AER EAR: A Matched Silicon Cochlea Pair With Address Event Representation Interface

Vincent Chan, *Student Member, IEEE*, Shih-Chii Liu, *Member, IEEE*, and André van Schaik, *Senior Member, IEEE*

***Abstract*—In this paper, we present an analog integrated circuit containing a matched pair of silicon cochleae and an address event interface. Each section of the cochlea, modeled by a second-order low-pass filter, is followed by a simplified inner hair cell circuit and a spiking neuron circuit. When the neuron spikes, an address event is generated on the asynchronous data bus. We present the results of the chip characterization and the results of an interaural time difference based sound localization experiment using the address event representation (AER) EAR. The chip was fabricated in a 3-metal 2-poly 0.5-μm CMOS process.**

***Index Terms*—Analog integrated circuits, neuromorphic engineering, silicon cochlea, sound localization.**

I. Introduction

FOR multi-chip neuromorphic systems, the address event representation (AER) interface has become the standard interface protocol in recent years. The AER approach allows us to model biological systems using discrete level (spikes) and continuous/analog time events to convey information, similar to the pulse code neural communication and processing systems found in living organisms. This representation is ideal for communicating sparse events from many sources using a narrow channel [1], [2].

AER allows multiple devices to share a common data bus by using a hand-shaking protocol that arbitrates between transmitters to determine which transmitter can access the bus. The next transmission is not permitted unless an acknowledgement is returned from the receiver. This method allows multiple transmitters to communicate with multiple receivers in a pseudo-parallel fashion. Since the events are generated asynchronously, random ordering of transmissions occurs, producing Poisson distributed event streams which, statistically, preserve the timing between events [3]. Furthermore, the event stream models the timing relationships of biological synaptic transmission very well.

In addition to allowing multiple sensors to communicate to multiple processors using a common bus, the AER protocol can be used to perform computation through the manipulation of the statistics and routing of events from the input stream. For example, spatial filtering on a receiver chip can be realised if each pixel in the transmitter array transmits a projection field to the receiver chip.

To date, the AER protocol has been used almost exclusively in vision chips (e.g., [3]–[10]) and some neural processors (e.g., [11]–[15]). The only silicon cochleae with an address event type representation that the authors are aware of are [16], [17] and [18], but these implementations use a small number of cochlea sections (9, 15, and 8 sections, respectively), and a nonstandard implementation of the AER protocol. Furthermore, in [17], only the zero-crossings of the output signal of each cochlear section were preserved.

In this paper, we present an all-purpose silicon cochlea chip with standard AER interface, designed for general neuromorphic applications including sound localization and visual-audio sensor fusion. We describe the building blocks of the system in Section II and then present measurements and mismatch analysis from the fabricated design in Section III. This data extends previous characterization results found in [19]. Section IV shows the results of two sound localization experiments with this chip conducted in a more controlled environment than [20]. This section is followed by a conclusion.

II. Implementation

A. Cochlea

The silicon cochlea used is identical to the one we have presented in [21] and is shown in Fig. 1(a). It is an improved version of the one presented in [22], which had already proven its use in various neuromorphic sound processing systems [23], [24]. First proposed by Lyon and Mead [25], the basic building block for the filters in this cochlear model is a second-order low-pass filter section show in Fig. 1(b). It is built with transconductance amplifiers operating in weak inversion. For input voltages smaller than about 60 $\mathrm{mV_{pp}}$, the amplifier can be approximated as a linear transconductor

$$I_{\mathrm{out}} = g_m(V_{\mathrm{in}+} - V_{\mathrm{in}-}) \tag{1}$$

where the transconductance g_m is given by

$$g_m = \frac{I_0}{2nU_T}. \tag{2}$$

I_0 is the bias current, n is the slope factor, and the thermal voltage $U_T = kT/q = 25.6\mathrm{mV}$ at room temperature. The bias current I_{cf} of amplifiers A1 and A2, controls the cutoff frequency of the filter, while amplifier A3 is biased by I_q. By changing the voltages V_e and V_q, one can control the ratio of I_q to I_{cf}, which sets the quality factor, Q, of the second-order

Manuscript received February 1, 2006, revised October 2, 2006. This work was supported in part by the Australian Research Council, and the Office of Naval Research under NICOP Grant N00014-05-1-0916. This paper was recommended by Guest Editor M. E. Zaghloul.

V. Chan and A. van Schaik are with the School of Electrical and Information Engineering at the University of Sydney, Sydney, NSW 2006, Australia (e-mail: andre@ee.usyd.edu.au).

S.-C. Liu is with the Institute of Neuroinformatics at the ETH/University of Zürich, Zürich, CH-8057, Switzerland.

Digital Object Identifier 10.1109/TCSI.2006.887977

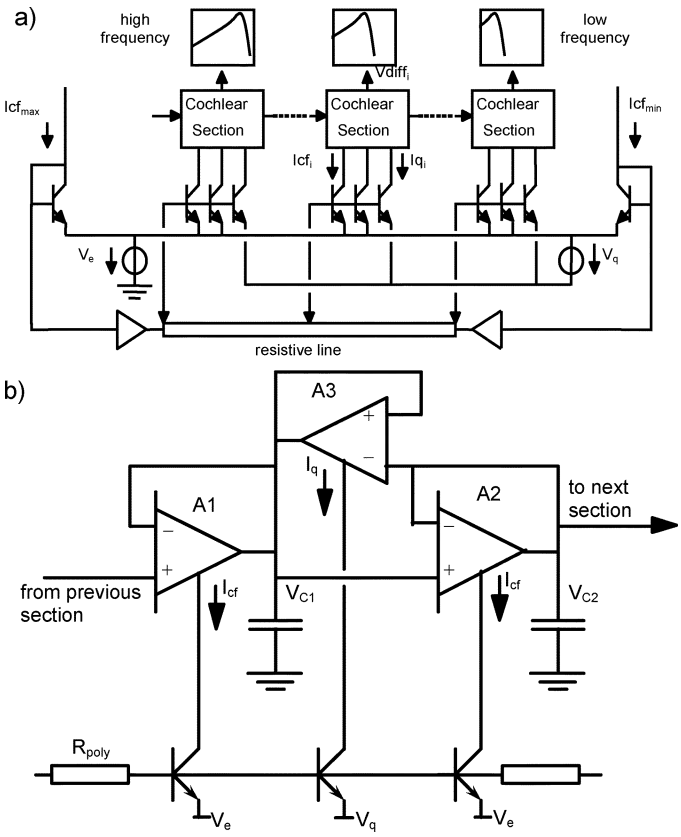

Fig. 1. (a) Silicon cochlea. (b) Single cochlear section. By biasing the transistors with a resistive line, we obtain the exponentially decreasing cutoff frequencies across the cochlea. Currents $I_{\mathrm{cf}\max}$ and $I_{\mathrm{cf}\min}$ set the cutoff frequencies of the first and the last section, respectively.

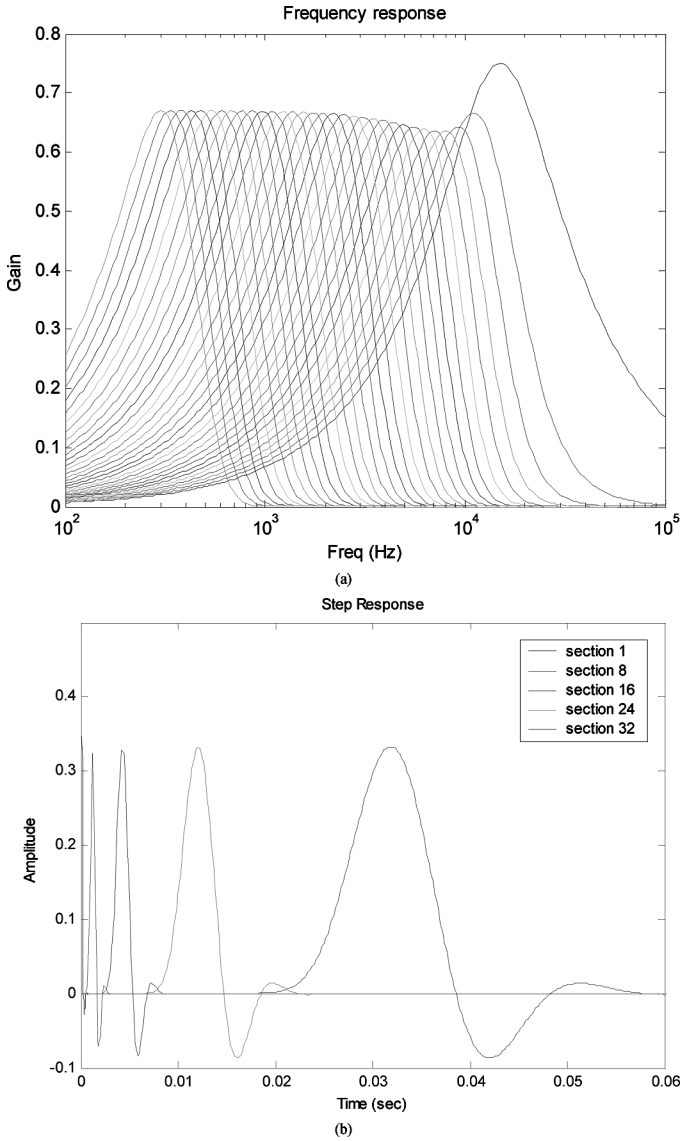

Fig. 2. (a) Theoretical frequency response . (b) Theoretical step response. $Q = 0.75$ in both cases.

section, with Q becoming infinity when $I_q = 2I_{\mathrm{cf}}$. If all three amplifiers in the circuit are identical, this second-order section may be stable for small signals, but will exhibit large signal instability due to slew-rate limitations [26]. This instability can be solved by using a transconductance amplifier with a wider linear input range in the forward path [26], allowing larger input signals of up to about 140 mVpp to be used.

Our silicon cochlea is implemented by cascading 32 of these second-order low-pass sections with exponentially decreasing cutoff frequencies. The exponential decrease is obtained by creating the bias currents of the second-order section with CMOS compatible lateral bipolar transistors (CLBTs), as proposed in [22]. A bandpass filtered output is obtained from each section by taking $V_{C1} - V_{C2}$ as the differential output signal.

