ABSTRACT

Title of dissertation: Estimation of Elevation and Azimuth in a Neuromorphic VLSI Bat Echolocation System

Hisham Abdalla, Doctor of Philosophy, 2009

Dissertation directed by: Professor Timothy K. Horiuchi
Department of Electrical and Computer Engineering

Auditory localization is an interesting and challenging problem; the location of the sound source is not spatially encoded in the peripheral sensory system as it is in the visual or somatosensory systems. Instead it must be computed from the neural representation of the sound reaching both ears. Echolocation is a form of auditory localization, however, an important distinction is that the sound being localized is an echo of the sound emitted by the animal itself. This dissertation presents a neuromorphic VLSI circuit model of a bat echolocation system. The acoustic cues that we use in our system are the binaural interaural level differences (ILDs) and the monaural spectral cues. We have designed an artificial bat head using 3D CAD software and fabricated it using a 3D printer. The artificial bat head is capable of generating the necessary acoustic cues for localization. We have designed and fabricated an ultrasonic cochlea chip with 16 cochlear filters and 128 spiking cochlear neurons (eight neurons per cochlear filter), the cochlear filters and neurons transform the analog input into a spike-based cochlear representation.

We have also designed and fabricated two feature extraction chips: a monaural spectral difference chip and a binaural ILD chip, that together can extract the localization cues from the spike-based cochlear representation. The monaural spectral difference chip consists of 240 spiking neurons; each neuron compares the activity of two cochlear filters within the same ear. The binaural ILD chip consists of 32 spiking neurons (two per cochlear filter) that model the processing that takes place in the lateral superior olive (LSO). We demonstrate that the spatiotemporal pattern of spiking outputs from the feature extraction chips can be decoded to estimate the direction (elevation and azimuth) of an ultrasonic chirp.

Estimation of Elevation and Azimuth in a Neuromorphic VLSI Bat Echolocation System

by

Hisham A. N. Abdalla

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:

Professor Timothy K. Horiuchi, Chair/Advisor
Professor Pamela Abshire
Professor Amr Baz
Professor Carol Espy-Wilson
Professor Shihab Shamma

DEDICATION

To my parents, family, and friends.

ACKNOWLEDGMENTS

I would like to express my deep and sincere gratitude to my advisor Dr. Timothy Horiuchi for his invaluable guidance and support during the course of this work. I have benefitted greatly from his knowledge, experience, and intuition. His understanding, mentorship, and trust were essential for the work in this dissertation. I have been very fortunate to have him as my advisor.

I would like to thank Dr. Pamela Abshire, Dr. Amr Baz, Dr. Carol Espy-Wilson, and Dr. Shihab Shamma for serving on my dissertation committee.

I would like to thank Dr. Francios Guimbretiere and Hyunyoung Song for their assistance in fabricating the artificial bat head and for the use of the ZCorp 3D printer and I thank Dr. Linda Schmidt for the use of the Dimension SST 3D printer. I thank Dr. Cynthia Moss, Dr. Murat Aytekin, and Dr. Kaushik Ghose from the Auditory Neuroethology Lab (Bat Lab) for the HRTF measurement apparatus, technical discussions and the big brown bat HRTF data. I thank Dr. Jonathan Simon and Dr. Jonathan Fritz for always being there to answer my questions. I thank Dr. Pamela Abshire, Nicole Nelson and Mark Dandin from the IBIS Lab for their assistance in GPIB programming and for photographing my chip. I thank my friends Dr. Moustafa Al Bassyiouni and Dr. Adel Elsabbagh for introducing me to the 3D CAD software SolidWorks which was used to design the artificial bat head. My special thanks goes to my friend Dr. Mohammed Fahmi

whom introduced me to this wonderful program, Scientific Word, which I used to write this dissertation.

I have been very fortunate to have the opportunity to work with my labmates Matt Cheely, Dr. Rock Shi, Tarek Massoud, and Chetan Bansal. I thank them for their support and assistance. I also thank Dr. Mounya El Hilali, Dr. Nima Mesgarani, and Dr. Steven David from the Neural Systems Lab. I thank Dr. Timothy Horiuchi, Kate McRoberts, Chetan Bansal, and Michael Kuhlman for proofreading parts of my dissertation. Any mistakes that remain are solely my responsibility.

Last but not least, I thank my parents who encouraged me to learn and who have always and continue to support me, and finally, I would like to acknowledge the understanding, patience, sacrifice, encouragement, and love of my wife.

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).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Auditory localization is an interesting and challenging problem; the location of the sound source is not spatially encoded in the peripheral sensory system as it is in the visual or somatosensory systems. Instead it must be computed from the neural representation of the sound reaching both ears. Echolocation is a form of auditory localization, however, an important distinction is that the sound being localized is an echo of the sound emitted by the animal itself.

1.2 Auditory Localization Cues

To unambiguously determine the direction of a sound source requires that a unique direction-dependent acoustic signature (*fingerprint*) be imprinted onto any arriving sound. Studies have shown that there are three major cues that are used in

auditory localization: interaural time differences (ITDs), interaural level differences (ILDs), and monaural spectral cues generated by the external ear (pinna). It is thought that ITDs and ILDs cues are the primary cues for horizontal sound localization (azimuth) whereas the primary cue for estimating elevation is the monaural spectral cue.

1.2.1 Interaural Level Difference (ILD)

The interaction of a sound wave with a head (obstacle) depends on the wavelength of the sound compared to the dimensions of the head. If the wavelength of sound is bigger than the size of the head, the sound wave diffracts around the head and the intensity difference between the two ears is small, however, if the wavelength of sound is smaller than the size of the head, the sound wave does not diffract as much resulting in an *acoustic shadow* and thus an intensity difference between the two ears. As a result, the frequencies at which the head is effective in shadowing is inversely proportional to the size of the head. This localization cue is referred to as the interaural level difference (ILD) or the interaural intensity difference (IID).

In humans, the frequencies at which the head can act as an effective acoustic shadow are above 2-3kHz [1] and the maximum ILD has been measured in human subjects to be approximately 20dB at 4kHz and 35dB at 10kHz [2][3]. In one study of the Mongolian gerbil *Meriones unguiculatus* where the separation between the two ears was approximately 12mm, the maximum ILD was measured to be 30dB at 30kHz [4]. In another study, the maximum ILD for the big brown

bat *Eptesicus fuscus* was measured to be 35dB at $45kHz$ [5].

1.2.2 Interaural Time Difference (ITD)

Before the end of the 19^{th} century, Rayleigh mathematically showed that for low frequencies, interaural intensity differences are small and hypothesized that interaural phase (time) differences may explain the localization acuity of humans [6]. The interaural time difference (ITD) cue, hypothesized by Rayleigh, stems from the fact that the speed of wave propagation is finite and that, in general, the sound source will not be equidistant to both ears. As a result, sound will generally arrive at both ears at slightly different times. Because this interaural time difference is related to the separation between the two ears, this cue scales directly with the size of the head. The "duplex theory of localization", attributed to Rayleigh, held that ITD and ILD are used for localization, and that the ITD cue dominates at low frequencies whereas the ILD cue dominates at high frequencies.

In the big brown bat, *Eptesicus fuscus*, because of its very small head, the maximum ITD is on the order of $50\mu s$ and it is believed that this cue is not being used for localization [7].

1.2.3 Spectral Cues

Rayleigh noted that because of the symmetry of the two ears, interaural intensity differences of pure tones would be insufficient to explain the ability to discriminate between front and back, and conjectured that "compound sounds" (such as the

human voice) may be modified by the external ear (pinna)[6]. Nevertheless, most of the early research on auditory localization focused on studying the intensity differences between the two ears as the cue for localization [8].

The role of the pinna in localization has received much more attention in the second half of the 20^{th} century (for a review of this work see [3]). Using real human ears and a rubber replica, Shaw and Teranishi measured the response of the external ear as a function of frequency and for different angles of incidence and showed that the pinna modifies the spectrum of the sound that enters the ear and that this modification is direction-dependent [9].

1.3 Echolocation

Bats and dolphins are known to use echolocation, however, not all bat species echolocate and contrary to common belief that bats have poor vision, some have excellent night vision [10].

Echolocating bats differ in their vocalizations; some bats emit a long (10-100ms) constant frequency (CF) sound followed by a frequency modulated (FM) sweep and are called long CF/FM bats, a second group emits a short (8-10ms) CF sound followed by a FM sweep and are called short CF/FM bats, and a third group emits just a FM sweep and are known as FM bats [10]. The intensity of echolocation sounds ranges from less than 60dB SPL (sound pressure level) to louder that 110dB SPL [11].

CF/FM bats have a large fraction of their cochlea dedicated to the frequency

range near the second harmonic of its CF component [12] and it has the ability to lower the frequency of its CF component so as to adjust the CF component of the echo (compensating for Doppler shifts in the returning echo) to fall within its range of maximum sensitivity. This enables the CF/FM bat to hunt in cluttered environments where it can distinguish between moving and stationary objects.

The bat that we are interested in modeling is the echolocating big brown bat *Eptesicus fuscus*. It emits wideband, ultrasonic, frequency modulated (FM) sounds. It uses echolocation for navigation as well as for detecting and capturing prey.

Behavioral studies have shown that *Eptesicus fuscus* has very sharp localization capabilities, its horizontal angle acuity is $\simeq 1.5^\circ$ [13], and its vertical angle acuity is $\simeq 3^\circ$ [14]. It has been shown that the big brown bat can manipulate the duration as well as the bandwidth of its frequency sweep [15] and can discriminate between artificial wing beat stimuli that flutter at different rates [16].

1.3.1 Echo Delay

If a target is at a distance r meters away, then the time $\triangle t$ it takes for an echo to return back to a bat is given by

$$\triangle t = \frac{2r}{c} \tag{1.1}$$

where c is the speed of sound in air ($340 m/s$). This time delay is approximately $5.8 ms/m$ (or $17 cm/ms$). For example, the time delay for a target located $5m$ away is $29ms$.

1.3.2 Atmospheric and Spreading Losses

The ultrasonic sound emitted from the bat undergoes three types of attenuation (losses) as it propagates in space: spreading loss, atmospheric absorption, and losses due to the reflective properties of the target.

The spreading loss is the loss of intensity due to the expansion of the sound in space. If the emitted sound is assumed to have a spherical wavefront, then the intensity $I(r)$ of the sound incident on a target can be written as

$$I(r) = \frac{I_o}{4\pi r^2} \tag{1.2}$$

where I_o is the intensity of the emitted sound and r is the range of the target.

The size of the target relative to the wavelength of the incident sound will affect how much energy is reflected back to the bat. If the target is small in size compared to the wavelength of sound, it can be treated as a point source and the reflected sound will be reradiated in a spherical wavefront, the total spreading loss in this case is given by $(I(r))^2$. It can be shown that spreading loss increases by 12dB for every doubling of the range. If the target is a large planar surface (e.g. a wall), then the return path is treated as if the incident sound continued to propagate in its initial direction but for twice the range of the target, i.e. the total spreading loss is given by $I(2r)$.

The absorption of sound propagating through the atmosphere is another source of attenuation that is frequency dependent and has been measured [17] to be 0.7dB/m (i.e. by a factor of 0.85/m) at 30kHz, and 2.7dB/m (factor of 0.54/m) at 90kHz. One study showed that the big brown bat *Eptesicus fuscus*

can detect insect-sized targets up to a range of 3-5 meters [18].

1.3.3 Intensity Compensation

As a bat closes in on its target the intensity of the received echoes increase as a result of the decreasing attenuation from both spreading losses and atmospheric absorption. The increase in intensity, if not compensated for, causes a reduction in the latency of auditory neurons, which may affect the range computation as well as the localization of sound. There appears to be agreement that the big brown bat *Eptesicus fuscus* compensates for spatially-dependent changes in echo intensity, however, there is a disagreement on how that is being done.

Kick and Simmons [19] tested the sensitivity of *Eptesicus fuscus* to sonar echoes at different time delays (ranges) by testing its ability to detect a phantom target at a specific range as a function of echo intensity and showed that the bat's echo detection threshold increases by approximately 11dB for each reduction in target range by a factor of two for echo delays ranging from $1ms$ to $6.4ms$ ($17cm$ to $110cm$) suggesting that this change in sensitivity acts as an automatic gain control stabilizing the encoded intensity of the echoes at the level of the cochlea input as the bat approaches its target.

Hartley [20] criticized the previous experiment suggesting that backward masking of the phantom target by the "real" echo from the loudspeaker itself may have resulted in the overestimation of the magnitude of the automatic gain control. Hartley conducted several experiments wherein he estimated that the automatic

gain control results in a reduction of sensitivity by 6 to 7dB per halving of target range and that *Eptesicus fuscus* reduced the intensity of the emitted sound by 6dB per halving of target range [21].

Interestingly, Simmons et al. [22] repeated the experiment by Kick and Simmons, however, they used an electronic target simulator rather than moving the loudspeaker to avoid backward masking. They showed that the bat's sensitivity increased by 11.5dB when the echo delay changed from $2.3ms$ to $4.6ms$.

In summary, one group of studies attributes the intensity compensation solely to changes in the bat's hearing threshold while another study shows that in addition to changes in hearing threshold the bat changes the intensity of the emitted sound.

1.4 Neuromorphic Modeling

Biological information-processing systems are different from digital computers in the way they execute their computation. For the most part, general purpose digital computers have an essentially fixed architecture for all computations and rely on different software programs to implement different algorithms to solve different problems. The computation is controlled by a central processor and the program is executed serially in a step-by-step fashion through fetch-execute cycles. On the other hand, in biological information processing systems, the architecture effectively defines the computation and thus it differs according to the desired computation. Neither the computations nor the memory used in the computation

is centralized and some of the computation is performed in parallel. This type of system is usually referred to as a "parallel distributed processing system". It should be noted that computations in the brain are neither entirely analog nor entirely digital but a mixture of both.

Neuromorphic (derived from the words "neuron" and "morphology") engineers aspire to incorporate biological information-processing architectures and algorithms into the design of engineering systems that are powerful (computationally), efficient (lower power consumption), and robust. Neuromorphic analog VLSI (Very Large Scale Integration) circuits use many of the elementary physical phenomena of microelectronic devices as computational primitives [23]. For example, the nonlinear (exponential) dependence of the drain-current on the gate-voltage in transistors operating in the subthreshold region is by itself a computational primitive, and is used in designing other computational blocks. Many of the neuromorphic-based analog VLSI circuits take advantage of the subthreshold region of operation of a MOS transistor where they can operate with currents many orders of magnitude smaller than the usual above-threshold region, which is the key to low power circuits. In 1998 Sarpeshkar et al. [24] made a comparison between an analog cochlea chip using subthreshold transistors and a digital cochlea (both an Application Specific Integrated Circuit (ASIC) cochlea and a microprocessor-based cochlea); the analog cochlea was two to five orders of magnitude more efficient than the digital implementations (estimated). Sarpeshkar also compared analog systems against digital systems and showed that for low-precision computations, analog computation is "cheaper" (in terms of power con-

sumption) than digital computation [25].

Because neuromorphic analog VLSI designs and neurons are both constrained by what can be implemented in a physical medium, it is strongly believed that they can provide insight into biological computations [26]. The range of applications for neuromorphic VLSI-based chips is huge (from prosthetic devices, such as retinal and cochlear implants, to ultra-low-power sensor networks, to space exploration).

1.4.1 Previous Neuromorphic VLSI Chips for Auditory Processing

There has been a wide range of neuromorphic VLSI circuit implementations modeling the computations that take place in biological auditory systems. In 1988 Lyon and Mead presented the first silicon cochlea chip [27]. Their cochlea design was followed by several improved designs [28][29][24]. Lazzaro and Mead presented an auditory localization chip that models the time-coding pathway in the barn owl's brainstem [30]. Lazzaro and Mead [31] and van Schaik [32] presented VLSI circuits for encoding auditory signals with neurons and hair cells. van Schaik [33] and Jones et al. [34] presented neuromorphic-based systems for periodicity extraction. Horiuchi and Hynna [35][36] and Shi and Horiuchi [37][38][39] have presented VLSI models for azimuthal sound localization in bats while Cheely and Horiuchi presented delay-tuned circuits for estimating range in a bat echolocation system [40].

1.4.2 Earlier Cochlea Designs

There have been many neuromorphic implementations, both analog and digital, of electronic cochlea. In 1988 Lyon and Mead presented the first silicon cochlea chip [27]. Their analog VLSI design consisted of a cascade of 480 second-order filter sections, the model is a unidirectional (feedforward) model where the output of each second order section feeds into the next section. The filters were operated in the subthreshold regime for low power consumption and the current supplied for all 480 stages was approximately $1\mu A$. Watts et al. [28] made several improvements to [27]. They increased the linear input dynamic range by adding source degeneration to the subthreshold transconductance amplifiers and improved the matching of the critical transistors. Their improved cochlea had 51 stages and consumed $11\mu W$. In both Lyon's and Watts' designs the average quality factor (Q) was less than unity. van Schaik et al. improved on the matching of the cascaded second-order design of Lyon and Watts by using Compatible Lateral Bipolar Transistors (CLBTs) [41] to bias the filters instead of the subthreshold MOS transistors [29]. Sarpeshkar et al. [24] designed a 117-stage cochlea, with a 61dB wide-dynamic-range, $10kHz$ bandwidth using an overlapping cascade structure. Their cochlea dissipated $0.5mW$ of power.

Watts [42] noted that the cascaded cochlea model had several drawbacks: the filters were not sharply tuned and that adding more second-order stages to obtain sharper responses increased the phase delay of the signal, voltage offsets (as a result of mismatch) at the output of each filter section may drive the DC op-

erating point of the cascade towards the power supply rails, and that the failure of one filter section would result in the failure of all following sections. Watts developed a two-dimensional *passive* resistor network that models the incompressible cochlear fluid thus allowing coupling between the adjacent filters [42]. The 2D resistive network connects to a one-dimensional array of non-connected basilar membrane circuits. Several researchers have extended Watts' work by implementing two-dimensional VLSI *active* models of the cochlea: Wen and Boahen proposed a novel mechanism called "active bidirectional coupling" (ABC) that models several structures of the Organ of Corti [43]. They demonstrated that their coupling mechanism was capable of transforming a passive cochlea into an active one [44][45], however, mismatch appears to have degraded their chip measurements. Hamilton et al. adopted a higher level approach by including an automatic quality factor control mechanism to their resonator design [46][47][48][49]. It should be noted that all previous cochlea chips operated in the human audible range ($<20kHz$).

In addition to the analog implementations there have been several digital VLSI implementations. In 1992 Summerfield and Lyon presented the first digital VLSI cochlea [50]. Their cochlea was a 71-stage ASIC implementation of a digital approximation to the Lyon/Mead analog cochlea model.

In 2000 Jones et al. [34] presented an FPGA-based cochlea (actually an entire digital neuromorphic pitch extraction system). Their cochlea consisted of a bank of 30 bandpass filters. This was followed by other FPGA designs [34][51]. Clarke et al. presented an FPGA-based ultrasonic cochlea to model the

bat cochlea and is the only VLSI implementation (that the author is aware of) that precedes the author's analog VLSI implementation of an ultrasonic cochlea [52][53].

1.4.3 Address-Event System

In neuromorphic VLSI chips, where it is desirable to have a large number of neurons (and synapses) and the physical number of chip pins is a limited resource, it is inefficient (and sometimes impossible) to dedicate a chip pin for every neuron. Alternatively, the address-event representation (AER) [54] is used where each neuron is assigned a unique binary address and all neurons share a single address bus.

The address-event system is similar to a time-division multiplexing system in that a number of neurons (users) share a single communication resource (the address bus), however, neurons are not assigned a specific time slot where they can broadcast their information on the address bus. Instead, when a neuron fires a spike a request signal is sent to an address-event arbiter. The address-event arbiter handles the requests from all neurons asynchronously, and decides which neuron can send its information on the address bus.

1.5 Statement of Proposed Problem

The problem that we propose to solve is an auditory localization problem: How could the echolocating bat estimate the direction (elevation and azimuth) of an

object from its ultrasonic echo. We propose the design and implementation of a neuromorphic, low-power, wideband, analog VLSI echolocation system that is capable of localizing an object from its echo. The proposed system will use interaural level difference cues and monaural spectral cues to simultaneously estimate both elevation and azimuth. The information processing computations will be spike-based and will process acoustic information in a manner similar to that taking place in the auditory system of the big brown bat, *Eptesicus fuscus*. Since this project is a part of a much larger system, (which is obstacle avoidance and navigation of an autonomous, low-speed, light-weight, flying vehicle) we believe that our neuromorphic analog VLSI design approach can satisfy two important requirements of this system: low-power consumption and light weight physical implementation.

The first task in this project is to design and fabricate an artificial bat head (and pinna) that is capable of generating the necessary acoustic cues for localization. The second task is to design and fabricate an ultrasonic cochlea that transforms the analog input into a spike-based representation. The third task is to design the feature extraction chips that can extract the binaural ILD and monaural spectral cues from the cochlea's spike-based representation and demonstrate that the output of these chips can be decoded to estimate the elevation and azimuth of the target simultaneously.

1.6 Previous Neuromorphic Localization Systems

There have been several neuromorphic bat localization systems. Shi and Horiuchi [37][38][39] presented a VLSI-based hardware system that focused on azimuthal localization using ILD cues. Their system used a narrowband ($\simeq 40kHz$) signal and, therefore, their system did not involve designing neither a pinna nor an electronic cochlea. Their work is similar to the work presented in this dissertation in that (apart from the front end) all of the signal processing is done using VLSI chips. Their system models the neurons of the lateral superior olive (LSO), the dorsal nucleus of lateral lemniscus (DNLL), and the inferior colliculus (IC).

Fountaine and Peremans [55] have recently presented a bat echolocation system that has many similarities to the work presented in this dissertation except that their processing is *software-based*. They used Aytekin's HRTF measurements [56] of the big brown bat *Eptesicus fuscus* to estimate the elevation and azimuth of a sound source (FM Sweep). In their system they used the first spike latency code to represent both monaural spectral cues and binaural ILD cues. Their cochlea design has 80 filters spanning the frequency range from $20kHz$-$80kHz$. Their system models the lateral superior olive (LSO) to extract the binaural ILD cues and they present a neural model for processing monaural spectral cues. Their model is a modified version of an existing model of spectral processing in the dorsal cochlear nucleus (DCN). They also model the neurons in the DNLL and IC.

1.7 Dissertation Organization

The dissertation includes eight chapters including the introductory chapter. Chapter 2 provides a brief overview of parts of the mammalian auditory system relevant to the localization problem. The chapter also addresses several questions: given that the duration of the entire FM sweep of the big brown bat *Eptesicus fuscus* is $\leqslant 20ms$, how are the localization cues encoded in the cochlear spiking output and where are they extracted ? Is the type IV neuron in the dorsal cochlear nucleus (DCN) that has been implicated in the processing of spectral localization cues in non-echolocating mammals still a valid model for the bat ?

Chapter 3 covers the design of the artificial bat head, the design of a microphone amplifier as well as measuring and analyzing the response of the artificial bat head to sounds from different directions. Chapter 4 focuses on the design and testing of the ultrasonic cochlea chip. Chapter 5 focuses on the design and testing of two feature extraction chips: a binaural interaural level difference (ILD) chip and a monaural spectral difference chip.

In Chapter 6, titled Localization, we demonstrate in *software*, that the measured responses from the artificial bat head contain enough information to be used for localization. We then show that the *hardware* system (consisting of the cochlea chips, spectral difference chips and ILD chips) can extract the relevant localization cues as shown in the software analysis and that the spatiotemporal pattern of spiking outputs from the feature extraction chips can be decoded to simultaneously estimate the elevation and azimuth of the sound source.

Chapter 7, which is not related to the localization project, presents earlier work of the author in the Ph.D. program, namely, the design and fabrication of a low-power analog VLSI envelope periodicity detector for acoustic detection problems [57].

In Chapter 8, the last chapter, we present the conclusions of the work presented in the dissertation, mention some of the limitations of the current system and suggest some future directions to extend this work.

Chapter 2

Mammalian Auditory System

2.1 Introduction

The mammalian auditory system is remarkable in its sound processing capabilities. It is remarkably sensitive and can handle sounds that change in intensity by 12 orders of magnitude (120dB)[58]. Our approach to solving the auditory localization problem involves processing the acoustic information in a manner similar to that of the echolocating big brown bat, *Eptesicus fuscus*. Therefore, we begin by reviewing parts of the mammalian auditory system involved in the localization task.

2.2 The Mammalian Ear

The basic structure of a mammalian ear is shown in Fig. 2.1. It can be divided into three parts: The external (outer), middle, and inner ear [59].

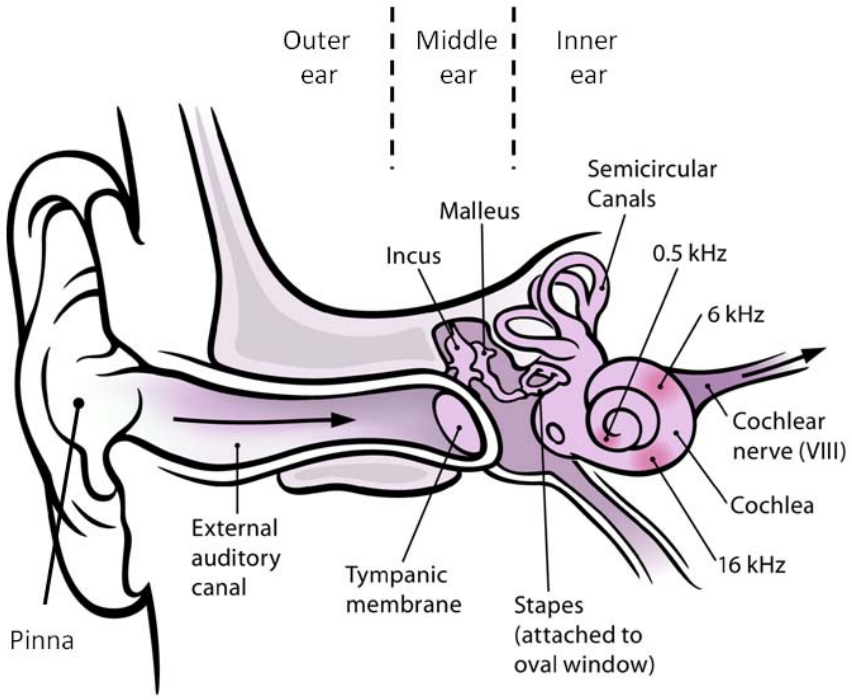

Figure 2.1: Illustration of a human ear. The ear can be divided into three sections: External ear (pinna), middle ear, and inner ear (cochlea). Reprinted from [60].

2.2.1 The External Ear

The pinna is part of the external ear. As the incident sound wave first encounters the pinna (as well as the head), it is filtered and funneled into the ear canal. The pinna acts as a direction-dependent linear filter. Its transfer function, referred to as the 'head-related transfer function' (HRTF), is measured by playing sounds from different directions and measuring the filtered sound close to the eardrum. The measured response is then compared against the free-field sound. HRTFs have been computed numerically for human-sized heads [61][62] [63][64], and have been measured experimentally in: humans [65], cats [66], bats [56] as well as in other animals. Figs. 2.2 and 2.3 show the measured HRTF of a cat. The HRTF in Fig. 2.2 is shown for three different elevations at a fixed azimuth whereas the HRTF in Fig. 2.3 is shown for three different azimuths at a fixed elevation. The important observation to notice in Fig. 2.2 is that there is a notch in the transfer function in the 5-18kHz region (first notch (FN) region) that changes systematically with elevation. In Fig. 2.3, a notch also exists in the spectrum that changes systematically with azimuth. There is also a change in the low-frequency gain with azimuth. Fig. 2.4 shows the contours of constant first notch (FN) frequency derived from the head related transfer function (HRTF) of the cat. The solid lines show iso-FN contours for the right ear, and the dashed lines are mirror image contours representing iso-FN contours for the left ear (assuming perfect symmetry between both ears). Rice et al. [66] proposed a localization scheme based on the knowledge of FN frequencies in the two ears; they called it the "binaural first

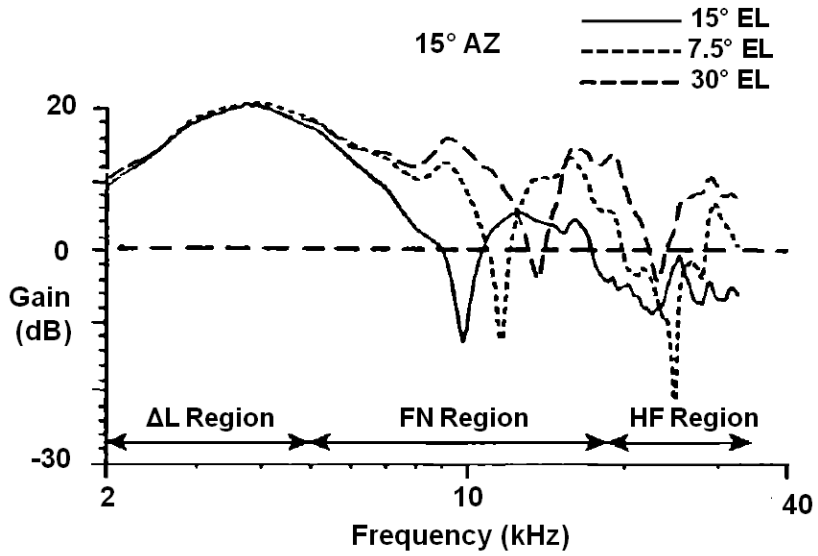

Figure 2.2: Head related transfer function (HRTF) of a cat at an azimuth of 15 degrees for three different elevations. Reprinted from [66] with permission from Elsevier.

notch direction code" or "BiFiND". This localization scheme allows simultaneous estimation of both azimuth and elevation. Similar to cats, Fig. 2.5 shows the iso-frequency contour lines (computed from the HRTF measured by Aytekin [56]) for the primary notch in the big brown bat *Eptesicus fuscus*. It should be noted, however, that spectral notches are not the only source of information for estimating elevation and azimuth; Aytekin observed that the interaural level difference (ILD) shows both azimuth and elevation dependence and proposed that *both* the ILD and spectral notches contribute to localization of sound source elevation and azimuth [56].

