# 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Kiang, 1980; Kim, 1984; Dallos, 1985) and the auditory brainstem nuclei (Carr and Konishi, 1988; Fujita and Konishi, unpublished).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

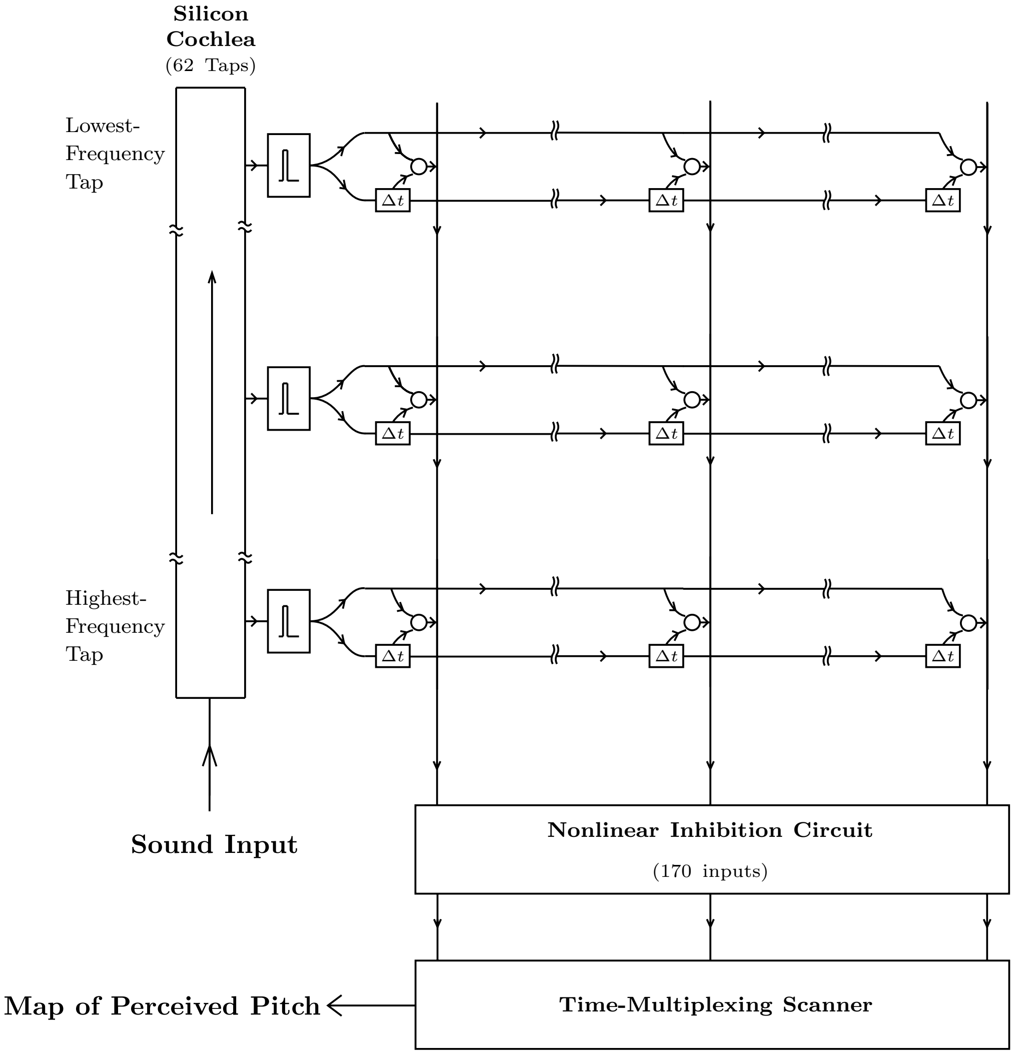
Silicon
Cochlea
(62 Taps)
Lowest-
Frequency
Tap
Highest-
Frequency
Tap
Δt
Sound Input
Nonlinear Inhibition Circuit
(170 inputs)
Time-Multiplexing Scanner
Map of Perceived Pitch

**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 "$\Delta t$," 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 $(f,\ \ 2f,\ \ 3f,\ \ 4f, \ldots,)$. 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.

