Biologically-based Auditory Signal Processing in Analog VLSI

John Lazzaro
UC Berkeley, EECS
Berkeley, CA, 94720

Abstract

This paper reviews many recent analog silicon implementations of computational models of neural auditory processing. These implementations are based on physiological and psychophysical knowledge about biological auditory systems.

1 Introduction

The questions "How do we hear?" and "How can we improve artificial hearing systems?" are rarely asked by the same researcher. Why is research that holistically addresses hearing systems so rare? In the vision community, a sizable group of researchers use both biological and artificial systems as motivations for their research. Biological motion control and robotic motion control are also considered in a unified framework by many researchers of movement.

The different computational environments of biological auditory systems and artificial hearing systems are a key reason for the separation between the disciplines. Artificial hearing systems are usually specified as sampled digital algorithms, and are usually implemented as programs on a small number of fast serial digital computers. The nature of the computing hardware has affected both the form and the content of algorithms. For example, both serial and shared-memory parallel digital computing architectures insure access to all memory locations in the system, and reward frequent accesses to the same memory location. Efficient algorithms in this architecture exploit these facts.

Likewise, the algorithms that are central to biological auditory processing also exploit the computational architecture of the brain. The cochlea, the sense organ of hearing, uses both electrical and micromechanical techniques to implement the first stage of auditory signal processing. Later processing stages exploit the computational properties of axons, dendrites, and synapses to efficiently perform non-linear time-domain processing of auditory signals. The algorithms are robust to the noisy nature of neural computation, and can be efficiently wired in two-dimensional planes and three-dimensional spaces.

Both artificial and biological auditory algorithms have evolved in harmony with their primary implementation medium. It is not surprising that implementing an algorithm from one system in the computational medium of the other system would be possible but often impractical. For example, a discrete Fourier transform (DFT) could be directly implemented with biological neurons, by forming bistable storage elements and logic gates from small groups of neurons, and then proceeding according to conventional logic design. A clever neural designer could use the three-dimensional wiring space of the brain to offset the slower response time of biological neurons, and leverage nonlinear properties of dendritic trees to perform fast reliable logic. But to our knowledge there are no DFT processors in the brain; the computational properties of the nervous system favor implementing algorithms radically different from those of the digital signal processing tradition.

Likewise, several researchers over the previous decade have implemented computational models of biological auditory processing, with the goal of incorporating these models into a speech recognition system [9] [1]. These projects have shown the promise of the biological approach, sometimes showing clear performance advantages over traditional methods [15]. However, these algorithms are very computationally costly if run on serial digital computers.

Analog integrated circuit technology is a computational medium that shares many of the strengths and weaknesses of neural circuitry [14]. Several research groups have combined analog integrated circuit technology and theories of biological audition, to create silicon microsystems that efficiently implement computational models of biological auditory processing. Some of these systems compute early representations of biological audition [10] [13] [3] [6] [7] [8] [12] [18]; other systems attempt to perform a particular task,

such as auditory localization [2] [5] or pitch perception [4].

The original publications of many of these systems are in the literature of neural networks and circuit design. In this article, I review these systems from a signal processing perspective.

2 Peripheral Processing

Auditory signal processing begins in the cochlea, the sense organ of hearing. The cochlea acts as a transducer, converting acoustic energy into the first neural representation of audition. The auditory nerve carries this representation to the brainstem; the nerve contains approximately 50,000 fibers. [16] includes a review of the structure and function of the cochlea.

Sound processing begins in the outer and middle ear, where the resonant properties of the ear canal and outer ear gently shape the spectrum of the sound signal. The bones of the middle ear couple the acoustic signal into the cochlea; these bones also implement a gain control mechanism that attenuates high-intensity sounds. This mechanism is under neural control.

Both mechanical and electrical processing occur in the cochlea. Sound energy is coupled into a mechanical traveling-wave structure, the basilar membrane, that 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.

Electrical signal processing in the cochlea occurs during the sensing of the mechanical motion of the basilar membrane. Inner hair cells occur at regular intervals along the basilar membrane, and convert basilar-membrane vibration into a graded electrical signal. 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. 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 may implement a stage of automatic gain control, exploiting the dynamics of synaptic-transmitter release. Auditory-nerve fibers are axons from spiral-ganglion neurons; these fibers present a neural representation of audition to the brain.