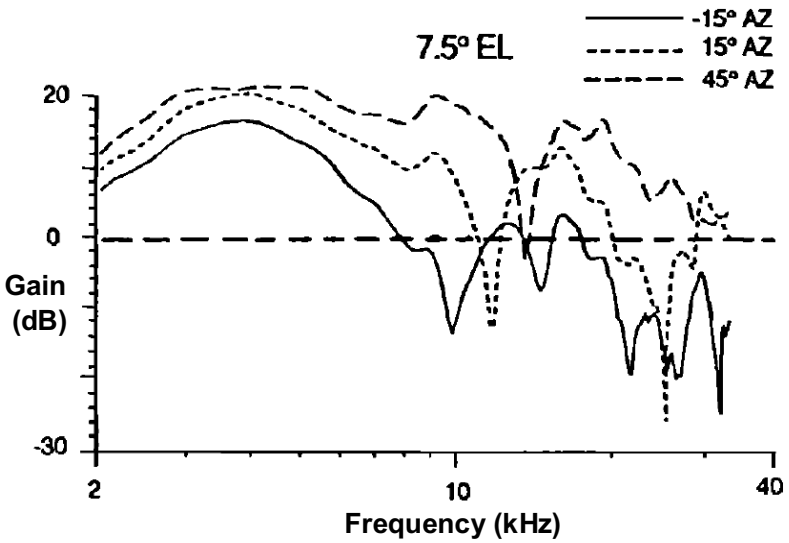

Figure 2.3: Head related transfer function (HRTF) of a cat at an elevation of 7.5 degrees for three different azimuths Reprinted from [66] with permission from Elsevier.

2.2.2 The Middle Ear

The middle ear is responsible for coupling the sound energy into the fluid filled cochlea (inner ear). The ear canal terminates with the *tympanic membrane* (ear drum). The eustachian tube equalizes the pressure on both sides of the eardrum. The chain of three linked bones, the *malleus*, *incus*, and *stapes* form the *ossicles*. The *malleus* makes contact with the *tympanic membrane* while the *stapes* makes contact with the oval window of the cochlea (at the basal end of the cochlea; the other end being referred to as the apical end). The pressure wave propagating in the ear canal sets the eardrum into motion. This vibration is mechanically transmitted by the *ossicles* to the fluid-filled cochlea [68]. In addition to lever action, the different areas of the *tympanic membrane* and the *stapes*, result in

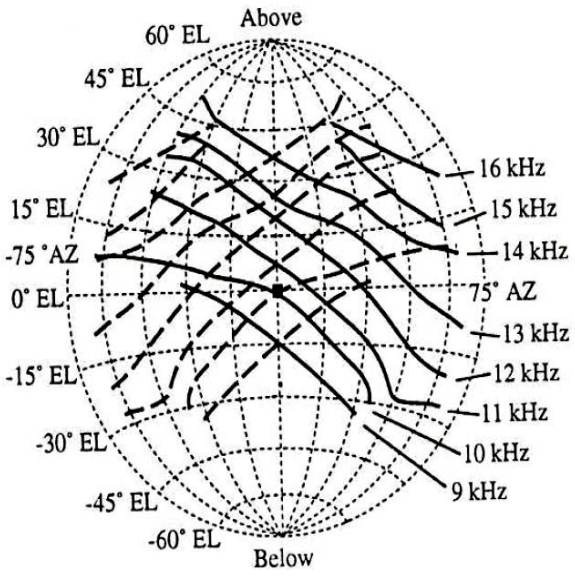

Figure 2.4: Contours of constant first notch (FN) frequency derived from the head related transfer function (HRTF) of a cat. Dotted lines show contours of constant azimuth or elevation. Solid lines show iso-FN contours for the right ear, and the dashed lines are mirror image contours (representing left ear) assuming perfect symmetry between both ears. Reprinted from [67], originally from [66] with permission from Elsevier.

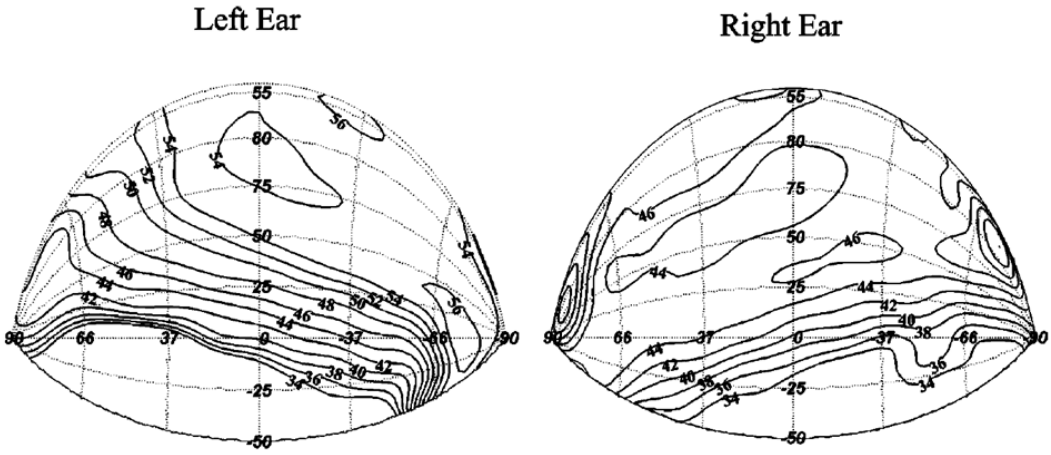

Figure 2.5: Iso-frequency lines for the primary spectral notch as a function of elevation and azimuth measured using the intact ears of a big brown bat. Reprinted with permission from [56]. Copyright 2004, Acoustical Society of America.

the *ossicles* acting as an impedance transformer ensuring efficient coupling of the sound energy into the cochlea. The bat's middle ear shows several specializations; the middle ear ossicles are small and stiff and the middle ear/cochlea is attached to the skull bones via ligaments [69]. The middle ear is also believed to play a role in the spatially-dependent intensity compensation described in the previous chapter. It has been shown in bats that the middle ear muscles contract at the time of vocalization [70], attenuating the signal reaching the inner ear, and that it takes around 5-8ms for the attenuation to decay to zero.

2.2.3 The Cochlea

The snail-shaped cochlea is located in the inner ear. The role of the cochlea is to transduce sound waves into electrical neural activity (action potentials or spikes)

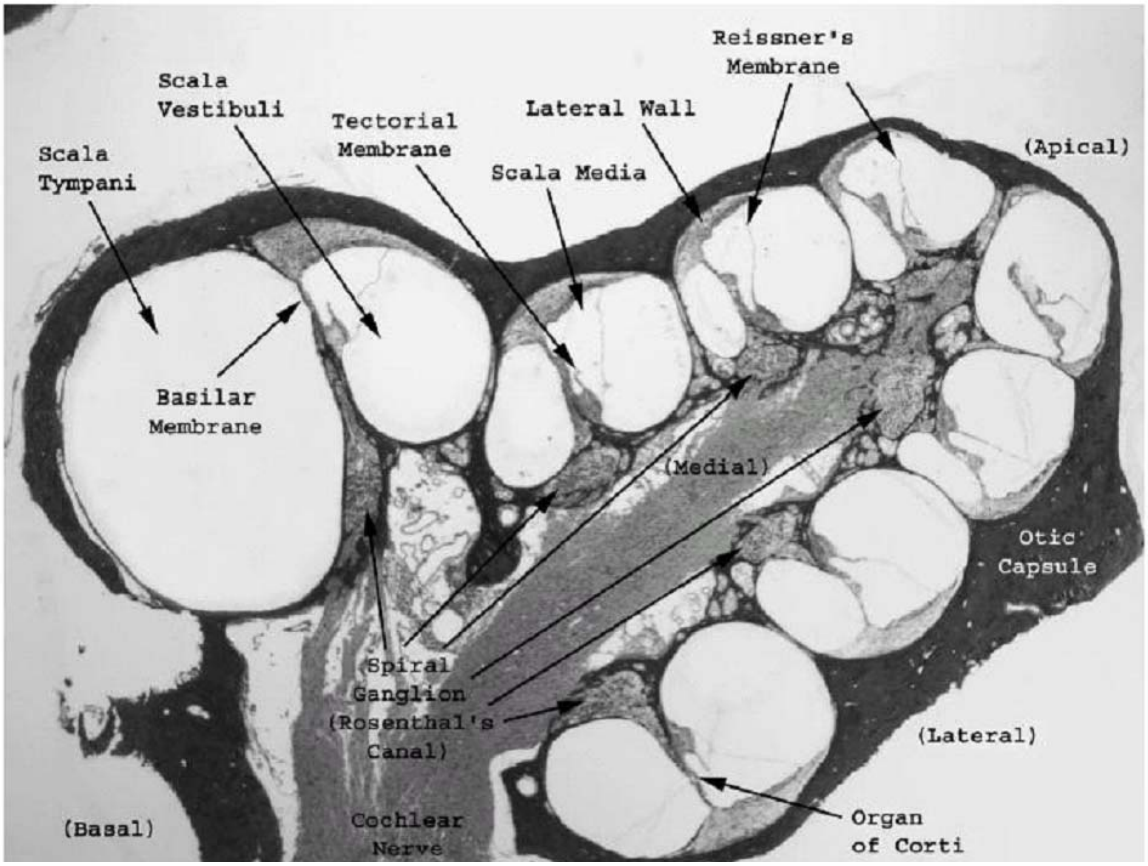

Figure 2.6: Micrograph of a cross section of a guinea pig cochlea. The cochlea has an external bony shell known as the otic capsule. Internally, it has three adjacent fluid-filled tubes (scalae): scala vestibuli (SV), scala media (SM), and scala tympani (ST). The organ of Corti is located in the scala media and rests on the basiliar membrane (BM), which separates the scala media from the scala tympani. The medial is also known as the modiolus. Reprinted from [71] with permission from Elsevier.

in the auditory nerve [71]. A cross section of a cochlea is shown in Fig. 2.6. The cochlea has an external bony shell known as the otic capsule. Internally, it has three adjacent fluid-filled tubes (scalae): scala vestibuli (SV), scala media (SM),

and scala tympani (ST). The scala tympani and scala vestibuli are filled with a high potassium solution called perilymph, whereas the scala media is filled with a high-potassium solution called endolymph. At the base of the cochlea there are two membrane-covered windows that open into the air-filled middle ear cavity. The round window is located in scala tympani, whereas the oval window (which is connected to the footplate of the *stapes*) is located in scala vestibuli [58]. The organ of Corti (shown in Fig. 2.7) is located in the scala media and rests on the basiliar membrane (BM), which separates scala media from scala tympani.

The organ of Corti, which spirals on the basilar membrane, contains both sensory and supporting cells [72]. Hair cells of the organ of Corti are the sensory cells of hearing [71]. There are two types of mechanoelectrical sensory hair cells in the organ of Corti: inner hair cells (IHC) and outer hair cells (OHC). Both types have hair bundles at their apical surface (tip) [72]. Each hair bundle has between 20 and 50 (or more) stereocilia protruding into the endolymphatic fluid. The length of the stereocilia increase along the length of the basilar membrane with the tallest stereocilia at the apical end of the cochlea [71]. In mammals, there are four rows of hair cells positioned longitudinally along the basilar membrane, three rows of outer hair cells and one row of inner hair cells. It has been estimated that, in humans, there are between 2800-4400 IHCs, $11,200$-$16,000$ OHCs and $31,400$ auditory nerve fibers [73]. The piston-like motion of the stapes sets up a pressure wave in the cochlear fluids. The pressure differences between the different scalae result in the displacement of the basilar membrane in the transverse direction. The stereocilia that protrude from the hair cells are deflected either directly by the

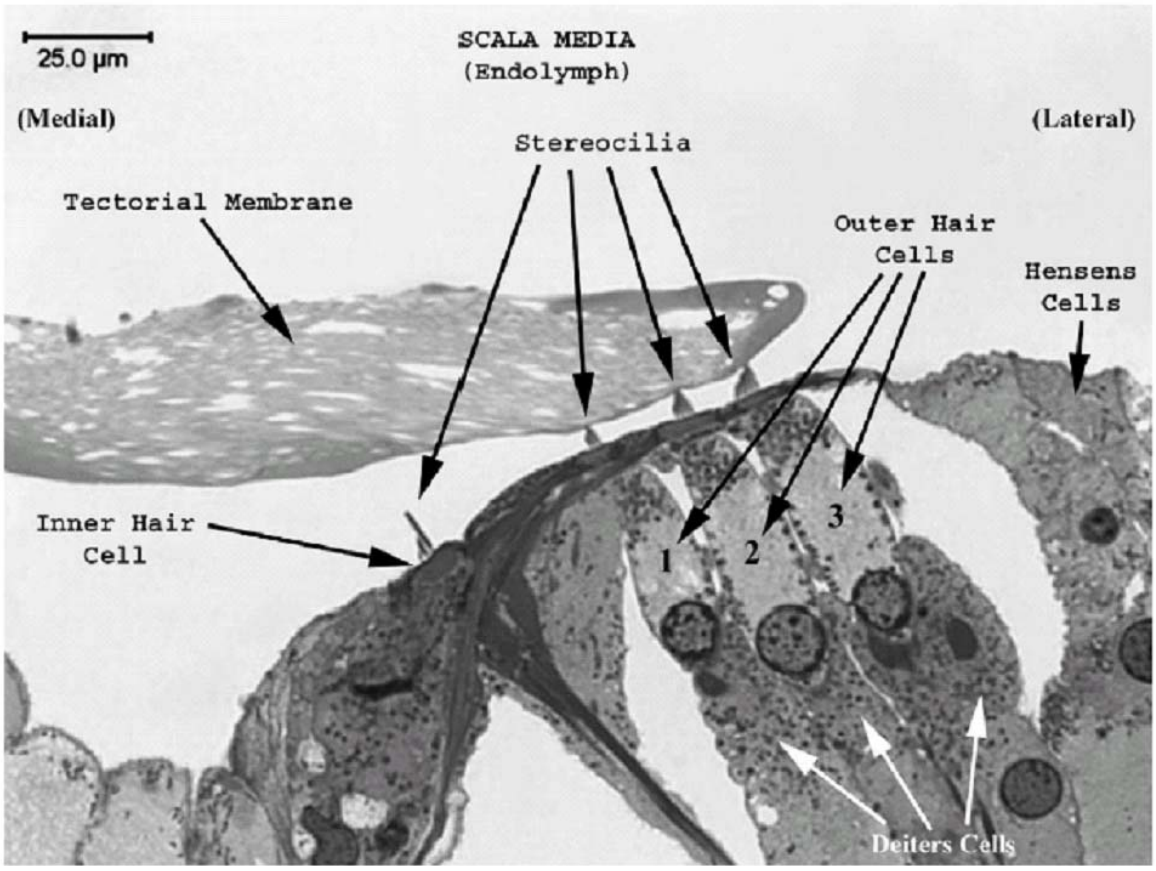

Figure 2.7: A cross section of the mole rat cochlea. Shown is the organ of Corti and the tectorial membrane. Notice that there are three outer hair cells and one inner hair cell and that the stereocilia of the outer hair cells make contact with the tectorial membrane while those of the inner hair cells do not. Reprinted from [71] with permission from Elsevier.

motion of the tectorial membrane (TM) (Fig. 2.7) as in the case of outer hair cells, or indirectly by the friction against the endolymphatic fluid. Modeling studies have indicated that the high-performance characteristics of the cochlea cannot be explained by a passive system. Therefore, active mechanisms (involving positive feedback) have been proposed to explain the cochlea's extraordinary sensitivity and sharp frequency selectivity [74]. The outer hair cells are known to take part in a positive feedback mechanism (the "cochlear amplifier") that increases the sensitivity of the basilar membrane responses at low stimulus levels [75]. There is a gradient in the width and thickness of the basilar membrane along its length (from base to apex). This produces a position-dependent frequency tuning, resulting in a spatially represented frequency analysis. The frequency at which the basilar membrane is most sensitive is called the characteristic frequency (CF) and is approximately given by:

$$\mathrm{CF} = A(10^{\alpha x} - \kappa) \tag{2.1}$$

where CF is in kHz, x is the distance measured from the apex normalized with respect to the length of the basilar membrane, and A is a species dependent constant. The constant α is the same for many cochlea (of different lengths) while the constant κ vary slightly across species [58]. The equation reveals that towards the basal part (high frequency) of the cochlea, where the first term dominates, the logarithm of CF is a linear function of position along the basilar membrane. As mentioned in Chapter 1, CF/FM bats have a large fraction of their basilar membrane tuned to the second harmonic of their CF component and have neurons

that are very sharply tuned. FM bats, however, are not known to have any such specializations [12].

2.3 The Auditory Nerve

In mammals, all auditory information enters the brain through the auditory nerve fibers. The auditory nerve fibers are the axons of the spiral ganglion neurons. The auditory nerve fibers travel in the central core of the cochlea, called the modiolus (see Fig. 2.6), towards the cochlear nucleus in the brain [73]. The cell bodies of the auditory nerve fibers comprise the spiral ganglion, which is located in Rosenthal's canal within the mediolus of the cochlea [76]. There are two types of afferent cochlear neurons, type I and type II. Each type I neuron innervates a single inner hair cell whereas type II neurons innervate more than one outer hair cell; however, a single inner hair cell is innervated by more than one type I neuron [73]. When an inner hair cell is deflected, ionic channels are activated and the depolarizing receptor potentials result in the generation of action potentials (**spikes**) in type I neurons [58]. The type II neurons represent only 5% of the spiral ganglion and their role remains largely unknown [77]. For the remainder of this chapter the term auditory nerve fiber will refer to the type I neuron.

One way of characterizing the response of an auditory nerve is through frequency-threshold tuning curves (or simply threshold-tuning curves), which is defined as the locus of intensity-frequency pairs for tones that generate a just-measurable increase in firing rate above the fiber's spontaneous rate [73]. The

lowest point on the threshold tuning defines the nerve's characteristic frequency (CF), which is physically determined by the position of the corresponding inner hair cell along the basilar membrane as described in the previous section. A measure of sharpness of the tuning curve that is commonly used is the Q10dB quality factor and is defined as [78]:

$$Q10\text{dB} = \frac{\text{CF}}{Bandwidth\ at\ 10\ \text{dB}\ above\ threshold} \tag{2.2}$$

The tuning curves have not been measured in the auditory nerve fibers of the big brown bat *Eptesicus fuscus*; however, Q10dB values have been estimated from neurons in the cochlear nucleus (CN) to range from 5 to 20 [79]. In another FM bat, *Myotis lucifugus*, the Q10dB for its auditory nerve fibers was found to be mostly less than 20 [80].

Another way of characterizing the response of auditory nerve fibers is through rate-intensity curves which are measured at the CF of the fiber and show how its spiking rate (discharge rate) changes with intensity. However, because of the short duration sounds in *Eptesicus fuscus*, rate coding of intensity appears to be an unlikely code.

Although more than one auditory nerve fiber innervates a single inner hair cell, they are not necessarily identical. Liberman [81] showed that (in cats) the threshold of an auditory nerve fiber is correlated with its spontaneous rate (SR) of activity (see Fig. 2.8); fibers with the highest SR had the smallest thresholds and that as the SR decreased the threshold of the fiber increased. In a later study Liberman showed that there is a high correlation between the SR of a fiber

and its diameter; the higher the SR the larger the diameter, and that there is a SR-based spatial segregation of auditory fiber terminals around the inner hair cell circumference [82]. He also showed that some degree of this spatial organization was maintained at least as far as the spiral ganglion [83]. Other studies have shown that auditory nerve fibers have distinct patterns of innervations to the cochlear nucleus [84].

2.3.1 Intensity Encoding in *Eptesicus fuscus*

In the big brown bat *Eptesicus fuscus*, the entire duration of its FM sweep is less than $25ms$ and can be as short as a few milliseconds, therefore, rate coding of intensity appears to be an unlikely code. Alternatively, intensity may be encoded in the latency of spikes as well as in the pattern of activation of auditory nerve fibers of different thresholds.

2.4 The Cochlear Nucleus

All afferent auditory nerve fibers synapse onto neurons in the cochlear nucleus (CN) [85]. The cochlear nucleus can be divided anatomically into three regions: the anteroventral cochlear nucleus (AVCN), the posteroventral cochlear nucleus (PVCN), and the dorsal cochlear nucleus (DCN). As the auditory nerve enters the cochlear nucleus it bifurcates into an ascending branch that innervates the AVCN and a descending branch that further branches to innervate both the PVCN and the DCN. This divergence of information is evidence of the parallel processing of

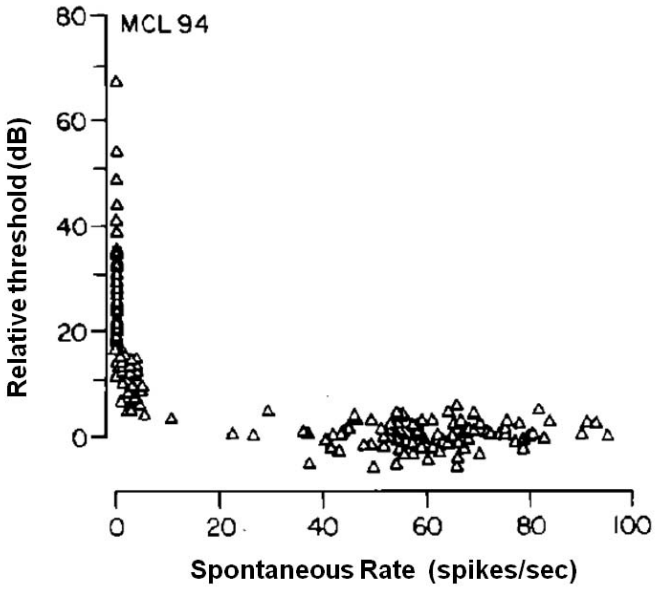

Figure 2.8: Relationship between the relative threshold of an auditory nerve fiber and its spontaneous rate (SR) as measured in a cat. The relative threshold is the difference between the threshold at CF and the average threshold of all fibers with similar CF having a SR >18 spikes/sec. Reprinted with permission from [81]. Copyright 1978, Acoustical Society of America.

the auditory information. In this section we will focus on the DCN and its role in spectral processing.

2.4.1 The Dorsal Cochlear Nucleus (DCN)

Three decades of research have implicated the dorsal cochlear nucleus in the processing of the spectral localization cues. Recall the presence of a spectral notch in the measured HRTF of a cat (see section 2.2.1) and that the frequency of the notch changes systematically with elevation and azimuth. There is both behavioral and physiological evidence that the DCN plays a role in processing this spectral feature.

Behavioral studies with cats [86] showed that when lesions were made that disrupted spectral processing in the DCN of cats, their orientation errors increased. One physiological study [87] showed that DCN neurons are more sensitive to spectral notches than VCN (Ventral Cochlear Nucleus) neurons. Another study [88] characterized the response of type O neurons in the central nucleus of the inferior colliculus (CNIC) that are a primary target of ascending DCN projections. Their findings support the hypothesis that DCN type IV neurons contribute to an auditory pathway that is uniquely specialized for processing directionally-dependent spectral features.

The DCN is a three-layered structure; the outermost layer is called the superficial (or molecular) layer, the second layer is called the pyramidal (or fusiform) cell layer, and the innermost layer is called the deep layer. The fusiform and gi-

ant cells are the principal cells of the DCN and their cell bodies are located in the fusiform layer. One of the characteristics of the DCN principal neurons (in unanesthetized cats) is their type IV response map (described below) which is significantly different from that of its auditory nerve input.

The response maps [89] (constructed by using pure tones) in Fig. 2.9 show the discharge rate of two different DCN neurons versus stimulus frequency for a series of fixed sound levels (the numbers to the right of each figure represent attenuation, so the sound level increases from bottom to top). The horizontal line in each plot represents the spontaneous discharge rate. Increases in rate above spontaneous are colored black showing excitatory response regions, whereas decreases in rate below the spontaneous rate are in gray showing inhibitory regions [90].

Fig. 2.9A shows the response map of a type IV DCN neuron. This neuron has a small excitatory area near threshold centered at its best frequency (BF), and for sounds with higher levels its response is inhibitory over a V-shaped region called the central inhibitory area (CIA). The characteristic features of a type IV map are the excitatory area near threshold (at its BF) and the CIA [90].

Fig. 2.9B shows the response map for a type II neuron in the DCN. It has a narrow V-shaped excitatory area centered on BF and an inhibitory surround. Two features characterize the type II neurons: they have no spontaneous activity, and they give weak or no response to broadband noise.

The response map of two different type IV neurons to broadband noise (BBN) and to noise notches is shown in Fig. 2.10. The response of the type IV

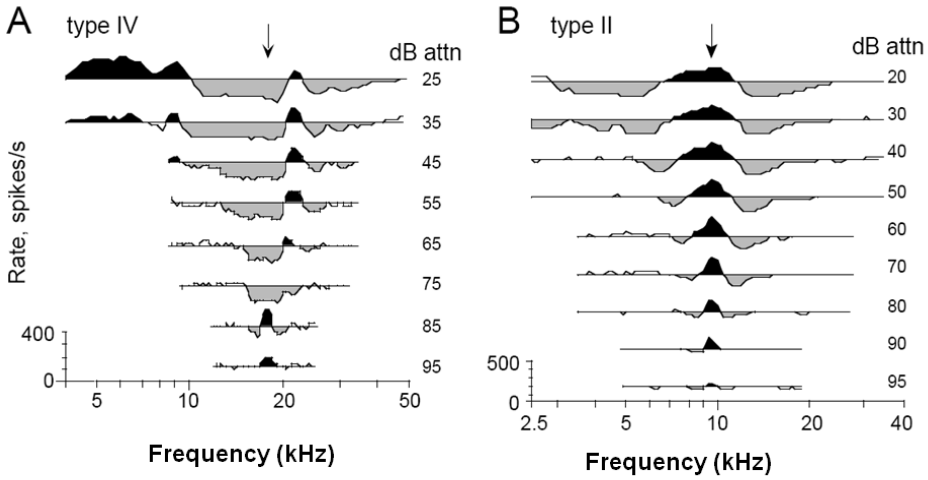

Figure 2.9: Response maps for two DCN neurons: type IV neuron (A), type II neuron (B). Arrows at the top point to the best frequency (BF) of the neuron. Because the type II neurons do not have spontaneous activity, its response map was constructed by presenting a low-level BF tone of fixed attenuation and frequency. Reprinted from [90] with kind permission of Springer Science+Business Media.

neuron to broadband noise cannot be predicted by the response map. For example, in Fig. 2.10B the response of the neuron to loud sounds (pure tones) is mostly inhibitory; however, Fig. 2.10D shows that the response to BBN is well above the level of spontaneous activity (shaded bars) whereas the response to tones has been inhibited. When a notch centered at the BF of the neuron was introduced into the BBN, the response of the neuron was reduced (see Figs. 2.10E-F). The inhibitory response at the BF of the neuron in the response map and the activity in response to broadband noise suggests the presence of an inhibitory connection to the type IV neuron that is inhibited by the broad band noise.

