

Periodicity detection and localization using spike timing from the AER EAR

Theodore Yu¹, Andrew Schwartz², John Harris³, Malcolm Slaney⁴, and Shih-Chii Liu⁵

¹School of Electrical and Information Engineering
University of San Diego
San Diego, USA
teyu@ucsd.edu

²Speech and Hearing Bioscience and Technology Program
Harvard-Massachusetts Inst of Technology
Cambridge, MA, USA
ozydingo@gmail.com

³School of Electrical and Computer Engineering
University of Florida
Gainesville, FL, USA
harris@cnel.ufl.edu

⁴Yahoo! Research
Santa Clara, CA, USA
malcolm@ieee.org

⁵Institute of Neuroinformatics
University of Zürich/ETH
Zürich, Switzerland
shih@ini.phys.ethz.ch

Abstract—We present a system consisting of a spiking cochlea chip and real-time event-based processing software that is able to discriminate between two sets of sounds based on their periodicity content. The periodicity measurements are computed from the spike timing information of asynchronous output spikes from the binaural spiking-cochlea chip. The chip consists of a matched pair of silicon cochlea with an address event interface for the output. Each section of the cochlea is modeled by a second-order low-pass filter followed by a simplified Inner Hair Cell circuit and a Spiking Neuron circuit. We show discrimination results using the periodicity measure for 2 classes of sound and preliminary localization results based on a discriminated sound.

I. INTRODUCTION

Periodicity information is an important alternative to spectral analysis because of its precision and repeatability. The biological cochlear filters are highly nonlinear, with bandwidths and gain changing based on the incoming sound level [1]. The location of the filter with the peak response changes, thus it is difficult for a perceptual system to estimate the spectrum from only the rate profile. Some authors [2] have suggested that neural circuits in the cochlea subtract the responses at two nearby locations to mark the location of the sharp high-frequency cutoff, which is more stable. But the phase of the signal, as a function of place on the cochlea, is also changing rapidly at this location, and it is difficult to know how this subtraction is implemented in a spiking network.

Instead, in this paper, we investigate the use of just timing information to recognize and localize sounds. Even as the cochlear filters change, peak excursions in a 100Hz cochlear channel due to a 100Hz input occur every 10ms. The timing information is preserved. Others have used this information to judge the relative time delay of a sound between two ears [3] and in this paper we use the same information to judge pitch and identify sounds. The same tolerance to imperfect filters useful to the auditory system is also useful in our silicon models.

Although we have 20 years of experience in designing silicon cochlea chips, it is only in recent years that we see cochlea chips that produce asynchronous spike outputs resembling outputs of the auditory nerve fibres. These spikes are transmitted using the Address Event Representation (AER) where each spike carries the identity of the sender. There are a handful of silicon cochleae with an Address Event type representation [4][5][6][7]. The AER EAR chip that we use in this work is an improved version over the prototype described by Chan et al. [3].

There are a couple of groups that have looked at aVLSI systems for extracting periodicity in sounds. The implementation from van Schaik [9] extracted periodicity information by ANDing the neuron outputs of bandpass filter channels that are spaced a period apart. The implementation from Abdalla and Horiuchi consists of an aVLSI chip which extracts the periodicity information directly from the output of the microphone [10]. In this work, we extract periodicity

from spike trains by using a system consisting of a spiking silicon cochlea (AER EAR) and an event-based software infrastructure (jAER) by using the spike timing information from the output spikes of the AER EAR [3]. The jAER software can process in real-time the spike events from AER chips/devices [11].

We use this periodicity information to discriminate between two classes of sounds, harmonic and inharmonic sounds or noise, independent of the speaker. In addition, this classification information can be used to selectively localize an auditory sound that falls into one of the two classes.