Two identical copies of the silicon cochlea are laid out side by side on the chip. In addition, we placed the CLBTs that bias the corresponding sections of both cochleae in a single well, that is, they share a common base. This placement was done to ensure optimal matching of the bias currents for the two cochleae.

Fig. 2(a) shows the theoretical frequency response of the output signal $V_{C1} - V_{C2}$ for all 32 sections, for a quality factor Q, set to 0.75. All sections have a -20-dB/dec slope for frequencies below the resonant frequency, and a much steeper roll-off for frequencies above the resonant frequency, except the first few sections. The steep roll-off is a result of cascading, where the frequency components above a section's cutoff frequency have already been reduced in gain by the previous sections, resulting in higher frequency selectivity compared to a second-order-section in isolation. The filters are evenly spaced on a log-frequency scale due to the exponentially decreasing cutoff frequency across the cochlea and each filter has a constant relative bandwidth.

Fig. 2(b) shows the theoretical step response of the channels for the same Q-value. The delays are inversely proportional to the resonant frequency of each section, thus the delays increase exponentially.

B. Inner Hair Cell

In the biological cochlea, the inner hair cells transduce vibration in the cochlea into a neural signal. This function is modeled by a novel inner hair cell circuit shown in Fig. 3. A transconductance amplifier transforms the differential cochlear output into a single ended current, to which a dc offset may be added using $V_{I_{\mathrm{off}}}$. The gain of the conversion can be set with V_{gain}. A current mirror rectifies the current signal before passing it through

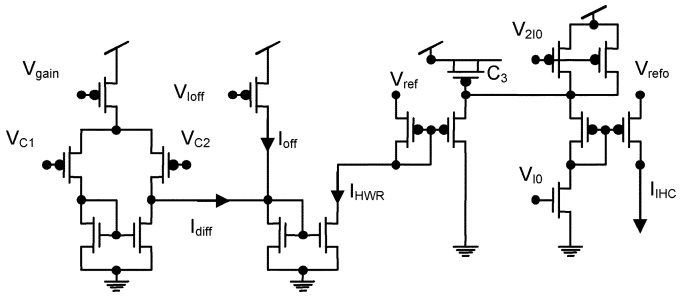

Fig. 3. The inner hair cell circuit.

a low-pass filter. This half wave-rectified current is as a first approximation given by

$$I_{\mathrm{HWR}} = \max(0, g_m(V_{C1} - V_{C2}) + I_{\mathrm{off}}) \tag{3}$$

where g_m is controlled by V_{gain} and I_{off} by $V_{I_{\mathrm{off}}}$.

The first-order low-pass filter is a log-domain circuit. Its transfer function in the Laplace domain is given by

$$\frac{I_{\mathrm{IHC}}}{I_{\mathrm{HWR}}} = \frac{G}{\mathrm{s}\tau + 1} \tag{4}$$

where G is given by

$$G = e^{(V_{\mathrm{refo}} - V_{\mathrm{ref}})/U_T} \tag{5}$$

and τ is given by

$$\tau = \frac{C_3 U_T}{I_0}. \tag{6}$$

The cutoff frequency of the low-pass filter can be controlled with V_{I0} and V_{2I0}, which generate bias currents I_0 and $2I_0$, respectively. C3, implemented as a pMOS capacitor, should have been implemented with an nMOS capacitor because the voltage is quite close to V_{dd}. Nonetheless, the circuit still operates correctly as the voltage swing is small and the pMOS capacitor acts almost linearly. The cutoff was set around 1 kHz as in the real inner hair cell, modeling the reduction in phase-locking observed on real auditory nerves at frequencies greater than 1 kHz. The two control signals V_{ref} and V_{ref0} are slightly below V_{dd} to allow the two pMOS transistors providing $2I_0$ to operate in saturation. Any voltage difference between V_{ref} and V_{ref0} will show up as a current gain depending exponentially on this difference as given in (5). In the results shown in this paper, both voltages were equal to 4.5 V, resulting in a gain of 1.

The biological inner hair cell exhibits adaptation to an ongoing stimulus, therefore it responds more strongly to the onset of stimulation than the sustained part of the stimulus and its response is suppressed temporarily after the offset of stimulation. This adaptation has been modeled in [27] but it was considered too complex and too large for inclusion on the current chip. However, we intend to include this in future versions.

C. Integrate-and-Fire Neuron for AER

The output current of the IHC (I_{IHC}) is passed through a current mirror (not shown) which is cascaded with V_{cas}, shown in Fig. 4. This point forms the boundary between the analog front-end and the digital AER circuits. The analog and digital circuits run on separate power supplies to reduce digital noise on the analog signals. The wells of the pMOS transistors in Fig. 4 are all tied to the digital V_{dd}.

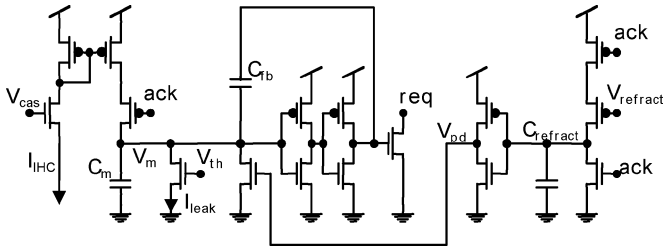

Fig. 4. AER spiking neuron.

When the acknowledge signal (ack) is low, I_{IHC} is integrated onto the membrane capacitor C_m of the neuron circuit, which has a leakage current I_{leak} controlled by V_{th}. The spike is generated by two inverters in series, with positive feedback through C_{fb}. When a spike is generated, the request line (req) is pulled low. A high pulse in acknowledgement (ack) resets the neuron after which it will enter a refractory period with a duration set by V_{refract}.

The AER protocol is described in [1]. When a neuron makes a request (by pulling the req signal low), an on-chip arbiter arbitrates between all neurons making a request, and sends off each neuron's address in sequence through two external handshaking signals. Once a neuron's address has been communicated off-chip, the arbiter sends an acknowledge to the neuron via the ack signal, which in turn resets the neuron. The AER circuitry used on our chip is adapted from [1] and has already been used on a number of chips.

D. AER EAR

The complete system consists of (A) two matched, 32-section silicon cochleae, followed by (B) simplified inner hair cell circuits and (C) spiking neurons and AER interface circuitry to generate and communicate auditory nerve spikes. It is fabricated in a 3-metal, 2-poly, 0.5-μm CMOS process with a die size of 2.7 mm by 2 mm. Fig. 5 shows the microphotograph of the chip.

III. Measurements

The cutoff frequencies of our second-order sections can be tuned to a wide range of frequencies, from 50 Hz to 50 kHz or higher. For the experiments presented here, we adjusted the bias voltages so that the first section is most responsive to a 15-kHz input frequency and the last (32nd) section is most responsive to 300 Hz. The quality factor was set to be close to 0.75 as shown in the simulations in Fig. 2. The offset current, (I_{off} in Fig. 3) as well as the leakage current at the neuron (I_{leak} in Fig. 4) were set to zero. Furthermore, V_{ref} and V_{ref0} (Fig. 3) were both set to 4.5 V, resulting in a low-pass filter gain of 1. These settings were chosen to simplify the interpretation of the measurement results.

In biology, the spike rate on the auditory nerve is limited to about 100 spikes per second, but a number of nerve fibres originate from the same IHC and many IHCs code for a similar

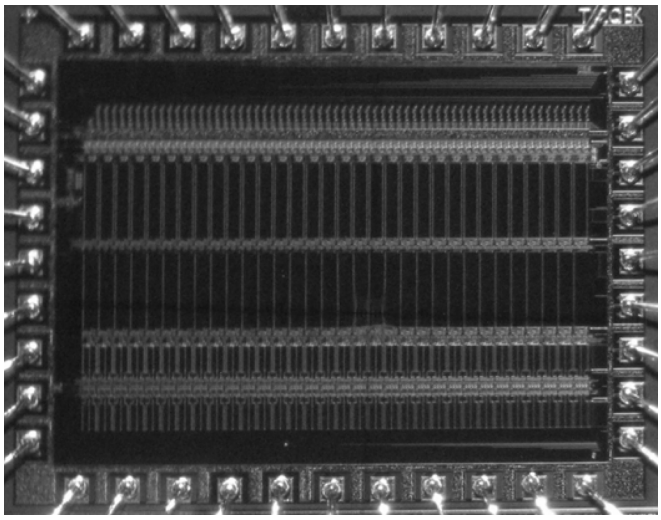

Fig. 5. Microphotograph of the AER cochlea.

frequency range. In our case we only have 32 filters for more than two octaves of input frequency range and we have implemented only one AER neuron per IHC. Therefore, to obtain enough spikes from each filter, we have given the IHC circuit a high gain by lowering V_{gain} (Fig. 3) and decreasing V_{refract} (Fig. 4) so that the maximum spike rate of the AER neuron is more than 10000 spikes per second. The cutoff frequency of the IHC low-pass filter is kept at 1 kHz, so that phase locking will be reduced only for input signals above 1 kHz.

A. Frequency Response

We have measured the frequency response for each filter section by measuring the differential voltage $V_{C1} - V_{C2}$ via an on-chip scanner. The sections are numbered from 1 (highest frequency) to 32 (lowest frequency). A 150-mV_{pp} sine wave with frequency varying from 200 Hz to 30 kHz was applied to the inputs of both cochleae and the peak-to-peak amplitude of the differential output signal $V_{C1} - V_{C2}$ was measured.

The gain, expressed as the ratio between input and output peak-to-peak amplitude, is shown in Fig. 6 as a function of frequency. The plot shows that the filters have approximately a constant relative bandwidth and maximum gain as predicted by theory. The gain across the 32 sections varies little between sections, but there is a consistent difference in gain between sections in the left cochlea and those in the right cochlea. On all the chips that we have measured, we found that the right cochlea systematically has a gain 3 to 4 dB higher than the left cochlea. We suspect that this is due to the large difference in the length of wiring from the second-order section to the scanner between the two cochleae. This difference will be corrected in future versions of the AER EAR .

