

Spike-Based Acoustic Signal Processing Chips for Detection and Localization

Hisham Abdalla

Electrical and Computer Engineering Dept
Institute for Systems Research
University of Maryland
College Park, Maryland, U.S.A
hisham@umd.edu

Timothy K Horiuchi

Electrical and Computer Engineering Dept
Institute for Systems Research
Neuroscience and Cognitive Science Program
University of Maryland
College Park, Maryland, U.S.A
timmer@umd.edu

Abstract—As a summary of the Ph.D. research of the first author we present two analog VLSI spike-based systems; one system is a model of the big brown bat’s (*Eptesicus fuscus*) echolocation system, the second system is a periodicity detector.

I. INTRODUCTION

The use of spiking, or pulsatile, signal representations in the nervous system has many different advantages and disadvantages. In this paper, we demonstrate the use of spike encoding of acoustic signals in two different contexts: ultrasonic intensity signaling and low-frequency time signaling. In particular, we explore how spike timing can play a role in the computations that are performed. First, we explore the use of spikes in the bat auditory system where the intensity of short acoustic sounds must be encoded. Second, we show the use of the spiking representation in the detection of a periodic sound. Portions of this work have been previously published [1-3].

II. ULTRASONIC LOCALIZATION

A. Modeling Bat Echolocation

Bat echolocation is a form of auditory localization. Echolocating bats, as in other mammals, determine the direction of sounds by the binaural and monaural spectral cues created by the direction-dependent filtering of the pinnae [4], head, and body. Unlike in other mammals, however, the sound being localized is an echo of the ultrasonic vocalization (‘ping’) emitted by the bat itself. This ultrasonic ping is a short-duration ($1ms - 20ms$), multi-harmonic chirp [5] that allows the bat to detect the range, direction, and shape of small flying insects and other objects. The use of short wavelengths ensures a strong reflection from small objects and allows the small head of the bat ($\sim 1.5cm$) to produce direction-dependent filtering. In bats, the short sound duration means that only a few spikes will be generated. To mimic and understand the bat’s echo localization abilities we have constructed an artificial bat head [3] that produces qualitatively similar,

This work was supported by a grant from the Air Force Office of Scientific Research (FA95500710446) and the National Science Foundation (CCF0347573) and by a contract with DARPA (Air Coupled Microsensors -0001400C0315).

direction-dependent, spectral features in the same ultrasonic frequency range used by the big brown bat (*Eptesicus fuscus*). An ultrasonic cochlea-like filter bank with 16 channels [3] was designed with moderate quality (Q) factor, and 128 spiking neurons convert these signals to spike trains. Using a simple pattern recognition scheme we demonstrate that the population of spiking neurons can be decoded to estimate the azimuth and elevation of ultrasonic chirps.

B. System Block Diagram

The system architecture is shown in Fig. 1. A speaker transmits a short ultrasonic frequency-modulated (FM) sweep. The sweep is acoustically filtered by the head-related transfer function (HRTF) and the microphones (inserted at end of the ear canals) transduce the acoustic signal into an electrical signal. This signal is then amplified and fed into the cochlea chip. The cochlea chip is a 1-dimensional array of 16 bandpass filters; the output of each filter is half-wave rectified and fed into 8 neurons with adjustable thresholds. A 2D address-event arbiter is used to transmit the spikes off of the chip.

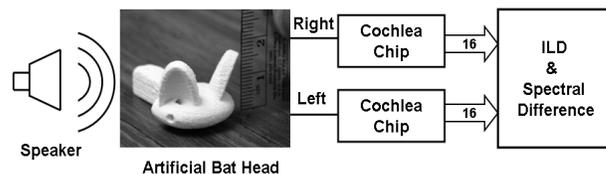


Fig. 1. System block diagram. A speaker emits an ultrasonic sweep. The acoustic signal is filtered by the head-related transfer function (HRTF). The two (right & left) microphones (inserted at the back of the head) generate the electrical signals that are amplified and stored (not shown). The signals are played back to the cochlea chip. A single chip was used to simulate both left and right cochleae. Two localization cues (binaural interaural level difference (ILD) cue and monaural spectral cue) were extracted from the spiking output to estimate the direction of the sound.

C. Artificial Bat Head

The artificial bat head, (see Fig. 1) was designed with SolidWorks™ software and fabricated using the ZCorp 310 3D printer. The head has an elliptical cross section with a height of $10mm$ and a width of $25mm$. The two pinnae are pointed outwards by 45 degrees and tilted forward by 20

degrees. The pinna cavity has a height of 12.6mm , a width of 9mm , and a depth of 2.5mm . The ear canals (holes inside the base of each pinna) lead to microphone (Knowles FG26163) mounting holes at the back of the head. Other mounting holes are for head positioning during characterization.