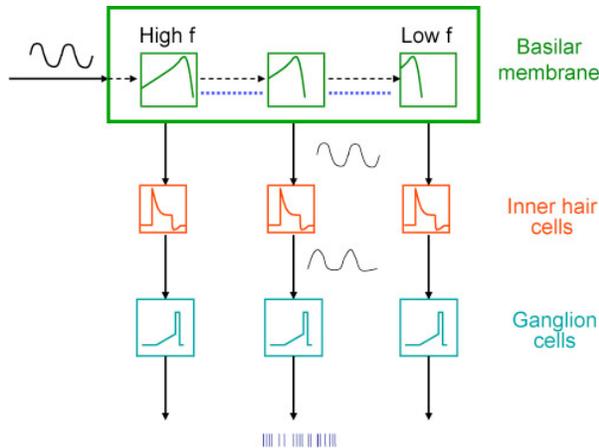


Figure 1. The block diagram of circuits of one of the 2 cochleas on the binaural AER EAR chip.

II. THE SILICON COCHLEA

Figure 1 shows the basic building blocks in the spiking cochlear chip. The incoming sound goes through a cascade of 32 bandpass filters for each cochlea of this chip with a range of exponentially decreasing cutoff frequencies usually tuned from around 100 Hz to 1 kHz.

A simplified Inner Hair Cell circuit rectifies and low passes the output of each bandpass filter before passing it to a ganglion cell circuit. The cut-off frequency of the Inner Hair Cell is set around 1 kHz, as in the real Inner Hair Cell. This low-pass filtering models the reduction in phase-locking observed on biological auditory nerve fibres at frequencies greater than 1 kHz. The outputs of the ganglion cell circuits are transmitted asynchronously on a common digital-address bus which carries the identity of the channel that produced the output spike. The time of the spike is coded implicitly in the event.

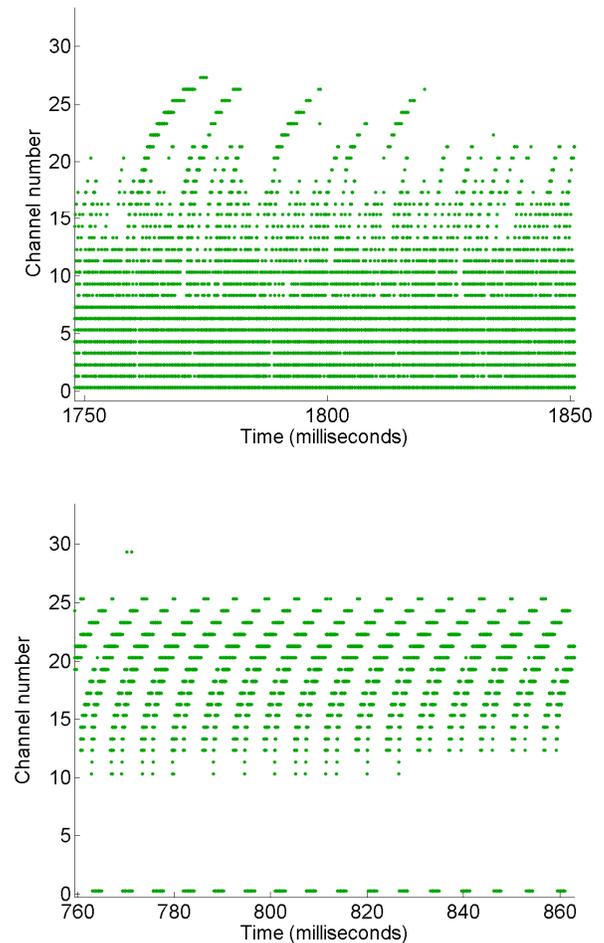


Figure 2. Spike responses from a single cochlea across 32 channels (y-axis) in response to (top) a “hiss” and (bottom) a “coo” from a speaker. Channel ‘0’ of this chip has a low threshold for spiking because of transistor mismatch.