The voltages V_{C1} and V_{C2} are buffered off-chip and read by a data acquisition board. The lack of on-chip buffering allows direct connections to the capacitor voltages in each second-order section, which we will make use of in Section III-C, but at the same time, it causes the node capacitance to significantly increase when the voltages are being read. This will artificially decrease the cutoff frequency of the section under test. A side effect of this is that the high frequency cutoff slope in Fig. 6 is not as steep as it should be, since the bandpass filters do not benefit from cascading until frequencies well above their reduced cutoff frequencies. This, however, is a measurement artefact, since in normal operation none of the cochlear sections are connected to the output pins through the scanner.

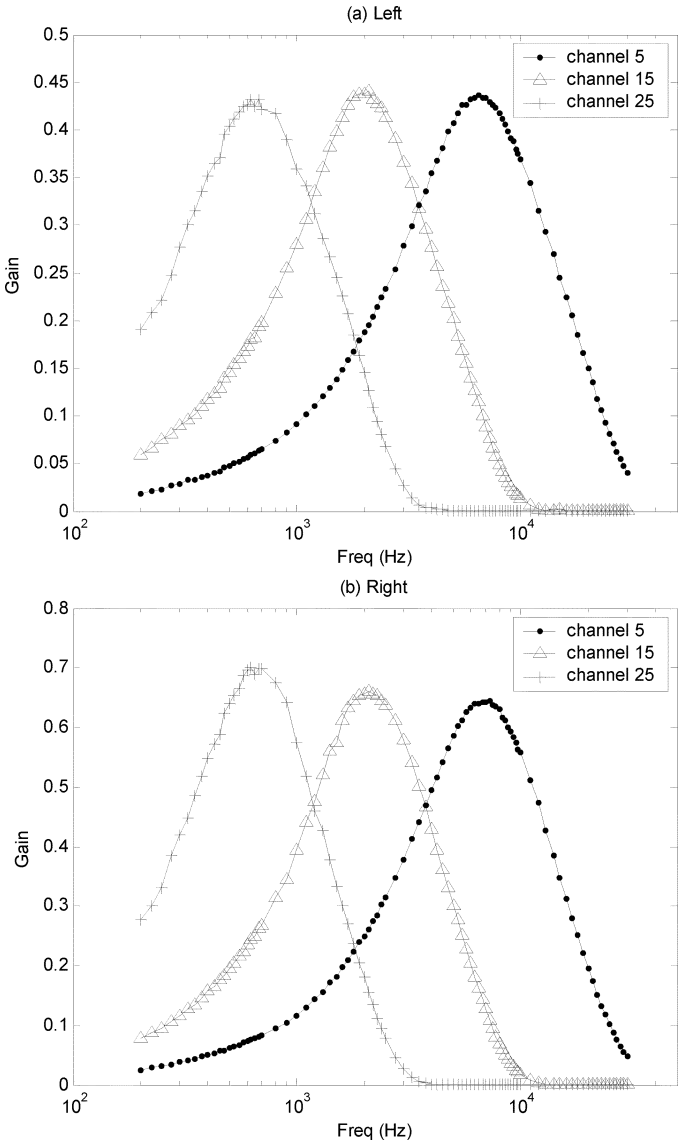

Fig. 6. Measured frequency response of $V_{C1} - V_{C2}$ at sections 5, 15, and 25 of (a) the left cochlea and (b) the right cochlea. The gain is very consistent within each cochlea, but the right cochlea has a gain systematically higher than that of the left cochlea.

B. Spike Rates and Cochleagram

At the inner hair cell (IHC), the differential signal ($V_{C1} - V_{C2}$) is converted to a current, half-wave rectified, and filtered with a low-pass filter with a cutoff frequency of about 1 kHz. Therefore, at input frequencies above a few kilohertz, the output is an almost constant current that is proportional to the magnitude of $V_{C1} - V_{C2}$. This current drives an integrate-and-fire neuron, producing spikes on the AER bus. The firing rates of the neurons in the same channels as in Fig. 6 are shown as a function

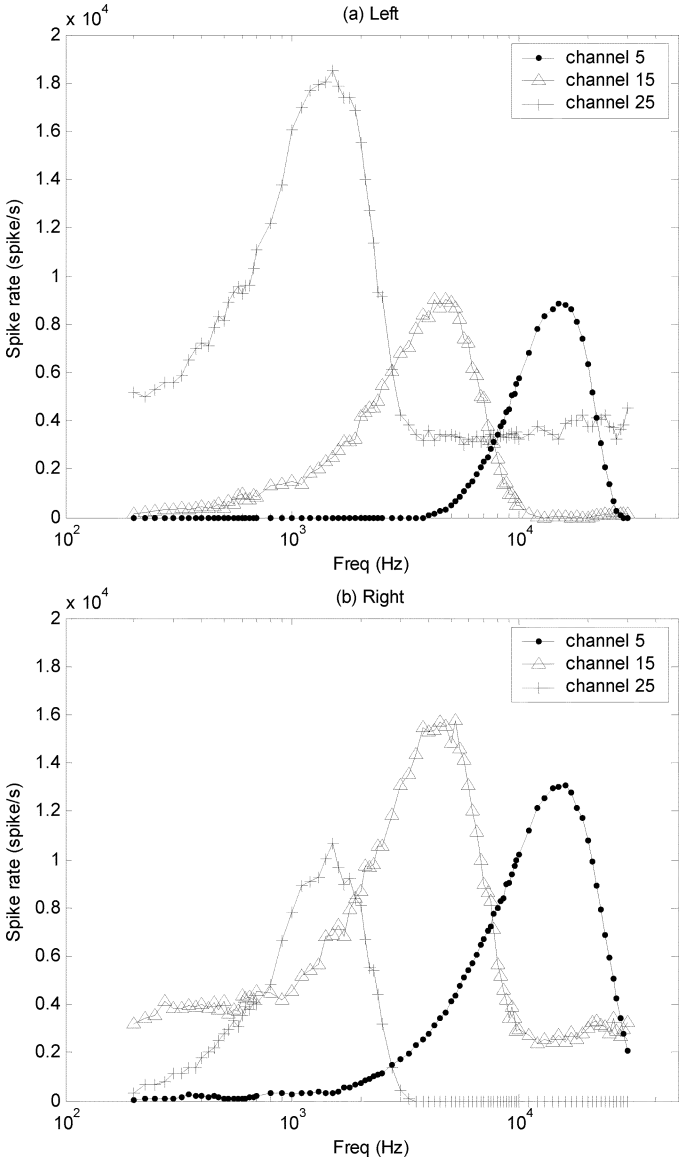

Fig. 7. Spike rates of neurons in selected channels. Offset and gain mismatch are evident.

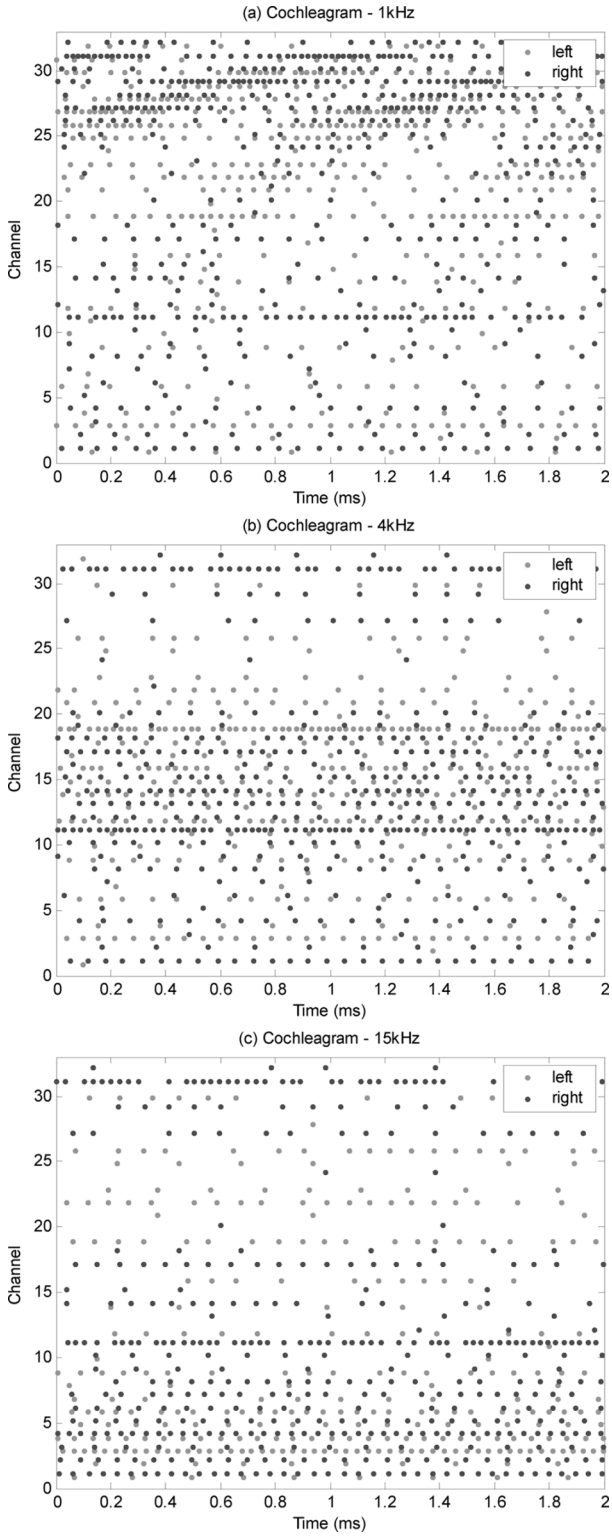

Fig. 8. Cochleagrams for different sinusoidal inputs. (a) Input is 1 kHz. (b) Input is 4 kHz. (c) Input is 15 kHz. Phase locking can be seen for the 1-kHz input, but it disappears at 4 kHz due to IHC low-pass filtering with $f_{\mathrm{cutoff}} = 1$ kHz.

of frequency in Fig. 7, demonstrating the frequency selectivity of the cochlea.