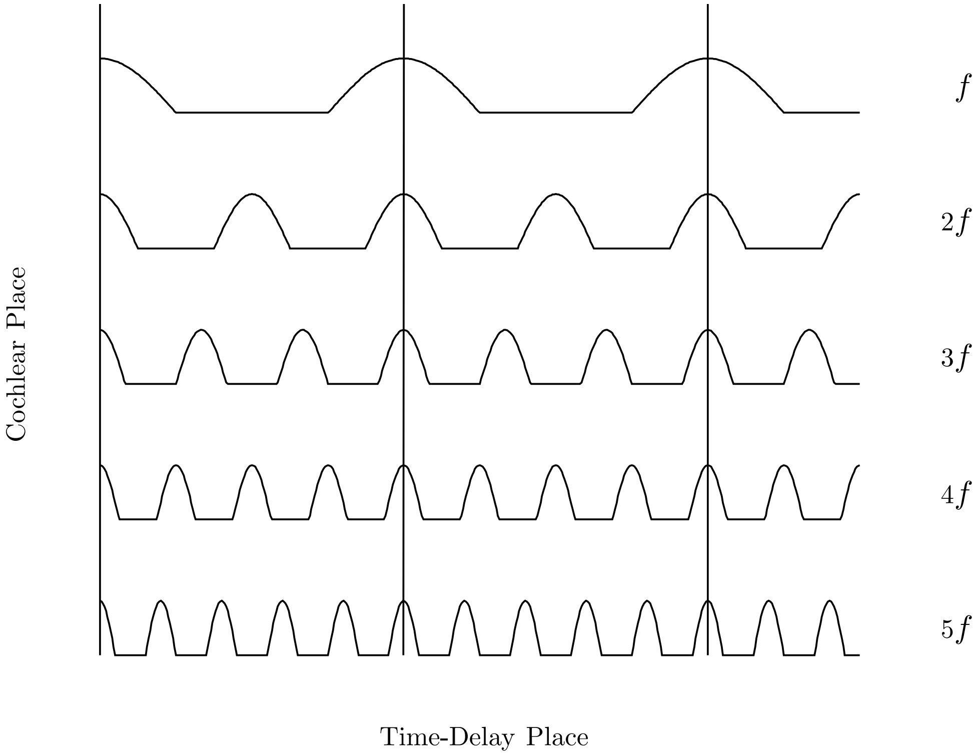

**Figure 5.2.** Analog representation of the signals in the delay lines across the chip, in response to a harmonic signal. Cochlear processing is idealized (fully resolved harmonic components, perfect half-wave rectification, no temporal smoothing). Peaks of activity in the horizontal direction coincide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

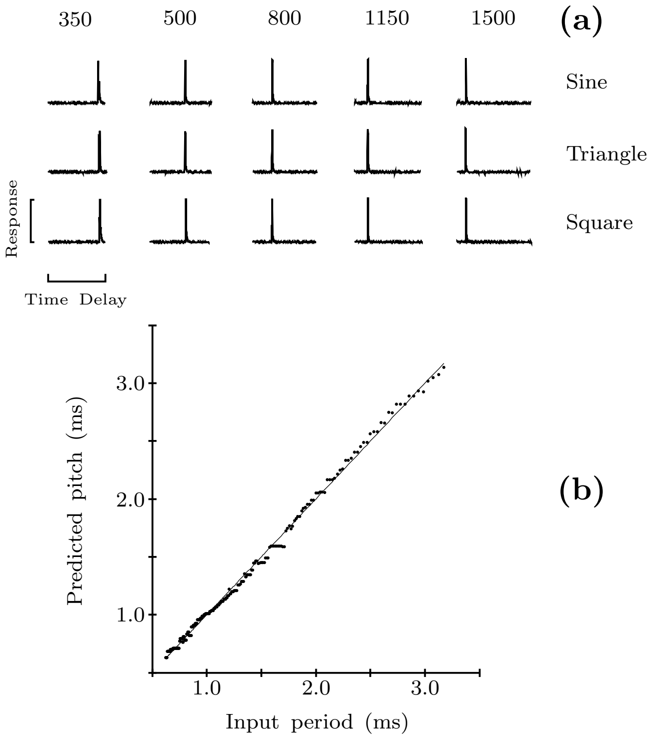

**Figure 5.3. a.** Maps of perceived pitch period from the chip, in response to sine, triangle, and square waves. Column numbers denote frequency in Hz. **b.** Plot showing map position of the neuron with maximum signal energy, for square waves of different frequency; ordinate axis is calibrated from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

