noise floor. In addition, chip response does not change when a lowpass-filtered white-noise signal, with a cutoff frequency above the fundamental of the stimulus, is added to the signal shown in Figure 5.4(b) (Licklider, 1954).

The stimulus in Figure 5.5(a) shows the sensitivity of the chip to harmonic patterns; to create the stimulus, we summed two synchronized pulse waveforms, of identical frequency f and relative phase difference ϕ. For $\phi = 180°$, the stimulus has a frequency of $2f$. For $\phi = 180° \pm \Delta$, the stimulus has a frequency of f, but the odd harmonics, especially the fundamental, are very weak, as shown

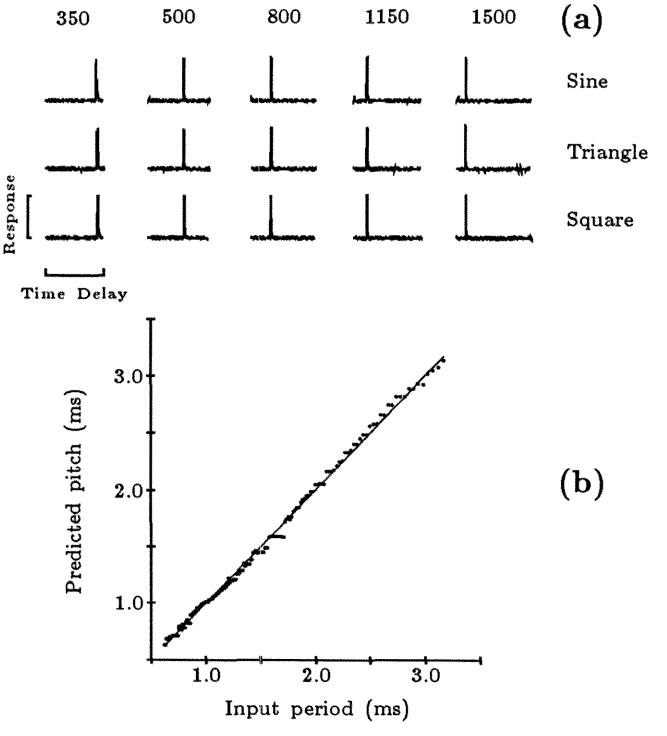


Figure 5.3. **a.** Maps of perceived pitch period from the chip, in response to sine, triangle, and square waves. Column numbers denote frequency in Hz. **b.** Plot showing map position of the neuron with maximum signal energy, for square waves of different frequency; ordinate axis is calibrated from data to indicate pitch period. Dots are data points; solid line shows best linear fit to the data.

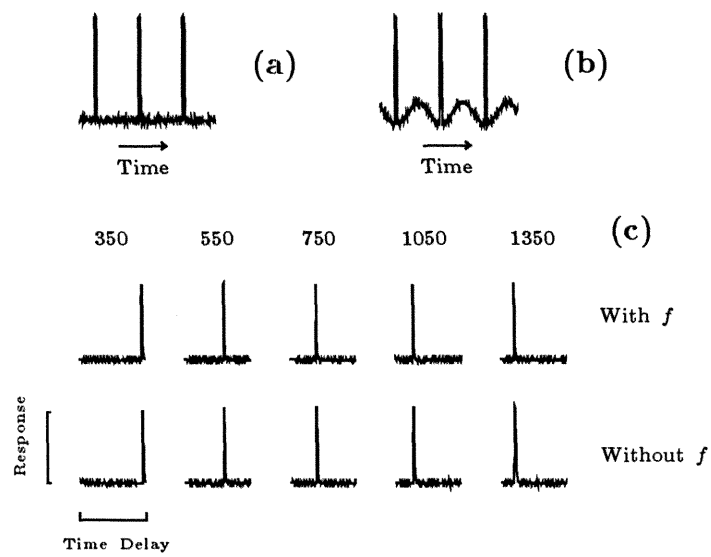


Figure 5.4. **a.** Narrow-pulse sound stimulus. **b.** Narrow-pulse sound stimulus, with canceled fundamental frequency (Schouten, 1940). **c.** Maps of perceived pitch period from the chip, in response to stimuli shown in parts a and b, at various frequencies; chip responds identically. Column numbers denote frequency in Hz.

in Figure 5.5(b). For remarkably small Δ, human subjects hear an octave pitch shift between the $\phi = 180°$ and $\phi = 180° \pm \Delta$ stimuli (Schouten, 1940). Figure 5.5(c) presents pitch maps from the chip for various ϕ, which also show the perceived octave pitch shift.