3 Silicon Cochleas

The first integrated circuit model of peripheral processing focused exclusively on the mechanical processing of the cochlea [10] [11]. This circuit is a one-dimensional physical model of the traveling-wave structure formed by the basilar membrane; the circuit also models the effect of the outer haircells on basilar membrane motion. A cascade of second-order sections with exponentially scaled time constants implement this model of basilar membrane motion. An analog, continuous-time circuit implementation of the model computes the pressure at selected discrete points along the basilar membrane in real time.

From a signal processing perspective, the silicon cochlea is a bank of low-pass filters; the response of each filter features a resonant peak and a sharp cutoff. The filter input and outputs are continuous-time voltage signals. The resonant frequencies of the output taps are exponentially spaced over several orders of magnitude. The delay through the silicon cochleas increases exponentially down the structure; the cascade structure enforces a monotonic increase of filter delay.

This cochlea circuit has been used in a complete model of the auditory periphery [3]. In this chip, each silicon cochlea output connects to a circuit that models the signal processing that occurs during inner hair cell transduction. The output from this circuit is a unidirectional current; this output connects to a circuit that models spike generation of the spiral ganglion cells. This circuit produces fixed-width, fixed-height spikes; these spikes are the the final output of the chip.

From a signal processing perspective, the inner hair cell circuit performs several operations. The circuit acts as a time differentiator; this action combines with the cochlea's resonant lowpass response to produce an asymmetric bandpass response. In addition, the circuit acts as a half-wave rectifier; the unidirectional current output produces zero current during half the

input waveform cycle. Finally, the circuit nonlinearly compresses the signal.

The spiral ganglion circuit integrates the input current signal until a threshold is reached. At this point a fixed-width, fixed-height spike is produced, and the integration variable is reset. Circuit noise adds a random offset to the integration process.

As a result, the position of each spike statistically encodes the shape of the input waveform. If many spiral ganglion circuits are fed the same input current signal, an average of the spiking outputs would reproduce the shape of the input signal. In addition, the mean rate of spikes over a period of time is a measure of the amplitude of the input current waveform.

Alternative silicon models of peripheral processing have been reported by several research groups [18] [8]. One published effort describes the implementation of a linear model of cochlear mechanics that extends easily to nonlinear modeling [18]. In this novel circuit, traveling waves propagate through a one-dimensional transmission line, implemented with a resistive network constrained by scaled impedances.

Two recent improvements to the inner hair cell and spiral ganglion cell circuits also have been reported. Both improvements aim to model the dynamic adaptation that occurs in the transduction process. One project is a detailed circuit model of current flow in an inner hair cell [8]; the other project is a simple model of dynamic adaptation during spike generation in spiral ganglion neurons [7].

4   Using the Models

Frequency-domain and time-domain issues are both important in systems that use a silicon model of the auditory periphery. Figure 1 shows a frequency-domain characterization of silicon auditory-nerve fiber. This data was taken by presenting sinusoidal signals at different frequencies and amplitudes as input, and noting the average firing rate from a spiking output [3]. Note that the filter bandwidth asymmetrically increases at higher amplitudes; the high-frequency filter cutoff remains fairly stationary, but the low-frequency filter cutoff shifts dramatically downward above 20 dB.

The wide bandwidth of silicon auditory nerve fibers at moderate intensities allows cycle-by-cycle timing information to be accurately preserved in the timing of individual spikes, as shown in Figure 2. This data was taken from a peripheral chip that modeled short-term adaptation [7]; this adaptation enhances changes
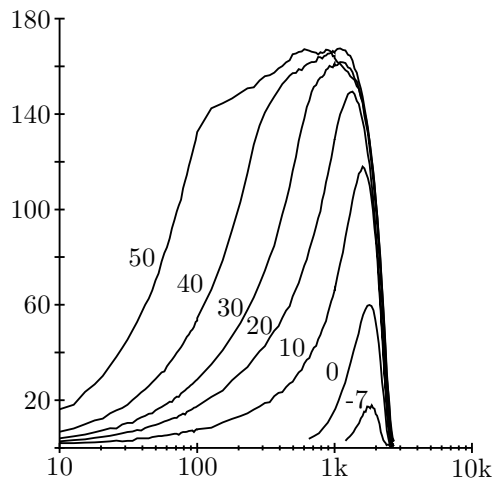

Figure 1: Frequency response of a silicon auditory nerve fiber. Y axis represents fiber activity (spikes/sec), X axis represents input frequency (Hz), labels denote input amplitude (dB).

in the envelope of the sound signal. These temporal properties are robust for high amplitude input signals.

The fundamental properties of silicon auditory nerve fibers, shown in Figures 1 and 2, are also seen in biological auditory nerve fibers. Unfortunately, these circuit models fall short of biological performance in several other functional aspects.