A 5ms hyperbolic FM frequency sweep ($120\text{kHz}-20\text{kHz}$) was played from a speaker at a distance of approximately 83cm from the bat head. Recordings were made at 703 different directions (19 elevations \times 37 azimuths). The speaker was scanned in the horizontal plane (azimuth) from -90 to 90 degrees in steps of 5 degrees and in the vertical plane (elevation) from -67.5 to 67.5 degrees in steps of 7.5 degrees. The right and left microphone signals were amplified and recorded. Shown in Fig. 2 is the spatial contour plot of the magnitude of the *Fast Fourier Transform (FFT)* of the sound received at the right ear at different frequencies. As the frequency increases the contour plot becomes more irregular.

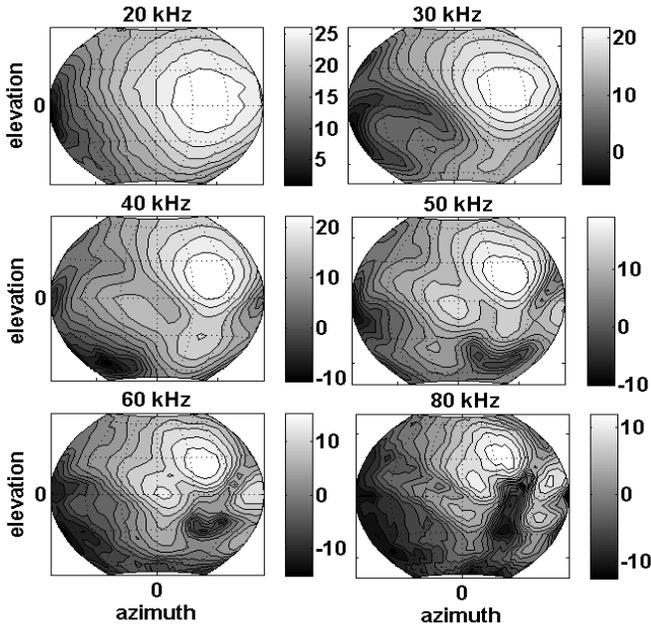


Fig. 2. Spatial contour plot of the magnitude (in dB) of the FFT of the sound received at the right ear at different frequencies. The grid lines have a separation of 30 degrees.

D. Cochlea Chip

Shown in Fig. 3 is the block diagram of the cochlea chip with spiking neurons. The cochlear filter bank consists of 16 bandpass filters. The filter's output is half-wave rectified by a p-type current mirror which is then mirrored to all eight integrate-and-fire neurons associated with this filter. The neuron circuit was designed to integrate charge and produce a pulse event when the membrane voltage exceeds the threshold. Neurons in the same column have the same threshold.

The cochlea chip was fabricated using a commercially-available $0.5\mu\text{m}$ 2-poly, 3-metal process and consumes approximately $36\mu\text{W}$. All testing was conducted at a supply voltage of 5V . The tuning curves for the cochlear filters are

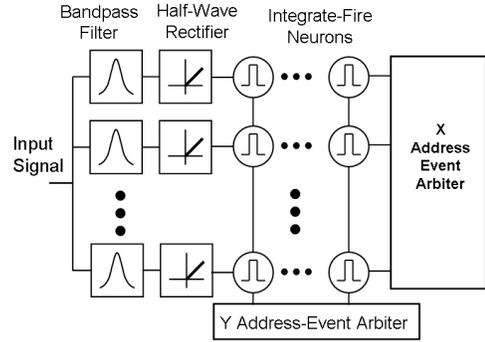


Fig. 3. Block diagram of the cochlea chip. The cochlea is a 1D array of 16 bandpass filters. The filter output voltage is converted to a current, half-wave rectified by a p-type current mirror, and mirrored to all eight integrate-and-fire neurons associated with that filter. The spikes are transmitted off of the chip using a 2D address-event arbiter.