A sum of three sinusoids, with arithmetically related frequencies $f_c - f_m$, f_c, and $f_c + f_m$, is a revealing pitch-perception stimulus; an amplitude-modulated sinusoid, with carrier frequency f_c and modulator frequency f_m, as shown in Figure 5.6(a), produces this spectral pattern. If f_c is equal to nf_m, where n is an integer, the three sinusoids form an integer-related series, and human subjects perceive a pitch equivalent to that of a sinusoid at the implied fundamental frequency f_m. If f_c is equal to $(n + \epsilon)f_m$, human subjects perceive, to a first order, a pitch equivalent to a sinusoid at the frequency f_c/n (de Boer, 1956). As postulated by de Boer (de Boer, 1956), the human perceptual system calculates a pseudoperiod of this near-harmonic stimulus. The chip response to varying f_c, shown in Figure 5.6(b), matches the first-order perception of human subjects.

If f_c is held constant and f_m is varied, human subjects, to a first order, perceive a pitch equivalent to that of a sinusoid with the frequency of the integral submultiple of f_c nearest to f_m (de Boer, 1956). The chip response to varying f_m, shown in Figure 5.6(c), matches the first-order response of human subjects, limited by the resolution of the output map.

Human perception of amplitude-modulated sinusoids has significant second-order properties. If f_m is held constant and f_c is varied, the perceived pitch is not described exactly by the expression f_c/n; the slope of the response is slightly greater than $1/n$. If f_c is held constant and f_m is varied, the perceived pitch is not exactly the integral submultiple of f_c nearest to f_m; the perceived pitch decreases slightly with increasing f_m (de Boer, 1956). As postulated by de Boer

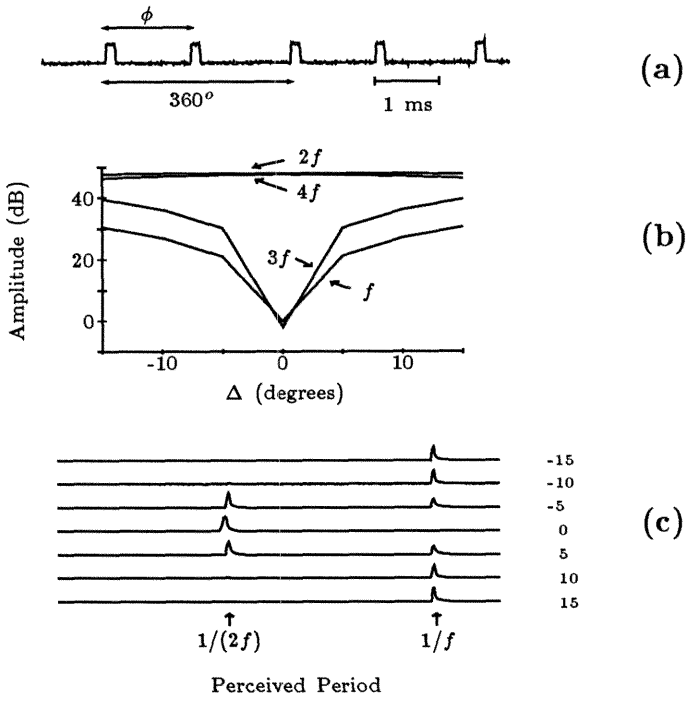


Figure 5.5. **a.** Sound stimulus created by the addition of two synchronized narrow-pulse waveforms, of identical frequency f and phase difference ϕ (Schouten, 1940). When $\phi = 180°$, stimulus frequency is $2f$; when $\phi = 180° \pm \Delta$, stimulus frequency is f. **b.** Measured relative strength of first four harmonics of stimulus shown in part a, in dB, as a function of Δ. The strength of the odd harmonics is at the noise floor for $\Delta = 0°$. **c.** Maps of perceived pitch period from the chip, in response to the stimulus, while Δ is varied; values of Δ are shown in the right column, in degrees.