It can be seen that there are variations in the spike rate offset, with some neurons spiking even when the input signal is well outside the pass-band. In fact, these neurons will spike in the absence of any input. There are also some variations in gain between the output of the cochlea and the spike rate of the corresponding neuron, but the overall frequency selectivity is preserved. The causes of this mismatch will be further investigated in Section III-C. Comparing the spike rates in Fig. 7 with the frequency responses at Fig. 6, we see that they peak at different frequencies for the same section in both cochleae and that the high frequency slopes are steeper in Fig. 7. This is consistent with the lowering of the cutoff frequency when measuring $V_{C1} - V_{C2}$ directly.

Fig. 8 shows a cochleagram, which is a plot of the spikes as a function of channel number versus time, for both the left (green) and the right (red) cochlea for three different input frequencies. Again, we see that some neurons fire significantly more than others. The fact that the right cochlea systematically has a 3 dB higher gain than the left cochlea also shows up in the average spike rates in Fig. 8.

In Fig. 8(a), we can see that phase locking of the spikes to the input signal is preserved for a 1-kHz input signal, as the spike density varies with a 1-kHz rate. This phase locking is due to

the 1-kHz cutoff frequency of the low-pass filter in the IHC and is similar to phase locking on the biological auditory nerve. For a 4-kHz and a 15-kHz input, this phase locking is absent.

C. Analysis of Mismatch – Offset and Gain Error

As mentioned in the previous section, there are two types of error that contribute to variations in the spike rate: the offset error and the gain error. In this section, we will analyse their causes. We can rewrite (3) to take into account the effect of mismatch in the circuits

$$I_{\mathrm{HWR}} = k_{\mathrm{IHC}} \max(0, g_m(V_{\mathrm{ac}} + V_{\mathrm{dc}} + V_{\mathrm{os}}) + I_{\mathrm{off}}) \quad (7)$$

where V_{ac} and V_{dc} are the ac and dc components of $V_{C1} - V_{C2}$, respectively, V_{os} is the input dc offset at the transconductance amplifier in the IHC, and k_{IHC} combines the gains of all the current mirrors in the IHC circuit (including those in the log domain filter).

For simplicity, we may ignore the effect of low-pass filtering at the IHC since we are only interested in the mismatch in the *average* spike rate, therefore $I_{\mathrm{IHC}} = I_{\mathrm{HWR}}$. Assuming a constant I_{IHC}, this current charges the membrane capacitor in the neuron linearly until the spiking threshold is reached, at which point the neuron fires. Assuming that the refractory period is very small compared to the inter-spike interval, the average spike rate is

$$R_{\mathrm{spk}} = \frac{I_{\mathrm{IHC}}}{C_m V_{\mathrm{thr}}} \quad (8)$$

where C_m is the membrane capacitance and V_{thr} is the spiking threshold of the neuron. The leakage current I_{leak} (Fig. 4) is not included in the equation as it is set to zero in all our measurements.

V_{dc} is a result of the combined dc offset at amplifiers A1 and A2 in the cochlear section Fig. 1(b). Like V_{os}, it is mainly due to threshold voltage mismatch of the input transistors of the transconductance amplifiers and can take either a positive or a negative value. Since I_{off} is normally turned off, the offset at each section is proportional to the sum $V_{\mathrm{dc}} + V_{\mathrm{os}}$ and the neuron will fire spontaneously if the sum is positive. If the sum is negative, the neuron will not fire until V_{ac} exceeds $|V_{\mathrm{dc}} + V_{\mathrm{os}}|$.

A measure of the offset in spike rates is shown in Fig. 9. For these measurements the inputs of both cochleae were shorted to an ac ground. These measurements show only sections with positive offsets, since those are the only sections that will generate spikes in the absence of an input signal. To determine the negative offsets of the other sections, we injected a bias current I_{off} (controlled via $V_{I_{\mathrm{off}}}$) into each section so that every channel spikes spontaneously. Since there will be unavoidable mismatch between the copies of I_{off} injected at each section, this will introduce another source of mismatch, but this additional mismatch is expected to be much smaller than the offsets we are trying to measure. The spike rates at all the channels are again measured and a constant rate R_{off} is subtracted to compensate for the additional spikes generated due to I_{off}. R_{off} is determined from the average increase in spike rates due to I_{off} among channels with positive offsets (shown in Fig. 10). The result, shown in Fig. 11(a), shows both positive and negative

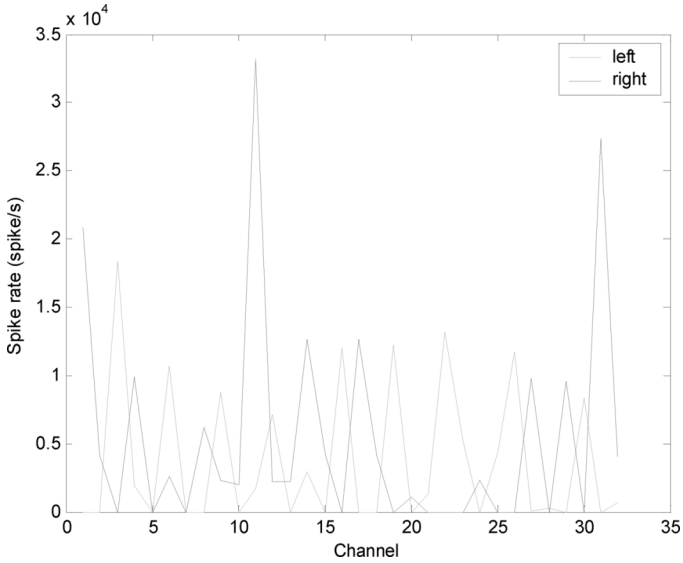

Fig. 9. Spontaneous spike rates at each channel for both the left and the right cochlea.

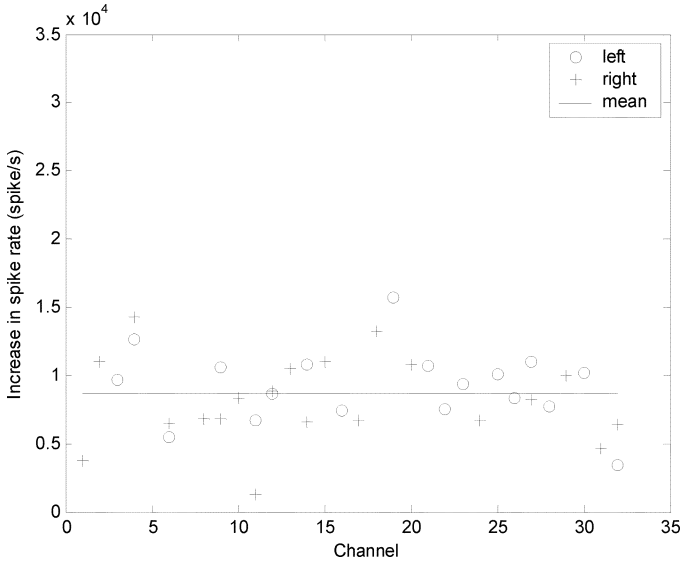

Fig. 10. Measured increase in spike rate due to the injected current I_{off}. This can only be measured for channels with positive offset. The mean, R_{off}, allows us to determine the negative offsets in Fig. 11.

offset values and these represent the combined offsets due to the cochlea (V_{dc}) and the IHC (V_{os} and I_{off}). Fig. 10, as well as the similarity of the positive offsets in both Fig. 9 and Fig. 11(a), confirm that the additional offset from injecting I_{off} is indeed smaller than the offset we are trying to observe.

To estimate the relative contributions of V_{dc} and V_{os} to the total offset, we measured the spontaneous spike rates when the input to the IHC circuit is shorted, thus removing V_{dc} and leaving us with only V_{os}. This is done by measuring one channel at a time through the scanner and shorting V_{C1} to V_{C2} with a wire off-chip. We use the same I_{off} and R_{off} as before to obtain the offsets in Fig. 11(b). By comparing Fig. 11(a) and (b), it can be seen that the spike rate offsets have been reduced. In fact, the variance in (a) is reduced by about 60% in (b), suggesting that V_{dc} (the dc offset at $V_{C1} - V_{C2}$) contributes more than half of the offset error of the entire system. Fig. 11(c) shows