III. PERIODICITY MEASUREMENT

We chose a database of “coo” and “hiss” sounds voiced by 12 speakers for our experiment. We recorded both the analog waveform from the microphone and the spike trains of the cochlea. Figure 2 shows exemplar responses of the 32 channels of the cochlea to these sounds as voiced by one speaker. The periodicity of the spike patterns in the “coo” is obvious while the spike patterns of the “hiss” do not show this regularity. To compute the periodicity from the spike times, we first calculated an all-order histogram of the interspike intervals (ISIs) of the spikes. Using only a first-order ISI histogram (that is, taking only the time difference between adjacent spikes in each channel) will not give the right period for the fundamental frequency because the low-frequency channels spike more than once per cycle while the high frequency channels spike once or none for a cycle. The

peaks in the histogram will not reflect the period of a cycle for low-frequency sounds.

For each speaker, we computed the ISI histogram of the spike responses to the two sets of sounds. We tried different order ISIs and found that the histogram of ISIs up to 7th order give noticeable peaks (Figure 3 shows an example). Including higher than 7th order ISIs does not change the histogram profile noticeably. The first peak under 1ms represents the ISIs of spikes from a single cycle and is ignored. The next peak reflects the pitch of the speaker and subsequent peaks represent the harmonics of the pitch frequency.

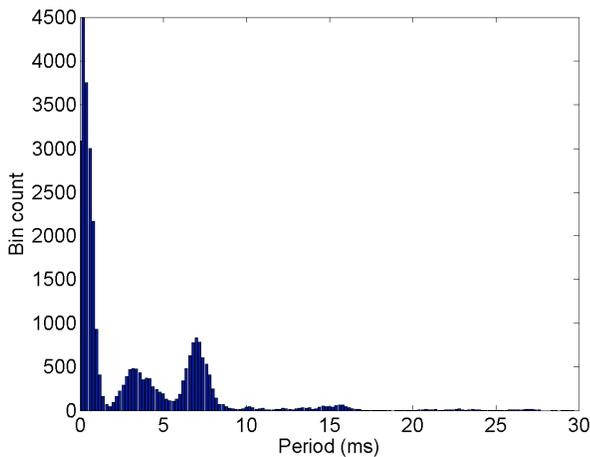


Figure 3. Histogram computed from 1st to 7th order ISIs of cochlear spikes from a single speaker voicing a “coo”. Peaks represent the harmonics in the speaker’s pitch except for the first peak which is due to the ISIs of spikes within a cycle.

To determine whether the pitch and harmonic frequencies of harmonic sounds such as “coo” extracted from spike trains are similar to the pitch information in the analog output from the microphone, we plotted the fundamental peak from the FFT of the analog waveform versus the fundamental frequency computed from the FFT of the ISI waveforms across all speakers (Figure 4). As seen, these points fall very close to the unity line, even with the response variations across the silicon cochlea frequency channels (Figure 2). Even in the case of a sound file where the speaker had varied his pitch in time while voicing the “coo” sound, the FFT of the extracted pitch from the analog waveform and the spike ISI histogram are almost equal (Figure 5).

IV. DISCRIMINATION

To discriminate between the two classes of sounds, we select a fundamental frequency that is around 100 to 200 Hz and we then look for multiples of this base frequency. Each detected frequency corresponds to the inverse of the period

of a local peak in the ISI histogram. A local peak in the ISI histogram meets two criteria: 1) it is significant because it contains a sufficient number of samples to meet a set threshold population level and 2) it is prominent because it contains a sufficient number of samples more than surrounding samples to distinguish it from the neighboring region. We label a segment of sound as harmonic if its ISI histogram has a local peak. For each 0.2s segment of the sound, we classified it as a “coo”, “hiss”, or “undecided”. Taking the majority of the hits in each of the 3 classes, we determined if the speaker was voicing a “coo” or a “hiss”. Using this approach and our database of 12 speakers, the “coo” was correctly identified in 10 speakers and the same was true of the “hiss”.