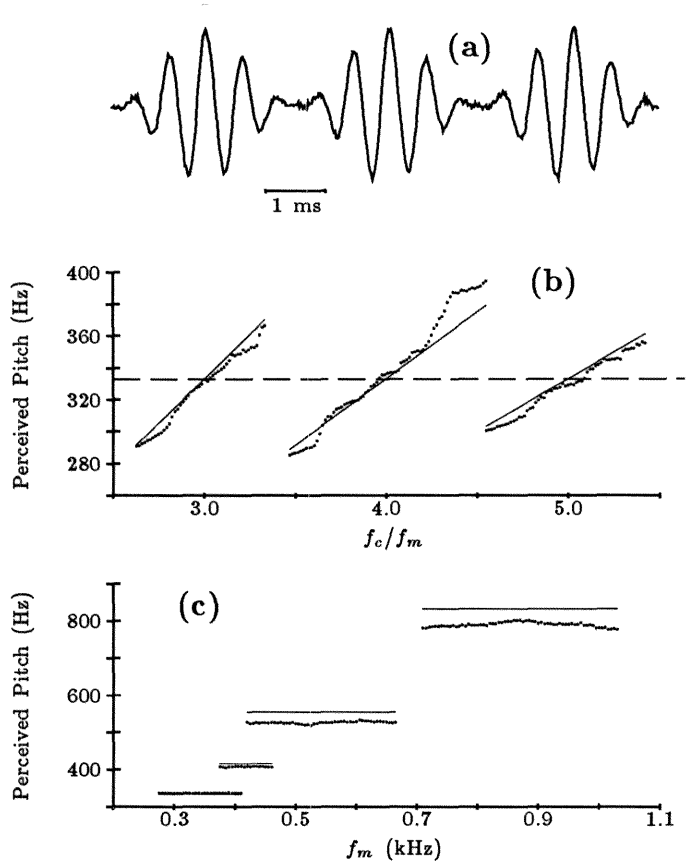


Figure 5.6. **a.** Amplitude-modulated sinusoid sound stimulus (de Boer, 1956). **b.** Plot showing center of energy of the chip map, in response to stimulus shown in part a, while the carrier frequency, f_c, is varied. Dotted line shows frequency of fixed modulation frequency f_m = 333 Hz. Dots are data points; solid lines show theoretical first-order human response, as explained in text. **c.** Plot showing center of energy of chip map, in response to the stimulus shown in part a, while the modulation frequency, f_m, is varied, with f_c = 1665 Hz. Dots are data points; solid lines show theoretical first-order human response, as explained in text.

(18), these second-order properties reveal the weighting of individual frequency components in the computation of the pseudoperiod by the human perceptual system. In the chip, the simple recognition algorithm does not support the relative weighting of frequency components; as a result, the responses depicted in Figure 5.6 do not show the second-order properties of the human perceptual system.

Human subjects perceive a faint, but distinct, pitch in response to a sum of two time-delayed, correlated noise signals (Fourcin, 1965); the period of the perceived pitch is equal to the time delay. This stimulus is relevant to auditory localization, as well as to pitch perception; the outer ear produces time-delayed replicas of incoming sounds, which encode the elevation angle of sound sources in mammals (Batteau, 1967). Output maps from the chip show a perceived pitch, in response to a bandpass-filtered sum of two time-delayed, correlated noise signals, as shown in Figure 5.7(a). As shown in Figure 5.7(b), the center of energy of the chip map varies linearly with time delay, in agreement with the linear characteristic of Figure 5.3(b). Like that of human subjects, the response of the chip to the noise stimulus is faint; to obtain the data in Figure 5.7, we decreased the integration time constant of the recognition algorithm, and time averaged the responses off-chip.