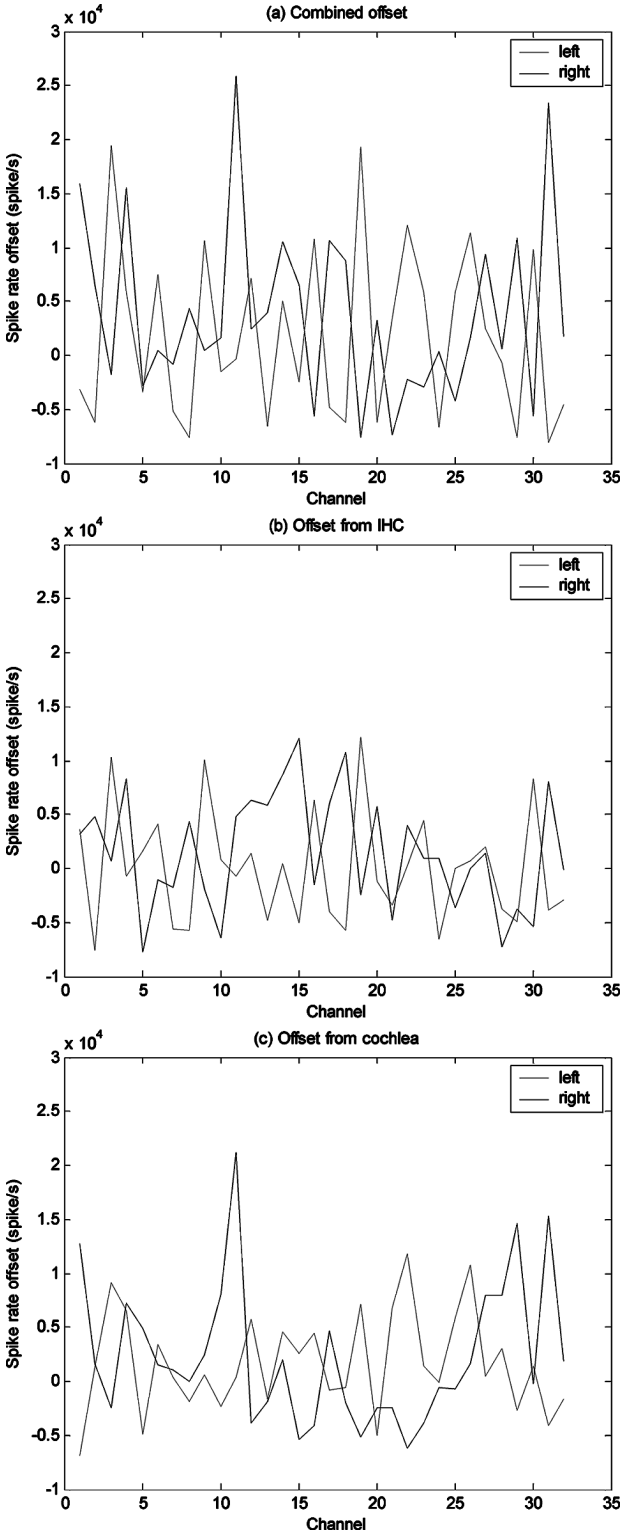

Fig. 11. (a) Total offset in spike rate at each channel. (b) Offsets with cochlear mismatch eliminated (shorting V_{C1} and V_{C2}). (c) Difference of (a) and (b), which represents the offset caused by the cochlea.

the result of subtracting the spike rate offsets in (b) from those in (a), representing the spike rate offset caused by $V_{\rm dc}$. As an additional verification step, we measured the dc offsets of $V_{C1} - V_{C2}$ directly via the scanner, as shown in Fig. 12. The measured dc offsets correspond well with the calculated spike rate offsets of Fig. 11(c).

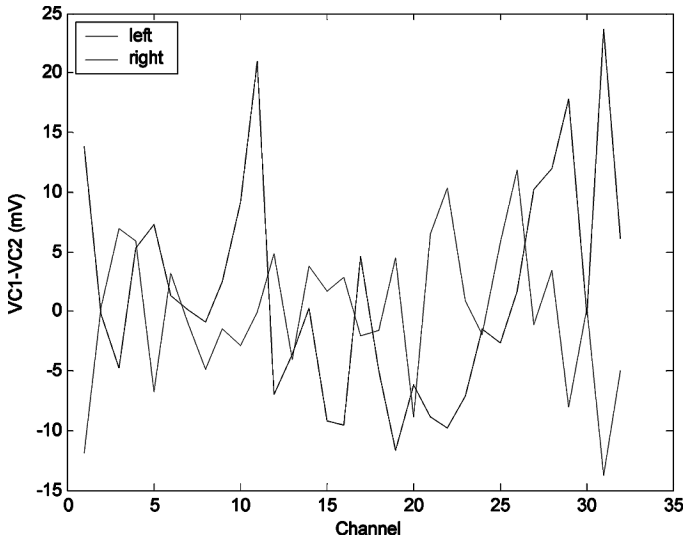

Fig. 12. Variation in dc offset at $V_{C1} - V_{C2}$ at both left and right cochleae. This offset is a major source of offset in spike rates.

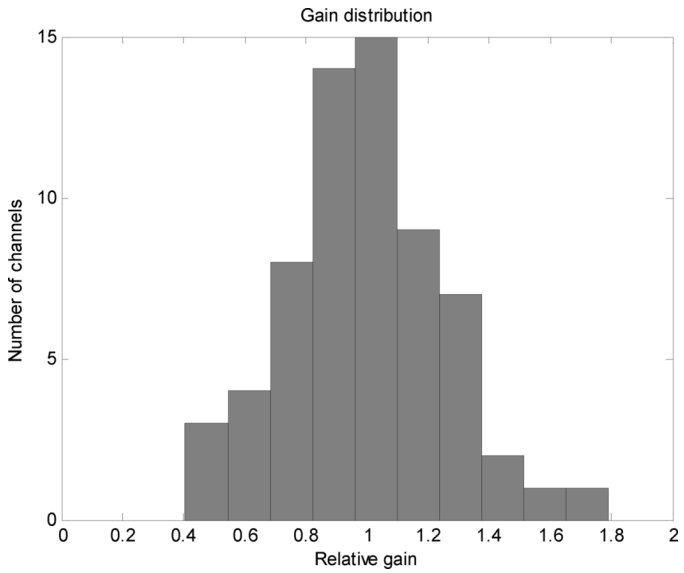

Fig. 13. Distribution of the gain of the channels.

Knowing the offset at each channel allows us to eliminate the offset by subtracting a constant value from the observed spike rate, but this still leaves us with a gain error. From (7) and (8), the overall gain depends on $g_m\, k_{\rm IHC}/C_m V_{\rm thr}$. All 4 values vary from channel to channel: g_m depends on the bias current $I_{\rm gain}$, the current mirror gain $k_{\rm IHC}$ varies due to threshold voltage variations and limited output resistance, and C_m, $V_{\rm thr}$ are slightly different for each neuron due to size mismatch at the capacitor and threshold voltage mismatch at the transistors, respectively.

Fig. 13 shows the distribution of the measured gains of all 64 channels (32×2), relative to the average. The gain ranges from 40% to 180% of the average, with a standard deviation of 26%. The gain error of each channel can be removed by normalization, that is, scaling the gain at each channel so that the gain at the resonant frequency equals one. Fig. 14 plots the spike rates of selected channels as a function of frequency after our attempts to remove both offset and gain variation, thus showing

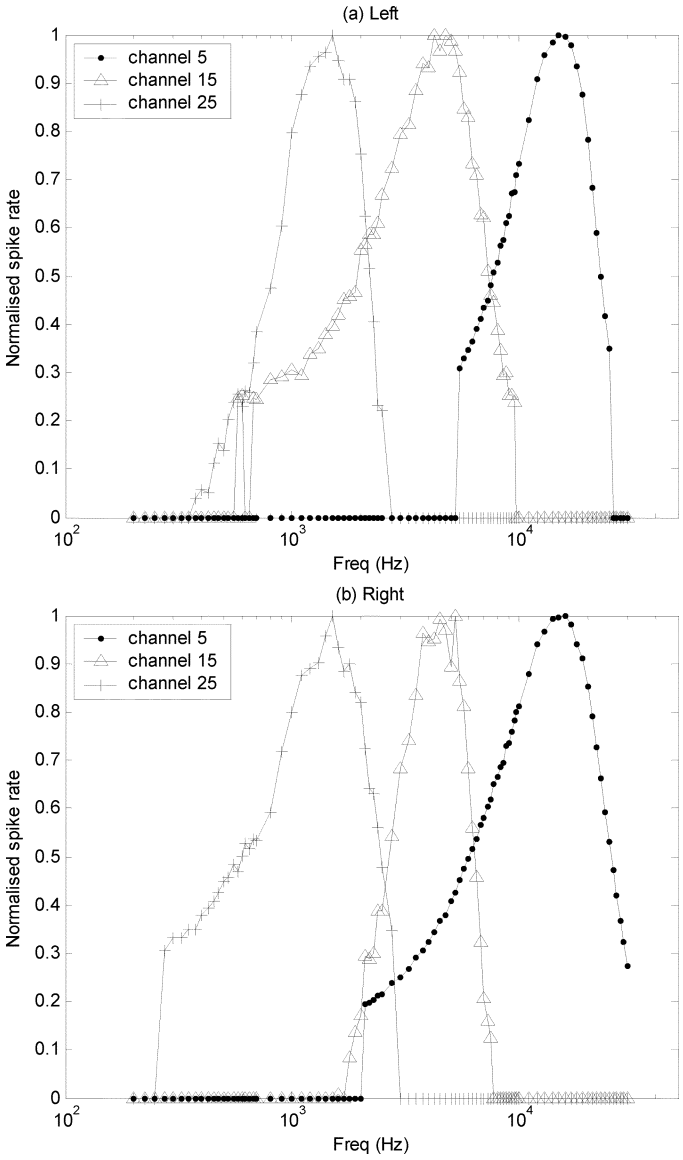

Fig. 14. Spike rates of neurons in selected channels of (a) the left cochlea and (b) the right cochlea after processing to remove offsets and gain errors.

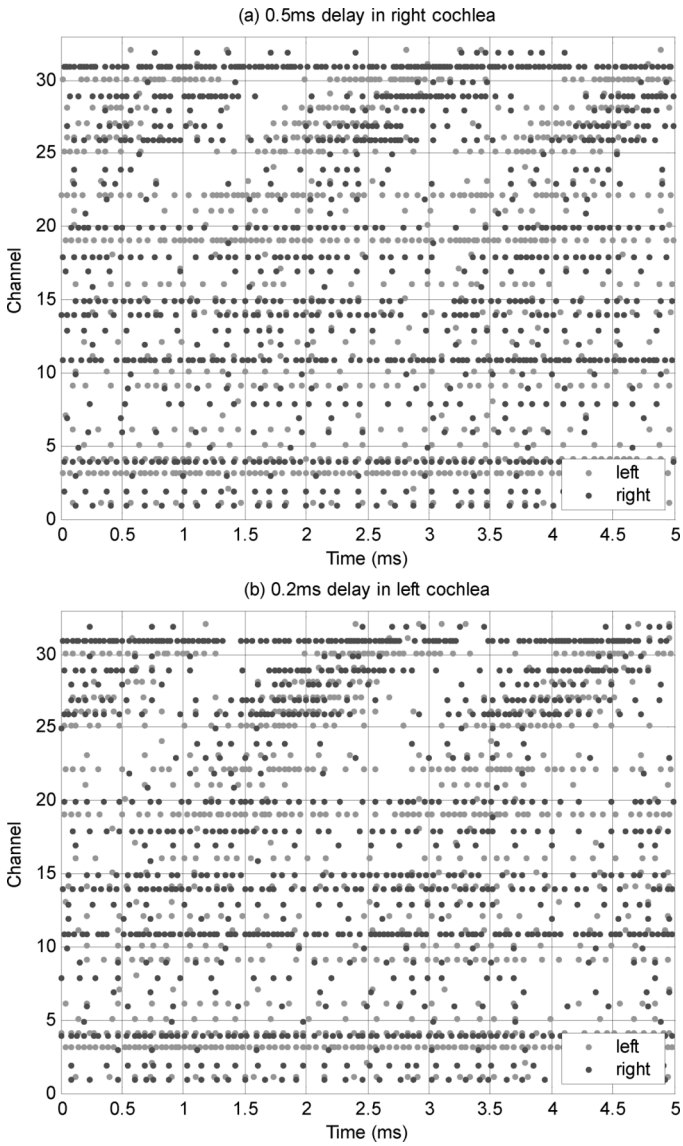

Fig. 15. Cochleagrams showing the delay in one channel relative to the other for a 500-Hz sine wave input. (a) Left leading right. (b) Right leading left.

that much of the mismatch could be removed in later processing stages through learning and adaptation.