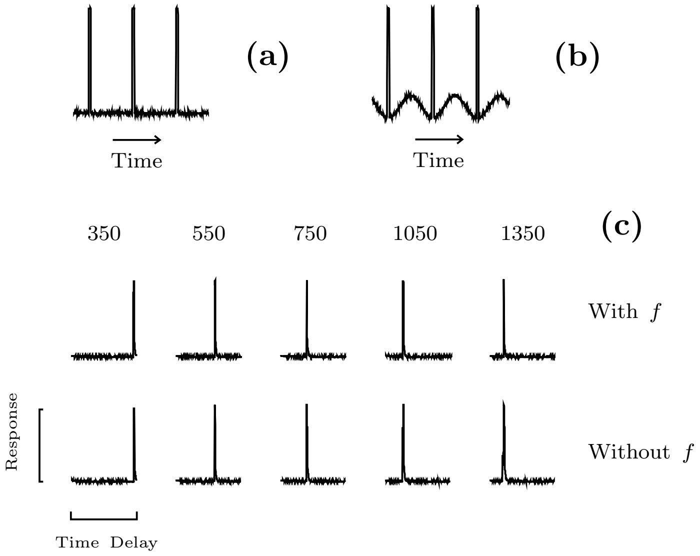

**Figure 5.4.** **a.** Narrow-pulse sound stimulus. **b.** Narrow-pulse sound stimulus, with canceled fundamental frequency (Schouten, 1940). **c.** Maps of perceived pitch period from the chip, in response to stimuli shown in parts a and b, at various frequencies; chip responds identically. Column numbers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

in Figure 5.5(b). For remarkably small $\Delta$, 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 $\phi$, 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

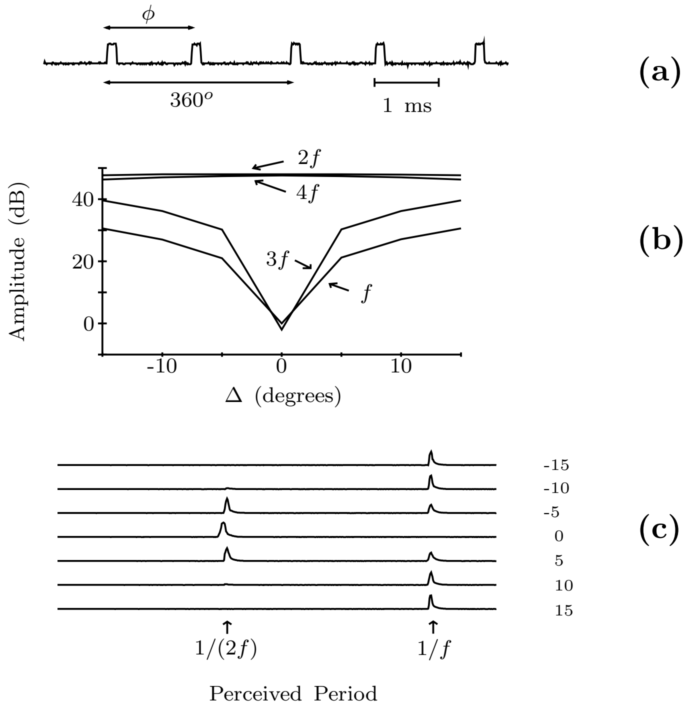

**Figure 5.5.** **a.** Sound stimulus created by the addition of two synchronized narrow-pulse waveforms, of identical frequency $f$ and phase difference $\phi$ (Schouten, 1940). When $\phi = 180^\circ$, stimulus frequency is $2f$; when $\phi = 180^\circ \pm \Delta$, stimulus frequency is $f$. **b.**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