5.3 Discussion

The chip output approximates human performance in response to a variety of classical pitch-perception stimuli. The major shortcoming of the chip is the inadequate modeling of the second-order properties of pitch perception of amplitude-modulated tones; this shortcoming is not a failure of neural autocorrelator structures as a representation, but rather is a property of

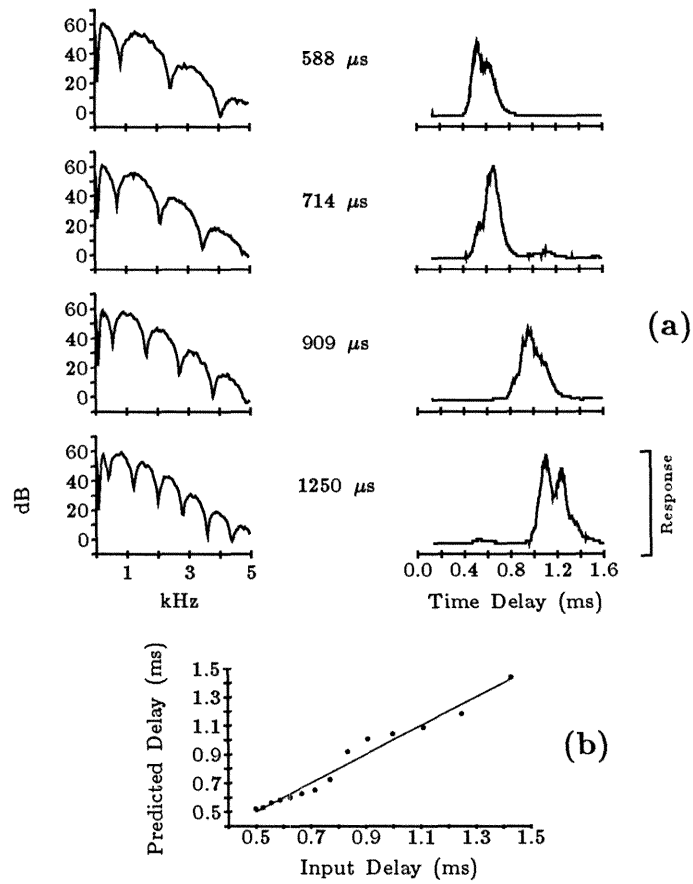


Figure 5.7. **a.** Plots showing maps of perceived pitch period from the chip, in response to a bandpass-filtered sum of two time-delayed, correlated noise signals (right). Filtering was kept constant for all plots. Centered numbers indicate time delay between noise signals. Left plots show the spectral content of the input stimulus (60-Hz filter bandwidth). **b.** Plot showing center of energy of chip map, in response to bandpass-filtered sum of two time-delayed, correlated noise signals, as a function of time delay. Ordinate axis is calibrated from data to indicate perceived delay. Dots are data points; solid line shows best linear fit to the data.

the chip's simple recognition algorithm, which does not support the relative weighting of frequency components in the recognition process.

Neural autocorrelation structures, at the appropriate time scale, are a natural initial representation for a variety of auditory tasks. Autocorrelation time delays of hundreds of microseconds match the time delays introduced by the outer ear to encode auditory localization information in the elevational plane. Time delays in the millisecond range support pitch-perception and complex sound-recognition tasks; time delays of hundreds of milliseconds might form a substrate for the perception of rhythms. As a result, there might be a number of autocorrelation structures in the auditory system, at different time scales. Faster delays probably use axonal delay lines, as do the localization structures of the barn owl (Carr and Konishi, 1988), whereas slower delays probably use neural circuits for delay units. The autocorrelation structures might form a logarithmic map of time, unlike our chip's linear map, to represent compactly many orders of magnitude of time delay.

In conclusion, the pitch-perception chip confirms the practicality of neural autocorrelation structures as a representation of pitch perception in auditory processing. The chip also demonstrates the utility of analog VLSI circuits as a modeling tool in computational neuroscience.

Chapter 6
Future Research

The projects described in this dissertation demonstrate the utility of analog VLSI technology as a modeling tool in auditory neuroscience. The dissertation does not, however, propose any new auditory theories. The auditory-localization and pitch-perception chips implement theories that were proposed by Jeffress (Jeffress, 1948) and Licklider (Licklider, 1951) several decades ago. The inner-hair-cell and spiral-ganglion-cell models presented in Chapter 2 are a codification of known properties of these cells.