A major shortcoming of the silicon peripheral model is limited dynamic range; this nonlinear aspects of Figure 1 is a direct result of limited dynamic range. Biological auditory nerve fibers, however, show very similar nonlinearity; both silicon and biological fibers encode about 30dB of intensity with mean spike rate, although temporal encoding in both representations continues at high intensity. Why is the limited dynamic range of the silicon model a deficit?

In the biological system, the spiking spiral ganglion neuron is the source of the limited dynamic range; the mechanical motion of the basilar membrane has a much wider dynamic range. As a result, each inner hair cell connects to about ten spiral ganglion neurons. Although most of these ten neurons are sensitive at low sound intensities and saturate at moderate intensities, a few are insensitive at low intensities, and encode sound intensity through mean spike rate at medium sound intensities. In this way, the limited dynamic range of spiral ganglion neurons is overcome.

In the silicon system, however, the silicon cochlea circuit and the inner-hair-cell circuit are the sources of limited dynamic range; the silicon cochlea circuit saturates at moderate sound intensities, and the ca-
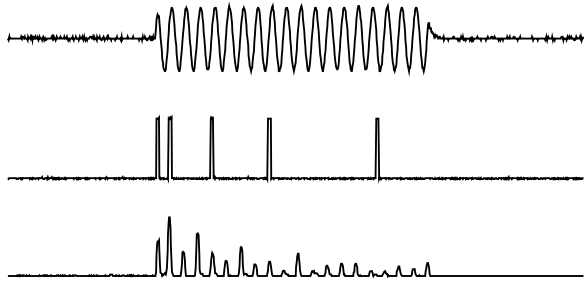
Figure 2: Temporal response of silicon auditory nerve fiber to a 1 kHz, 20 ms tone burst. Top trace shows tone burst input, middle trace shows a sample response from a fiber, bottom trace shows averaged output of 64 responses to tone bursts.

pability of inner-hair-cell circuit to sense small signals is also limited. These two constraints result in the 30dB dynamic range of the silicon periphery model. Expanding the dynamic range of these circuits is an important research issue.

Is the limited dynamic range of silicon cochleas a flaw that prevents their use as a component in larger auditory systems? Although the mean-rate encoding of sound intensity has a 30dB dynamic range, shown by the low-frequency cutoff in Figure 1, the high-frequency cutoff in Figure 1, as well as the temporal coding of Figure 2, are features with a dynamic range of 60dB. Biological models that primarily rely on these features of cochlear representation are well suited to analog VLSI implementation using present silicon cochleas.

Fortunately, many biological models implicitly assume a low dynamic range of mean-rate encoding in the entire auditory nerve. The importance of mean rate encoding in biological audition is a controversial issue in auditory neuroscience. Many theorists believe the temporal encoding shown in Figure 2 is of primary importance in auditory processing, assigning mean-rate encoding a secondary role. These theorists note that the mean firing rate of most auditory fibers are saturated for loud sounds, leaving a small number of fibers to encode the sound. This fact is counterintuitive with practical experience and psychophysical experiments: louder sounds are easier to locate and understand than softer sounds [1].

These theorists have developed models of auditory processing that do not require a large dynamic range of mean-rate encoding. These models have inspired most of the integrated circuits that use silicon cochleas as

components. As the dynamic range of silicon cochleas improves, silicon models of auditory theories that require mean-rate encoding with a large dynamic range will also become practical.

5   Correlation Models

Temporal correlation of spike patterns from auditory nerve fibers is a common theme in models of higher function in biological audition. This processing method does not require a large dynamic range of sound encoding, and is realizable with biological neurons; implementation of these algorithms typically requires only time-delay elements and spike-coincidence circuits.

General correlation algorithms often expand the one-dimensional cochlear representation into a two-dimensional representation. In the new representation, one axis represents cochlear place, whereas the other axis represents correlation time delay. This two-dimensional representation directly encodes an aspect of the auditory signal, by extracting information from the cycle-by-cycle time structure of the cochlear representation.

Computing the running autocorrelation function for the cochlear representation, for many linearly-spaced correlation time delays, produces a two-dimensional correlation representation that encodes periodicity pitch information [9]. Computing the running cross-correlation function of the left and right cochlear representations, for many linearly-spaced correlation time delays, produces a two-dimensional correlation representation that encodes auditory localization information [16]. The cross-correlation of two cochlear representations, without any explicit time delays, also produces a two-dimensional representation that encodes auditory localization information [17]; this representation uses the implicit time delay of traveling-wave propagation on the basilar membrane of each cochlea.