A model that was put forward to explain the observed results for both neurons is shown in Fig. 2.11 [91]. In this model, type II and IV neurons interact with each other and a third neuron to shape each other's responses. The type IV neuron is modeled as having a best frequency (BF) (determined by its auditory nerve fiber input) slightly above that of the type II neuron. Also, because of its spontaneous activity, the threshold of the type IV neuron is lower than that of type II. This explains the excitatory response of the type IV neuron observed at low stimulus levels (see Fig. 2.9), where the type II neuron's higher threshold renders it incapable of inhibiting the type IV neuron. However, as the stimulus level increases, the type II neuron inhibits the type IV neuron. The wideband inhibitor neuron (labeled WBI in Fig. 2.11) is weakly activated by tones but can be activated by a broadband stimulus. When activated, the wideband inhibitor "shuts off" the type II neuron and its (type II) inhibition to the type IV neuron; therefore, if a stimulus exists at the BF of the type IV neuron its response will

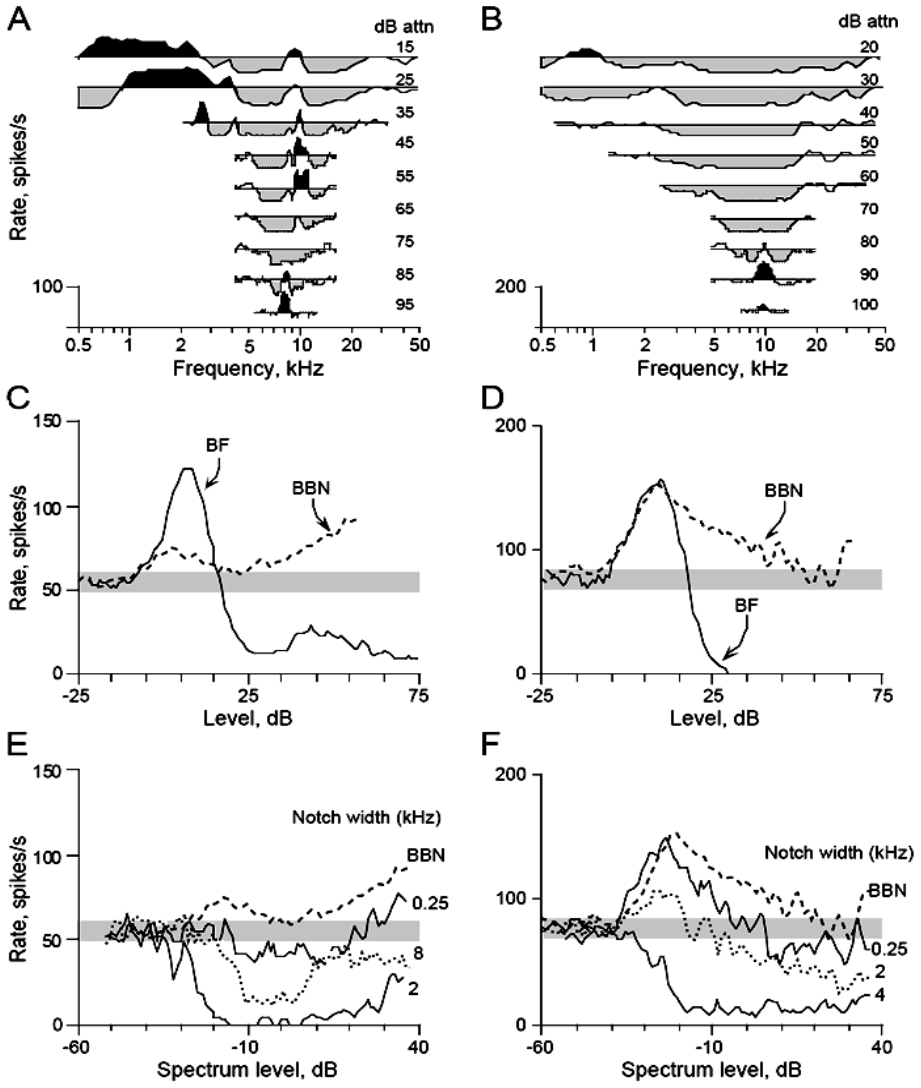

Figure 2.10: Response maps and responses to broadband noise for two type IV neurons in the DCN. Each column shows data from one neuron. A,B) Response maps. C,D) Discharge rate vs. sound level for 200ms BF-tone bursts and noise bursts. The shaded bars show the range of spontaneous rates. E,F) Response of the neuron to notch noise (broadband noise with a notch centered on BF). The bandwidth of the notches is shown to the right of the curve. Reprinted from [90] with kind permission of Springer Science+Business Media.

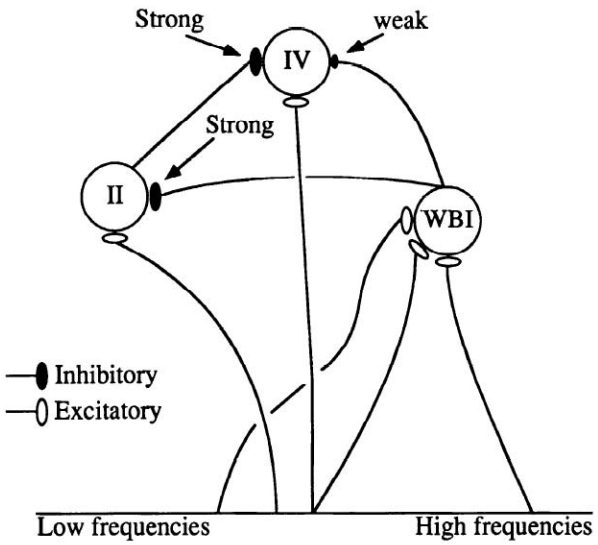

Figure 2.11: Model of the type IV DCN neuron. Used with permission from [91].

be excitatory, if not, its response will be inhibitory giving the type IV neuron its notch detecting behavior as observed in Fig. 2.10.

In summary, it has been shown that the type IV DCN neurons respond to frequency notches centered at their best frequency (BF) by decreasing their firing rate below their spontaneous firing rate.

DCN of *Eptesicus fuscus*

The dorsal cochlear nucleus of the big brown bat *Eptesicus fuscus* is similar to that in other mammals in that it has a laminated structure; however, it has a poorly laminated fusiform cell layer [92]. The implication of this observation is not known. It should be noted that the response maps in Figs. 2.9 and 2.10 were obtained for sounds that have a $200ms$ duration while the sound duration for the entire FM sweep of *Eptesicus fuscus* is an order of magnitude shorter. For the

bat to detect a reduction in firing below its spontaneous rate during this short sweep would require the type IV neuron to have a spontaneous rate close to its theoretical upper limit of 1000 spikes/sec (i.e. once every millisecond assuming a $1ms$ refractory period). This appears to be an unrealistic and energy inefficient solution.

It should be noted that in a more recent study [93], the majority (65%) of type IV DCN neurons in cats were shown to exhibit spectral edge sensitivity responding by a peak in their discharge rate when a rising spectral edge is aligned near their BF while others (11%) responded by excitation to notches centered on their best frequency.

2.5 Superior Olivary Complex (SOC)

The superior olivary complex is an auditory brainstem structure that contains several nuclei that have been implicated in the processing of binaural localization cues, namely, the medial superior olive (MSO), the lateral superior olive (LSO), and the medial nuclei of the trapezoid body (MNTB). The MSO has been implicated in the processing of the interaural time difference (ITD) cue while the LSO has been implicated in the processing of the interaural level difference (ILD) cue as shown in Fig. 2.12. The figure shows that the MNTB receives an excitatory input from the contralateral AVCN transforming it into an inhibitory output that synapses onto a LSO neuron.

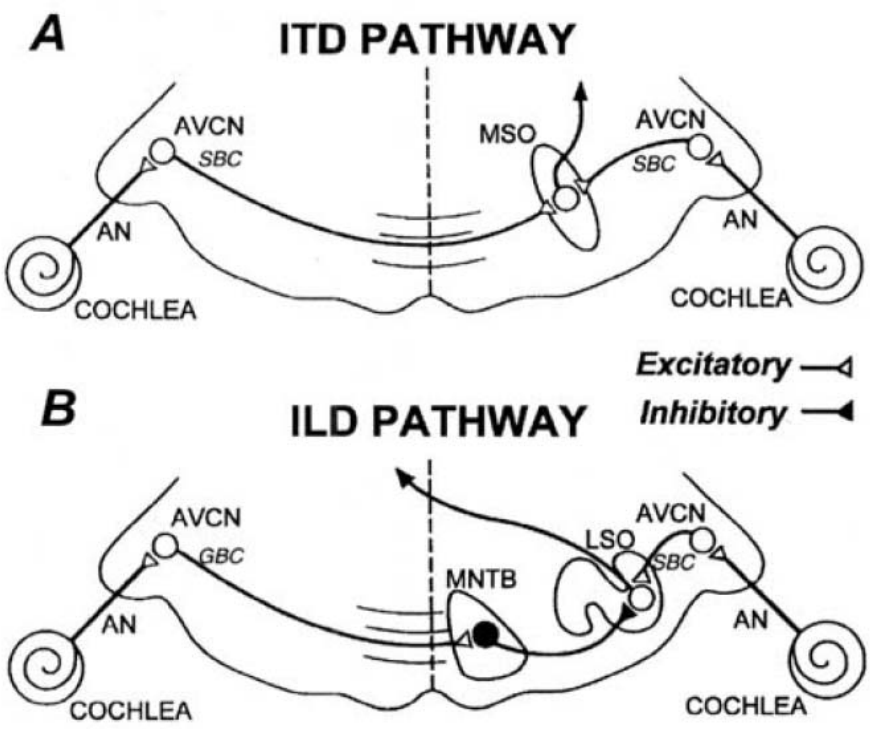

Figure 2.12: Schematic diagram of the two parallel pathways in the superior olivary complex (SOC) for encoding ITDs in the MSO (A) and ILDs in the LSO (B). The MSO receives bilateral excitation from the spherical bushy cells of the AVCN while the LSO receives excitation from the spherical bushy cells of the ipsilateral AVCN and inhibition from the globular bushy cells of the contralateral AVCN via the ipsilateral MNTB. Reprinted from [1] with kind permission of Springer Science+Business Media.

2.5.1 Medial Superior Olive (MSO)

The structure and function of the MSO varies among mammals. In mammals with good low-frequency hearing the MSO neurons receive binaural excitation from the spherical bushy cells located in the AVCN (as shown in Fig. 2.12A) and there is an over representation of low frequencies in the MSO. It has been shown that these MSO neurons are sensitive to interaural time differences (ITDs).

Unlike mammals with low frequency hearing, the bat MSO neurons are tuned to frequencies in the ultrasonic range. In one study of the big brown bat *Eptesicus fuscus*, the best frequencies of MSO neurons ranged from 11 to $79kHz$ with the majority of neurons tuned between 20 and $40kHz$. Half of the MSO neurons of *Eptesicus fuscus* were found to be monaural (receiving both excitation and inhibition from the contralateral ear) while the other half were binaural. Some of the binaural neurons received excitation from the contralateral ear and inhibition from the ipsilateral ear while others received excitation from both ears [94]. In the case where MSO neurons receive inhibition from the contralateral ear, the output of the contralateral AVCN goes through the ipsilateral MNTB which then provides the inhibition to the MSO neurons.

Many hypotheses have been put forward to explain the role of the MSO in bats [12]. One hypothesis that is related to the localization problem is that the large population of monaural MSO neurons may extract the monaural spectral cues caused by the external ear as described in section 2.2.1. Another hypothesis is that binaural MSO neurons might be involved in an ITD-like processing scheme

if the response latencies are taken into account [95].

2.5.2 Lateral Superior Olive (LSO)

Similar to other mammals, the bat LSO receives excitation from the ipsilateral ear and inhibition from the contralateral ear as shown in Fig. 2.12B. This arrangement has been implicated in the processing of interaural level (intensity) differences. Fig. 2.13 is an example of an ILD (IID) function for an LSO neuron measured in a mustache bat [96]. The excitatory stimulus at the ipsilateral ear was fixed at 40dB SPL while the inhibitory input at the contralateral ear (x-axis) was varied from 0-60dB SPL. The ILD defined as the difference between the contralateral (inhibitory) intensity and the ipsilateral (excitatory) intensity is also displayed along the x-axis. Also shown on the right side are the spike rasters corresponding to points (open circles) on the ILD function. The smallest ILD that results in a complete suppression of spikes is called the *ILD of complete inhibition*. Stimuli were $20ms$ tones presented at the cell's characteristic frequency and were repeated 20 times and the total number of spikes for all 20 presentations are shown on the y-axis.

Fig. 2.14 illustrates the effect of stimulus duration on the ILD function; as the duration of the sound decreases the number of spikes decreases but the ILD of complete inhibition does not change. The model put forward by Reed and Blum [97] for encoding azimuth assumed that LSO neurons are arranged in columns with each column having a different ILD of complete inhibition. Assuming that

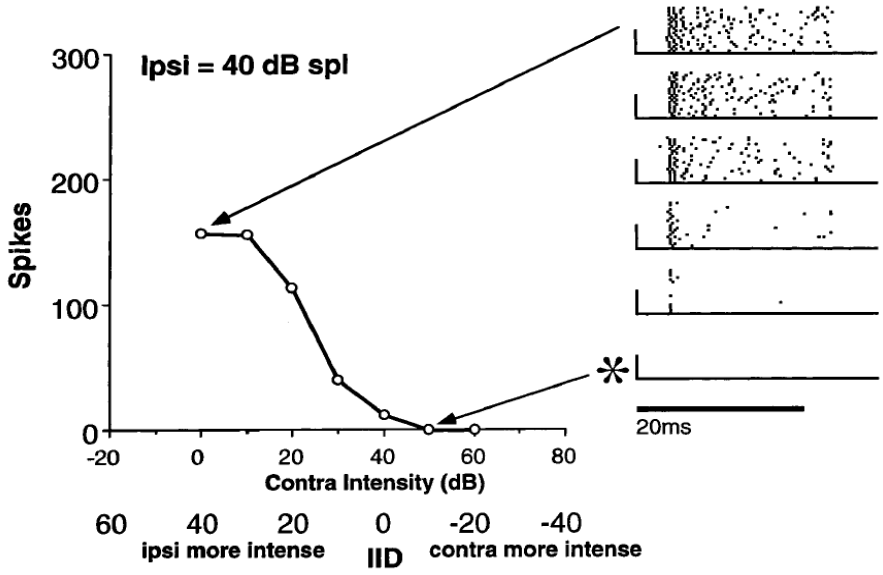

Figure 2.13: Interaural Intensity (Level) Difference (IID or ILD) function measured from a bat LSO neuron. The excitatory stimulus at the ipsilateral ear was fixed at 40dB SPL while the inhibitory input at the contralateral ear (x-axis) was varied from 0-60dB SPL. The ILD is also displayed along the x-axis. Shown on the right are the spike rasters corresponding to points (open circles) on the ILD function. The smallest ILD that results in a complete suppression of spikes is called the ILD of complete inhibition. Stimuli were $20ms$ tones presented at the cell's characteristic frequency and were repeated 20 times. The total number of spikes for all 20 presentations are shown on the y-axis. Used with permission from [96].

the columns are arranged in order of their ILD of complete inhibition, azimuth is encoded as a function of ILD by the location along the LSO column where the inhibitory input exceeds the excitatory input (i.e. the first column that has been completely suppressed).

Consider the response of the LSO neuron to a single presentation of sound (as opposed to summing the number of spikes from 20 identical presentations), and recall that in *Eptesicus fuscus* the FM sweep can be as short as a few milliseconds. Therefore, a reasonable approximation is to treat the output of the LSO neuron as a binary output shown as the dashed curve in Fig. 2.14. The LSO neuron is off for ILDs $\geqslant$ ILD of complete inhibition and is on otherwise. This approximation does not alter the encoding scheme proposed by Reed and Blum.

The LSO model proposed by Reed and Blum hypothesized that the ILD of complete inhibition is determined by the difference between the thresholds of the excitatory and inhibitory inputs; however, this hypothesis was tested [96] in the mustache bat and it was found that for the majority of LSO neurons the ILD of complete inhibition did not correspond to the difference in thresholds of the excitatory and inhibitory inputs. By assuming that there are latency differences between the excitatory and inhibitory inputs at the input to the LSO neurons, they suggest that the differences in thresholds is the major factor in determining the ILD of complete inhibition but that latency differences also plays a role.

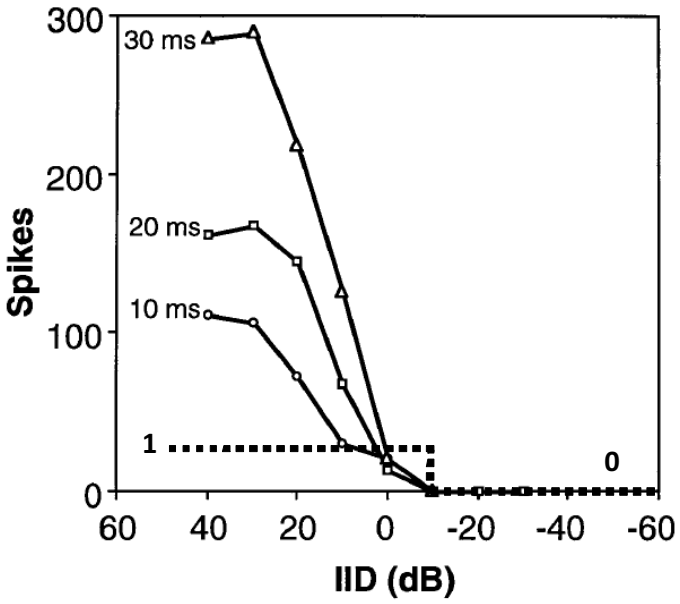

Figure 2.14: Effect of changing the duration of the stimulus on the IID (ILD) function. Notice that the ILD of complete inhibition does not change. The dashed curve (not drawn to scale) is an approximation of the ILD function for short sounds and for a single presentation of the stimulus. Used with permission from [96].

2.6 Conclusion

In this chapter we reviewed parts of the mammalian auditory system involved in the localization task. In Chapter 4 we present an electronic ultrasonic cochlea that models the cochlear filtering, half-wave rectification of the inner hair cells, and the auditory nerve fibers. In Chapter 5 we present two feature extraction chips that can extract the localization cues from the spiking output of the cochlea chips: a binaural ILD chip and a monaural spectral chip. The binaural ILD chip models the processing that takes place in the LSO neurons while the architecture of the monaural spectral chip that will be presented bears some similarity with the structure of the MSO.

Chapter 3

Artificial Bat Head

3.1 Introduction

To unambiguously localize sounds, a unique direction-dependent acoustic signature ("fingerprint") must be imprinted onto any arriving sound. The acoustic signature of a sound originating from a specific direction depends on the shape and size of the pinnae and head. Bat pinnae come in different shapes and sizes; one study showed that for some bats there is a correlation between the size of the pinna and the frequency of its vocalization while for other bats the correlation is not significant (for a detailed review see [5]). The bat that we are interested in modeling is the big brown bat *Eptesicus fuscus*. In this chapter we present an artificial bat head that was designed using 3D CAD software and fabricated using a 3D printer to generate the different acoustic cues for our localization system. We measured the response of the artificial bat head from 703 directions in the frontal hemisphere and illustrate that the bat head generates monaural spectral

Figure 3.1: Picture of the big brown bat *Eptesicus fuscus*.

cues and binaural ILD cues.

3.2 Artificial Bat Head

Fig. 3.1 shows a picture of the big brown bat *Eptesicus fuscus*. In one study [5], the height and width of the pinna for *Eptesicus fuscus* were measured to be $13mm$ and $9.5mm$ respectively.

The artificial bat head, shown in Figs. 3.2 and 3.3, was designed with the 3D CAD software SolidWorks^{TM} (for the design steps we used, see Appendix A) and fabricated using a ZCorp 310 3D printer. The same bat head design was also fabricated (not shown) using a Dimension SST 3D printer which produced a smoother texture, however, inserting the microphones required widening the holes at the back of the head and it was easier to manipulate the "ZCorp" head. 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 mounting holes at the back of the head (see Fig. 3.3).

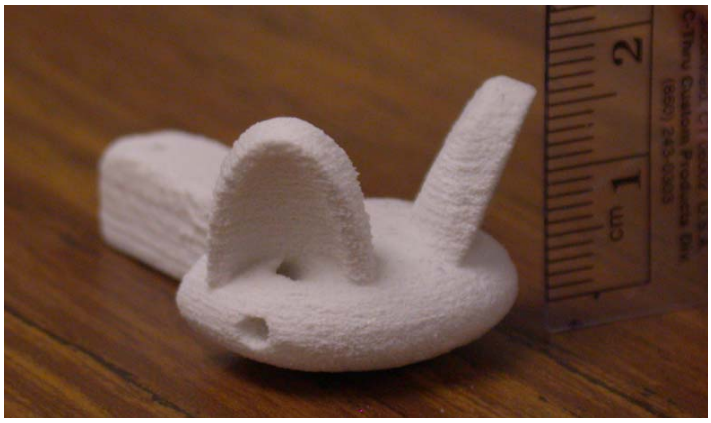

Figure 3.2: The artificial bat head fabricated using a 3D printer. The ear canals (holes inside the base of each pinna) lead to microphone mounting holes at the back of the head. Other mounting holes are for head positioning during characterization.

3.3 Ultrasonic Speaker and Microphones

A Tucker-Davis Technologies (TDT) electrostatic speaker (shown in Fig. 3.4) was used to transmit the ultrasonic FM sweep. The frequency response of the speaker is shown in Fig. 3.5. Two Knowles FG-26163 (formerly FG-6163) microphones were inserted at the back of the artificial bat head. Although the FG-26163 microphone is exactly the same size as the FG-3453 used in measuring the Head Related Transfer Function (HRTF) of real bats [56] and the FG3329 used in audio applications [98], it has greater attenuation in the lower audio frequency range

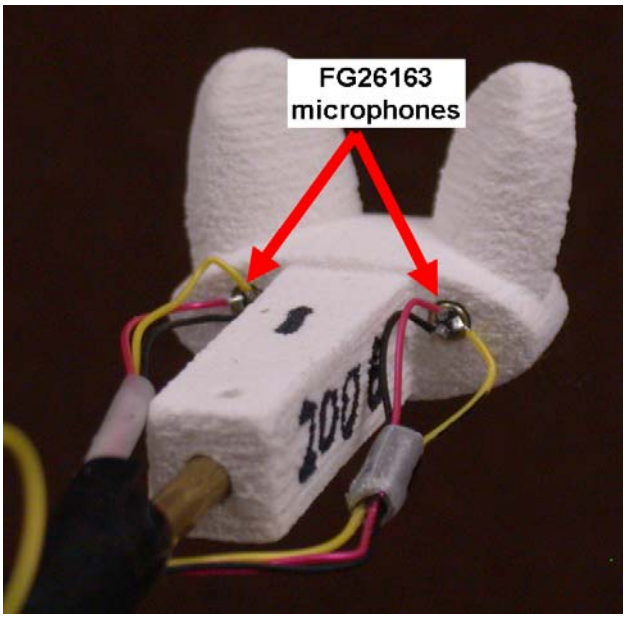

Figure 3.3: Picture of the artificial bat head showing the two microphones (Knowles FG-26163) inserted at the ends of the two ear canals.

(see Fig. 3.6). Shown in Fig. 3.7 is a simplified schematic the FG-26163 Knowles microphone. The microphone uses a supply voltage (Vdd) of 1.5 volts and has a built in source resistance of approximately $22k\Omega$. The microphone requires a DC power supply current of approximately $20\mu A$. Fig. 3.3 shows the two microphones inserted at the back of the artificial bat head.

3.4 Signal Conditioning

Shown in Fig. 3.7 is a simplified schematic of the FG-26163 Knowles microphone followed by three external signal conditioning stages. The first stage is responsible for amplifying the microphone's output in the frequency range from

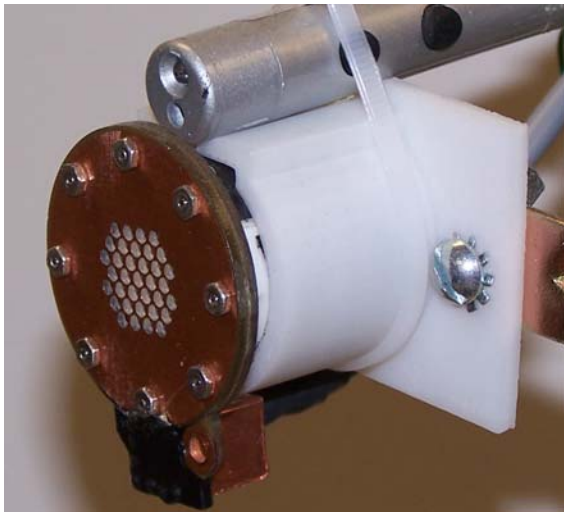

Figure 3.4: Electrostatic speaker (Tucker Davis Technologies). The speaker was inserted into a plastic holder (fabricated using a Dimension SST 3D printer). The holder was designed with a tilted groove so that a laser pointer mounted on top of the speaker points to the bat head placed at the center of the measurement sphere.

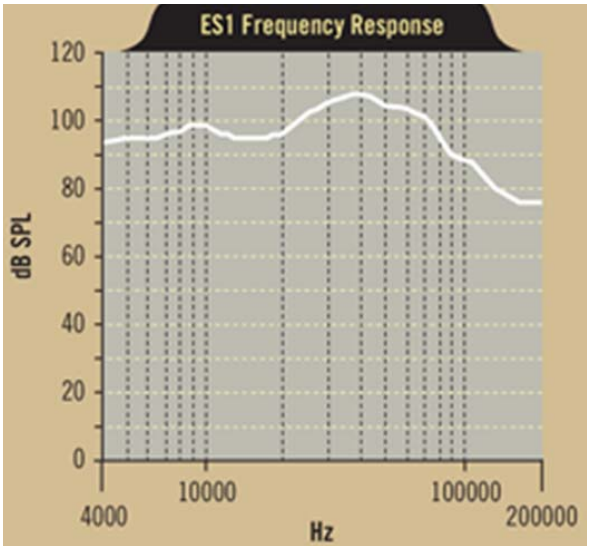

Figure 3.5: Frequency response of the electrostatic speaker (ES1: Tucker-Davis Technologies) used to transmit the ultrasonic FM sweeps. Reprinted with permission from [99].

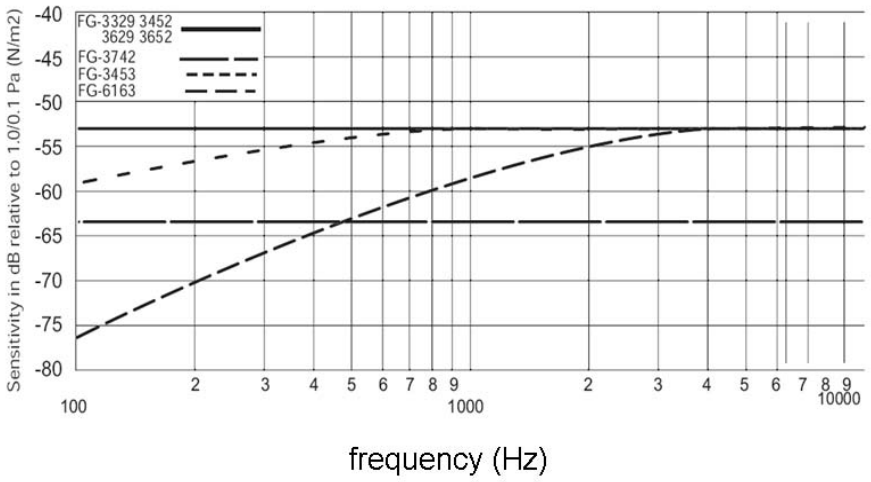

Figure 3.6: Frequency response of the Knowles FG-series microphones. Compared to the FG-3329 and FG-3452 microphones, The FG-6163 (now FG-26163) has greater attenuation for frequencies $< 3kHz$. Notice that the frequency response plot stops at $10kHz$. Reprinted with permission from [100].

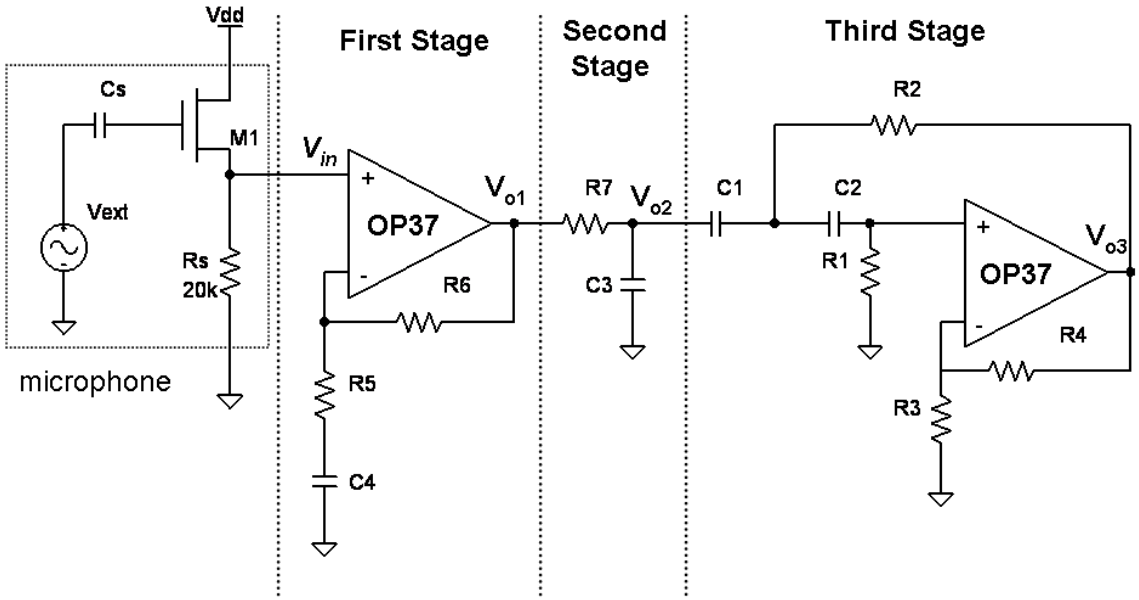

Figure 3.7: Signal conditioning. A simplified schematic of the microphone is shown on the left. The first stage provides a flat gain over the frequency range of interest, the RC circuit ($R5$,$C4$) is a high pass filter (cutoff frequency $\simeq 3kHz$) that reduces the gain of the first stage to unity at DC. The second stage is a lowpass filter (cutoff frequency $\simeq 339kHz$). The third stage is a second order Sallen-Key highpass filter used to compensate for the attenuation of the speaker and microphones at higher frequencies.

$20kHz$-$100kHz$. The gain (G) is determined by the ratio

$$G = \frac{R_5 + R_6}{R_5} \tag{3.1}$$

The RC circuit (R_5 and C_4) is a highpass filter with a cutoff frequency of:

$$f_{_{HPF}} = \frac{1}{2\pi R_5 C_4} \tag{3.2}$$

reducing the gain of the first stage to unity at DC. The transfer function of the first stage is given by:

$$H_1(s) = \frac{V_{o1}(s)}{V_{in}(s)} = \frac{1 + s(R_5 + R_6)C_4}{1 + sR_5C_4}. \tag{3.3}$$

The design values for the first stage are:

R_5	R_6	C_4
560Ω	$67.8k\Omega$	$100nF$

which result in:

$$\begin{aligned} G &= 41.6\text{dB} \\ f_{_{HPF}} &= 2.85kHz. \end{aligned}$$

The second stage is a lowpass filter that attenuates the spectrum at frequencies greater than $100kHz$, its transfer function is given by

$$H_2(s) = \frac{V_{o2}(s)}{V_{o1}(s)} = \frac{1}{1 + sR_7C_3} \tag{3.4}$$

The lowpass cutoff frequency ($f_{_{LPF}}$) is given by

$$f_{_{LPF}} = \frac{1}{2\pi R_7 C_3} \tag{3.5}$$

The design values for the second stage are:

$R7$	$C3$
$1k\Omega$	$470pF$

which result in:

$$f_{_{LPF}} = 339kHz.$$

Shown in Fig. 3.8 is the measured transfer function of the first two stages.