This conservative approach to the dissertation was deliberate, and is in harmony with a well-known computer-science dictum: Do not let your research project depend on another research project. Through the projects described in this dissertation, my co-workers and I wanted to answer the question: Is analog VLSI technology an appropriate medium for auditory modeling? This dissertation answers this question affirmitively; if we had attempted, and failed, to invent new models of auditory function, we would have left our primary question unanswered. There is another question for the future, which this chapter explores: What are the interesting, yet tractable, auditory questions to ask using analog VLSI technology? We devote each section of the chapter to a different topic.

6.1 The Cochlea: Frequency-Specific Automatic Gain Control

As detailed in Chapter 2, the cochlea contains outer hair cells, which are active electromechanical elements. Outer hair cells reduce the damping of the passive basilar membrane, and allow weaker signals to be heard. Axons from higher brain centers innervate the outer hair cells; these centers might

dynamically vary the local damping of the cochlea, providing frequency-specific automatic gain control (Kim, 1984).

The circuit model of auditory-nerve response, described in Chapter 2, includes outer-hair-cell function, but does not model dynamic variation of outer-hair-cell motility by higher brain centers. Adding a frequency-specific automatic-gain-control system to the silicon model of auditory-nerve response is a tractable project in analog VLSI technology. The control system should independently vary the damping of local regions of basilar-membrane circuitry, in response to sound input. The control circuit must maintain the stability of the basilar-membrane circuit; although human ears are occasionally oscillatory (ringing ears), chronic oscillations are classified as a symptom (tinnitus) of auditory diseases.

6.2 The Barn Owl: Learning Elevational Space

As described in Chapter 4, the barn owl uses interaural time differences to encode azimuthal position in auditory space, and interaural intensity differences to encode elevational position. The interaural intensity difference that corresponds to a particular spatial position is frequency dependent; computing a spatial map from interaural intensity difference requires knowledge about the spectral characteristics of the head of the animal.

A silicon model of elevation coding in the barn owl would require several levels of processing. A possible model would begin with two circuit models of auditory-nerve response (Chapter 2), followed by processing to enhance the intensity information in the auditory-nerve response; this enhancement processing would model the function of the nucleus angularis (Sullivan and Konishi, 1984). Binaural information would converge onto neurons that compute interaural intensity differences in different frequency regions, modeling neuronal

responses in the nucleus ventralis lemnisci lateralis pons posterior (VLVp) (Manley *et al.*, 1988). To complete the computation, a learning network would correlate patterns of interaural intensity difference at different frequencies, to form a representation of elevation space. The output of this network would model the elevational response of the computed map of auditory space in the ICx, shown in Figure 4.1 in Chapter 4.

6.3 The Mustache Bat: Active Sonar Processing

The mustache bat, *Pteronotus parnelli rubiginosus*, uses active sonar for capturing flying insects, and for short-term navigation. The bat emits ultrasonic signals, and processes the resulting echos, to sense the environment. Suga and co-workers (Suga, 1984) have characterized a variety of neural maps in the bat auditory cortex, which systematically process different aspects of the sonar signal.

Analog VLSI technology is an appropriate tool for modeling these maps. The bat processes sound frequencies in excess of 100 kHz; analog VLSI circuits can process sounds at these frequencies. The peripheral auditory system of the bat is specialized for sonar processing; these specializations could be incorporated into the silicon model of auditory-nerve response (Chapter 2). Time-correlation techniques used in the pitch-perception (Chapter 5) and auditory-localization (Chapter 4) chips, in modified form, could form the basis for silicon models of certain maps in the bat auditory cortex.

6.4 Bats and Owls: Sensorimotor Feedback

When localizing a sound, the barn owl aligns its head to the source of the sound in a stereotyped manner (Knudsen and Konishi, 1979). Designing an electromechanical model of the alignment motion of the owl, using a motor-driven head with two microphones and the silicon model of auditory localization (Chapter 4), is a tractable project. Similar electromechanical systems modeling alignment to a visual stimuli have been built (DeWeerth and Mead, personal communication).