These representations easily map to analog VLSI implementation, using coincidence circuits and monostable delay circuits to implement the correlation primitives, and inhibitory circuits to process the final representation. All three representations have been successfully implemented in analog VLSI. A silicon model of pitch perception [4] generates a two-dimensional running autocorrelation function, and estimates pitch from this representation using a simple algorithm. A silicon model of azimuthal auditory localization [2] generates the Jeffress two-dimensional cross-correlation function, and estimates interaural

time delay from this representation using a simple algorithm. These single-chip systems each contain over 100,000 transistors.

These two systems estimate properties of auditory signals using simple algorithms on a single chip, and output only the estimate, not the two-dimensional correlation representation. In contrast, a different silicon model of binaural hearing [13] generates a two-dimensional cross-correlation function, using implicit cochlear delays, and outputs this two dimensional representation for display on a video monitor. Another silicon system generates and displays a two-dimensional auto-correlation function, using CCD technology [12].

Other auditory algorithms use temporal correlation to compute new one-dimensional representations of audition. Several algorithms use temporal correlation for spectral enhancement, transforming the nonlinear frequency representation of Figure 1 into an amplitude-independent spectral representation [1]. A spectral enhancement algorithm of this nature has been recently implemented in analog VLSI [6].

Acknowledgements

References

[1] S. Greenberg, "The ear as a speech analyzer," *J. Phonetics,* vol 16, pp. 139-149, 1988.

[2] J. P. Lazzaro, C. Mead, "Silicon models of auditory localization," *Neural Computation,* vol. 1, pp. 47–57, 1989.

[3] J. P. Lazzaro, C. Mead, "Circuit models of sensory transduction in the cochlea," in *Analog VLSI Implementations of Neural Networks,* Mead, Ismail, (eds.), Norwell, MA: Kluwer, pp. 85-101, 1989.

[4] J. P. Lazzaro, C. Mead, "Silicon models of pitch perception," *Proc. Natl. Acad. Sci. USA,* vol 86, pp. 9597–9601, 1989.

[5] J. Lazzaro, C. Mead, "Silicon models of auditory localization," in *An Introduction to Neural and Electronic Networks,* Zornetzer, Davis, Lau (eds.), New York: Academic Press, pp. 158–174, 1990.

[6] J. P. Lazzaro, " A silicon model of an auditory neural representation of spectral shape," *IEEE Journal Solid State Circuits,* 26: 772–777, 1991.

[7] J. P. Lazzaro, "Temporal adaptation in a silicon auditory nerve," in *Advances in Neural Information Processing Systems 4,* San Mateo, CA: Morgan Kaufmann Publishers, 1992.

[8] W. Liu, A. Andreou, M. Goldstein, "Analog VLSI implementation of an auditory periphery model," in *Proceedings 25th Annual Conference on Information Sciences and Systems,* Baltimore, MD, 1991.

[9] R. F. Lyon, "Computational models of neural auditory processing," in *IEEE ICASSP,* 1984.

[10] R. F. Lyon, C. Mead, "An analog electronic cochlea," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 36, pp. 1119–1134, 1988.

[11] R. F. Lyon, "Analog Implementations of Auditory Models," *DARPA Workshop on Speech and Natural Language,* Morgan Kaufmann Publishers, San Mateo CA, 1991.

[12] R. F. Lyon, "CCD correlators for auditory models," in *IEEE Asilomar Conference on Signals, Systems, and Computers*, 1991.

[13] C. A. Mead, X. Arreguit, J. P. Lazzaro, "Analog VLSI models of binaural hearing," *IEEE Journal of Neural Networks,* **2**: 230–236, 1991.

[14] C. Mead, *Analog VLSI and Neural Systems.* Reading, MA: Addison-Wesley, 1989.

[15] Y. Muthusamy, R. A. Cole, M. Slaney, "Speaker-independent vowel recognition:spectrograms versus cochleagrams," in *IEEE ICASSP*, pp. 533-536, 1990.

[16] J. O. Pickles, *An introduction to the physiology of hearing,* San Diego, CA: Academic Press, 1988.

[17] S. A. Shamma, N. Shen, P. Gopalaswarmy, " Stereausis: Binaural processing without neural delays," *J Acoust. Soc. Am,* **86:** 989–1006, 1989.

[18] L. Watts, R. Lyon, C. Mead, "A bidirectional analog VLSI cochlear model," *Advanced Research in VLSI, Proceedings of the 1991 Santa Cruz Conference,* C. Sequin (ed.), Cambridge, MA: MIT Press, pp. 153–163, 1991.