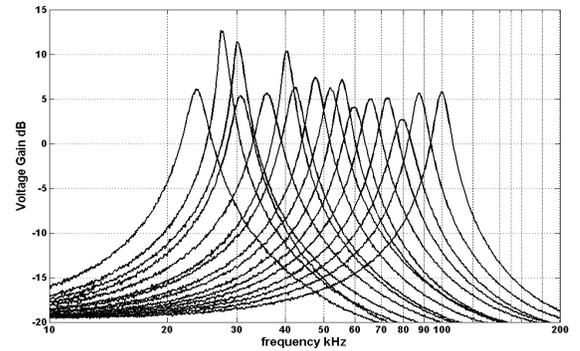


Fig. 4. Cochlear frequency tuning curves: The filters have a mean quality factor (Q) of 11.7 and a standard deviation of 3.2.

shown in Fig. 4. In Fig. 5 we show the spike output in response to a 5ms hyperbolic sweep.

E. Data Analysis

To demonstrate that the HRTF and cochlear spike encoding can be used to estimate azimuth and elevation of a sound source (echo), a binary encoding of the binaural cochlear response was used. The right and left cochleae each consist of sixteen filters. Let $R(i)$ and $L(i)$ denote the total number of spikes generated by the i^{th} right and left cochlear filters respectively. Two cues were extracted: an interaural level difference (ILD) cue, and a monaural spectral difference cue. The binary ILD cues were extracted by evaluating the following logical expressions:

$$\begin{aligned} R(i) - L(i) &> 0, & i = 1, 2, \dots, 16 \\ L(i) - R(i) &> 0, & i = 1, 2, \dots, 16 \end{aligned}$$

The logic states "true" and "false" were assigned the values of +1 and -1 respectively. The binary monaural spectral cues were extracted by evaluating the following logical expressions:

$$\begin{aligned} R(i+1) - R(i) &> 0, & i = 1, 2, \dots, 15 \\ L(i+1) - L(i) &> 0, & i = 1, 2, \dots, 15. \end{aligned}$$

The result is a 62 bit code (using +1 and -1) for each direction. For the 703 directions (see section II-C), 689 different codes

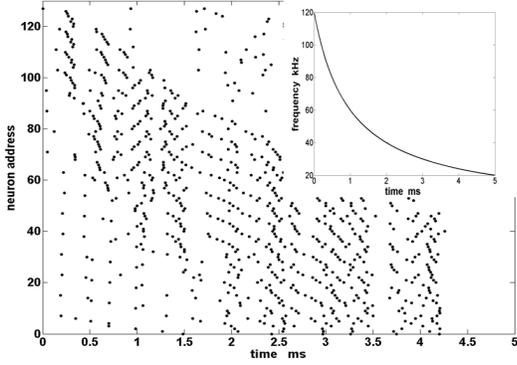


Fig. 5. Spike Raster: Response to a 5ms hyperbolic sweep (see inset). Spike timing is relative to the first spike. Neurons 0-7 correspond to the filter with the lowest center frequency, neurons 120-127 correspond to the filter with the highest center frequency.

were observed. A spatial correlation plot can be constructed by cross-correlating a code from a given direction with codes from all other directions. The maximum correlation is 62 for identical codes and decreases by 2 for every 1 bit difference. Fig. 10 shows spatial correlation plots for nine example directions. The key observation is that codes that differ by only a few bits will code for neighboring directions.

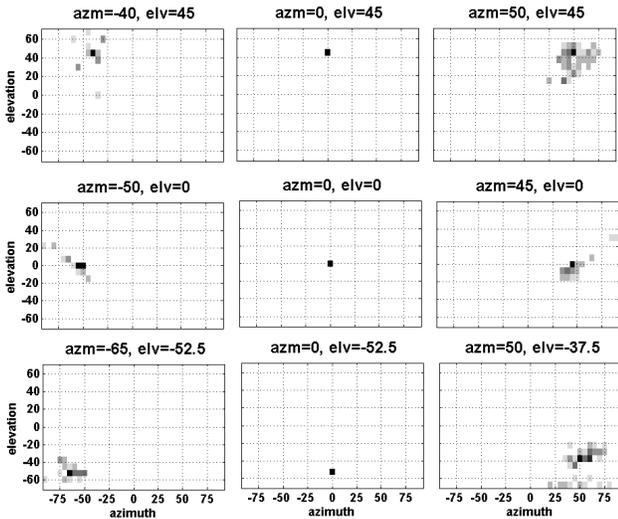


Fig. 6. Spatial correlation plots for nine directions (azm=azimuth, elv=elevation). The code from a given direction is cross-correlated with codes from all other directions. The darker the color, the greater the correlation. Only the correlations for the 7 closest codes (i.e. correlation ≥ 48) are shown to emphasize the neighborhood relationships.