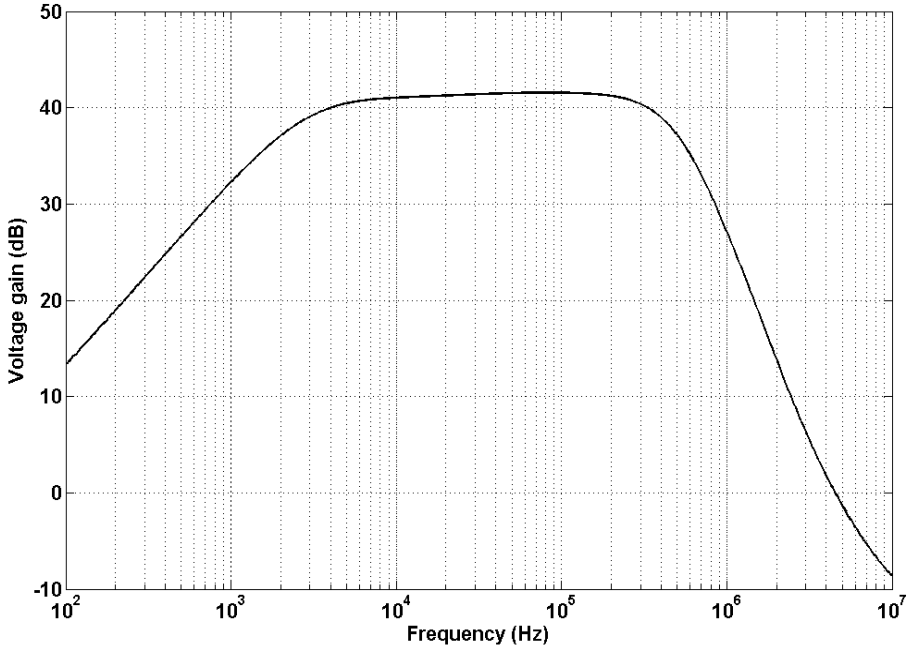

Figure 3.8: Combined frequency response of the first two stages of the microphone amplifier.

The third stage is a second order Sallen-Key highpass filter used to compensate for the attenuation of the speaker and microphones at higher frequencies. The goal is to use the 40dB/*decade* slope in the stopband of the filter as a compensation filter. The transfer function of this stage (see Appendix B) including

the effect of the open-loop gain of the operational amplifier is given by,

$$H_3(s) = \frac{V_{o3}(s)}{V_{o2}(s)} = \frac{(k//A)sR_1R_2C_1C_2}{s^2R_1R_2C_1C_2 + s(R_2C_2 + R_2C_1 + R_1C_2(1 - k//A)) + 1} \quad (3.6)$$

where k is a gain factor, A is the open-loop gain of the operational amplifier, and the term $k//A$, which shows up in both the numerator and denominator, is defined as

$$k//A = \frac{kA}{k + A} \quad (3.7)$$

which resembles the expression for the equivalent resistance of two parallel resistors k and A. To reduce the effect of the open-loop gain of the operational amplifier, the gain factor k must be chosen such that

$$k < \frac{A(j\omega_{\max})}{10} \quad (3.8)$$

where $\omega_{\max}$ is the maximum frequency of interest. If this condition is satisfied then the transfer function may be written as

$$H_3(s) = \frac{ksR_1R_2C_1C_2}{s^2R_1R_2C_1C_2 + s(R_2C_2 + R_2C_1 + R_1C_2(1 - k)) + 1} \quad (3.9)$$

The transfer function in equation 3.9 can be written as

$$H_3(s) = \frac{ks}{s^2 + s(\frac{\omega_o}{Q}) + \omega_o^2} \quad (3.10)$$

where ω_o is the cutoff frequency and Q is the quality factor. Setting equations 3.9 and 3.10 to be equal we get:

$$\begin{aligned} \omega_o &= \frac{1}{\sqrt{R_1R_2C_1C_2}} \\ Q &= \frac{\sqrt{R_1R_2C_1C_2}}{R_2C_2 + R_2C_1 + R_1C_2(1 - k)} \end{aligned} \quad (3.11)$$

To simplify the design procedure, let $C_1 = C_2 = C$, $R_1 = R$, and $R_2 = mR$, where m is a scalar, then

$$f_o = \frac{\omega_o}{2\pi} = \frac{1}{2\pi RC\sqrt{m}} \tag{3.12}$$

and

$$Q = \frac{\sqrt{m}}{2m + 1 - k} \tag{3.13}$$

The design procedure is to specify the cutoff frequency f_o, the quality factor Q, and the gain term k. First the scalar m is computed from 3.13, and then RC is computed from 3.12. The gain factor k is used to find R_3 and R_4.

The initial design specification was to have a second order highpass filter with a gain of 0dB at $10kHz$ and a gain of 40dB at $100kHz$; the two frequencies should lie in the filter's stopband region. Initially, the OP27 operational amplifier was used in the design. However, it has an open-loop gain of 40dB at $100kHz$ (8 MHz gain-bandwidth product), so the OP37 operational amplifier (the uncompensated version of the OP27) which has a $63MHz$ gain-bandwidth product was chosen to replace it. The cutoff frequency f_o was chosen to be $200kHz$ with a gain k of 38dB (80) and a quality factor Q of 0.75. The capacitor C was chosen to be $100pF$. The design values are:

R_1	R_2	R_3	R_4	C_1	C_2
$1.2k\Omega$	$52.7k\Omega$	$100k\Omega$	$8k\Omega$	$100pF$	$100pF$

Fig. 3.9 shows the simulated and measured response of the third stage of the

microphone amplifier. The magenta line is the open-loop gain (A) of the OP37 operational amplifier. The red line is the term $k//A$ as defined in equation 3.7. The light blue line is the simulated (in MATLAB^{TM}) transfer function assuming an ideal operational amplifier (infinite gain). The black line is the MATLAB simulated transfer function considering the finite open-loop gain and the dark blue line is the measured response. Fig. 3.10 shows the measured transfer function of the (right and left) three-stage microphone amplifier. The maximum difference between the right and left channels in the frequency range of interest is 2.3dB at $85kHz$. Fig. 3.11 shows the combined frequency response of the TDT speaker and the Knowles FG-26163 microphone.

3.5 Hyperbolic FM Ultrasonic Sweep

The ultrasonic sound ("ping") used is a hyperbolic frequency-modulated (FM) sweep. The instantaneous frequency for the hyperbolic FM sweep is given by:

$$f_{inst}(t) = \frac{a}{1+bt} \tag{3.14}$$

where a and b are constants.

For a sinusoidal signal $y_{hyper}(t)$

$$y_{hyper}(t) = \sin(\phi(t)) \tag{3.15}$$

the instantaneous frequency is given by:

$$f_{inst}(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{3.16}$$

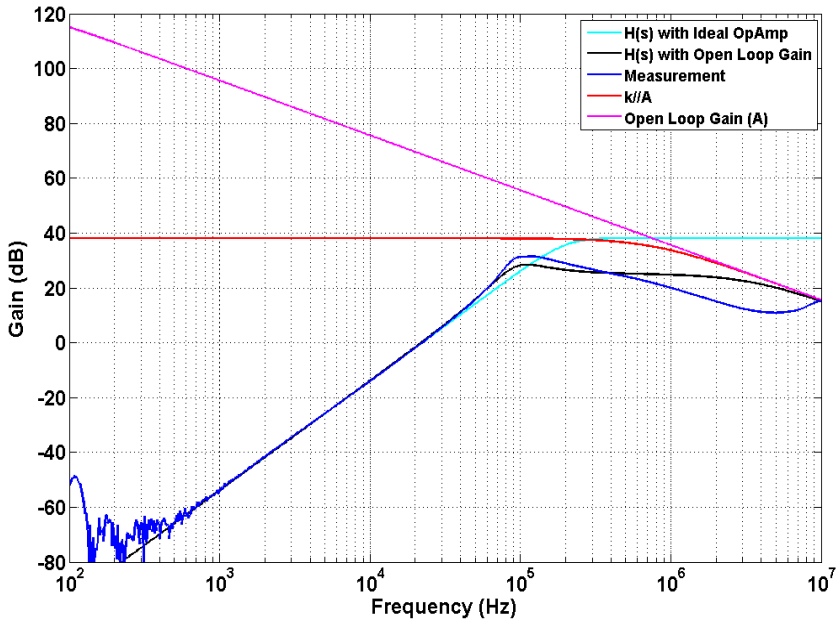

Figure 3.9: Transfer function of the third stage. The magenta line is the open-loop gain (A) of the OP37 operational amplifier. The red line is the term $k//A$ as defined in equation 3.7. The light blue line is the simualted (in MATLAB^{TM}) transfer function assuming an ideal operational amplifier (infinite gain). The black line is the simulated transfer function considering the open-loop gain and the blue line is the measured response.

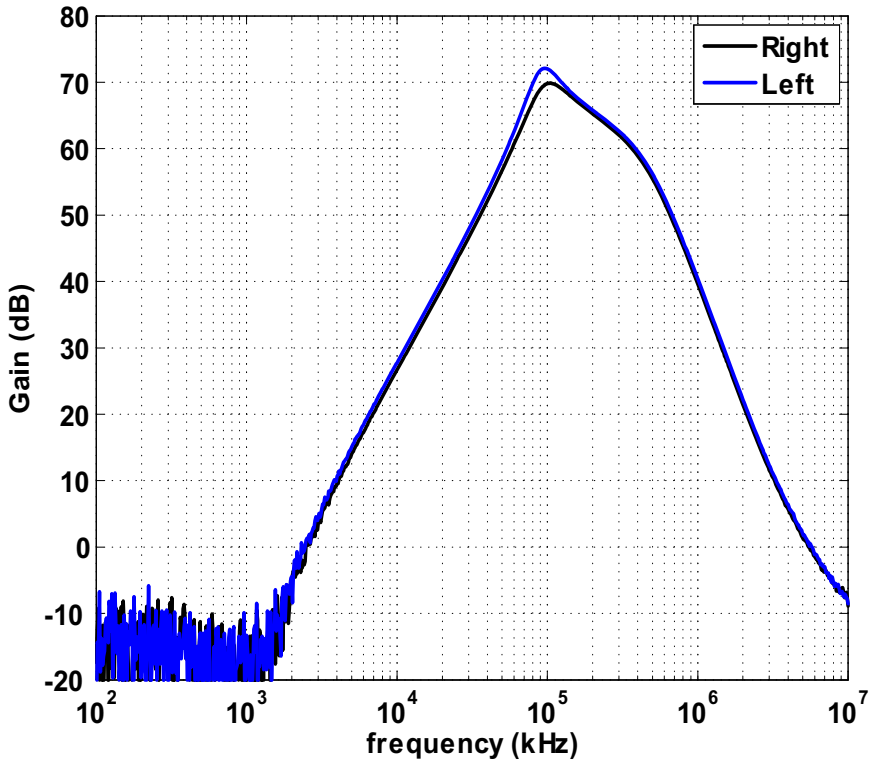

Figure 3.10: Measured transfer function of the (right and left) three-stage microphone amplifier. The maximum difference between the right and left channels in the frequency range of interest is 2.3dB at $85kHz$.

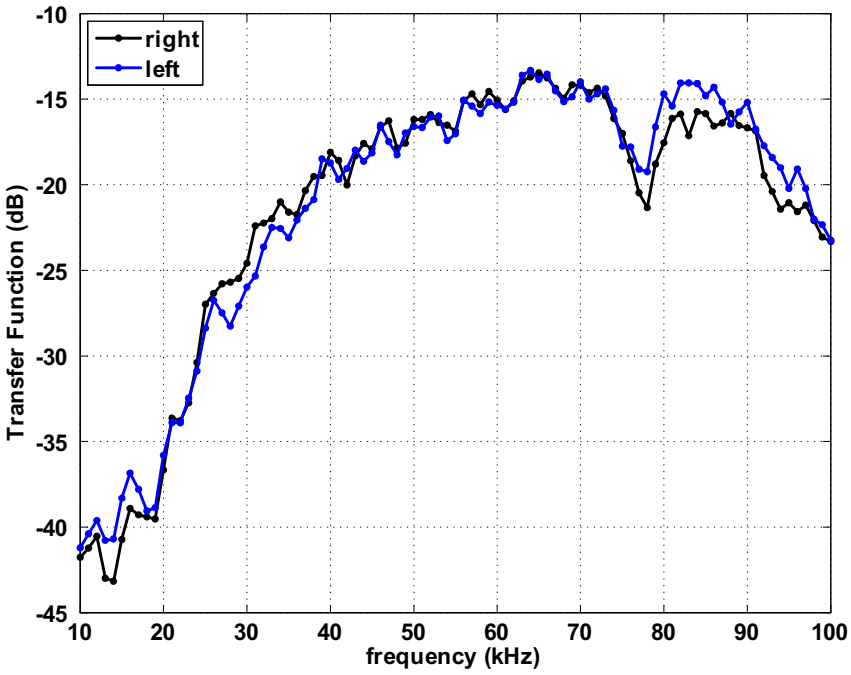

Figure 3.11: Transfer function of the system formed by the speaker, microphones, and the 3-stage external microphone amplifier. The input was $2ms$ pulses of pure tones and the output was averaged over 20 repetitions. It is not clear what is the source of the frequency dip between $70kHz$ and $80kHz$.

Equating 3.14 and 3.16

$$\frac{1}{2\pi}\frac{d\phi(t)}{dt} = \frac{a}{1+bt} \tag{3.17}$$

Integrating to find $\phi(t)$

$$\begin{aligned} \phi(t) &= 2\pi a \int \frac{1}{1+bt} dt \\ &= \frac{2\pi a}{b} \ln(1+bt) \end{aligned} \tag{3.18}$$

Substituting 3.18 in 3.15

$$y_{hyper}(t) = \sin(\frac{2\pi a t}{b} \ln(1+bt)) \tag{3.19}$$

If the hyperbolic sweep has a duration T, start frequency $F_{start,}$ and stop frequency F_{stop} it can be shown that:

$$\begin{aligned} a &= F_{start} \\ b &= \frac{1}{T}(\frac{F_{start}}{F_{stop}} - 1). \end{aligned} \tag{3.20}$$

The parameters for the FM sweep are

$$\begin{aligned} T &= 5ms \\ a &= 120kHz \\ b &= 18kHz \end{aligned}$$

Fig. 3.12 shows the instantaneous frequency of the hyperbolic sweep as a function of time. Notice that more than half of the entire duration of the signal is spent in the last $20kHz$ of the sweep. The time-domain waveform for the FM sweep was shaped to compensate for some of the attenuation in the speaker, microphone, filter system (as seen in Fig. 3.11).

Fig. 3.13 shows the time-domain waveform for the hyperbolic sweep; the upper plot shows the original hyperbolic sweep having a constant amplitude throughout the entire duration of the signal while the lower plot shows the modified (shaped) version. Fig. 3.14 shows the spectrum of both the original and modified hyperbolic sweep.

The artificial bat head was placed at the center of a sphere (see below) and the modified hyperbolic FM sweep was transmitted from the TDT electrostatic speaker placed approximately $83cm$ away from the bat head. Fig. 3.15 shows the magnitude spectrum of the sound received at the right and left microphones after being amplified by the 3-stage amplifier and before inserting the microphones into the ear canals of the artificial bat head.

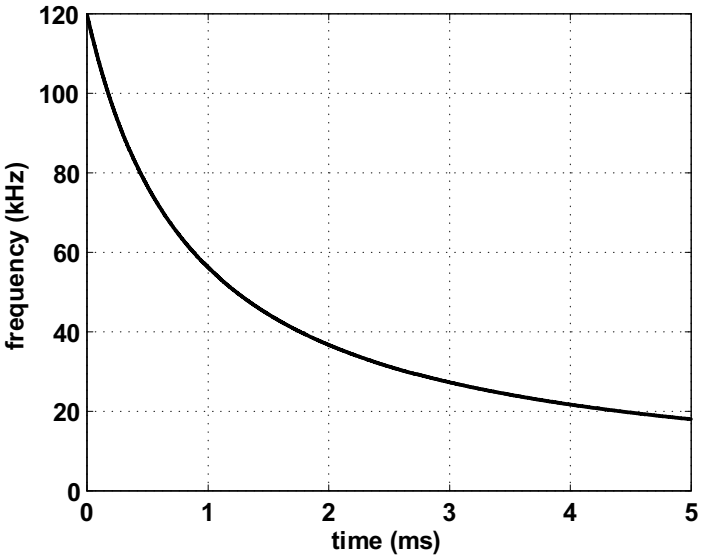

Figure 3.12: Hyperbolic frequency modulated (FM) sweep. The instantaneous frequency changes from $120kHz$ to $18kHz$ in $5ms$.

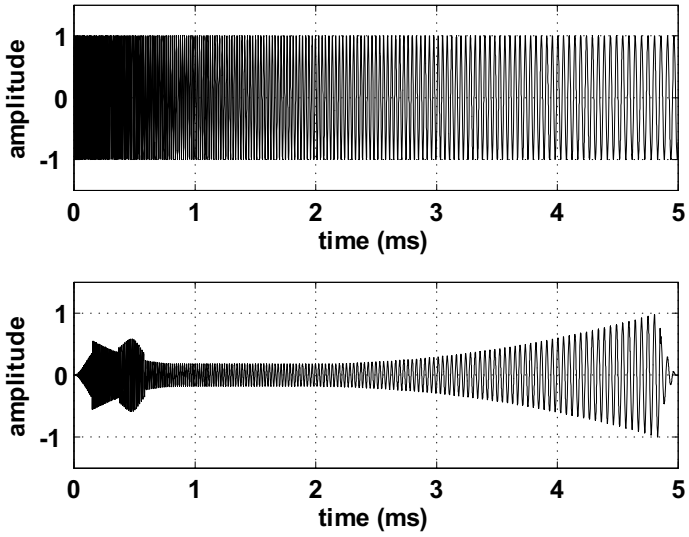

Figure 3.13: Hyperbolic sweep: time domain waveform. The top plot shows the original hyperbolic sweep having a constant amplitude throughout the entire duration of the signal. The lower plot shows the modified version.

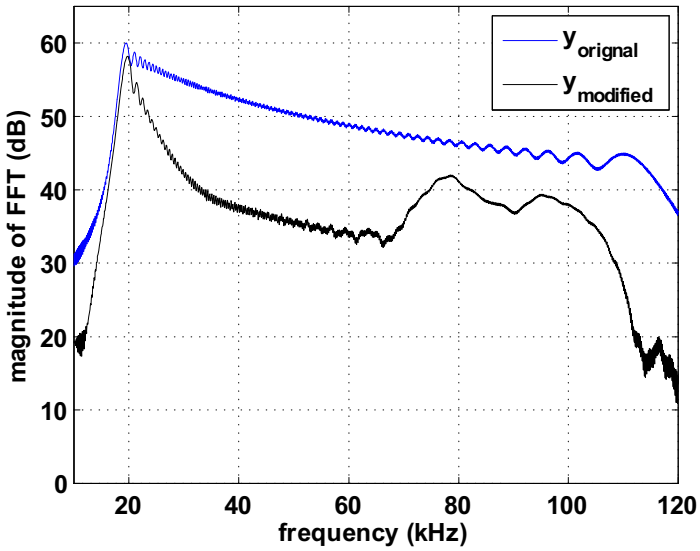

Figure 3.14: Magnitude spectrum of the original and modified (shaped) hyperbolic sweep.

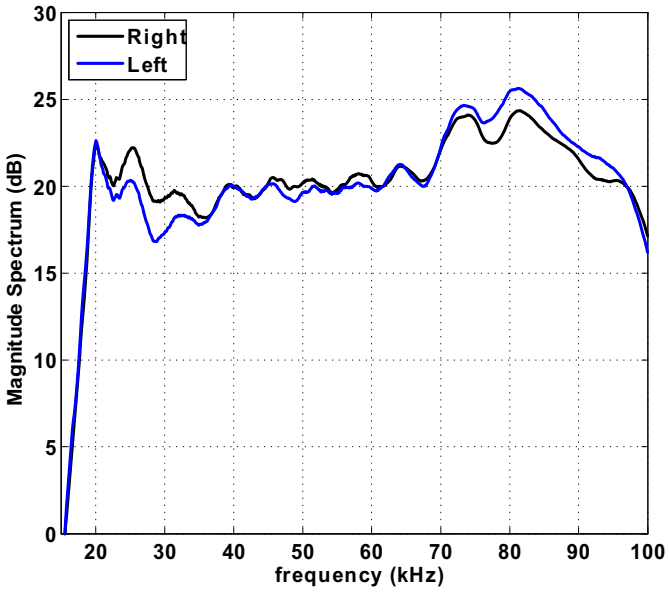

Figure 3.15: Magnitude spectrum of the sound received at the right and left microphones (before inserting the microphones into the ear canals) after being amplified by the 3-stage amplifier.

3.6 Elevation and Azimuth

Fig. 3.16 shows how azimuth and elevation are defined. In this plot, the bat head is assumed to be centered at the origin (**O**) facing the $+ve$ x-axis and the xz plane is the mid-sagittal plane. The target (**T**) has the spherical coordinates (r,θ,ϕ) where r is the range, θ is the elevation and ϕ is the azimuth. In our localization problem we are only interested in determining the direction (i.e. elevation and azimuth) of our sound source (target). It can be shown that any point (**T**) in space can be uniquely mapped onto the surface of a sphere by finding the intersection of the line **OT** with the surface of the sphere, therefore, without loss of generality it can be assumed that $r = 1$. The elevation (θ) is the angle between the xy plane and the line connecting the origin to the target (**T**). Targets with the same elevation θ lie on a cone whose vertex is the origin, axis is the z-axis, and whose generatrix makes an angle θ with the xy plane. The azimuth (ϕ) is defined as the angle (in the xy plane) between the x axis and the line connecting the origin to the projection (**P**) of the target on the xy plane. Targets with the same azimuth ϕ lie on a plane that makes an angle ϕ with the xz plane. Using the unit sphere as a compact representation of the entire space, targets with the same elevation lie on a circle parallel to the xy plane while targets with same azimuth lie on a circle whose plane makes an angle ϕ with the xz plane.

Fig. 3.17 shows the sampled directions from which the ultrasonic sounds were emitted, the artificial bat head being placed at the center of the sphere. The frontal hemisphere was sampled in azimuth from 90° to -90° in steps of 5 degrees

and in elevation from -67.5° to 67.5° in steps of 7.5 degrees for a total of 703 directions (37 azimuth x 19 elevation). Notice that the sampled directions are not uniformly spaced on the sphere; as we move away from the equator, the radius of the constant elevation circle becomes smaller and smaller and eventually becomes zero at the poles of the sphere. This means that the north (south) pole will have an elevation of 90 (-90) degrees and all possible azimuth values. An alternative definition for azimuth is shown in Fig. 3.18. The new (modified) azimuth ϕ_m is defined as the angle between the line **OT** and the xz plane. Using this definition, directions with the same azimuth lie on a circle whose plane is parallel to the xz plane and whose center lies on the y-axis. It can be shown that the relation between the azimuth ϕ_m and the length (**d**) of the projection of **T** to the xz plane is given by:

$$d = \sin(\phi_m) \tag{3.21}$$

The importance of this relation lies in the fact that d is a measure of how far **T** is from the mid-sagittal plane (xz plane), i.e. from the plane of symmetry of the bat head. Fig. 3.19 illustrates another difference between the two definitions of azimuth. The top plot shows that ϕ and θ are independent, i.e. knowing ϕ does not provide any information about θ, however, the lower plot shows that ϕ_m and θ are correlated. The implication in the context of the localization problem is that interaural level difference (ILD) depends on the position of the sound source relative to the plane of symmetry (i.e. the mid-sagittal plane), therefore, knowing the azimuth ϕ_m may assist in determining the elevation of the target.

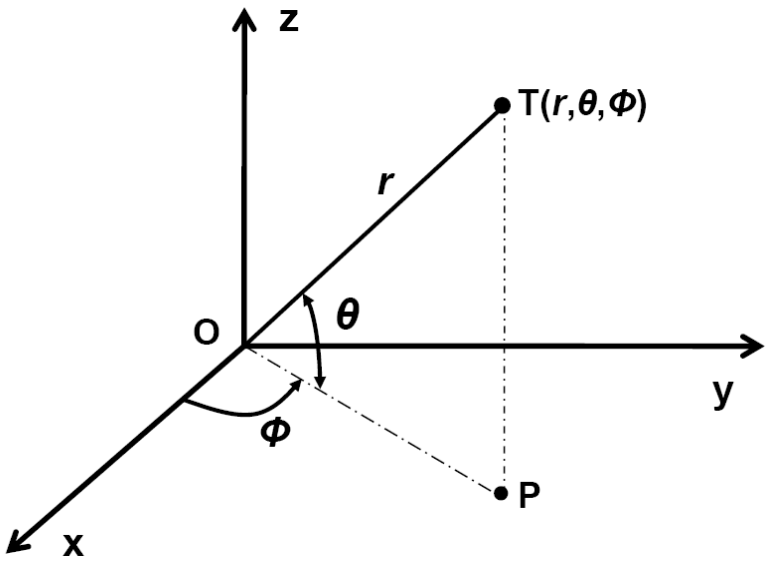

Figure 3.16: Definition of the azimuth (ϕ) and elevation (θ). The target (**T)** has the spherical coordinates (r,θ,ϕ) where r is the range, θ is the elevation and ϕ is the azimuth. The azimuth (ϕ) is the angle (in the xy plane) between the x-axis and the line **OP,** where **P** is the projection of the target on the xy plane. The elevation (θ) is the angle between the xy plane and the line **OT**.

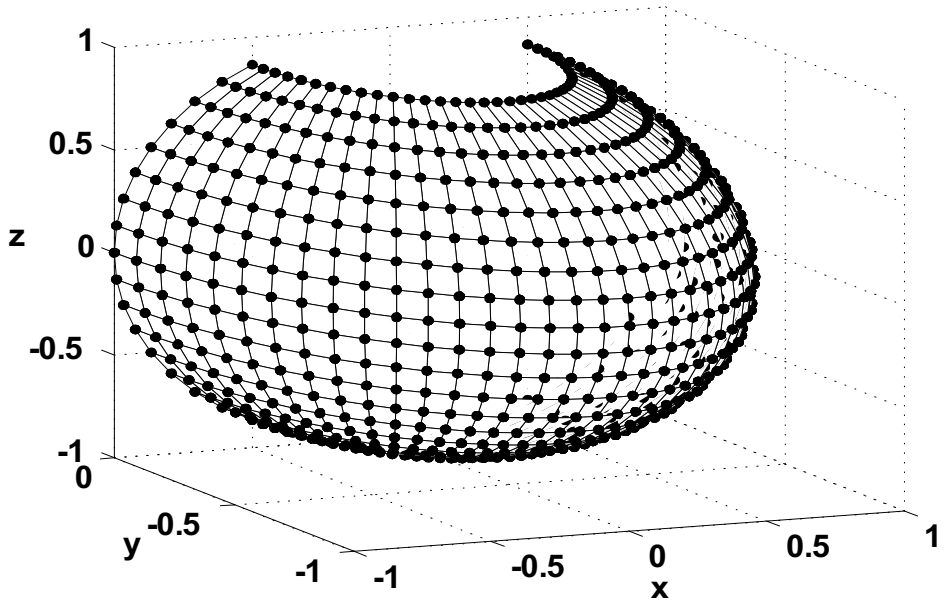

Figure 3.17: Sampled sound directions. The frontal hemisphere was sampled in azimuth from 90° to −90° in steps of 5 degrees and in elevation from −67.5° to 67.5° in steps of 7.5 degrees for a total of 703 directions.

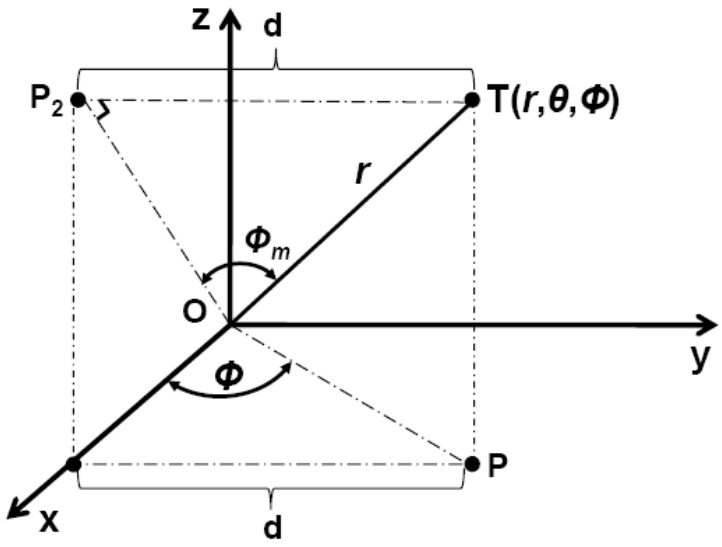

Figure 3.18: Modified azimuth (ϕ_m). The modified azimuth ϕ_m is defined as the angle between the line **OT** and the xz plane. **P** is the projection of the target on the xy plane while $\mathbf{P_2}$ is the projection of the target on the xz plane. The relation between ϕ_m and the distance (d) of T from the xz plane is d=sin(ϕ_m) under the assumption that r=1.