IV. Sound Localization

In this section, we investigate the possibility of using the AER EAR for sound localization. In the first instance, we are interested in how well the AER spikes preserve timing information for the extraction of interaural time differences. For these experiments, we have therefore increased the cutoff frequency of the low-pass filter in the IHC to a biologically unrealistic value of 50 kHz in order to observe phase-locking of the AER output at all audio frequencies. With such a high cutoff frequency the filter in the IHC circuit simply acts as a variable gain current mirror at audio frequencies. The gain of this mirror is exponentially proportional to $V_{\rm ref} - V_{\rm refo}$, which we have set, as in the previous experiment, to be equal (4.5 V) to give a gain of one. No offset or gain correction is applied to the output of the cochlea in the following experiments.

In the first experiment, we played signals from a stereo sound card directly to the inputs of the two cochleae. Both channels played the same sound except that we delayed the signal in one channel relative to the other to simulate an interaural time delay (ITD), that is, the difference in time of arrival of the sound to the two ears because one ear is closer to the source than the other. Two examples of the resulting spike trains in each of the cochlear channels for different ITDs are shown in Fig. 15.

We extract the ITD from the spike trains as follows. We perform cross-correlation of the spike trains between all corresponding channels of the left and right cochleae using a 10-ms window. From the average correlation across the 32 cochlear channels, we use the delay at which the maximum correlation occurs to represent the ITD. An example of the average cross-correlation curve is shown in Fig. 16. The length of the window is chosen so that it is short enough to allow rapid updates and fast response to stimulus, but long enough to obtain a good ITD estimate. For example, for a 500-Hz pure

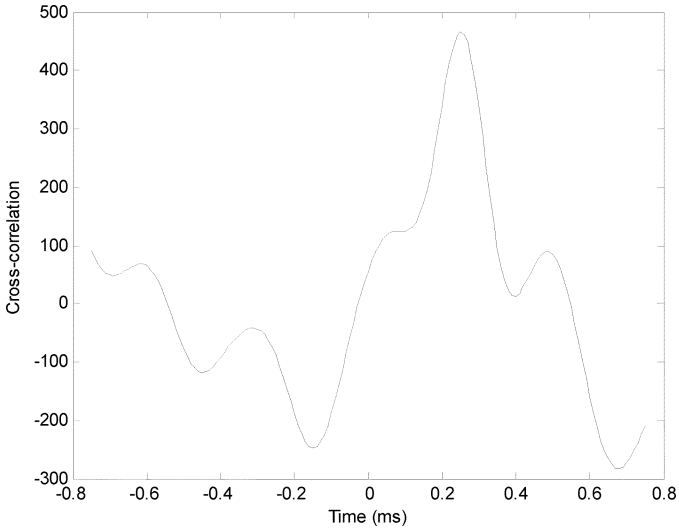

Fig. 16. Average cross-correlation of the spike trains in the left cochlea and those in the right cochlea. In this example, the maximum occurs at 0.25 ms, which will be used as the calculated ITD.

tone stimulus, a 10-ms window gives us five cycles of the signal to determine its phase accurately.

Fig. 17 shows calculated delay versus actual delay, when the input signal is (a) white noise with a flat spectrum from 75 Hz to 24 kHz, and (b) a 500-Hz pure tone. At each delay position, five calculations are performed, each 20 ms apart in the middle of the 0.5-s signal. Calculations based on white noise inputs are more accurate than those based on pure tone inputs as spikes from all cochlear channels contribute to the delay estimate for the broad band white noise stimulus. For the 500-Hz tone, only the low frequency channels carry phase information as the high frequency channels are not responsive to 500 Hz, as shown in the cochleagrams in Fig. 15. Therefore, when the higher frequency channels are removed from the calculation, accuracy improves. This is evident by comparing Fig. 17(b) with Fig. 17(c), which shows less variation. Overall, the results in this experiment show that the AER cochlea is able to preserve ITD information accurately.

In a second experiment, we attempt to extract ITD in a more realistic reverberant environment. We recorded sound using a pair of microphones inserted into the ears of a dummy head located at the center of a small room. The dummy was rotated from -90° to $+90^\circ$ in 5° steps, where 0° represents the dummy facing the loudspeaker straight ahead (See Fig. 18). The recorded sounds were played to the left and the right cochlea through a soundcard as in the first experiment. Fig. 19 shows the outputs of the cochleae at 2 different positions.

We calculate the ITD from the spikes using the same procedure as in the previous experiment. Fig. 20 shows the calculated ITD as the position of the source varies from -90° (left) to 90° (right). The theoretical maximum delay occurs at $\pm 90^\circ$, and for low frequency signals, the curve should be a sine function of angle [28]. However, as it can be seen in Fig. 20, the maximum occurs at $\pm 60^\circ$ in the measurements. This is because the presence of echoes in a reverberant environment causes the phase to vary depending on the geometry on the room and the exact position in the room where it is measured,

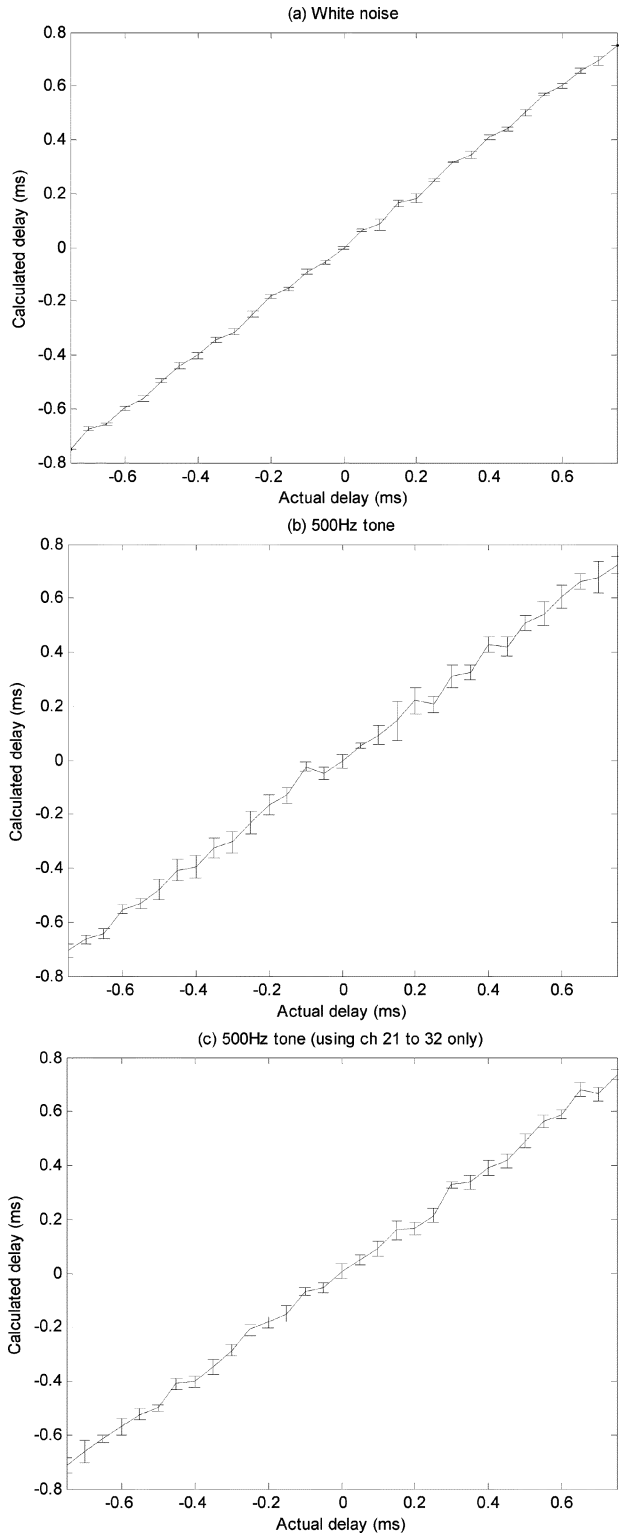

Fig. 17. Calculated delay versus actual delay based on cross-correlation of spikes from left and right cochleae. Five calculations are performed for each actual delay, each using a different section of the spike trains. The mean and standard deviation at each delay are shown.

rather than depending solely on the distance of the microphones from the source. In fact, we have noticed during the course of the experiment that even the simple opening or closing of a door in the room will affect both the interaural time and interaural level differences.

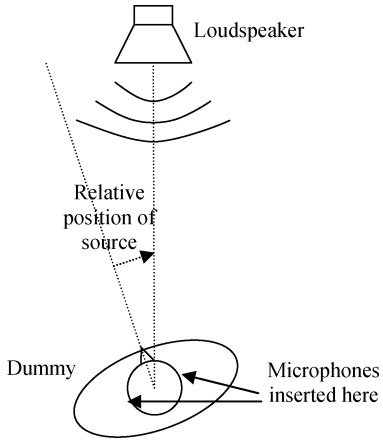

Fig. 18. Setup of experiment 2. The source is fixed but the dummy is free to rotate from $-90°$ to $90°$. Microphones are inserted into its ears to record sound at each position.