A more challenging project involves electromechanical modeling of bat insect capture. A flying moth-chaser, linking a sonar transceiver, a radio-controlled airplane, and silicon models of neural maps in the bat auditory cortex, would be a spectacular display of analog VLSI processing. Radio control would be used for takeoff and landing; once in flight, the airplane would seek out a flying object of a specified shape, size, and speed, and would follow the object at a safe distance. This project clearly would not be a trivial undertaking, but it would be possible to complete with our current technology.

6.5 Human Perception: Modeling Binaural Phenomena

Human binaural perception includes phenomena other than auditory localization of sounds. Binaural processing enhances our ability to detect signals in noise: for example, binaural hearing aids the comprehension of speech against a background of other speech (the cocktail-party effect). Several binaural signal-detection theories have been advanced (Jeffress, 1972; Durlach, 1972). Constructing analog VLSI models of these theories is a tractable project.

Binaural perception includes a wealth of interesting auditory illusions, which raise interesting psychological and physiological questions about audition

(Tobias, 1972; Deutsch, 1981). The systems described in Chapters 2, 4, and 5 are appropriate substrates for analog VLSI projects that explore these curious binaural phenomena.

6.6 Speech Recognition: The Final Frontier

Analog VLSI systems have two different missions in the speech-recognition community. The projects described in this dissertation offer real-time biological models of early auditory processing, for integration into existing speech-recognition architectures. In addition, analog VLSI technology is an appropriate implementation medium for neurally inspired speech-recognition algorithms. The projects described in this dissertation form a substrate for speech-oriented neural maps in analog VLSI technology.

Such maps could represent interesting characteristics of the speech signal; alternatively, these maps could represent the muscle movements necessary to produce the input speech signal (Liberman and Mattingly, 1989). Using the techniques of Chapter 4 and Chapter 5, these maps could process temporal as well as spectral aspects of the speech signal, as encoded by the silicon auditory-nerve model (Chapter 2). The pitch-perception chip (Chapter 5) computes the fundamental frequency of a speech signal in real time; the chip could serve as a component in a speech-recognition system that uses fundamental-frequency knowledge to create robust neural maps of interesting characteristics of the speech signal.

6.7 Conclusions

This chapter has presented interesting, tractable projects in auditory modeling, using analog VLSI technology. The previous chapters detail completed VLSI auditory models. Transforming project ideas into completed systems is a rewarding experience; we invite interested researchers to verify this claim.

Building analog VLSI chips has never been easier. The MOS Implementation Service (MOSIS), administered by the University of Southern California Information Sciences Institute (USC ISI), offers the fabrication of prototype VLSI chips to researchers in North America, for as little as $500 per project. Computer-aided design tools for personal computers and workstations are available at nominal costs from universities, and from innovative companies. Practical texts exist for learning the craft of analog VLSI design (Mead, 1989).

To become a new branch of computational neuroscience, analog VLSI modeling needs more participants, from both the engineering and the scientific communities. The territory is vast, and largely unexplored. The rewards are great for those who simply press forward.

References

Batteau, D. W. (1967). The role of the pinna in human localization. *Proc. Roy. Soc.* (London B) **158**: 158–180.

de Boer, E. (1956). Pitch of inharmonic signals. *Nature* (London) **178**: 535–536.

Carr, C. E. and Konishi, M. (1988). Axonal delay lines for time measurement in the owl's brainstem. *Proc. Natl. Acad. Sci.* **85**: 8311–8315.

Dallos, P. (1985). Response characteristics of mammalian cochlear hair cells. *J. Neurosci.* **5**: 1591–1608.

Deutsch, D. (1981). The octave illusion and auditory perceptual integration. In Tobias, J. V. (ed). *Hearing Research and Theory,* Vol. I. New York: Academic Press, pp. 99–142.

Durlach, N. I. (1972). Binaural signal detection: Equalization and cancellation theory. In Tobias, J. V. (ed). *Foundations of Modern Auditory Theory*, Vol. II. New York: Academic Press, pp. 371–462.

Evans, E. F. (1982). Functional anatomy of the auditory system. In Barlow, H. B. and Mollon, J. D. (eds), *The Senses.* Cambridge, England: Cambridge University Press, p. 251.