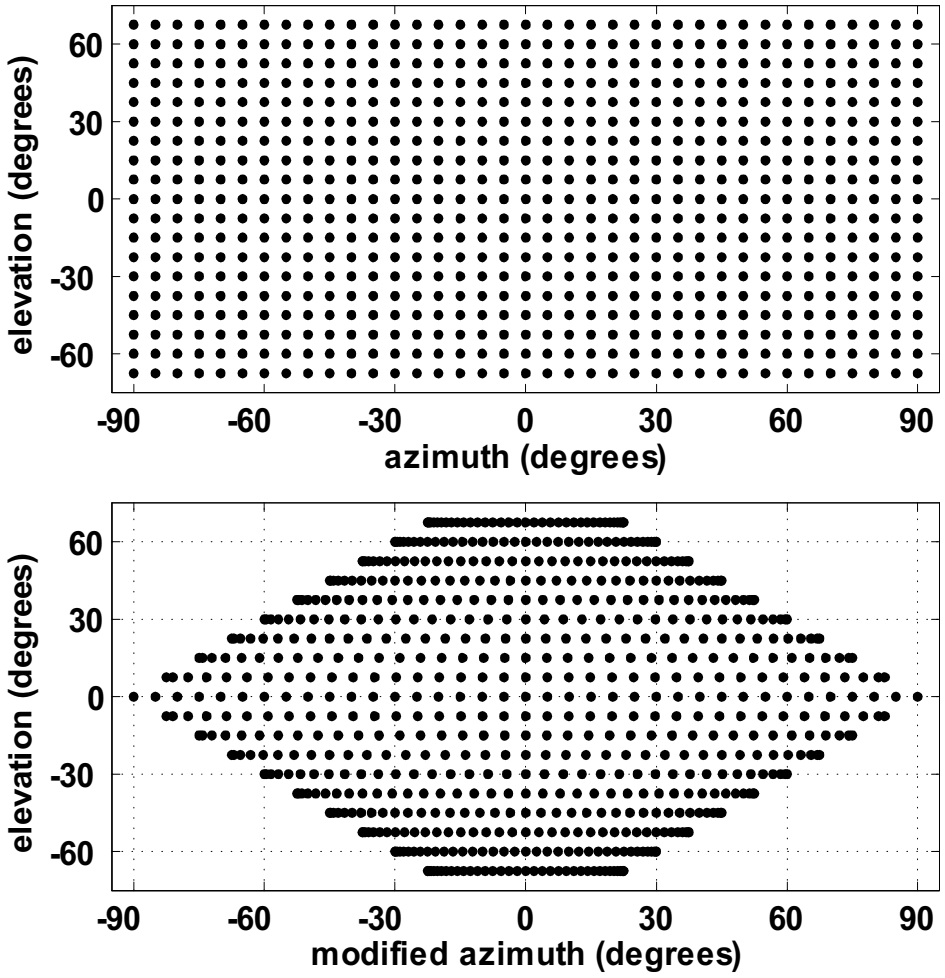

Figure 3.19: Azimuth (ϕ) and modified azimuth (ϕ_m) versus elevation. The upper plot shows that ϕ and θ are independent while the lower plot shows that ϕ_m and θ are correlated.

3.7 Test Results

A $5ms$ hyperbolic FM frequency sweep ($120kHz - 20kHz$) was played from a speaker at a distance of approximately $83cm$ from the bat head. Recordings were made at 703 different directions (19 elevations x 37 azimuths). The speaker was scanned in the horizontal plane (azimuth) from $90°$ to $-90°$ in steps of 5 degrees and in the vertical plane (elevation) from $-67.5°$ to $67.5°$ in steps of 7.5 degrees. The right and left microphone signals were amplified and recorded using a TDS5104 Tektronix oscilloscope. The sampling rate was set at 1M samples/sec and the waveforms were averaged 50 times.

Figs. 3.20 and 3.21 show the spatial contour plot of the magnitude (in dB) of the FFT of the sound at eight different frequencies as measured at the left and right ears respectively. Notice that the colorbars are different and that the size of the *spatial receptive field* (above a given threshold) becomes smaller as the frequency increases, i.e. hearing becomes more directional as frequency increases, this has also been observed in measurements of many bat species [5]. It is also observed that as the frequency increases, the shape of the spatial receptive field becomes more complex and a notch appears in the spectrum for some the of higher frequencies. As is desirable, the two measurements are qualitatively symmetrical.

Shown in Fig. 3.22 is the spectrum of the sound received at the right ear at a fixed azimuth of 45 degrees and from different elevations. It is observed that the notches in the spectrum disappear for higher elevations. Shown in Fig. 3.23 is the spectrum of the sound received at the right ear at a fixed elevation of -15

degrees and from different azimuths.

The *frequency axis* is defined as the direction for which the magnitude spectrum is the greatest for a particular frequency. Figs. 3.24 and 3.25 show how the frequency axis changes with elevation and azimuth respectively; as the frequency increases the frequency axis rises in elevation and moves towards the mid-sagittal plane. This trend has also been observed in measurements of *Eptesicus fuscus* [101][56].

Fig. 3.26 shows the spatial contour plot of the magnitude (in dB) of the interaural level difference (ILD) at eight different frequencies (ILD=Right-Left). It is observed that at the lowest frequency ($20kHz$) the change in ILD is relatively smooth (compared to the higher frequencies) as the sound source moves from the extreme right to the extreme left. It is also observed that for some of the higher frequencies (e.g. $70kHz$) there are directions where sounds from the contralateral (i.e. the other side of the head) hemisphere are louder than sounds from the ipsilateral hemisphere.

Fig. 3.27 shows the ILD at $20kHz$ as a function of azimuth (ϕ) for different elevations. It is observed that as elevation of the sound source decreases (increases) below (above) the equator, the maximum ILD decreases. This is attributed to the fact that as the sound source move away from the equator, the radius of the constant-elevation circle becomes smaller bringing the sound source closer to the plane of symmetry (as shown in Fig. 3.17) where the ILD should be equal to zero. As a result, directions that have the same azimuth ϕ but different elevations are not equidistant to the plane of symmetry (mid-sagittal plane). Fig. 3.28 shows

the same ILD data as in Fig. 3.27 but as a function of the modified azimuth (ϕ_m). The figure illustrates that the ILD correlates better with the modified azimuth ϕ_m, this is attributed to the fact that ϕ_m is related to the distance from the plane of symmetry as discussed in the previous section.

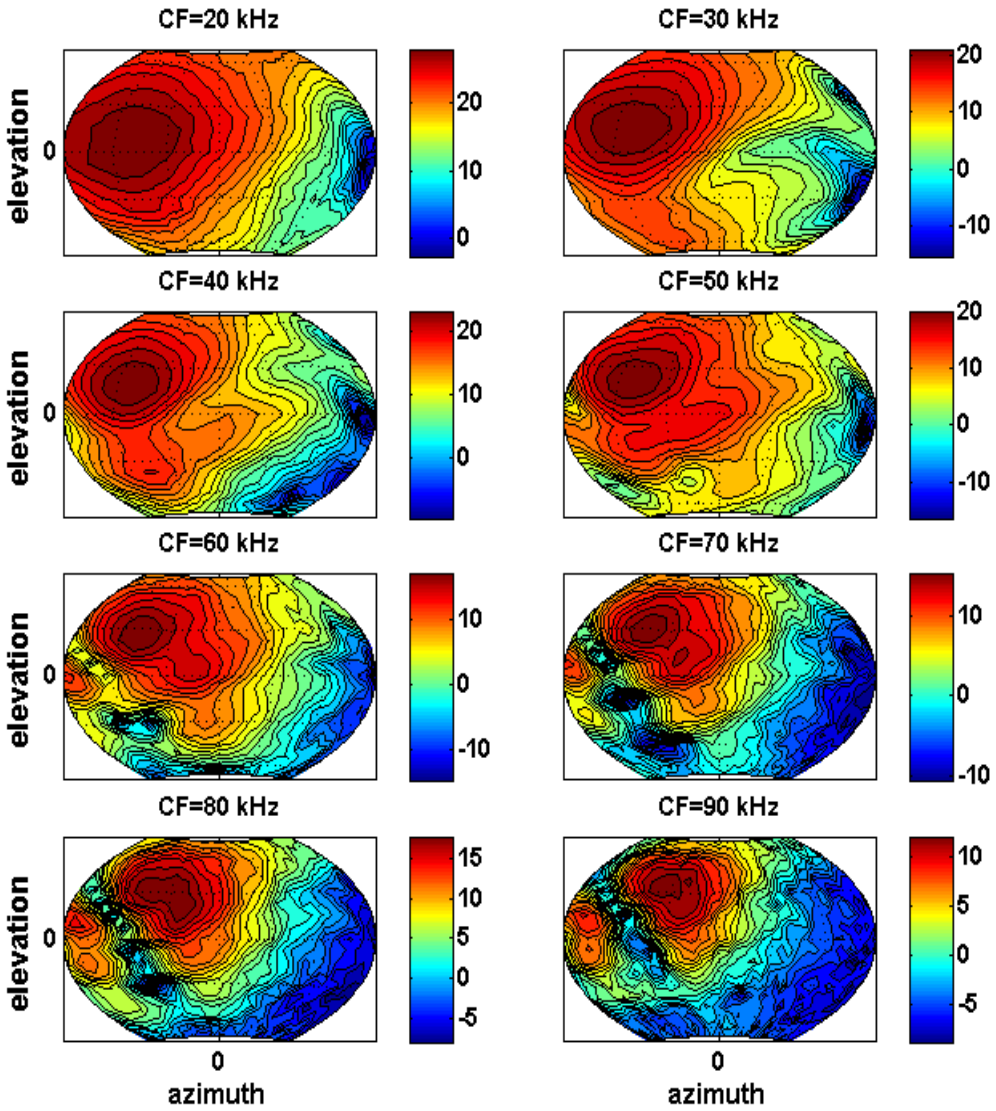

Figure 3.20: Spatial contour plot of the magnitude (in dB) of the FFT of the sound received at the left ear at different frequencies (CF). The grid lines have a separation of 30 degrees. Notice that the colorbars are different.

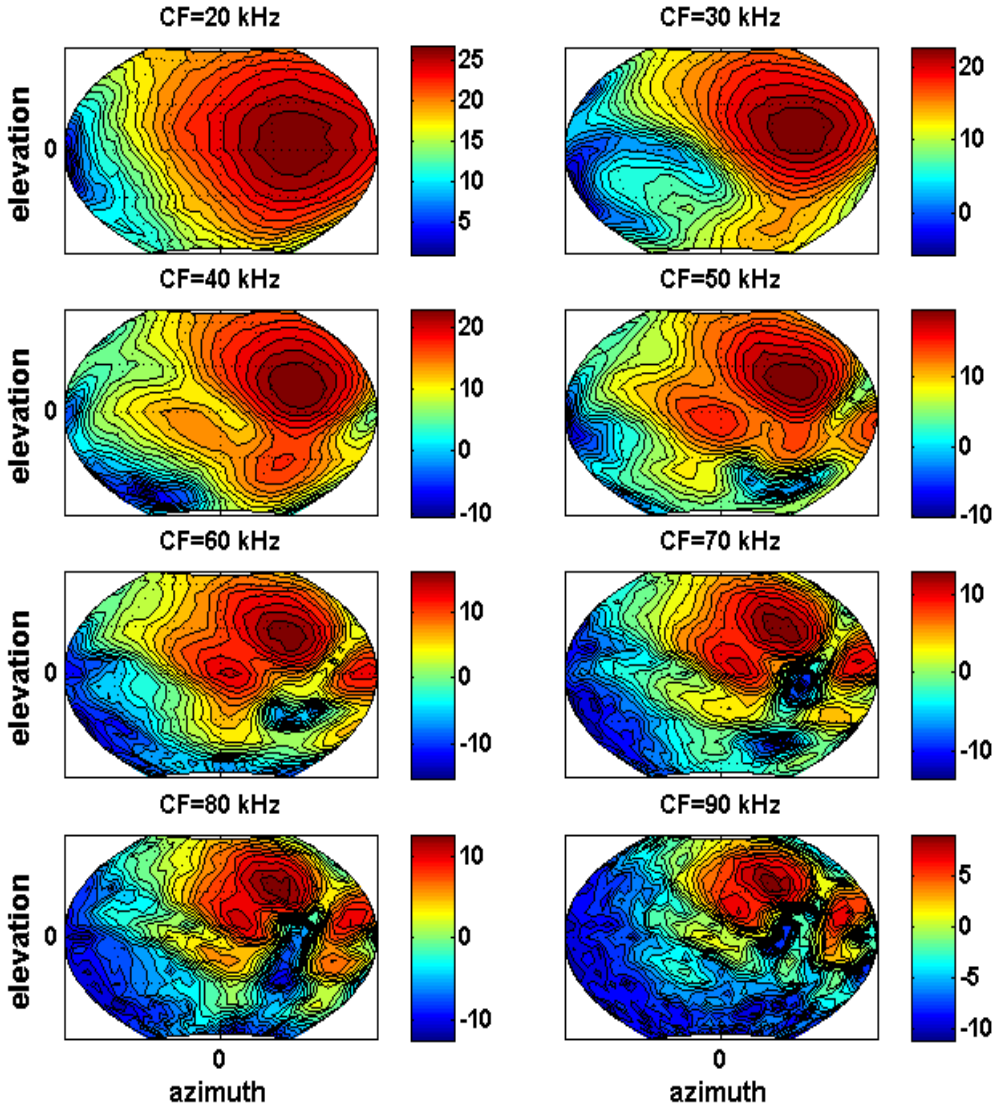

Figure 3.21: Spatial contour plot of the magnitude (in dB) of the FFT of the sound received at the right ear at different frequencies (CF). The grid lines have a separation of 30 degrees. Notice that the colorbars are different.

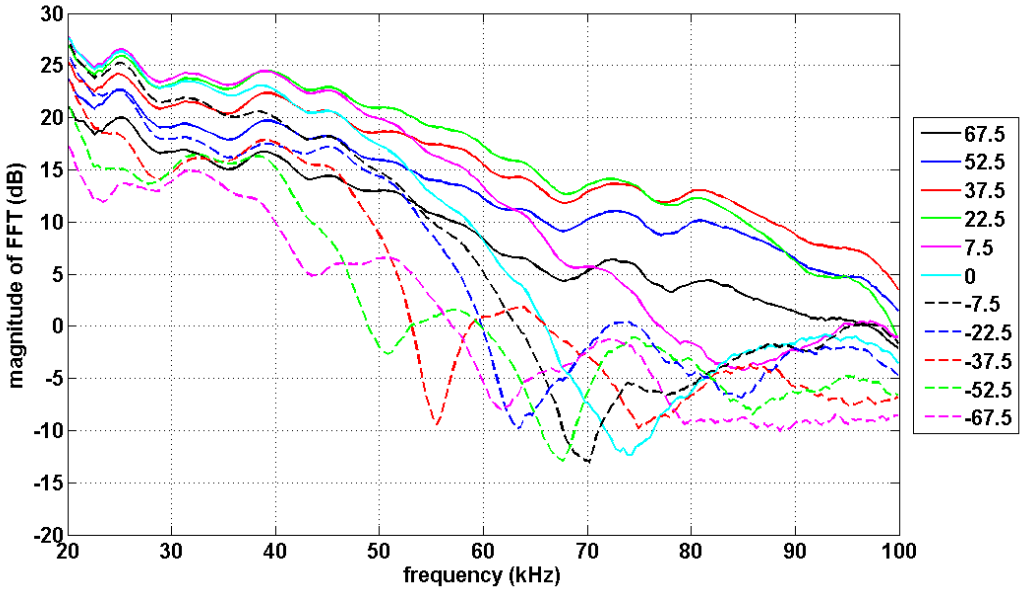

Figure 3.22: Magnitude spectrum of sound measured at the right ear at a fixed azimuth of 45 degrees (ipsilaterally) and for different elevations.

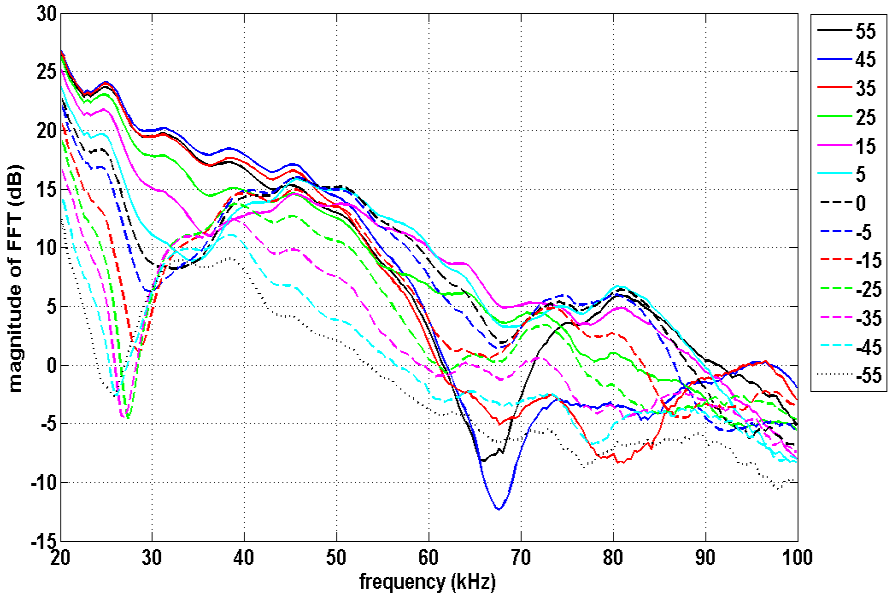

Figure 3.23: Magnitude spectrum of sound measured at the right ear at a fixed elevation of -15 degrees and for different azimuths.

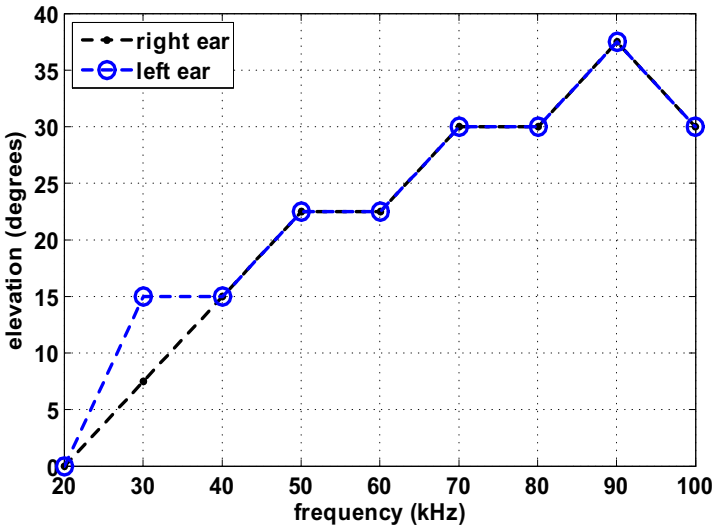

Figure 3.24: The *frequency axis* is defined as the direction for which the magnitude spectrum is the greatest for a particular frequency. The plot shows the elevation of the *frequency axis* for different frequencies. As as the frequency increases, the *frequency axis* rises in elevation.

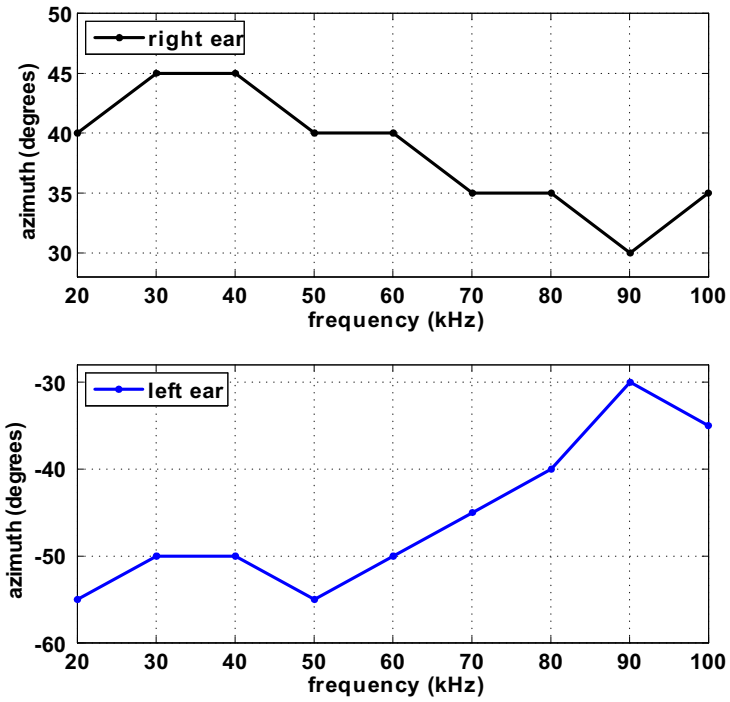

Figure 3.25: The plot shows the azimuth of the frequency axis for different frequencies. As the frequency increases the *frequency axis* moves towards the mid-sagittal plane.

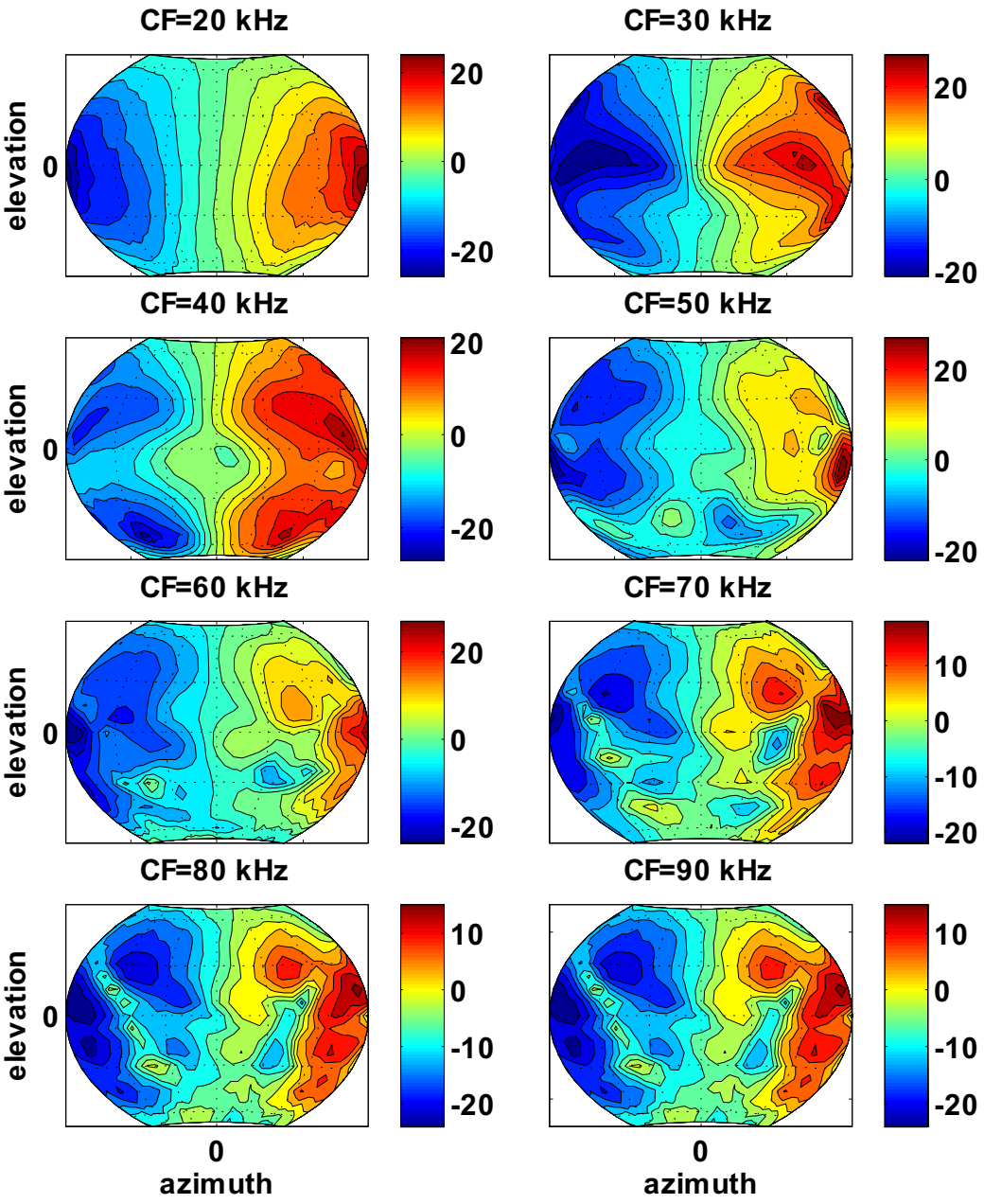

Figure 3.26: Spatial contour plot of the interaural level differences (ILD) in dB at different frequencies (CF). Notice that the colorbars are different. The grid lines have a separation of 30 degrees.

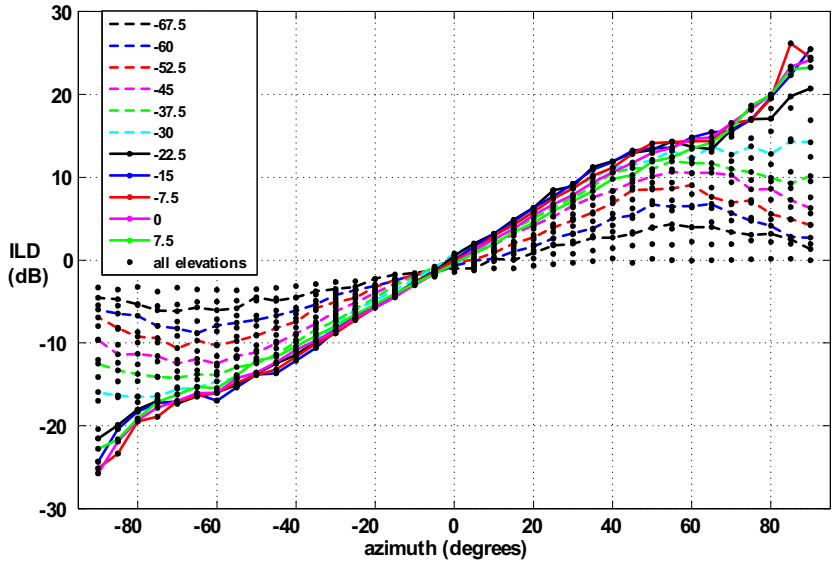

Figure 3.27: ILD at $20kHz$ vs. azimuth (ϕ) for different elevations. To reduce clutter, the ILD data points for elevations greater than 7.5 degrees were drawn without any connecting lines. Notice that as the elevation decreases below the equator the maximum ILD decreases. This is attributed to the fact that as the sound source moves away from the equator the radius of the constant-elevation circles decrease bringing the sound source closer to the plane of symmetry where the ILD should be equal to zero.

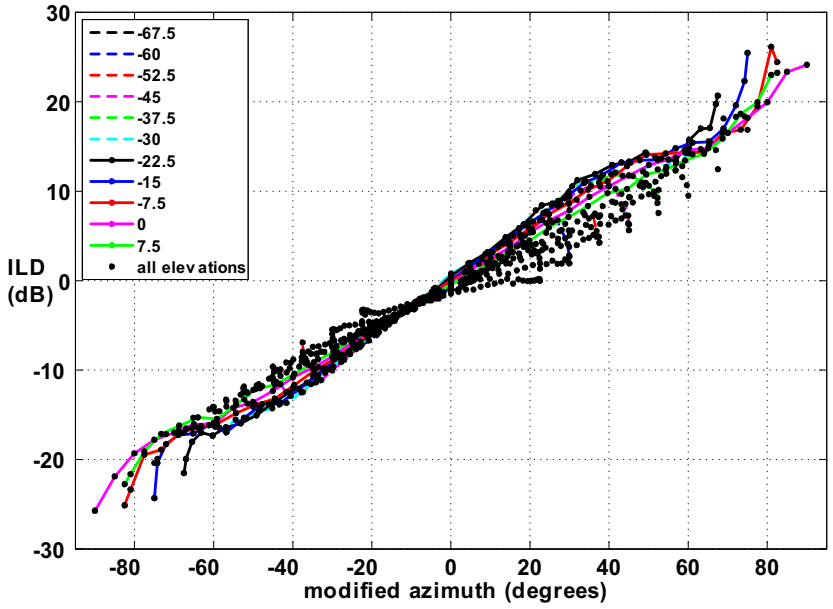

Figure 3.28: ILD at $20kHz$ vs. modified azimuth (ϕ_m) for different elevations. This is the same data as in Fig. 3.27. This plot shows the improved correlation between the ILD and ϕ_m which is attributed to the fact that ϕ_m is linked to the distance of the sound source from the plane of symmetry (see text).

3.8 Conclusion

In this chapter we have presented an artificial bat head that was designed using 3D CAD software and fabricated using a 3D printer. We measured the response of the bat head to ultrasonic hyperbolic FM sweeps and illustrated that the bat head generates both monaural spectral cues and binaural ILD cues. In Chapter 6 we will show that there is sufficient information in the measured responses to estimate the elevation and azimuth of the sound source.

Chapter 4

Ultrasonic Cochlea Chip

4.1 Introduction

The electronic cochlea is the first stage of processing of the ultrasonic echoes. Modeled after the mammalian cochlea, it transforms the analog acoustic sound from one ear into a parallel set of discrete amplitude, spike-based representation outputs.

As discussed in Chapter 1, Lyon's original electronic cochlea [27] and many improved implementations [28][29] were based on the cascaded filter model, using second-order filter stages and were tuned to operate in the human audible range ($<20kHz$). Our cochlea design is based on a parallel array of bandpass filters and are tuned to operate in the ultrasonic frequency range $20kHz$-$100kHz$. An integrate-and-fire neuron circuit converts the membrane potential into a spiking representation. Finally, the address-event representation (AER) is used to encode the address of the firing neuron. We first begin by discussing the cochlear filter.