III. PERIODICITY DETECTOR

A. System Block Diagram

Voices and engines typically have a strong harmonic acoustic signature that drives most frequency domain voice and engine detection algorithms. In smart sensor networks where power consumption is a major constraint, having an ultra-low power “intelligent” technique of turning on the

sensor to check for changes in the surrounding environment could dramatically reduce power consumption. While low-power cochlea chips with inter-frequency comparisons [6] can be used to detect unknown harmonic signals, a dedicated time domain approach would likely result in even lower power consumption. Goldberg et. al. [7] demonstrated a low power time-domain approach based on autocorrelation. In our system, however, we use a time-domain-based approach that utilizes the observation that the periodicity of the time-domain envelope occurs at the fundamental frequency.

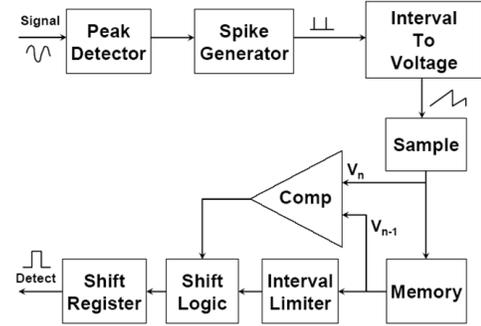


Fig. 7. System block diagram of the peak-periodicity detection chip.

Fig. 7 shows the block diagram of the circuit. A microphone signal, externally amplified and lowpass filtered, is sent into the chip to a peak-detector/spike-generator circuit. The ‘spike generator’ produces a voltage spike at the peaks of the signal envelope. The ‘Interval to Voltage’ stage linearly converts (via a voltage ramp) the interspike intervals into a voltage. This interval voltage is sampled (V_n) and compared (‘Comp’ block) with the previous interval (V_{n-1}) that is stored in ‘Memory’. The sampled interval is subsequently transferred to memory in the period between spikes. Because we are only interested in a limited range of frequencies for a particular application, a “valid interval” signal is generated by the ‘Interval Limiter’ circuit. Consecutive valid intervals that match generate a “hit” pulse that is used to advance a shift-register-based counter. Output taps on the shift register indicate when three-consecutive hits and five-consecutive hits have occurred. When non-matching intervals are encountered, the counter is reset. Fig. 8 shows the timing diagram for the first four stages.

In the presence of noise, there is an increased probability of getting more than one spike at the peak of the signal. The solution to this problem was to include a refractory period in the ‘spike generator’ circuitry. The effect of including a refractory period on the probability density function (*pdf*) of the interspike interval can be seen in Fig. 9. With a $20Hz$ sinusoidal input, the *pdf* of the interspike interval should be centered at $50ms$, however, in the absence of a refractory period more than spike is generated at the peak of the signal resulting in short interspike intervals followed by a long one.

B. Test Results

The chip [2] was fabricated in a commercially-available 2-poly $1.5\mu m$ CMOS process and occupies an area of about

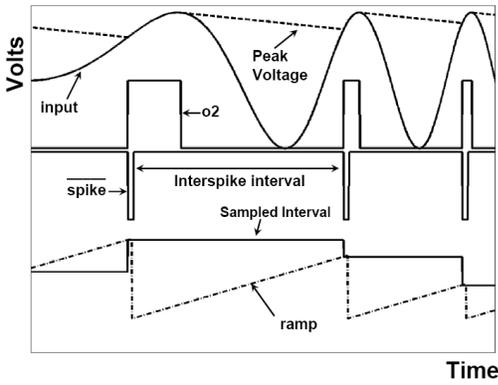


Fig. 8. Timing diagram of the first four stages: The peak voltage (dotted line) follows the peaks of the signal. o_2 is an internal signal of the peak detector. The rising edge of o_2 marks the onset of a peak in the envelope and is used to generate a spike. The ramp (dotted line) converts the interspike interval into a voltage. Sampled interval is the voltage representation of the current interspike interval. ramp is then reset to encode a new interval.

0.242 square mm. The chip was operated at a supply voltage of 3V and consumes less than $1.8\mu W$. It consists of 215 transistors and 9 capacitors. The analog processing blocks of the circuit account for 92 of the transistors; the logic and shift register blocks account for the rest. The response of the chip to the speech utterance "zero" is shown in Fig. 10, the three-consecutive hit output is shown to trigger in the voiced segments of the utterance and not in the noisy onset.