Fourcin, A.J. (1965). The pitch of noise with periodic spectral peaks. In *Proceedings 5th International Congress on Acoustics, Leige*, Vol. Ia, B 52.

Fujita I., and Konishi, M. (unpublished). The role of GABAergic inhibition in coding of interaural time difference in the owl's auditory system.

Geisler, C.D. and Greenberg, S. (1986). A two-stage nonlinear cochlear model posseses automatic gain control. *J. Acoust. Soc. Am.* **80**: 1359–1363.

Goblick, T.J., Jr., and Pfeiffer, R. R. (1969) Time domain measurements of cochlear non-linearities using combination click stimuli. *J. Acoust. Soc. Am.* **46**: 924–938.

Goldberg, J. M. and Brown, P. B. (1969). Response of binaural neurons to dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization. *J. Neurophysiol.* **32**: 613–636.

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am.* **54**: 1496–1516.

Greenberg, S. (1988). The ear as a speech analyzer. *J. Phonetics* **16**: 139–149.

Helmholtz, H. L. F. (1895). *Sensations of Tone.* Ellis, A. J. (Trans.) New York: Longmans Green, pp. 49–65.

Jeffress, L. A. (1948). A place theory of sound localization. *J. Comp. Physiol. Pyschol.* **41**: 35–39.

Jeffress, L. A. (1972). Binaural signal detection: Equalization and cancellation theory. In Tobias, J. V. (ed). *Foundations of Modern Auditory Theory*, Vol. II. New York: Academic Press, pp. 349–368.

Kiang, N. Y.-s, Watenabe, T., Thomas, E.C., and Clark, L.F. (1965). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve.* Cambridge, MA: MIT Press.

Kiang, N. Y.-s (1980). Processing of speech by the auditory nervous system. *J. Acoust. Soc. Am.* **68**: 830–835.

Kim, D. O. (1984). Functional roles of the inner- and outer-haircell subsystems in the cochlea and brainstem. In Berlin, C. I. (ed), *Hearing Science.* San Diego, CA: College-Hill Press, pp. 241–262.

Knudsen, E. I. and Konishi, M. (1978). A neural map of auditory space in the owl. *Science* **200**: 795–797.

Knudsen, E. I., Blasdel, G. G., and Konishi, M. (1979). Sound localization by the barn owl measured with the search coil technique. *J. Comp. Physiol.* **133**: 1–11.

Knudsen, E. I. and Konishi, M. (1979). Mechanisms of sound localization in the barn owl (*Tyto alba*). *J. Comp. Physiol.* **133**: 13–21.

Knudsen, E. I. and Konishi, M. (1980). Monaural occlusion shifts receptive field locations of auditory midbrain units in the owl. *J. Neurophysiol* **44**: 687–695.

Konishi, M. (1986). Centrally synthesized maps of sensory space. *TINS* **4**: 163–168.

Lazzaro, J. P., Ryckebusch, S., Mahowald, M. A., and Mead, C. (1988). Winner-take-all networks of O(n) complexity. In Tourestzky, D. (ed), *Advances in Neural Information Processing Systems 1.* San Mateo, CA: Morgan Kaufmann Publishers, pp. 703-711.

Lazzaro, J. P. and Mead, C. (1989a). Silicon models of auditory localization. *Neural Computation* **1**: 41–70.

Lazzaro, J. P. and Mead, C. (1989b). Circuit models of sensory transduction in the cochlea. In Mead, C. and Ismail, M. (eds), *Analog VLSI Implementations of Neural Networks.* Norwell, MA: Kluwer Academic Publishers, pp. 85-101.

Liberman, A. M. and Mattingly, I. G. (1989). A specialization for speech perception. *Science* **243**: 489–494.

Licklider, J. C. R. (1951). A duplex theory of pitch perception. *Experientia* **7**: 128–134.

Licklider, J. C. R. (1954). "Periodicity pitch" and "place pitch." *J. Acoust. Soc. Am.* **26**: 945 (A).

Licklider, J. C. R. (1959). Three auditory theories. In Koch, S. (ed), *Pyschology: A Study of a Science,* Vol 1. New York: McGraw-Hill, pp. 94–144.