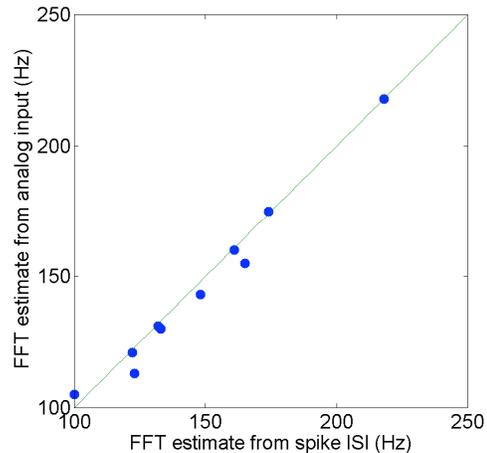


Figure 4. Data set of the 12 speakers (circles) showing that the fundamental frequency for a steady pitch computed from the FFT of the analog waveform vs the fundamental frequency computed from the ISI histogram.

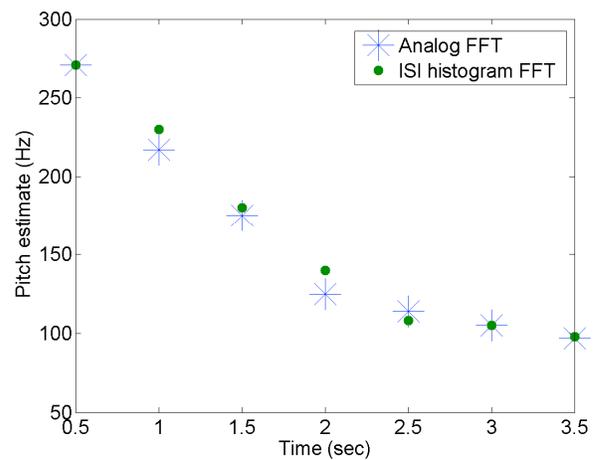


Figure 5. Data set of a speaker showing the correspondence for a time-varying pitch between the frequency computed from the FFT of the analog waveform and the frequency computed from the ISI histogram.

V. SOUND LOCALIZATION

We were able to use the outputs of the binaural cochlea, each with input from its own microphone, to predict the horizontal location of the sound source. When a sound wave travels from a source to the two microphones, there is a difference in the travel time to each microphone that is visible in the recorded waveforms. This time difference can be seen in the phase difference of the spikes between the two cochleas.

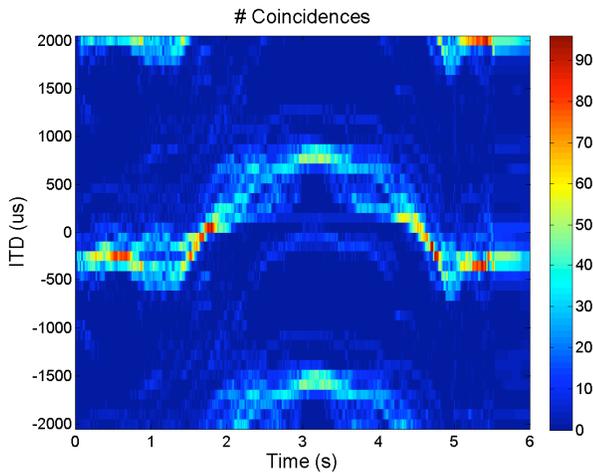


Figure 6. Localization data where a speaker voicing the “coo” sound moves continuously in time from the far end of one microphone to the other microphone and back again. The ITD is computed from the spike trains of the binaural cochlea.

We fed the outputs of matching channels from each cochlea chip to a correlation algorithm that counts the number of occurrences at a range of inter-spike delays. This algorithm was also implemented in jAER. If the spike outputs from corresponding channels are exactly in phase with each other (for example, being fed input from the same microphone), the algorithm gives a spike at 0us, and have no output for any other inter-spike delay. For real signals the data is somewhat noisier, but there is still an observable peak at 0us delay. When a sound comes from one side of the microphone pair, the spikes from the chip with the closest microphone will lead the other, and there will be a peak in the correlation algorithm’s output at the corresponding delay.