4.2 Cochlear Filter

The ultrasonic cochlea chip has gone through several design changes and implementations, however, the common aspect about all of the different implementations of the cochlea chip is that the cochlear filter bank is based on the filter whose circuit block diagram is shown in Fig. 4.1. This architecture, implemented earlier by Furth [102] and Abdalla and Horiuchi [52] is a tunable high-Q bandpass filter.

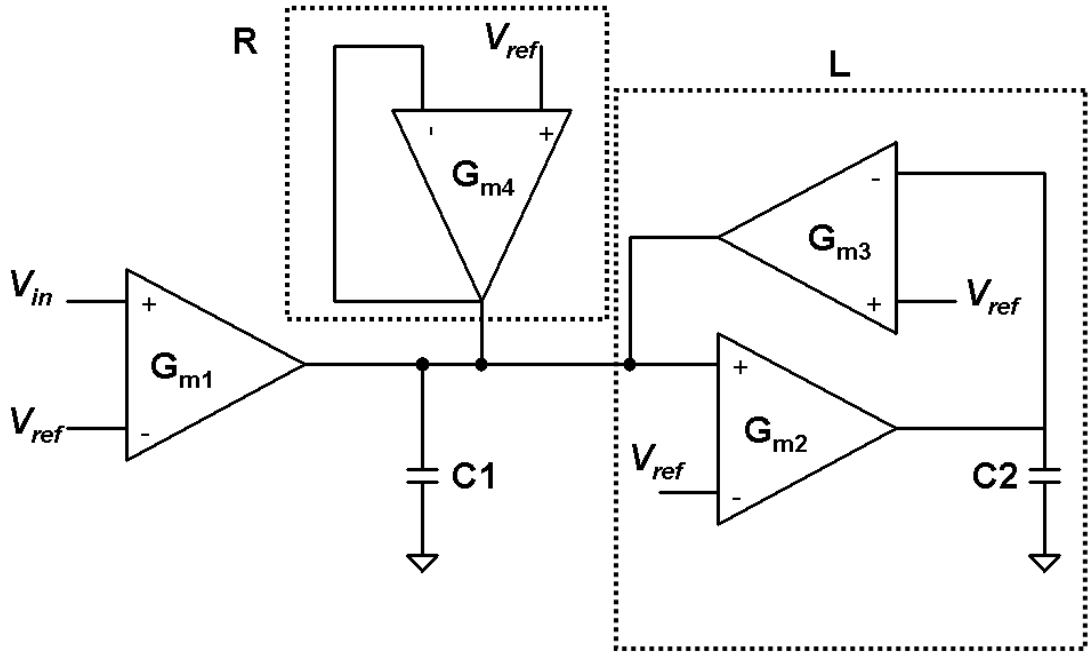

Figure 4.1: The gyrator-based RLC bandpass filter. V_{ref} sets the DC level of the external input signal. G_{mi} is the transconductance of the i^{th} transconductance amplifier. The follower connected G_{m4} can be modeled as a resistor **R**, while G_{m2} and G_{m3} along with C2 can be modeled as an inductor **L**.

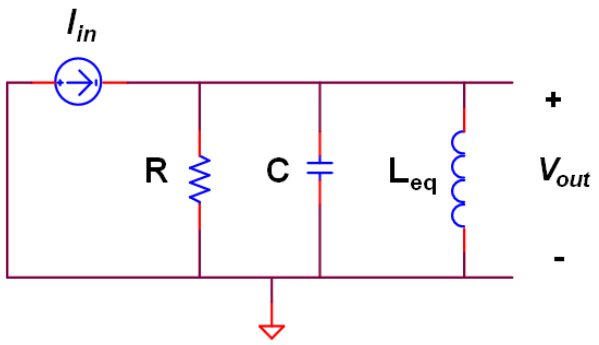

Figure 4.2: The circuit block diagram shown in Fig. 4.1 can be modeled as a parallel RLC filter. Leq is the equivalent inductor formed by the gyrator circuit.

This circuit can be modeled as a parallel RLC filter as shown in Fig. 4.2 where:

$$\begin{aligned} I_{in} &= G_{m1}V_{in} \\ R &= \frac{1}{G_{m4}} \\ C &= C1 \\ L_{eq} &= \frac{C2}{G_{m2}G_{m3}} \end{aligned} \tag{4.1}$$

The output of the filter is given by:

$$V_{out} = \frac{I_{in}}{Y} = \frac{G_{m1}V_{in}}{G_{m4} + sC1 + \frac{1}{sL_{eq}}} \tag{4.2}$$

The transfer function of this filter can be written as:

$$H(s) = \frac{V_{out}}{V_{in}} = \frac{Ks}{s^2 + s\left(\frac{\omega_o}{Q}\right) + \omega_o^2} \tag{4.3}$$

where:

$$\begin{aligned} \omega_o &= \frac{\sqrt{G_{m2}G_{m3}}}{C}, \\ Q &= \frac{\sqrt{G_{m2}G_{m3}}}{G_{m4}}, \\ K &= \frac{G_{m1}}{C}. \end{aligned} \tag{4.4}$$

where ω_o is the center frequency of the filter, Q is the quality factor, K is a gain factor, G_{mi} is the transconductance of the i^{th} transconductance amplifier ($i = 1, 2, 3,$ and 4), and $C1 = C2 = C$. Among the several cochlea chips that have been fabricated, one difference lies in the implementation of this cochlear filter.

4.3 Cochlea Chips

The different implementations of the cochlea chips will be referred to as LeqC2, LeqC3, ... and LeqC7 (Leq refers to the equivalent inductance). We first begin by describing LeqC2, one of the earliest cochlea chips designed in the Computational Sensorimotor Systems Lab (CSSL).

4.3.1 LeqC2

The block diagram of the cochlea chip LeqC2 shown in Fig. 4.3 is one of the early cochlea chips designed in the Computational Sensorimotor Systems Lab (CSSL) [52]. The chip was designed by Prof. Timothy Horiuchi and the testing and mathematical analysis was done by the author. The chip has two cochlear filterbanks. Each cochlear filterbank is modeled as an array of ten second-order bandpass filters

whose center frequencies are logarithmically spaced, modeling that of a biological cochlea. This array of filters covers the frequency range of $20kHz$-$100kHz$. The outputs are then full-wave rectified and converted to currents that are injected into an integrate-and-fire neuron. The output spikes generate fixed-value inward or outward current pulses on the address lines to represent the address of the neuron. A current sense amplifier was used (external to the chip) to convert the currents into a voltage to read out the neuron's address. No address-event arbiter was used in this chip.

The schematic shown in Fig. 4.4 is the implementation of the cochlear filter used in the chip LeqC2. All ten cochlear filters share the same voltage V_{ref} which sets the DC level of the external input signal. The filters also share the same voltages V_{fbk} and V_{offset}. The voltage V_L was generated using a uniform polysilicon resistor ladder. The two end voltages (V_{L1} and V_{L2}) of the ladder were externally set using potentiometers.

During the initial stages of testing the cochlear filters, the voltage V_{offset} was generated using a potentiometer. As the voltage V_L was changed to set the filters' center frequencies, it was observed that V_{offset} changed also. It was therefore necessary to actively hold the voltage V_{offset} at a fixed potential using an external voltage follower. The transconductance amplifiers were operated in the transistor's subthreshold region of operation. The transconductance of the

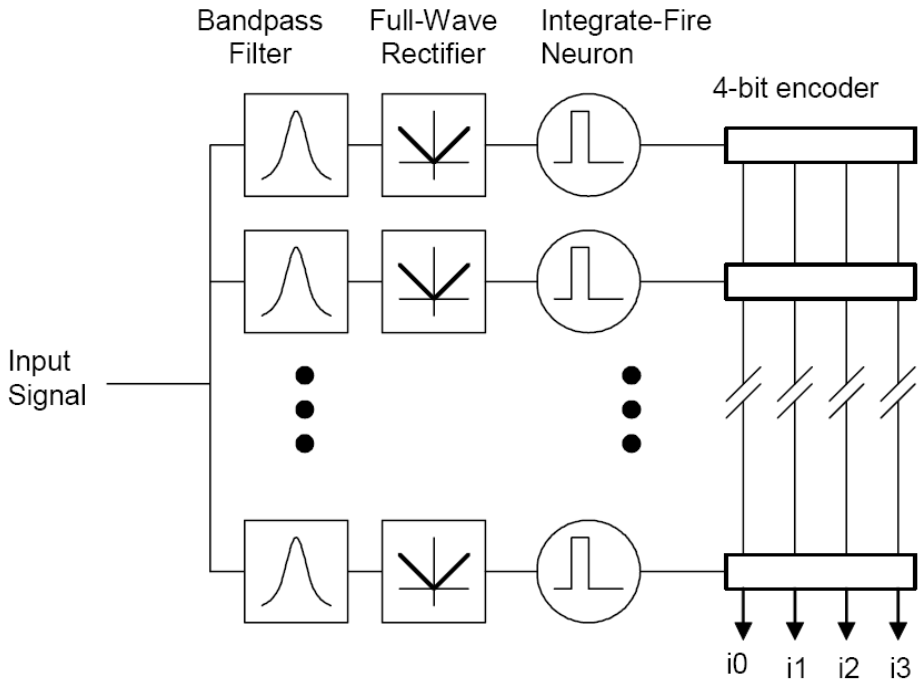

Figure 4.3: Block diagram of the ultrasonic cochlea chip (LeqC2). The input signal is provided in parallel to all filters. The gyrator-based bandpass filters are created with exponentially increasing center frequencies. The outputs are then full-wave rectified and converted to currents that are injected into an integrate and fire neuron. The output spikes generate fixed-value inward or outward current pulses on the address lines to represent the address of the neuron.

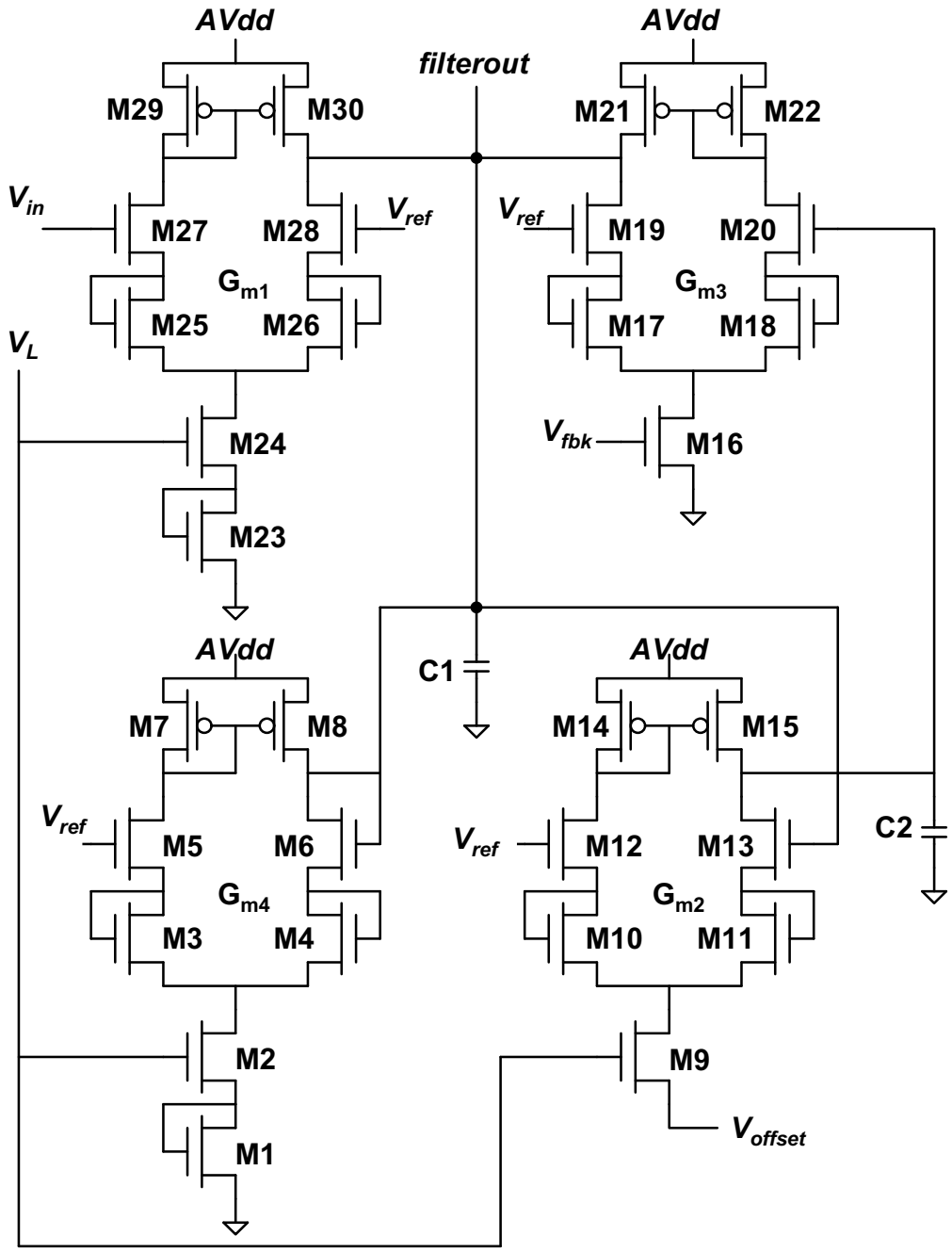

Figure 4.4: Implementation of the individual cochlear filter in the cochlea chip LeqC2. All ten filters in the array share the same V_{offset}, V_{fbk}, and V_{ref}. V_L is generated using a uniform polysilicon resistor ladder. The bulk terminal for all NMOS (PMOS) transistors is connected to ground (Vdd).

four amplifiers is given by:

$$\begin{aligned} G_{m1} &= \frac{\beta_1 I_o}{2}\exp(\beta_1 V_L), \\ G_{m2} &= \frac{\beta_1 I_o}{2}\exp(\beta_2 V_L - \alpha V_{offset}), \\ G_{m3} &= \frac{\beta_1 I_o}{2}\exp(\beta_2 V_{fbk}), \\ G_{m4} &= \frac{\beta_1 I_o}{2}\exp(\beta_1 V_L), \end{aligned} \tag{4.5}$$

where

$$\begin{aligned} \beta_1 &= \frac{\kappa^2}{(1+\kappa)V_T}, \\ \beta_2 &= \frac{\kappa}{V_T}, \\ \alpha &= \frac{1}{V_T} \end{aligned} \tag{4.6}$$

where κ is the body effect [103], V_T is the thermal voltage, and I_o is the zero-bias current. The center frequency ω_o can be written as

$$\begin{aligned} \omega_o &= \frac{1}{C}\frac{\beta_1 I_o}{2}\exp(\frac{\beta_2}{2}(V_{fbk}+V_L) - \frac{-\alpha}{2}V_{offset}) \\ &= \frac{I_o}{C}\exp(-\frac{1}{2V_T}V_{offset})\exp(\frac{\kappa}{2V_T}(V_L+V_{fbk})) \end{aligned} \tag{4.7}$$

The quality factor, Q can be written as

$$\begin{aligned} Q &= \exp(\frac{\beta_2}{2}(V_{fbk}+V_L) - \frac{-\alpha}{2}V_{offset} - \beta_1 V_L) \\ &= \exp(-\frac{1}{2V_T}V_{offset})\exp(\frac{1}{2V_T}(\kappa V_{fbk} - \frac{V_L}{1+\kappa}) \end{aligned} \tag{4.8}$$

Equations 4.7 and 4.8 show that Q and ω_o are coupled through V_L. Since all filters share the same V_{offset} and V_{fbk}, it is the voltage V_L (set by the resistive ladder) that allows the filters to have different center frequencies. If the filterbank is to

span the frequency range from $20kHz$ to $100kHz$ then, from equation 4.7, the voltages V_{L1} and V_{L2} must be set such that

$$\frac{\omega_o^{10}}{\omega_o^1} = \frac{100}{20} = \exp(\frac{\kappa}{2V_T}(V_{L2} - V_{L1}) \tag{4.9}$$

where ω_o^1 and ω_o^{10} are the center frequencies of the first and last cochlear filters respectively. V_{L2} sets ω_o^{10}, while V_{L1} sets ω_o^1. Equation 4.9 can be rewritten as

$$\begin{aligned} (V_{L2} - V_{L1}) &= \frac{2V_T}{\kappa} \ln 5 \\ &\approx 119mV \end{aligned} \tag{4.10}$$

where $V_T \approx 25.8mV$ and a value of 0.7 was chosen for κ. From equation 4.8

$$\frac{Q^{10}}{Q^1} = \exp(\frac{-1}{(1+\kappa)2V_T}(V_{L2} - V_{L1})) \tag{4.11}$$

where Q^1 and Q^{10} are the quality factors of the first and last cochlear filters respectively. Equation 4.11 predicts that all filters will have a different Q (since each filter has a different V_L) and that the change is systematic such that the filters with a higher center frequency will have a lower Q compared to the filters with a lower center frequency. Substituting 4.10 in 4.11 results in

$$\begin{aligned} \frac{Q^{10}}{Q^1} &= \exp(\frac{-1}{(1+\kappa)\kappa} \ln 5) \\ &\approx 0.26 \end{aligned} \tag{4.12}$$

This results predicts that the quality factor will change by a factor of approximately four for a filterbank spanning the frequency range $20kHz - 100kHz$.

4.3.2 LeqC3

The implementation of the cochlear filter in LeqC2 suffered from several drawbacks. One drawback is that it is difficult to change the quality factor of a filter without changing its center frequency. Another drawback is that, theoretically, the quality factor is not constant across the filter array. A third, but minor, drawback is that the filters require that the bias voltage V_{offset} be actively held using an external voltage follower.

Modified Cochlear Filter

Equation 4.8 reveals that if $\beta_1 = \beta_2$, then the quality factor (Q) would be dependent on the difference between V_{fbk} and V_L. The implementation of the transconductance amplifiers G_{m2} and G_{m3} were modified from the earlier design shown in Fig. 4.4 so that all four transconductance amplifiers have the same implementation. V_{offset} was connected to ground, transistor M9 (in Fig 4.4) was replaced by transistors M9 and M10 in Fig. 4.5, and transistor M16 (in Fig 4.4) was replaced by transistors M17 and M18 in Fig. 4.5. The schematic of the modified cochlear filter is shown in Fig. 4.5. The transconductances of the modified cochlear filter can be written as:

$$\begin{aligned} G_{mi} &= \frac{\beta_1 I_o}{2} \exp(\beta_1 V_L), \qquad for\ i = 1, 2, \text{ and } 4, \\ G_{m3} &= \frac{\beta_1 I_o}{2} \exp(\beta_1 V_f), \end{aligned} \tag{4.13}$$

The center frequency (ω_o) of the modified cochlear filter can now be written as:

$$\omega_o = \frac{1}{C} \exp(\beta_1 (\frac{V_{fbk} + V_L}{2})), \tag{4.14}$$

And the quality factor (Q) of the modified filter can be written as

$$Q = \exp(\beta_1(\frac{V_{fbk} - V_L}{2})). \tag{4.15}$$

This modified design has been implemented in the cochlea chips: LeqC3, LeqC4, LeqC5, and LeqC6. The bias voltages V_{fbk} and V_L are generated using two uniform polysilicon resistive ladders (see Fig. 4.8) and are set such that the difference $(V_{fbk} - V_L)$ remains constant for each filter thus achieving a constant-Q filterbank whereas the average voltage $(V_{fbk} + V_L)/2$ increases linearly resulting in the center frequencies being exponentially spaced to mimic the frequency spacing in the bat cochlea. In addition to the modified cochlear filter, the full-wave rectifier in LeqC2 was replaced with a half-wave rectifier that models the receptor potentials seen in inner hair cells.

4.3.3 LeqC4

Fig. 4.6 shows the block diagram of the cochlea chip LeqC4 (as well as LeqC5, LeqC6, and LeqC7). The number of cochlear filters was increased to 16, with the same filter implementation as in LeqC3, and the number of neurons associated with each cochlear filter was increased from one to eight for a total number of 128 neurons.

Each neuron has a voltage follower that can be selected using an on-chip two-dimensional scanner to monitor its membrane voltage. Each cochlear filter has a voltage follower that can be selected to measure its output (*filterout*).

The filter output voltage (*filterout*) is converted to a current, half-wave recti-

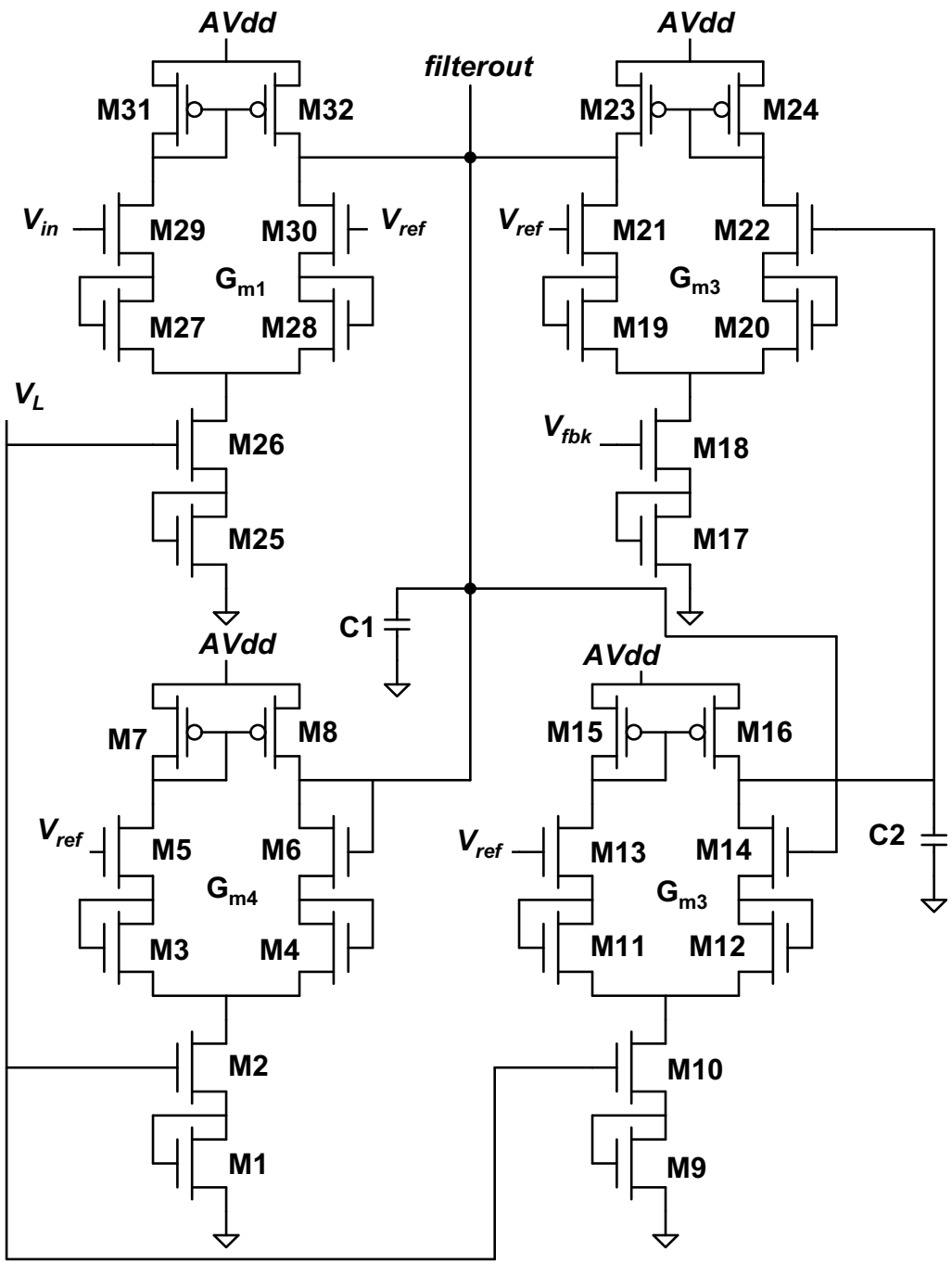

Figure 4.5: Modified cochlear filter. Compared to the implementation shown in Fig. 4.4, V_{offset} has been connected to ground and the diode connected transistors M9 and M17 were added to make $\beta_1 = \beta_2$, thus making the quality factor Q dependent on the voltage difference $V_{fbk} - V_L$.

fied by a p-type current mirror, and mirrored to all eight integrate-and-fire neurons associated with that filter.

All neurons in a given column have the same threshold. In this cochlea chip, five potentiometers were used to set the eight neuron thresholds. One potentiometer provided the voltage bias for one threshold and was completely separate from the rest of the thresholds. Four potentiometers were used as the end voltages for two resistor ladders providing the bias voltages for the remianing seven thresholds. The neuron spikes are transmitted off of the chip using a two-dimensional address-event arbiter.

In the cochlea chip LeqC3 it was observed that some of the neuron fired spikes spontaneously (i.e. with no external input). This was attributed to the non-zero DC current provided to the neuron due to the mismatch in the DC level of the filter output. In section 4.4 we describe an attempt to prevent neurons from firing spikes spontaneously.

4.3.4 LeqC5, LeqC6, and LeqC7

The implementation of the cochlear filter bank in LeqC5 and LeqC6 is the same as in LeqC3 and LeqC4. In LeqC7, however, a slight modification was made to the cochlear filter; the source-degenerated bias transistors M1, M2, M9, M10, M17, M18, M25, and M26 for the transconductance amplifiers shown in Fig. 4.5 were replaced by single transistors (M1, M8, M15, and M22), as shown in Fig. 4.7. To prevent neurons from firing spikes spontaneously, a separate bias was put in

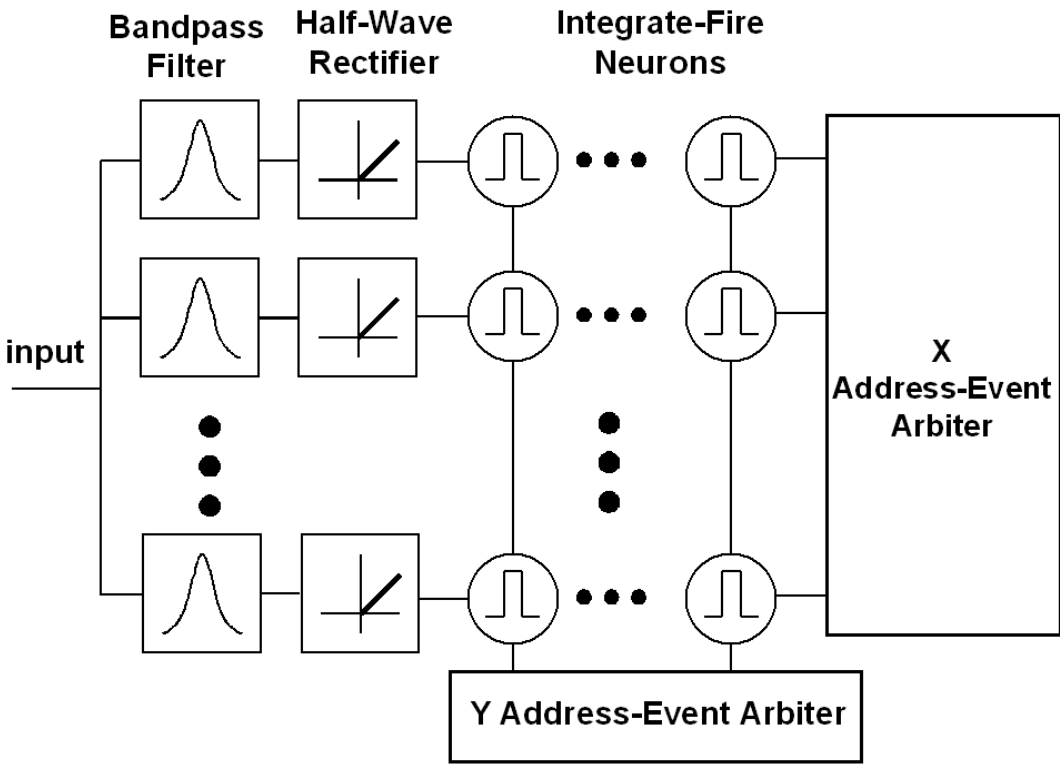

Figure 4.6: Block diagram of the cochlea chips LeqC4, LeqC5, LeqC6, and LeqC7. The cochlear filterbank is a 1D array of 16 bandpass filters. The filter output voltage is converted to a current, half-wave rectified by a p-type current mirror, and mirrored to all eight integrate-and-fire neurons associated with that filter. All neurons in a given column have the same threshold. The spikes are transmitted off of the chip using a 2D address-event arbiter.

place for the reference voltage of the half wave rectifier (see section 4.4) so as to be lowered until no spontaneous firing was observed. The design and layout of cochlea chip LeqC6 is nearly identical to that of LeqC5, the only difference is in how the threshold of the neuron is set and will be discussed in section 4.5. Since it was established that the chip was completely functional, in LeqC7, all but one of the voltage followers used to monitor the membrane voltages were removed and replaced by isolation blocks to increase the isolation between neurons and the vertical scanner was also removed. Five potentiometers were also used to set the eight neuron thresholds, however, in this chip a single uniform resistor ladder provided the bias voltages for five of the eight thresholds The remaining three thresholds were biased by three potentiometers.