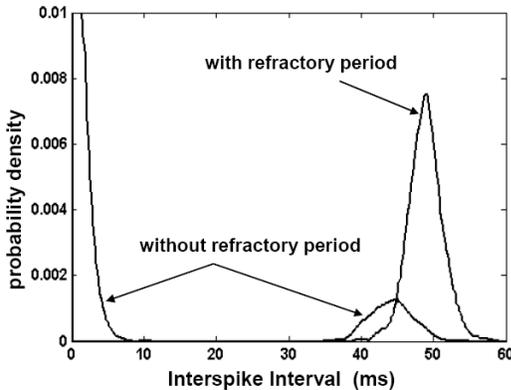


Fig. 9. Effect of the refractory period ($T_{refr} = 11.8ms$) on the pdf of interspike intervals in the signal plus noise case. The signal is a 20Hz sinusoid (SNR= 15dB).

IV. DISCUSSION

The implementation of analog VLSI spike-based systems using subthreshold CMOS circuits could form the basis of extremely low-power systems for use in many different applications.

In the bat modeling system we have constructed an artificial bat head that produces direction-dependent spectral cues in the ultrasonic frequency range where bat echolocation operates. We have also designed and tested an ultrasonic cochlea and we encode the spectrotemporal response with a population of spiking neurons. Using a binary encoding of the population, a

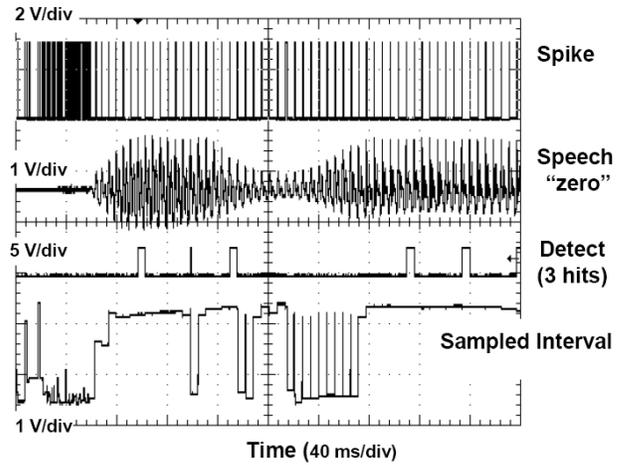


Fig. 10. Response to the speech utterance "zero". No detection during "noisy" onset. Detects the voiced segments.

mapping between the population response and different directions was demonstrated. The binary encoding of the binaural cochlea response in its current form, however, is obviously simplified and will not be completely invariant with changes in amplitude but demonstrates that sufficient information exists in the coding to localize in elevation and azimuth using binaural spectral and intensity cues. By utilizing all of the rich information in the individual neurons and their spike timing, we predict significant improvements in amplitude invariance. In the periodicity detection system we demonstrated the use of the spiking representation in the detection of a periodic sound, and illustrated how the incorporation of the biological-inspired refractory period improves the detection of periodic signals in the presence of noise.

ACKNOWLEDGMENT

We thank Francois Guimbretiere and Hyunyoung Song for their assistance in fabricating the bat head and for the use of the ZCorp 3D Printer. We thank Cynthia Moss and Murat Aytekin for the HRTF measurement apparatus.

REFERENCES

- [1] H. Abdalla and T.K. Horiuchi, "An ultrasonic filterbank with spiking neurons", in *Proc. IEEE Int. Symp. on Circuits and Systems*, vol. 5, pp. 4201-4204, 2005.
- [2] H. Abdalla and T.K. Horiuchi, "An Analog VLSI Low-Power Envelope Periodicity Detector", *Circuits and Systems I, IEEE Transactions on*, vol. 52, pp. 1709-1720, 2005.
- [3] H. Abdalla and T.K. Horiuchi, "Binaural Spectral Cues for Ultrasonic Localization", in *Proc. IEEE Int. Symp. on Circuits and Systems*, pp. 2110-2113, 2008.
- [4] M. Aytekin, E. Grassi, M. Sahota, C.F. Moss, "The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation," *J. Acoust. Soc. Am.*, vol. 116, pp. 3594-3605, 2004.
- [5] A. N. Popper and R. R. Fay, "Hearing by Bats," in *Springer Handbook of Auditory Research*, vol. 5. New York: Springer-Verlag, 1995.
- [6] A. van Schaik, "An Analog VLSI Model of Periodicity Extraction in the Human Auditory System," *Analog Integrated Circuits and Signal Processing*, vol. 26, no. 2, pp. 157-177, Feb. 2001.
- [7] D.H. Goldberg, A.G. Andreou, et. al., "A Wake-Up Detector for an Acoustic Surveillance Sensor Network: Algorithm and VLSI Implementation," *Int. Symp. on Information Processing in Sensor Networks*, pp. 134-141, April 2004.