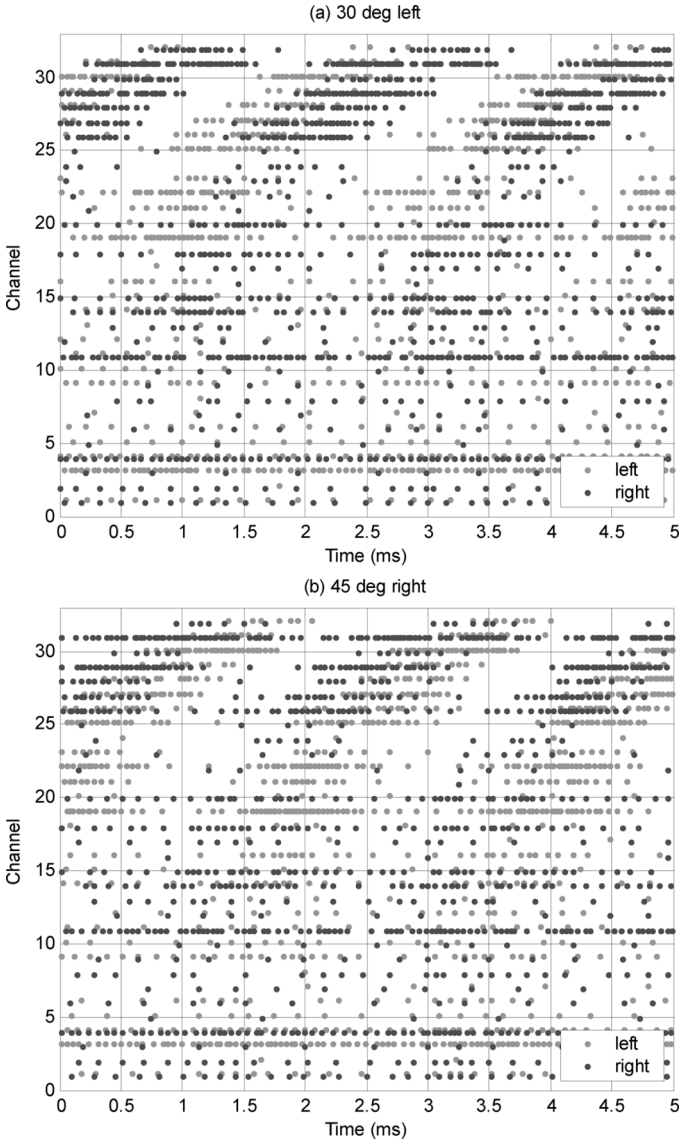

Fig. 19. Cochleagrams. A 500-Hz tone is played at two different positions relative to the dummy in a reverberant environment.

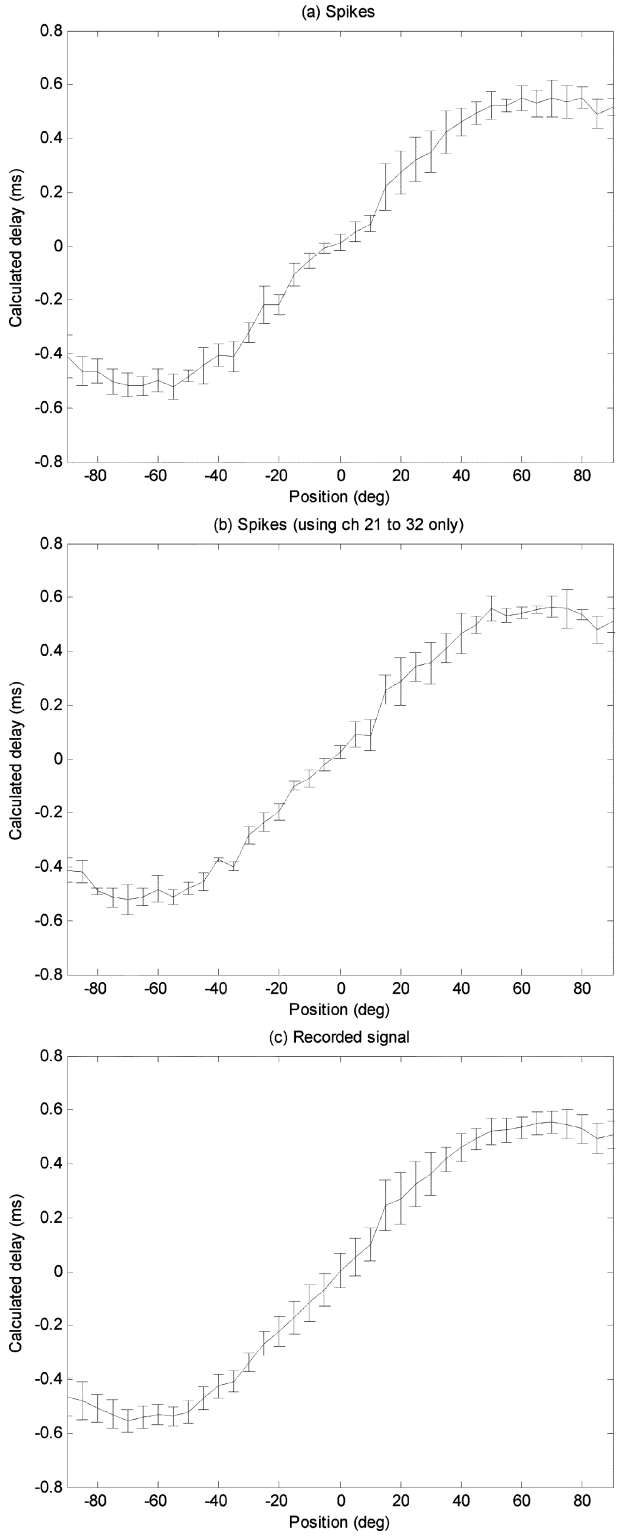

Fig. 20. ITD computed by cross-correlation of (a) the spike trains from all the channels, (b) the spike trains in channel 21 to 32 only, and (c) the recorded signals. The source is a 500-Hz pure tone. The plots show the mean and standard deviation at each position based on five calculations.

Just as in the previous experiment, when the input is a 500-Hz pure tone, computing ITD using only the low frequency channels can improve accuracy slightly, as shown in Fig. 20(b). Fig. 20(c) is computed based on direct cross-correlation between the left and the right channels recorded by the microphones without processing by the cochleae. Comparing

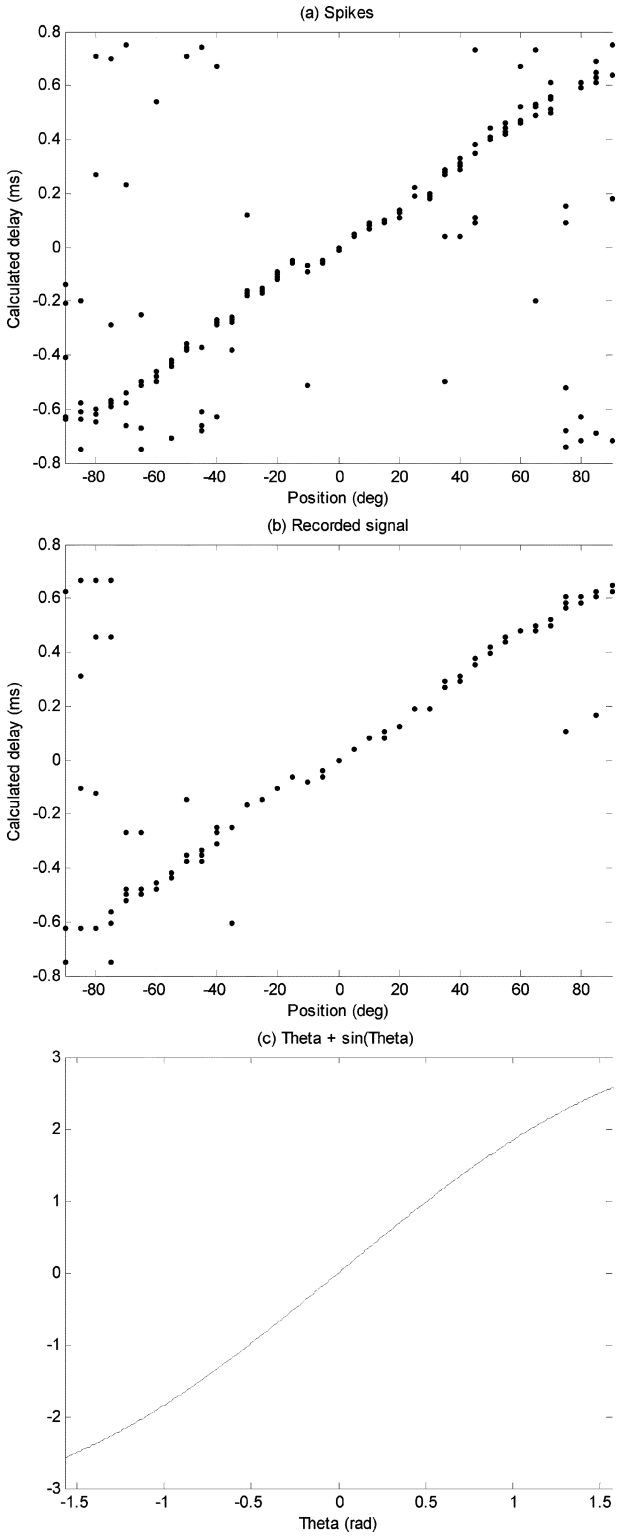

Fig. 21. ITD calculation for a white noise input, computed from the cross-correlation of (a) spike trains and (b) the recorded signals. (c) A plot of the function $\theta + \sin\theta$.

Fig. 20(c) with Fig. 20(a) and (b) shows that the use of a cochlea does not degrade timing information in the signal as ITD can still be extracted from the spikes accurately. Furthermore, even in such a simple setting with only a single sound source, filtering the recorded sounds with the cochlea and selecting the appropriate channels for cross correlation Fig. 20(b) does offer a slight advantage for extracting the ITD. This may be more pronounced in the case of simultaneous sound sources, or other more complex sound environments.