4.4 Half-Wave Rectifier

The half-wave rectifier models the function of the inner hair cells. The circuit implementation of the half-wave rectifier is shown in Fig. 4.9 where *filterout* is the voltage output of the cochlear filter. The differential pair circuit compares the voltage *filterout* against the reference voltage V_{ref_hwr}. The voltage V_{b_hwr} controls the bias current I_b of the half-wave rectifier. As *filterout* increases above V_{b_hwr}, the current I_b is steered into the left branch (M3, M5, and M6) and eventually V_{mirror} is pulled up to *AVdd.* As *filterout* drops below V_{b_hwr}, I_b is steered into the right branch (M2, M4, and M8) and the voltage V_{mirror} encodes the magnitude of the current flowing through M8. This voltage is mirrored to

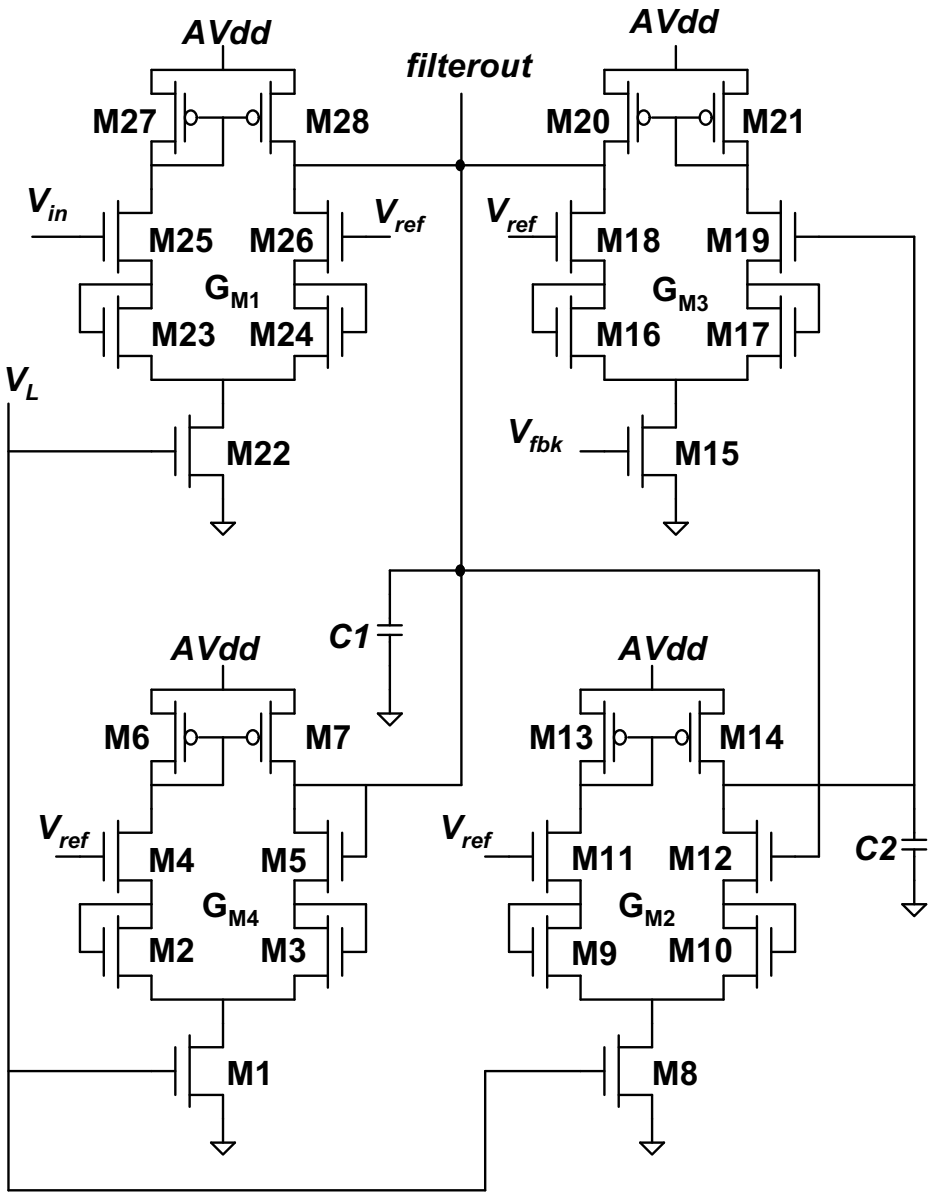

Figure 4.7: Implementation of the cochlear filter in the cochlea chip LeqC7.

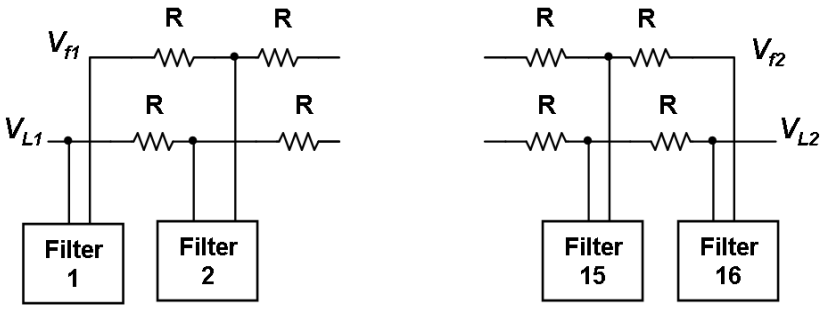

Figure 4.8: Biasing the cochlear filters. Two uniform polysilicon resistor ladders were used to provide the bias voltages V_{fbk} and V_L for each cochlear filter stage. The four end voltages V_{f1}, V_{f2}, V_{L1}, and V_{L2} were externally set.

all eight neurons associated with this filter. Because the output of the half-wave rectifier was not monitored on the chip, we show in Fig. 4.10 a *simulation* of the half-wave rectifier circuit. A $70kHz$ sinusoidal input with a $3V$ DC voltage offset was applied to the gate of M5. V_{ref_hwr} was set to $3V$. As the input voltage increases above the reference voltage, the bias current I_b is steered into the left branch and no current flows in M8. However, as the input voltage decreases below the reference voltage, I_b is steered into the right branch, M6 and M7 begin to turn off, and current flows into M8. Therefore, current is mirrored to the neurons during the negative half-cycle of the input.

In the cochlea chip LeqC3, V_{ref_hwr} was set to equal V_{ref}, the DC voltage offset of the cochlear filters; however, it was observed that some of the neurons fired spikes spontaneously (i.e. with no external input). This was attributed to the mismatch in the DC level of the cochlear filter output *filterout* which results in some neurons receiving excitatory current and causing them to fire spontaneously.

This is a potential problem, for if the spontaneous firing rate is too high, the address-event arbiter may become "locked" and be unresponsive to the rest of the neuron array. Fig. 4.11 illustrates one attempt (implemented in LeqC4) to avoid this problem. A lowpass filter was used to extract the DC offset of the cochlear filter and then to use the extracted voltage as the reference voltage of the half-wave rectifier. Measurements from the chip, however, showed no improvement. It should be noted, however, that both the output of the filter and the lowpass filter were measured through two different followers which may have different voltage offset. The alternative approach was to have the capability to independently set V_{ref_hwr}. Although this approach does not eliminate mismatch, it allows V_{ref_hwr} to be set so as to prevent any neuron from spontaneously firing spikes.

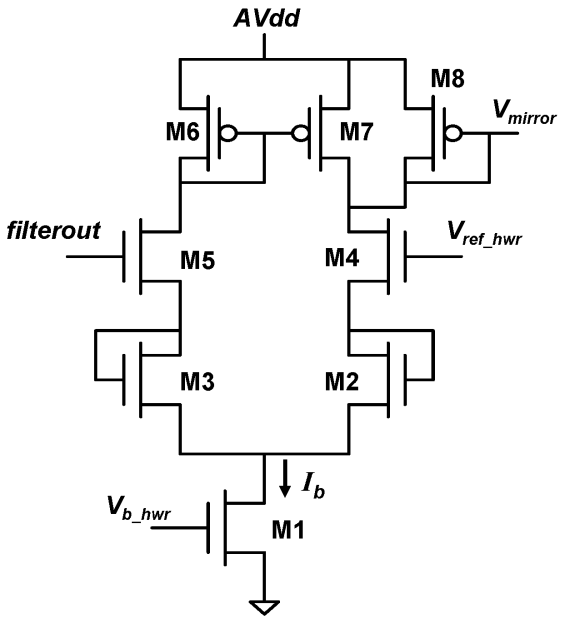

Figure 4.9: Schematic of the half-wave rectifier. The filter's output voltage *filterout* is converted to a current and half-wave rectified using the M8 transistor.

In an earlier design, the filter's voltage output (*filterout*) was connected to the negative input (gate of M4) of the differential pair, while V_{ref_hwr} was connected to the positive input (gate of M5), allowing current to flow through M8 during the positive cycle of *filterout*. However, as the half-wave rectifier was turned on, it was observed that the filter's tuning curve shifted to a lower frequency. This was attributed to the Miller effect, where the rapid change in the drain voltage of M4 (V_{mirror}) was capacitively coupled back into *filterout*. The design was, therefore, modified as shown in Fig. 4.9.

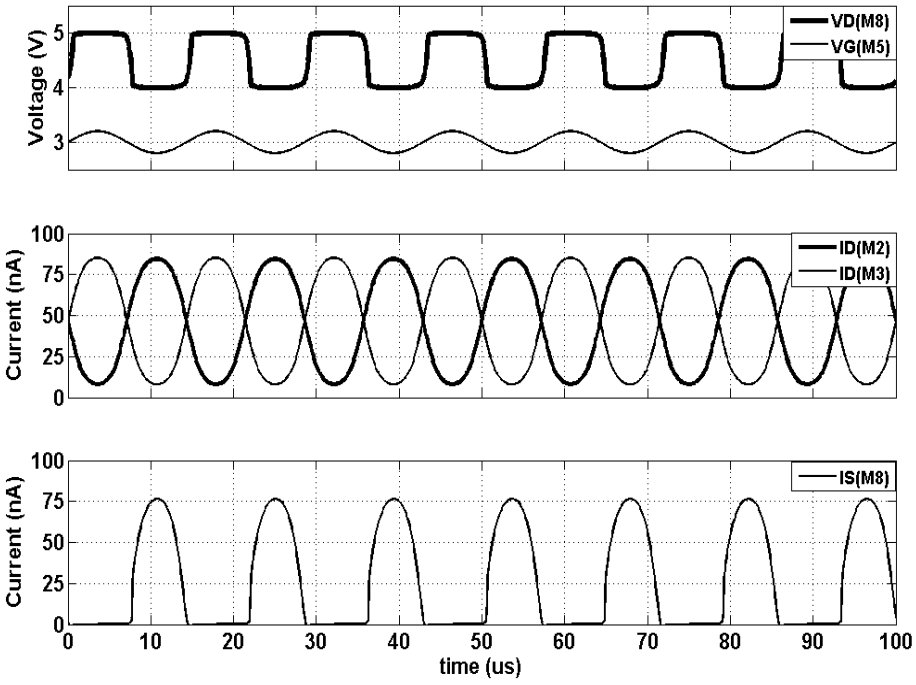

Figure 4.10: Simulation of the half-wave rectifier circuit. The input is a $70kHz$ sinusoid with a $200mV$ amplitude and a $3V$ DC voltage offset. Current flows through M8 during the negative half cycle of the input.

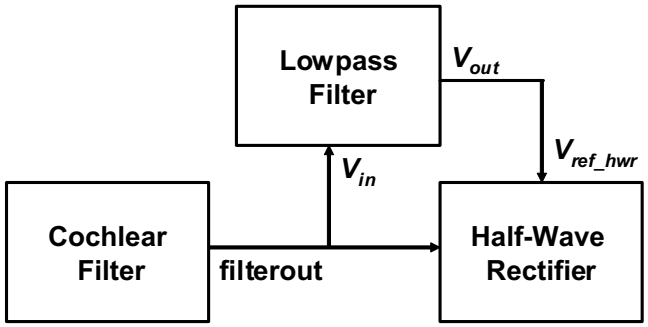

Figure 4.11: DC offset compensation. The cochlear voltage output, *filterout,* was low-pass filtered to extract its DC level and use that voltage as the reference voltage in the half-wave rectifier.

4.5 Neuron

The neuron whose schematic is shown in Fig. 4.12 was designed to integrate charge and produce a pulse event (spike) when the membrane voltage (V_{mem}) exceeds the comparator threshold (V_{cmp}). The neuron receives excitatory current (I_{exc}) from the half-wave rectifier connected to a specific cochlear filter. The neuron has a leak current I_{Leak}. All neurons in the same column have the same threshold.

In the cochlea chips LeqC4 and LeqC5, all neurons have the same leak current (I_{Leak}) and the voltage V_{cmp} was varied to model different neuron thresholds. The drawback of this design is that V_{cmp} does not serve as a true threshold and as long as I_{exc} is greater than I_{Leak}, the neuron will eventually fire a spike.

In the cochlea chips LeqC6 and LeqC7, the design was modified so that all neurons have the same comparator voltage V_{cmp} and different leakage currents are used to model individual neurons' thresholds.

The circuit implementation of the neuron is shown in Fig. 4.13. The voltage V_{mirror} is the output of the half-wave rectifier shown in Fig. 4.9. This voltage generates the excitatory current I_{exc} via M18. Transistors M20 through M28 constitute the voltage comparator; it compares the membrane voltage V_{mem} (C_{mem} is the membrane capacitance) against the comparator threshold V_{cmp}. The threshold of the neuron is determined by the voltage V_{th} which generates the leakage current I_{Leak} through M16. V_{refr} controls the duration of the refractory period of the neuron. Transistors M8 through M11 constitute the positive feedback circuit that provides hysteresis to ensure that a full spike is generated. Transistors M1 through M7 and M12 through M15 are for interfacing with the address-event system and for resetting and controlling the refractory period. To avoid having voltage glitches on the power supply rails from coupling back into the analog circuitry (due to high speed switching currents), a separate power supply rail ($DVdd$) was used for the neurons and the address-event arbiter.

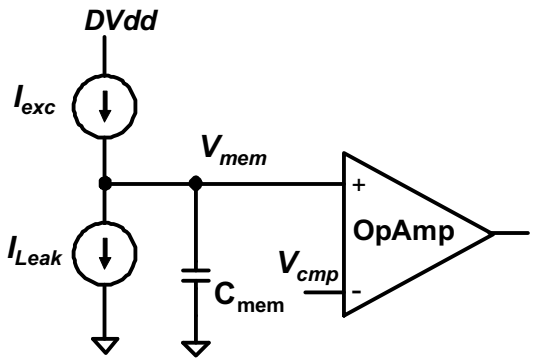

Figure 4.12: Schematic of the neuron circuit.

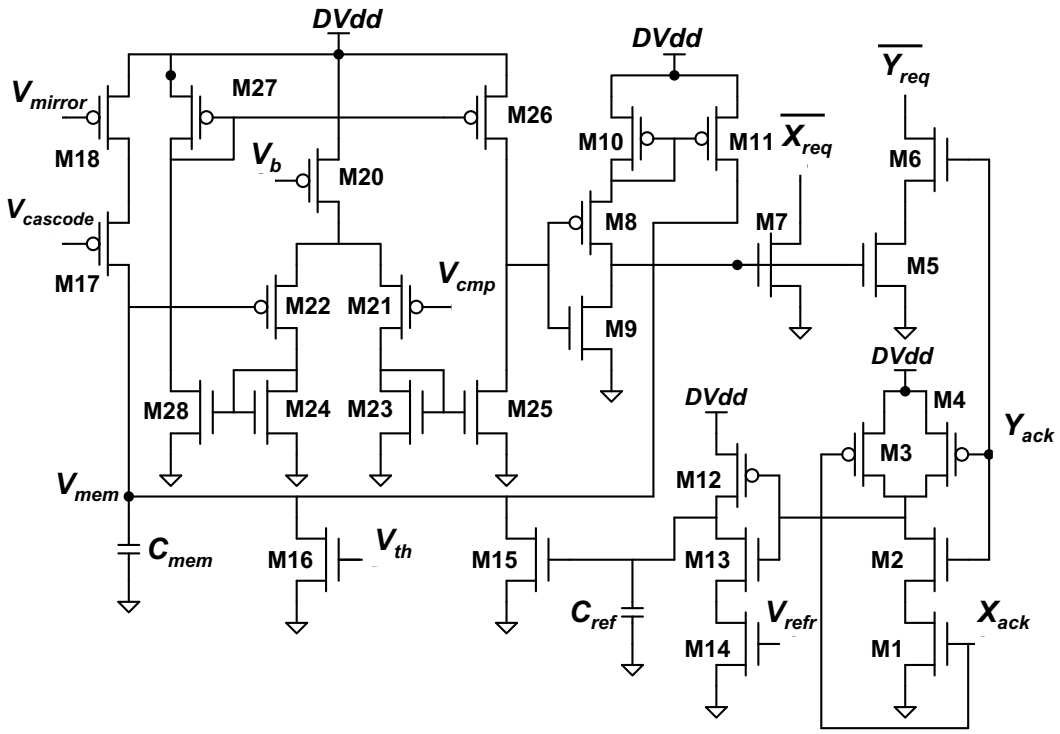

Figure 4.13: Implementation of the neuron circuit. The half-wave rectified current is mirrored into the neuron via the voltage signal V_{mirror}. The current is integrated on the membrane capacitance C_{mem}. A spike is generated when the membrane voltage V_{mem} exceeds the comparator voltage V_{cmp}.

4.6 Layout

The photomicrograph of the cochlea chip LeqC6 is shown in Fig. 4.14. Two dummy filters were added on each end of the cochlear filterbank layout to reduce edge mismatch effects. Isolation blocks (substrate contacts connected to $0V$ and well contacts connected to Vdd) were placed between the analog and digital circuitry to prevent switching noise from coupling back into the analog circuitry. The padframe was also split into two parts. The upper part provides $AVdd$ to the analog circuitry, and the lower part provides $DVdd$ for the digital circuitry and the address-event system. The followers that were used to monitor the membrane potentials for all 128 neurons were removed from the cochlea chip LeqC7 and replaced by isolation blocks to increase isolation between neurons.

4.7 Test Results

All chips were fabricated using a commercially-available $0.5\mu m$ 2-poly, 3-metal process. All testing was conducted at a supply voltage of $5V$. For all cochlea chips, the voltage V_{ref} (at a value of $3V$) sets the DC offset of the input signal.

4.7.1 Cochlear Tuning Curves

Fig. 4.15 shows the tuning curves for the cochlear filters in the cochlea chip LeqC2 (chip #2). The filter bias voltages were set at: V_{L1}=1.293, V_{L2}=1.153, V_{fbk}=0.559, and V_{offset}=0.472. The ten filters have a mean quality factor (Q) of 5.5 with a standard deviation of 0.77 and the mean of their peak magnitude is 2.4dB with a

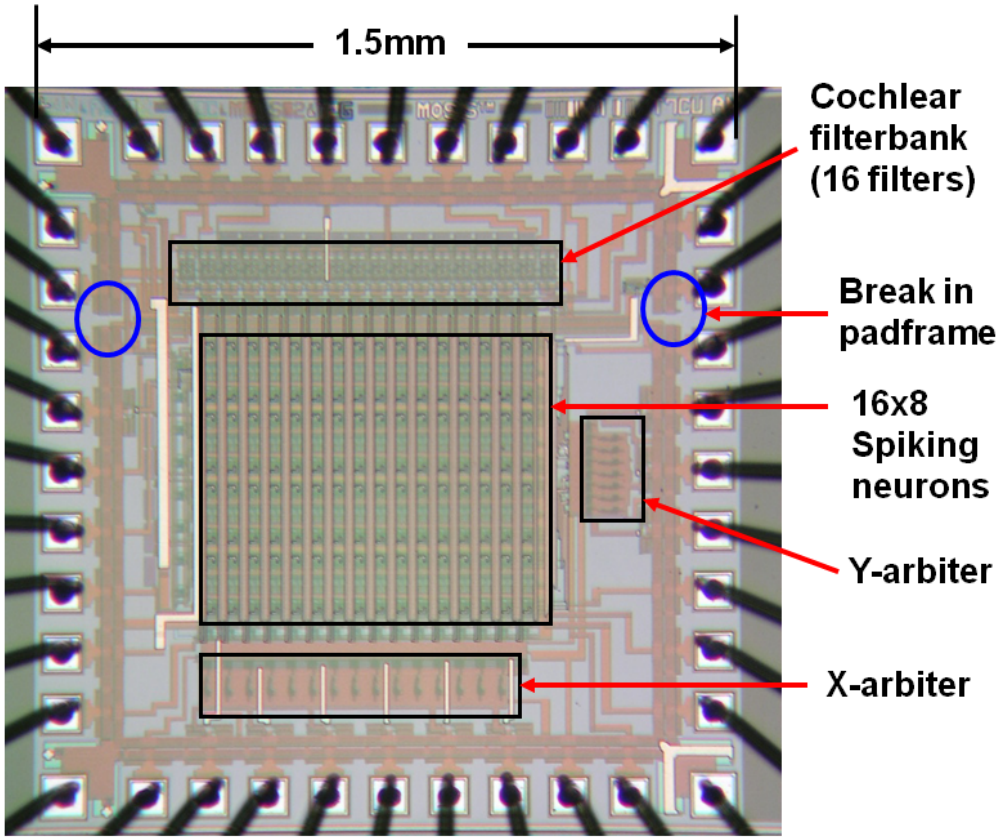

Figure 4.14: Photomicrograph of the cochlea chip LeqC6. The chip was fabricated using the AMI 0.5μm, 3-metal, 2-poly CMOS process. The chip is 1.5mm by 1.5mm. Notice that the padframe was split into two parts (upper and lower) to separate the analog biasing from the digital biasing. Isolation blocks were also placed between the analog and digital circuitry to prevent switching noise from coupling back into the analog circuitry.

standard deviation of 1.4.

The tuning curves shown in Fig. 4.16 are also from the cochlea chip LeqC2 but from a different chip (chip #3) fabricated in the same chip run. The filter bias voltages were set at: V_{L1}=1.261, V_{L2}=1.118, V_{fbk}=0.572, and V_{offset}=0.464. The filters have a mean quality factor (Q) of 7.4 with a standard deviation of 1.18 and the mean of their peak magnitude is 2.0dB with a standard deviation of 0.94.

Although equation 4.12 predicts that that quality factor of the filter with the highest center frequency (CF) should be 1/4 that of the filter with the lowest CF, this has not been observed. Fig. 4.15 shows that the peak gain increases across the filter array, while in Fig. 4.16, no such trend is seen. The effect of mismatch can clearly be seen in Fig. 4.16, where two of the filters essentially have the same center frequency ($30.7kHz$).

Fig. 4.17 shows the quality factor Q (on the log scale) as a function of V_{offset} for the cochlea chip LeqC2. The slope of this line, predicted by equation 4.8, is $\frac{-1}{2V_T}$, which approximately equals -19.4; the measured slopes for chips 2, 3, and 4 were measured to be -27.5,-27.1, and -26.8 respectively. It is not clear why the measurement deviates from the predicted value. Also notice that the slope increases as V_{offset} is decreased below $460mV$.

Fig. 4.18 shows the tuning curves for the 16 (modified) cochlear filters in the cochlea chip LeqC7 (chip #3). The 16 filters have a mean quality factor (Q) of 6.0 with a standard deviation of 1.32. The mean of the peak gain is 5.1dB with a standard deviation of 1.4.

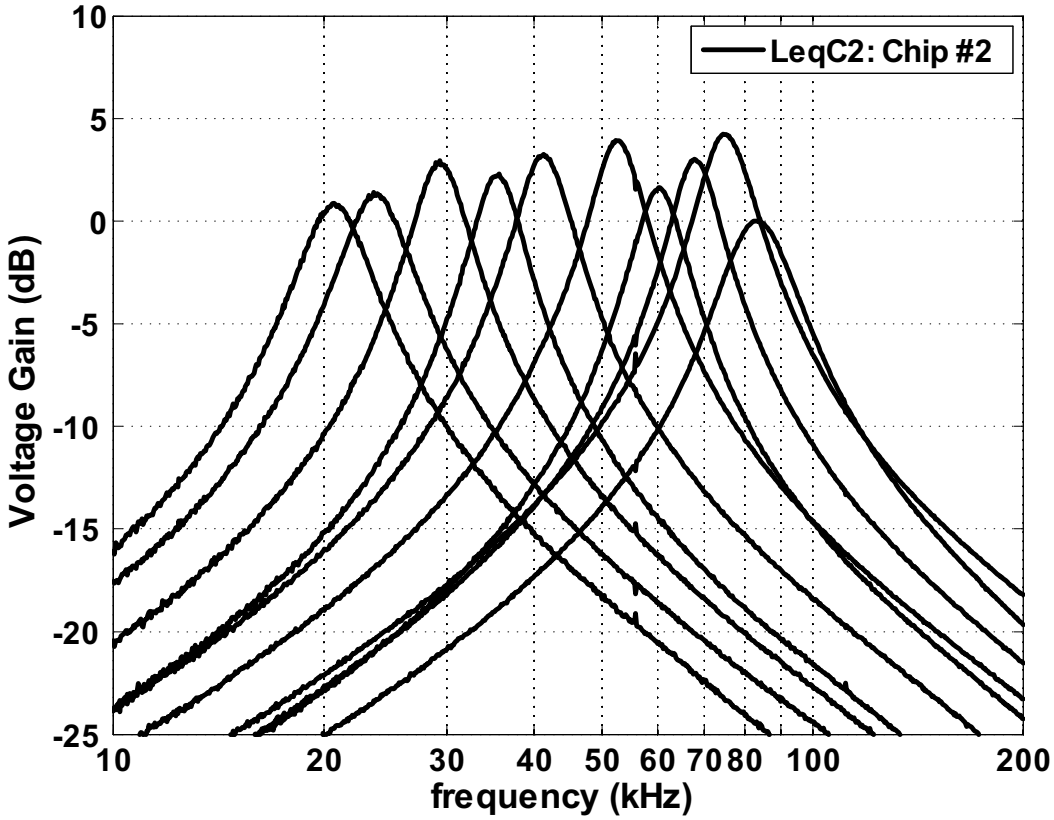

Figure 4.15: Frequency tuning curves for the cochlea chip LeqC2 (chip #2). The ten filters have a mean quality factor (Q) of 5.5 with a standard deviation of 0.77. The mean of the peak gain is 2.4dB with a standard deviation of 1.4.

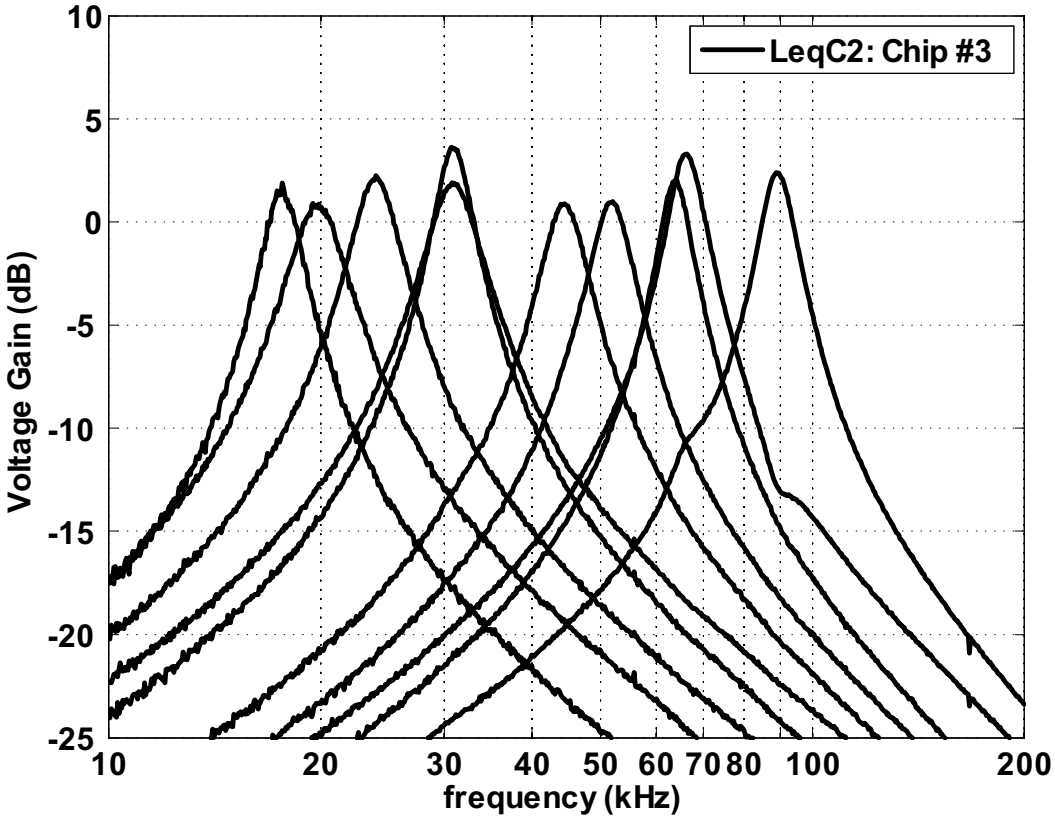

Figure 4.16: Frequency tuning curves for the cochlea chip LeqC2 (chip #3). The filters have a mean quality factor (Q) of 7.4 with a standard deviation of 1.18. The mean of the peak gain is 2.0dB with a standard deviation of 0.94.