We combine information across frequency channels to estimate the sound’s location. A naïve approach simply weights each frequency channel equally. However it is also possible to use available information about the stimulus, for example the fundamental frequency and harmonic composition of the sound as determined by the periodicity measurement described in Section IV, to assign more weight to frequencies that are known to be dominated by the stimulus of interest. This can serve to reduce the effect of the corruption of phase information by background noise and competing sound sources.

VI. CONCLUSION

In this work we demonstrated a system that uses the timing of spike outputs from a binaural silicon AER cochlea to determine the harmonicity of a sound. We present data that shows that the harmonicity information in the spike trains was compatible with the information in the original analog waveform, even with the variance in the ISIs across the different frequency channels of the silicon AER EAR. This information can be used for distinguishing the sex of the speaker [10] or to distinguish between two classes of sounds. This harmonicity computation can be combined with a localization module which uses the interaural time difference information in the spike trains from the binaural cochlea. The subsequent system consisting of the silicon AER cochlea and the jAER program can detect the location of a particular class of sounds in real-time. This approach is important because it shows the temporal information can be used for perceptual experiments, even in the face of imperfect cochlear filters.

ACKNOWLEDGMENT

We acknowledge Tobi Delbruck for help with jAER, and the Telluride Neuromorphic Workshop for providing the infrastructure to execute this project.

REFERENCES

- [1] M. Slaney and R. F. Lyon, “On the importance of time: A temporal representation of sound,” *Visual Representations of Speech*, Martin Cooke, Steve Beet, Malcolm Crawford (editors). J. Wiley, New York, 1993, pp. 95–116.
- [2] S. Shamma, “Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve,” *J Acoust Soc Am*, vol. 78, no. 5, 1985, pp.1622–32.
- [3] V. Chan, S-C. Liu, and A. van Schaik, “AER EAR: A matched silicon cochlea pair with address event representation interface,” *IEEE Transactions on Circuits and Systems I*, vol. 54, no. 1, 2007, pp. 48–59.
- [4] J.P. Lazzaro and J. Wawrzynek, “A multi-sender asynchronous extension to the address-event protocol,” *Proceedings 16th Conference on Advanced Research in VLSI*, 1995, pp. 158–169.
- [5] N. Kumar, W. Himmelbauer, G. Cauwenberghs, and A. G. Andreou, “An analog VLSI chip with asynchronous interface for auditory feature extraction,” *IEEE Transactions On Circuits And Systems—II: Analog and Digital Signal Processing*, vol. 45, no. 5, May 1998, pp. 600–606.
- [6] H. Abdalla and T. Horiuchi, “An ultrasonic filterbank with spiking neurons,” *Proceedings of IEEE International Circuits and Systems 2005*, vol. 5, 2005, pp. 4201–4204.
- [7] B. Wen and K. Boahen, “A 360-channel speech preprocessor that emulates the cochlear amplifier,” *IEEE International Solid-State Circuits Conference*, Digest of Technical Papers, pp. 556–557, IEEE Press, 2006.
- [8] A. van Schaik, E. Fragnière, and E. Vittoz, “Improved silicon cochlea using compatible lateral bipolar transistors,” *Advances in Neural Information Processing Systems 8*, edited by D. Touretzky, et al., MIT Press, Cambridge MA, 1996, pp. 671–677.
- [9] A. van Schaik, “An analog VLSI model of periodicity extraction in the human auditory system,” *Analog Integrated Circuits & Signal Processing*, vol. 26, 2001, pp. 157–77.
- [10] H. Abdalla and T. Horiuchi, “An analog VLSI low-power envelope periodicity detector,” *IEEE Transactions on Circuits and Systems I*, vol. 52, no. 9, 2005, pp. 1709–1720.
- [11] jAER Sourceforge Open Source Project.
<http://jaer.wiki.sourceforge.net>.