We repeated the second experiment using a white noise stimulus. Fig. 21(a) shows the ITD extracted from the AER output, while Fig. 21(b) shows the ITD extracted from the recorded noise signals directly. The close resemblance of the two curves demonstrated once again that timing information is preserved by the cochleae. The poorer performance of white noise (indicated by higher error rate) compared to the excellent performance with white noise in the first experiment Fig. 17(a) is due to the low signal-to-noise ratio at the far ear. The high frequency components of the noise are attenuated significantly at the far ear by the shadowing of the head, and the low frequency components are less energetic than in the pure tone case because energy is spread evenly across the spectrum for the white noise stimulus. An additional factor is that the echoes in the reverberant environment may cause different ITDs in different bands of the spectrum, thereby causing more noisy estimates of the overall ITD. It is of note that the calculated ITD, both directly from the recorded signals and from the AER EAR outputs, is not a sine function of position as in the 500-Hz pure tone case. This is because at higher frequencies the signal diffracts around the head and the actual path-length difference will become directly proportional to $\theta + \sin\theta$ if one assumes a circular cross section of the head [28]. Fig. 21(c) plots the function $\theta + \sin\theta$ and our results here seem to match the theory very well. Overall, results from these 2 experiments suggest that timing information is well preserved by the AER cochlea and ITD can be easily extracted from the spike trains.

V. Conclusion

We have presented our AER EAR – a pair of matched silicon cochleae with AER interface – which models the early stages of the mammalian auditory system. We show measurements of the frequency responses of the cochleae and spike rates of the channels. From an analysis of the mismatch present in the circuit, we found that the dc offsets at the cochleae are the major source of offset error. Despite the mismatch, frequency selectivity is preserved and it is possible to remove the offset and gain errors in later processing stages. Finally, we demonstrated the cochleae's capability in sound localization by successfully extracting ITD from the AER cochleae in both ideal and reverberant environments. A future version of this AER cochlea will focus on reducing the variations and increasing the frequency selectivity of each channel, as well as including more faithful IHC and spiking neuron models.

References

[1] K. A. Boahen, "Communicating neuronal ensembles between neuromorphic chips," in *Neuromorphic Systems Engineering*, T. S. Lande, Ed.. Boston, MA: Kluwer, 1998, pp. 229–259.

[2] ——, "Point-to-point connectivity between neuromorphic chips using address events," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 47, no. 4, pp. 416–434, Apr. 2000.

[3] E. Culurciello, R. Etienne-Cummings, and K. Boahen, "High dynamic range, arbitrated address event representation digital imager," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 6–9, 2001, vol. 3, pp. 505–508.

[4] M. Barbaro, P.-Y. Burgi, A. Mortara, P. Nussbaum, and F. Heitger, "A 100 × 100 pixel silicon retina for gradient extraction with steering filter capabilities and temporal output coding," *IEEE J. Solid-State Circuits*, vol. 37, no. 1, pp. 160–172, Jan. 2002.
[5] K. A. Boahen, "A retinomorphic vision system," *IEEE Micro.*, vol. 16, pp. 30–39, 1996.
[6] C. M. Higgins and S. A. Shams, "A biologically inspired modular VLSI system for visual measurement of self-motion," *IEEE Sens. J.*, vol. 2, no. 6, pp. 508–528, Dec 2002.
[7] G. Indiveri, "Modeling selective attention using a neuromorphic analog VLSI device," *Neur. Comp.*, vol. 12, pp. 2857–2880, 2000.
[8] G. Indiveri, A. M. Whatley, and J. Kramer, "A reconfigurable neuromorphic VLSI multi-chip system applied to visual motion computation," in *Proc. 7th Int. Conf. Microelectronics for Neural, Fuzzy and Bio-Inspired Systems (Microneuro'99)*, Los Alamitos, CA, 1999.
[9] S.-C. Liu, J. Kramer, G. Indiveri, T. Delbrück, T. Burg, and R. Douglas, "Orientation-selective aVLSI spiking neurons," *Neural Netw.*, vol. 14, no. 6/7, pp. 629–643, 2001.
[10] T. Serrano-Gotarredona, A. G. Andreou, and B. Linares-Barranco, "AER image filtering architecture for vision-processing systems," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 46, no. 9, pp. 1064–1071, Sep. 1999.
[11] S. R. Deiss, R. J. Douglas, and A. M. Whatley, "A pulse-coded communications infrastructure for neuromorphic systems," in *Pulsed Neural Networks*, W. Maass and C. M. Bishop, Eds. Boston, MA: MIT Press, 1999, pp. 157–178.
[12] D. H. Goldberg, G. Cauwenberghs, and A. Andreou, "Analog VLSI spiking neural network with address domain probabilistic synapses," in *Proc. 2001 IEEE Int. Symp. Circuits Syst.*, May 6–9, 2001, vol. 3, pp. 241–244.
[13] S.-C Liu and R. Douglas, "Temporal coding in a network of silicon integrate-and-fire neurons," *IEEE Trans. Neural Netw.*, vol. 15, no. 5, pp. 1305–1314, Nov. 2004.
[14] A. van Schaik, "Building blocks for electronic spiking neural networks," *Neural Netw.*, vol. 14, no. 6–7, pp. 617–628, 2001.
[15] P. Venier, A. Mortara, X. Arreguit, and E. A. Vittoz, "An integrated cortical layer for orientation enhancement," *IEEE J. Solid-State Circuits*, vol. 32, no. 1, pp. 177–185, Jan. 1997.
[16] J. P. Lazzaro and J. Wawrzynek, "A multi-sender asynchronous extension to the address-event protocol," in *Proc. 16th Conference on Advanced Research in VLSI*, 1995, pp. 158–169.
[17] N. Kumar, W. Himmelbauer, G. Cauwenberghs, and A. G. Andreou, "An analog VLSI chip with asynchronous interface for auditory feature extraction," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 45, no. 5, pp. 600–606, May 1998.
[18] H. Abdalla and T. Horiuchi, "An ultrasonic filterbank with spiking neurons," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 23–26, 2005, vol. 5, pp. 4201–4204.
[19] V. Chan, A. van Schaik, and S.-C. Liu, "Spike response properties of an AER EAR," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 21–24, 2006, ISCAS.2006.1692721.
[20] A. van Schaik and S.-C. Liu, "AER EAR: A matched silicon cochlea pair with address event representation interface," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 23–26, 2005, vol. 5, pp. 4213–4216.
[21] A. van Schaik and S. Shamma, "A neuromorphic sound localizer for a smart MEMS system," *Anal. Integr. Circuits Signal Process.*, vol. 39, pp. 267–273, 2004.
[22] A. van Schaik, E. Fragnière, and E. Vittoz, "Improved silicon cochlea using compatible lateral bipolar transistors," in *Advances in Neural Information Processing Systems*, D. Touretzky, Ed. *et al.* Cambridge, MA: MIT Press, 1996, vol. 8, pp. 671–677.
[23] A. van Schaik, "An analog VLSI model of periodicity extraction in the human auditory system," *Anal. Integr. Circuits Signal Process.*, vol. 26, pp. 157–77, 2001.
[24] A. van Schaik and R. Meddis, "Analog very large-scale integrated (VLSI) implementation of a model of amplitude-modulation sensitivity in the auditory brainstem," *J. Acoust. Soc. Amer.*, vol. 105, pp. 811–821, 1999.
[25] R. F. Lyon and C. A. Mead, "An analog electronic cochlea," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 7, pp. 1119–1134, Jul. 1988.
[26] L. Watts, D. A. Kerns, R. F. Lyon, and C. A. Mead, "Improved implementation of the silicon cochlea," *IEEE J. Solid-State Circuits*, vol. 27, no. 5, pp. 692–700, May 1992.
[27] A. van Schaik and A. McEwan, "An analog VLSI implementation of the meddis inner hair cell model," *EURASIP J. Appl. Signal Process.* , pp. 639–648, Jun. 2003.
[28] G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Amer.*, vol. 62, no. 1, pp. 157–167, 1977.

Vincent Chan (S'05) received the B.E. degree in computer engineering in 2003 from the University of Sydney, Sydney, Australia, where he is currently working towrd the Ph.D. degree in electrical engineering.

His research interests include artificial vision, sound localization, and sensor fusion in analog VLSI.

Shih-Chii Liu (M'02) received the B.S. degree in electrical engineering from Massachusetts Institute of Technology, Cambridge, MA, the M.S. degree in electrical engineering from University of California, Los Angeles, and the Ph.D. degree in computation and neural systems program from California Institute of Technology, Pasadena, in 1983, 1988 and 1997, respectively.

She is currently an Überassistent at the Institute of Neuroinformatics, ETH/UNIZ in Zurich, Switzerland. Her research interests encompass computational neuromorphic models of visual processing in biological systems.

André van Schaik (M'00–SM'02) received the M.Sc. degree in electrical engineering from the University of Twente, Enschede, The Netherlands, i and the Ph.D. degree in electrical engineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland n 1990 and 1998, respectively.

He is a Reader in Electrical Engineering in the School of Electrical and Information Engineering at the University of Sydney, Sydney, Australia and an Australian Research Fellow. His research focuses on two main areas: neurmorphic engineering and spatial audio. He has authored or co-authored more than 50 papers in these two research areas and inventor of more than 25 patents. He is the Director of the Computing and Audio Research Laboratory at the University of Sydney and a founder of VAST Audio Pty. Ltd.

Dr. van Schaik is a member of the EPSRC college, a board member of the Institute of Neuromorphic Engineering and a member of the Audio Engineering Society. He is the Chairman of the Sensory Systems technical committee of the IEEE's Circuits and Systems Society and an Associate Editor of the IEEE TRANS. CIRCUITS AND SYSTEMS—I: REGULAR PAPERS. He is a member of the Analog, BioCAS, and Neural Network Technical Committees of the IEEE Circuits and Systems Society.