Lyon, R. F. (1984). Computational models of neural auditory processing. In *Proceedings, 1984 IEEE ICASSP,* San Diego, CA (Mar. 1984).

Lyon, R. F. and Mead, C. (1988a). An analog electronic cochlea. *IEEE Trans. Acoust., Speech, Signal Processing* **36**: 1119–1134.

Lyon, R. F. and Mead, C. (1988b). *Cochlear Hydrodynamics Demystified.* Caltech Computer Science Technical Report, Caltech–CS–TR–88–4, California Institute of Technology, Pasadena, CA, February 1988.

Mahowald, M.A., and Delbruck, T.I. (1989). Cooperative stereo matching using static and dynamic image features. In Mead, C. and Ismail, M. (eds), *Analog VLSI Implementations of Neural Networks.* Norwell, MA: Kluwer Academic Publishers, pp. 181-206.

Mead, C. (1989). *Analog VLSI and Neural Systems.* Reading, MA: Addison-Wesley.

Manley, G. A., Koppl, C., and Konishi, M. (1988). A neural map of interaural intensity difference in the brain stem of the barn owl. *J. Neurosci.* **8**: 2665-2676.

Moiseff, A. and Konishi, M. (1981). Neuronal and behavioral sensitivity to binaural time differences in the owl. *J. Neurosci.* **1**: 40–48.

Moiseff, A. and Konishi, M. (1983). Binaural characteristics of units in the owl's brainstem auditory pathways: Precursors of restricted spatial fields. *J. Neurosci.* **3**: 2553–2562.

Rhode, W. S. (1971). Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique. *J. Acoust. Soc. Am.* **49**: 1218–1231.

Rhode, W. S., Geisler, C.D., and Kennedy, D.T. (1978). Auditory nerve fiber response to wide-band noise and tone combinations. *J. Neurophysiol.* **41**: 692–704.

Robles, L., Rhode, W. S., and Geisler, C.D. (1976) Transient response of basilar membrane measured in squirrel monkeys using the Mossbauer effect. *J. Acoust. Soc. Am.* **59**: 926–939.

Rose, J.E., Hind, J.E., Anderson, D. J., and Brugge, J. F. (1971). Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey. *J. Neurophysiol.* **34**: 685–699.

Sachs, M. B. and Abbas, P. J. (1974) Rate versus level functions for auditory-nerve fibers in cats: Tone-burst stimuli. *J. Acoust. Soc. Am.* **56**: 1835–1847.

Schouten, J. F. (1940). The residue, a new component in subjective sound analysis. *Proc. Kon. Acad. Wetensch. (Neth.)* **43**: 356–365.

Suga, N. (1984). The extent to which biosonar information is represented in the bat auditory cortex. In Edelman, G. M., Gall, W. E., and Cowan, W. M. (eds), *Dynamic Aspects of Neocortical Function.* New York: Wiley, 1984.

Sullivan, W. E. and Konishi, M. (1984). Segregation of stimulus phase and intensity coding in the cochlear nucleus of the barn owl. *J. Neurosci.* **7**: 1787–1799.

Takahashi, T. T. and Konishi, M. (1988a). Projections of the cochlear nuclei and nucleus laminaris to the inferior colliculus of the barn owl. *J. Compar. Neurol.* **274**: 190–211.

Takahashi, T. T. and Konishi, M. (1988b). Projections of the nucleus angularis and nucleus laminaris to the lateral lemniscal nuclear complex of the barn owl. *J. Compar. Neurol.* **274**: 221–238.

Tobias, J. V. (1972). Curious binaural phenomena. In Tobias, J. V. (ed). *Foundations of Modern Auditory Theory*, Vol. II. New York: Academic Press, pp. 463–486.

Wagner, H., and Konishi, M. (unpublished).

Wightman, F. L. (1973). The pattern-transformation model of pitch. *J. Acoust. Soc. Am.* **54**: 407–416.

Yin, T. C. T and Kuwada, S. (1984). Neuronal mechanisms of binaural interaction. In Edelman, G.M., Gall, W.E., and Cowan, W.M. (eds), *Dynamic Aspects of Neocortical Function.* New York: Wiley, 1984.