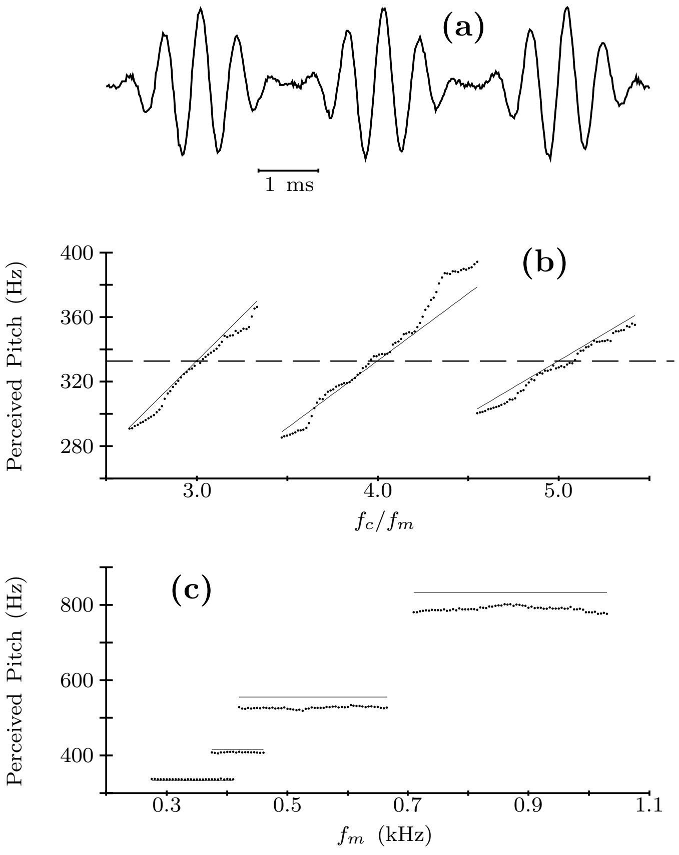

**Figure 5.6.** **a.** Amplitude-modulated sinusoid sound stimulus (de Boer, 1956). **b.** Plot showing center of energy of the chip map, in response to stimulus shown in part a, while the carrier frequency, $f_c$, is varied. Dotted line shows frequency of fixed modulation frequency $f_m$ = 333 Hz. Dots 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

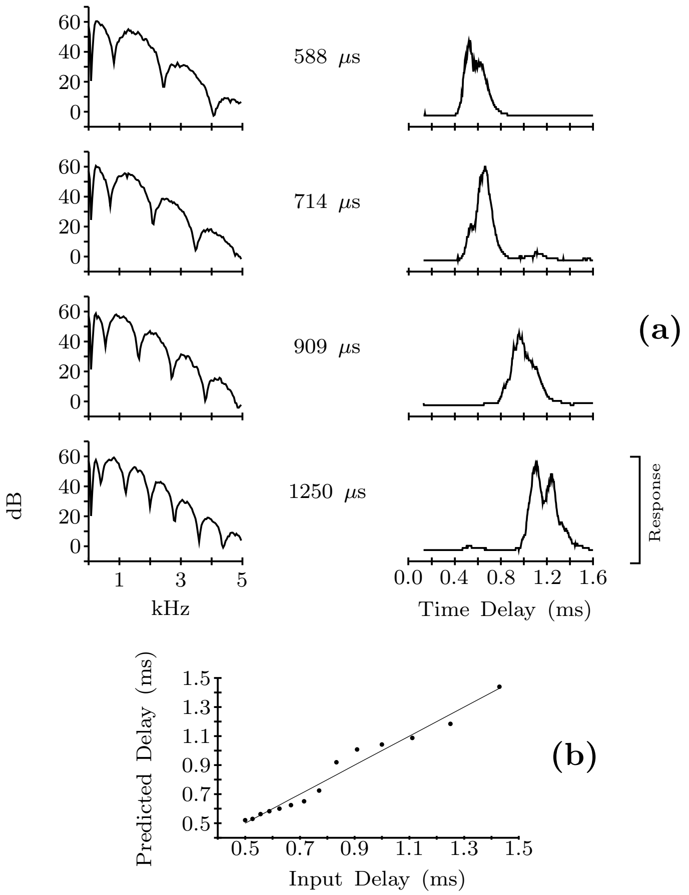

**Figure 5.7. a.** Plots showing maps of perceived pitch period from the chip, in response to a bandpass-filtered sum of two time-delayed, correlated noise signals (right). Filtering was kept constant for all plots. Centered numbers indicate time delay between noise signals. Left plots show the spectral 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.