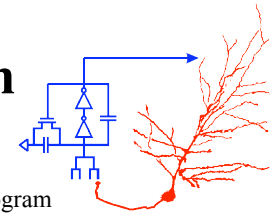# Binaural Spectral Cues for Ultrasonic Localization

## Hisham Abdalla[1,2] Timothy K. Horiuchi[1,2,3]

[1]Electrical and Computer Engineering, [2]Institute for Systems Research, [3]Neuroscience and Cognitive Sciences Program

Introduction

To support our ongoing work in modeling bat echolocation, an artificial bat head was designed and fabricated using a 3D printer, an ultrasonic cochlea-like filter bank with 16 channels has been designed with moderate quality (Q) factor, and 128 spiking neurons convert these signals to spike trains. A two-dimensional address-event arbiter is used to transmit these spikes off of the chip. We demonstrate that the population of spiking neurons can be decoded to estimate azimuth and elevation of ultrasonic chirps. This chip was fabricated in a commercially-available 0.5 μm CMOS process.
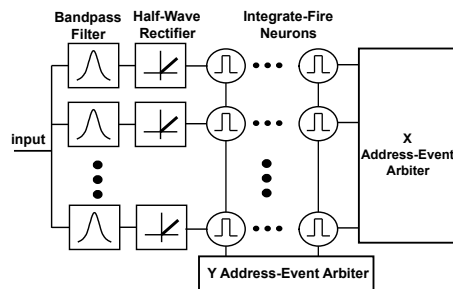
System Block Diagram
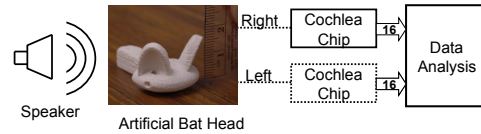
A speaker emits an ultrasonic sweep. The acoustic signal is transformed by the head-related transfer function (HRTF). The two (right & left) microphones (inserted from the back of the head) generate the electrical signals that are amplified and stored (not shown). The signals are played back to the cochlea chip. A single chip was used to simulate both left and right cochleae.

Artificial Bat Head

The artificial bat head was designed with SolidWorks™ software and fabricated using a 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 FG6163) mounting holes at the back of the head. Other mounting holes are for head positioning during characterization.

Cochlea Chip

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.
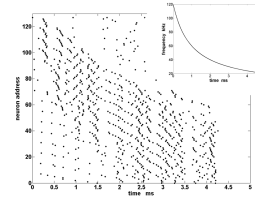
Chip Testing Results

Cochlear Filter Tuning Curves
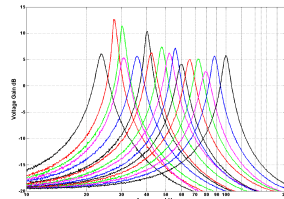
Cochlear frequency tuning curves: The filters had a mean quality factor (Q) of 11.7 and a standard deviation of 3.2.

**Neural Spikes (Spike Raster)**

Spike Raster: Response to a 5ms FM hyperbolic sweep (see inset). Spike timing is relative to the first spike.

HRTF Measurement

A 5ms hyperbolic FM frequency sweep (120 kHz-20 kHz) was played from a speaker at a distance of approximately 83cm from the bat head. 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.
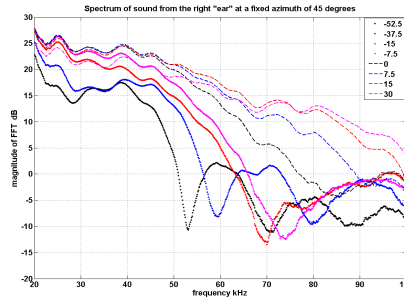
Spectrum of sound from the right "ear" at a fixed azimuth of 45 degrees and different elevations. As the elevation increases, the spectral shape changes significantly.
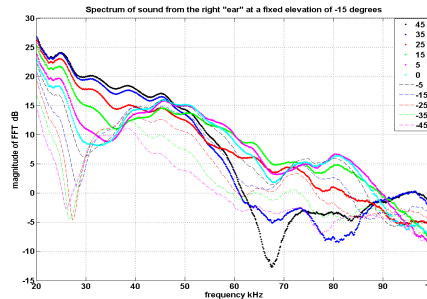
Spectrum of sound from the right "ear" at a fixed elevation of -15 degrees and different azimuths.
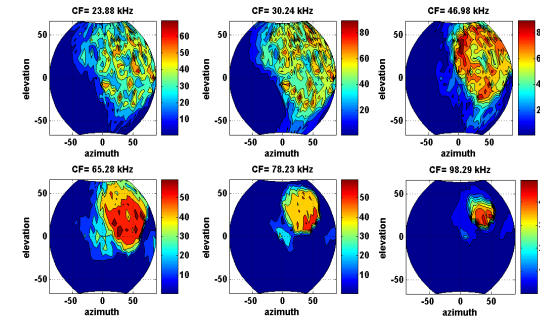
Receptive fields: The receptive fields of six filters with different center frequencies (CF). The color code indicates spike count (sum of all spikes for a given filter). Note that the color bars are different. Filters with higher CF have a more localized response.

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* cochleae response was used. Let R(i) denote the total number of spikes generated by the ith right cochlear filter and L(i) that of the ith left cochlear filter. 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:

$$R(i) - L(i) > 0, \qquad i = 1,2,...16$$
$$L(i) - R(i) > 0, \qquad i = 1,2,...16$$

where i is the filter index. The logic state "true" was assigned a value of +1 while the logic state "false" was assigned a value of -1. The binary monaural spectral cues were extracted by evaluating the following logical expressions:

$$R(i+1) - R(i) > 0, \quad i = 1,2,...15$$
$$L(i+1) - L(i) > 0, \quad i = 1,2,...15$$

The result is a 62 bit code (using ±1) for each direction. Recordings were made at 703 (19 elevations x 37 azimuths) different directions. For the 703 directions, 689 different codes 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.
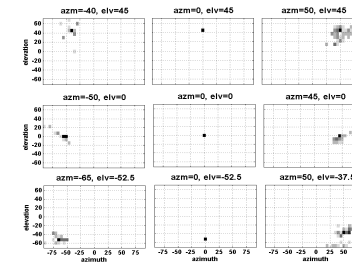
Spatial correlation plots for nine different directions where azm denotes azimuth and elv denotes elevation. The 62 bit code for a given direction is cross-correlated with codes from all other directions. The darker the color the greater the correlation. The maximum correlation is 62 for identical codes and decreases by 2 for every 1 bit difference. Only the correlations for the 7 closest codes (i.e. correlation ≥48) are shown to emphasize the neighborhood relationships.