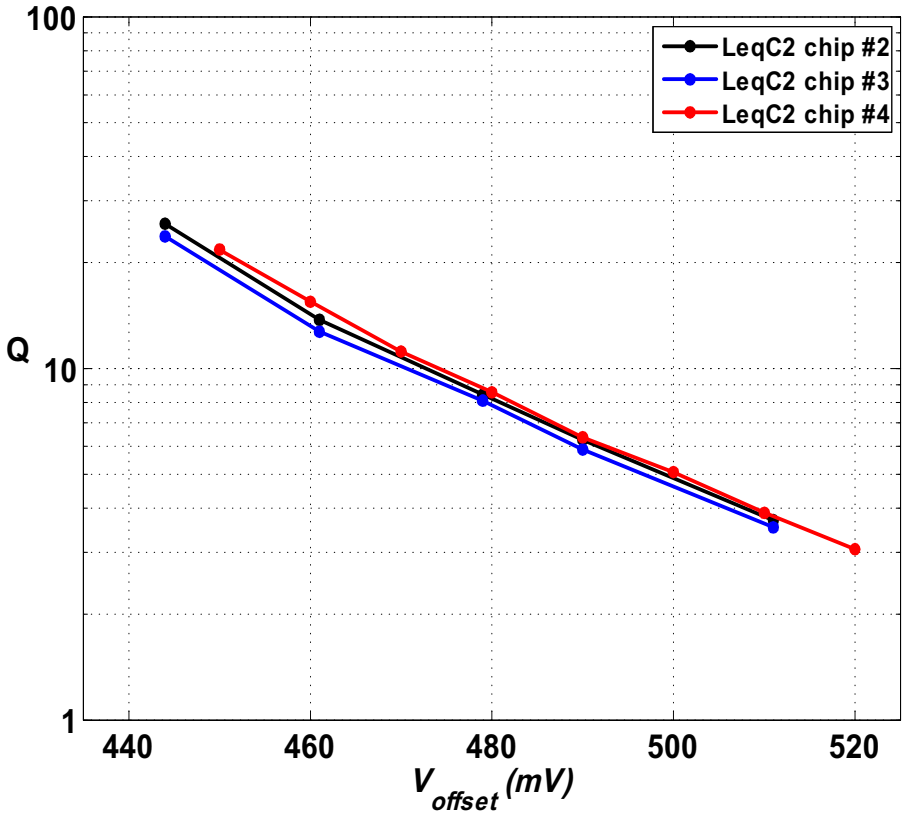

Figure 4.17: Q as a function of V_{offset} in the cochlea chip LeqC2.

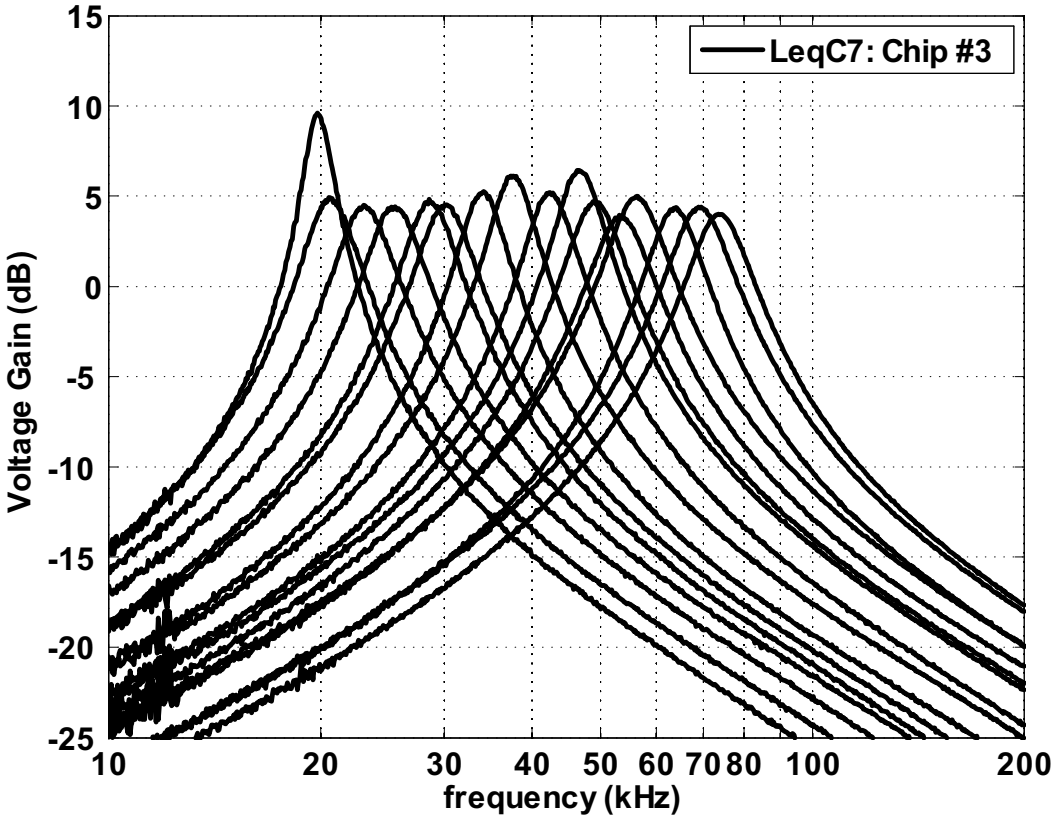

Figure 4.18: Frequency tuning curves for the cochlea chip LeqC7 (chip #3). The 16 filters have a mean quality factor (Q) of 6.0 with a standard deviation of 1.32. The mean of the peak gain is 5.1dB with a standard deviation of 1.4.

4.7.2 Mismatch

Q-mismatch

It has been observed that as the quality factor Q increases, the mismatch in Q across the filter array also increases. Fig. 4.19 shows measurements from 5 cochlea chips (LeqC4) fabricated within the same chip run. The quality factor Q is plotted as a function of delta (delta=V_{fbk}-V_L) which was varied from $0mV$ to $240mV$ in steps of $20mV$. The black dots represent the measurements from 80 individual filters (5 chips x 16 filters/chip) while the solid lines represent the average across a single chip. The inset shows the same data but focuses on the first $80mV$. Notice that as delta increases so does the spread between the individual filters' Q measurements and across the chip averages.

Note that equation 4.15 predicts that

$$V_{fbk}=V_L \Longrightarrow Q = 1$$

however, the measurements show that when delta=0 (i.e. V_{fbk}=V_L) the cochlear filters have a mean Q of 1.371 with a standard deviation (σ) of 0.073 (see inset). There are two potential sources of mismatch that may have contributed to the observed mismatches: mismatch in the resistor ladders that provide biasing for the filters and transistor mismatch. Recall that the bias voltages V_{fbk} and V_L are set using two uniform resistor ladders whose end voltages are V_{f1}, V_{f2}, V_{L1}, and V_{L2} (see Fig. 4.8). By setting V_{f1}=V_{f2} and V_{L1}=V_{L2} all 16 filters are forced to have the same bias voltages and any possible mismatch from the resistor ladders would have been circumvented. Also recall that the center frequency of the filter

is a function of (V_{fbk}+V_L) (equation 4.14), so forcing the filters to have the same V_{fbk} and V_L results in the filters having the same center frequency (except for mismatch).

Fig. 4.20 shows the quality factor (Q) of the cochlear filters as a function of the voltage difference delta=(V_{fbk}-V_L) measured from the cochlea chip (LeqC7). The black circles represent the mean Q for all 16 filters measurements when all filters were forced to have the same bias voltages. The four circles at a fixed value of delta represent measurements at different center frequencies (i.e. changing V_{fbk} and/or V_L but keeping it fixed across the array), while the solid line represents the mean Q when the individual filters were tuned to different frequencies (normal operation). It is observed that even when the (potential) resistor ladder mismatch was eliminated, the quality factor deviated from unity at delta=0. Therefore, it appears that the mismatch in Q is a result of transistor mismatch.

As a side note, recall that the filter implementation in LeqC4 and LeqC7 is identical except for the removal of the source degeneration transistor pairs for the bias transistors and their replacement with single transistors in LeqC7. The effect of this change can been seen by comparing Figs. 4.19 and 4.20 and noting that, in LeqC7, a smaller delta is required to obtain the same Q.

To explain why Q-mismatch increases as V_{fbk}-V_L increases, we first introduce the lognormal distribution.

The lognormal distribution

If X is a normal (Gaussian) random variable with mean μ_X and standard deviation σ_X, then the transformation $Y = e^X$ results in a lognormal random variable Y with a mean μ_Y given by:

$$\mu_Y = e^{(\mu_X + \frac{\sigma_X^2}{2})} \tag{4.16}$$

and a standard deviation σ_Y given by:

$$\begin{aligned} \sigma_Y &= e^{(\mu_X + \frac{\sigma_X^2}{2})} \sqrt{(e^{\sigma_X^2} - 1)} \\ &= \mu_Y \sqrt{(e^{\sigma_X^2} - 1)} \end{aligned} \tag{4.17}$$

It is important to note that μ_Y is not only a function of μ_X but also a function of σ_X. If

$$\sigma_X^2 << 1 \tag{4.18}$$

then

$$e^{\sigma_X^2} \approx 1 + \sigma_X^2 \tag{4.19}$$

and the standard deviation of Y can be written as:

$$\sigma_Y = \mu_Y \sigma_X \tag{4.20}$$

which is the same as that of linearizing the transformation $Y = e^X$ about the point $X = \mu_X$.

The subthreshold current for a NMOS transistor is given by:

$$I = I_o e^{\frac{\kappa V_{gb}}{V_T}} (e^{\frac{V_{sb}}{V_T}} - e^{\frac{V_{db}}{V_T}}) \tag{4.21}$$

Since the bulk terminal is usually grounded ($V_b = 0$) for NMOS transistors, and assuming that the transistor is operated in saturation ($V_{ds} > 100mV$), the current can be written as:

$$I = I_o e^{\frac{\kappa V_g - V_s}{V_T}} \tag{4.22}$$

It will be assumed that all sources of mismatch can be lumped into a single parameter X such that

$$I = I_o e^{\frac{\kappa V_g - V_s + X}{V_T}} \tag{4.23}$$

For the cochlea chip LeqC7, Q can be written as:

$$\begin{aligned} Q &= \frac{\sqrt{G_{m2}G_{m3}}}{G_{m4}} \\ &= \frac{\sqrt{e^{\frac{\kappa V_L + X2}{V_T}} e^{\frac{\kappa V_f + X3}{V_T}}}}{e^{\frac{\kappa V_L + X4}{V_T}}} \\ &= \exp(\beta_1(\frac{V_{fbk} - V_L}{2}) + X) \end{aligned} \tag{4.24}$$

where $\beta_1 = \frac{\kappa}{V_T}$, and

$$X = \frac{1}{V_T}(\frac{1}{2}(X2 + X3) - X4) \tag{4.25}$$

where $X2$, $X3$, and $X4$ are the noise parameters for the bias currents of the transconductance amplifiers G_{m2}, G_{m3}, and G_{m4} respectively. Therefore, even if the voltage difference V_{fbk}-V_L is equal to zero, the mean of Q will not be equal to unity. If $X2$, $X3$, and $X4$ are assumed to be independent with zero mean and variance σ^2, then

$$\begin{aligned} \mu_X &= 0 \\ \sigma_X^2 &= \frac{2\sigma^2}{(V_T)^2} \end{aligned} \tag{4.26}$$

The data shown in Fig. 4.21 is identical to that of Fig. 4.19 but shown on a logarithmic scale. The dashed red line p1 is the quality factor predicted by equation 4.15 with $\kappa = 0.7$. The dashed black line (p2) is the same as p1, but an offset of 0.315, corresponding to $\frac{\sigma_X^2}{2}$ in equation 4.16, was added to take into account the deviation observed at delta=0. The dashed blue line (p3) is similar to p2 but for $\kappa = 0.8$. The value of σ corresponding to this offset is $14.5mV$. It is clear that adding the offset at delta=0 has improved the estimation of Q. Fig. 4.22 shows the standard deviation of Q (σ_Q) as a function of Q as measured from eight cochlea chips (five chips are from LeqC4 and three are from LeqC2). The line p1 shows σ_Q as a function of Q as predicted by equation 4.17 with $\sigma_X^2 = 0.63$. The line p2 is similar to p1 but with σ_X^2 an order of magnitude smaller. It is not clear why equation 4.17 fails to accurately predict the relation between σ_Q and Q. It may be that the assumption of lumping all mismatch into a single parameter X is not an accurate one.

DC offset mismatch

As discussed in section 4.4, it was observed in the cochlea chip LeqC3 that some of the neurons fired spikes spontaneously (i.e. with no external input). This background firing rate was attributed to the mismatch in the DC level of the cochlear filter output *filterout*. Table 4.1 shows the standard deviation (in mV) of the DC level of *filterout* as measured from several cochlea chips. The standard deviation is measured across all filters within the same chip. Table 4.2 shows the range (maximum-minimum) of the DC level of *filterout* within each chip. In the

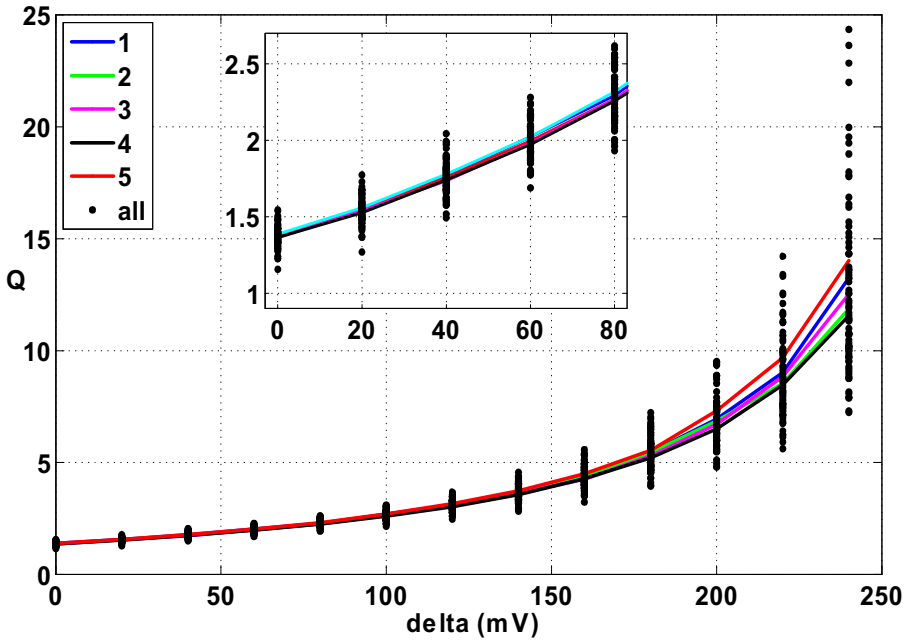

Figure 4.19: Quality factor (Q) of the cochlear filters as a function of the voltage difference delta=V_{fbk}-V_L as measured from five cochlea chips (LeqC4) fabricated in the same chip run. The black dots represent Q from individual filters while the solid color lines represent the mean value of Q for a single chip at a specific value of delta. The inset is the same data but focuses on the region delta<$80mV$. Notice that when delta=0, $Q \neq 1$.

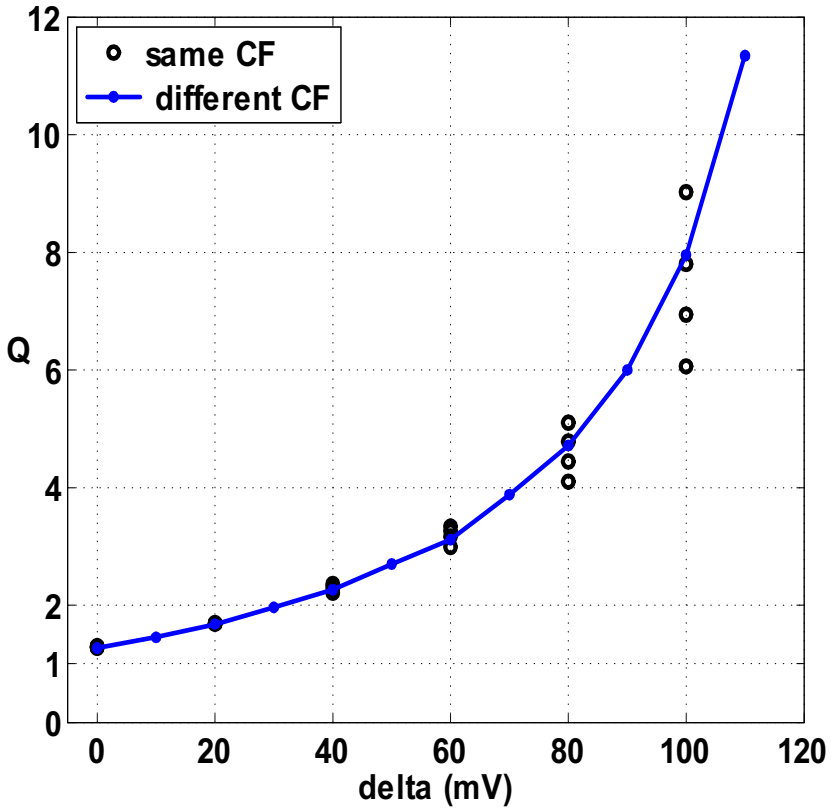

Figure 4.20: Quality factor (Q) as a function of delta=($V_{fbk} - V_L$) as measured from the cochlea chip LeqC7 (chip #3). CF is the center frequency of the cochlear filter.

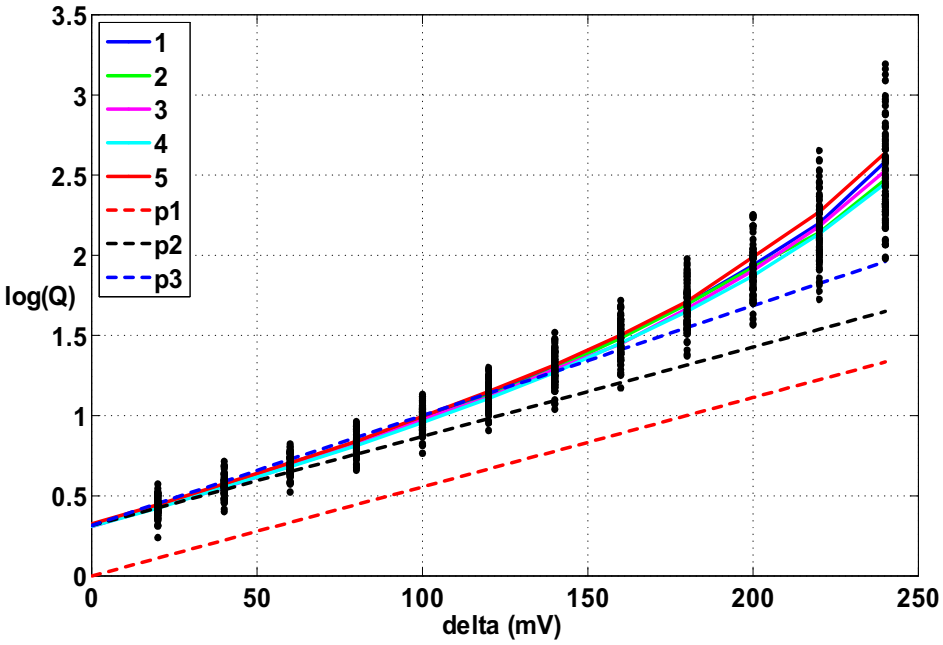

Figure 4.21: log(Q) versus delta. This plot shows the same data as in Fig. 4.19 but on a log scale. The dashed red line (p1) shows log(Q) as predicted by equation 4.15 with κ=0.7. The dashed black line (p2) is the same as p1 but with an offset of 0.315. The dashed blue line (p3) is the similar to p2 but for κ=0.8.

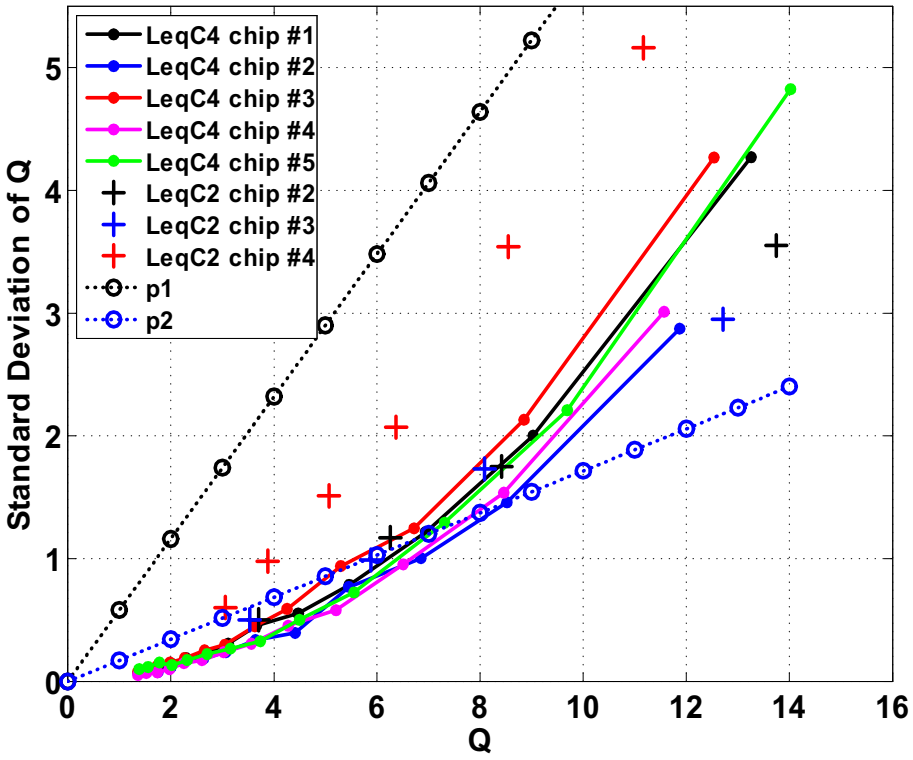

Figure 4.22: Standard deviation of the quality factor Q as a function of Q as measured from eight cochlea chips (five chips are from LeqC4 and three are from LeqC2). The line p1 shows σ_Q as a function of Q as predicted by equation 4.17 with $\sigma_X^2 = 0.63$. The line p2 is similar to p1 but with σ_X^2 an order of magnitude smaller.

Table 4.1: Standard Deviation (in mV) of the DC Level of the filter's output.

version\chip #	1	2	3	4	5
LeqC2	10	13	15	16	12
LeqC3	16	13	17	17	12
LeqC4	7	9	10	9	10
LeqC5	x	9	x	x	x
LeqC6	11	10	x	x	x
LeqC7	14	13	x	x	x

Table 4.2: Range (maximum-minimum) of the DC level of the filter's output (in mV).

version\chip #	1	2	3	4	5
LeqC2	33	45	41	49	28
LeqC3	49	46	55	49	37
LeqC4	23	32	35	32	41
LeqC5	x	29	x	x	x
LeqC6	45	37	x	x	x
LeqC7	60	52	x	x	x

cochlea chips LeqC5, LeqC6, and LeqC6, the reference voltage V_{ref_hwr} is typically set $30mV$ below V_{ref} to prevent neurons from spontaneously firing spikes. The boxes with the symbol x indicate chips where the DC level has not been measured.

4.7.3 Spike Rasters

Fig. 4.23 shows spike rasters for neurons in the cochlea chip LeqC7 in response to a $5ms$ hyperbolic FM sweep ($120kHz$ to $18kHz$). The filter index 16 corresponds to the filter with highest center frequency. There are eight neurons per filter. The neurons have different thresholds; **th1** corresponds to the smallest threshold while **th8** corresponds to the highest threshold. Notice that neurons with higher thresholds have longer latencies.

Fig. 4.24 shows the spike latency versus peak gain (in dB) measured at the center frequency ($\simeq 20kHz$) of the cochlear filter for four cochlear neurons in response to the recorded sound from the artificial bat head. The four neurons are connected to the same cochlear filter but with different thresholds. In this plot, th1 corresponds to the neuron with the lowest threshold while th4 corresponds to the neuron with the highest threshold. The plot shows that as the sound intensity increases the spike latency decreases so that the intensity of the sound is encoded in the latency of the spikes. The plot also shows that for a given intensity, as the neuron's threshold increases the latency of its spikes increases. It is also observed that as the neuron's threshold increases the minimum intensity at which the neuron can fire a spike also increases.

4.7.4 Power Consumption

Fig. 4.25 shows the power supply current of the cochlear filter bank as a function of the bias voltage V_L (or V_{fbk}) as measured from the cochlea chip LeqC5. In

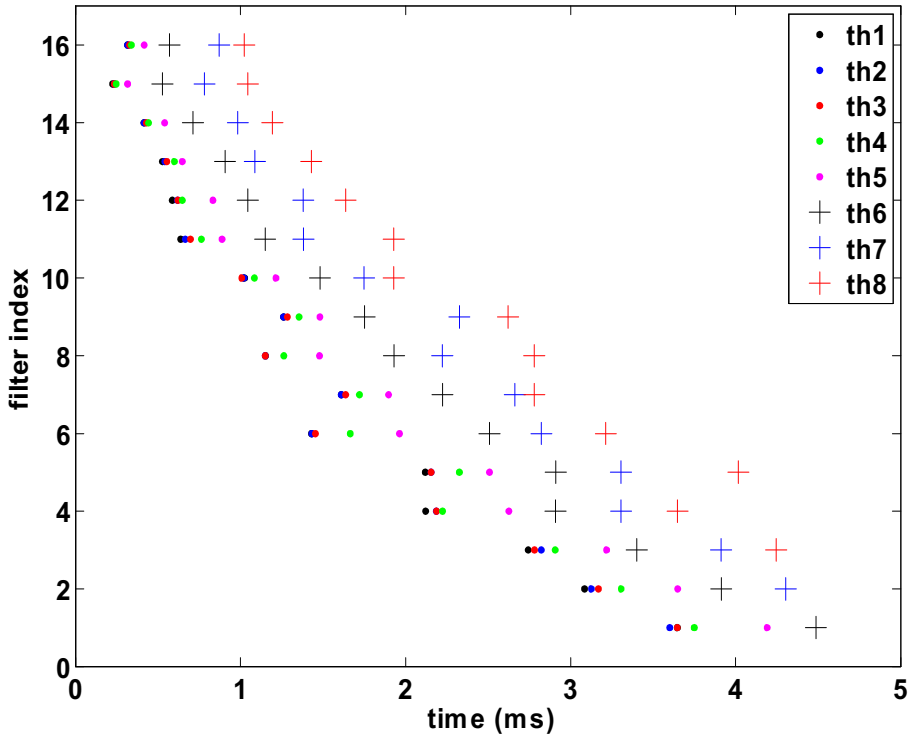

Figure 4.23: Spike raster plot for neurons in the cochlea chip LeqC7 in response to a $5ms$ hyperbolic FM sweep ($120kHz$ to $18kHz$). Filter index 16 corresponds to the cochlear filter with the highest center frequency. There are eight neurons per filter. The neurons have different thresholds; **th1** corresponds to the smallest threshold while **th8** corresponds to the highest threshold. Notice that neurons with higher thresholds have longer latencies.

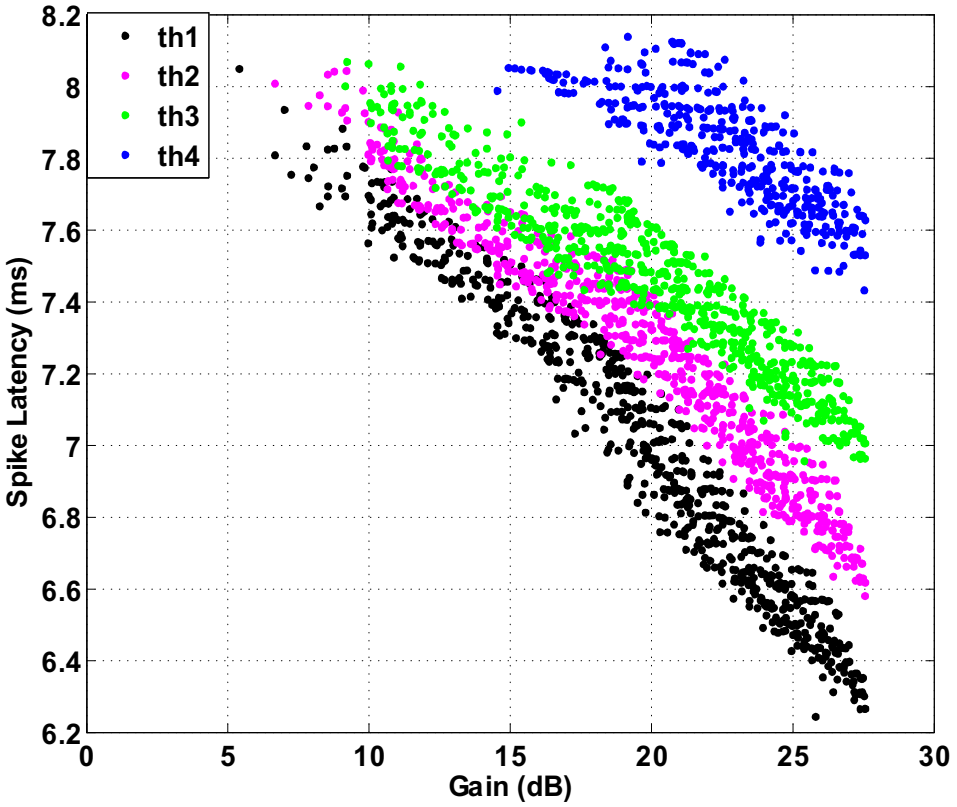

Figure 4.24: Spike latency versus peak gain (measured at the center frequency of the cochlear filter) for four cochlear neurons in response to the recorded sounds from the artificial bat head. The four neurons are connected to the same cochlear filter but with different thresholds. In this plot, th1 corresponds to the neuron with the lowest threshold while th4 corresponds to the neuron with the highest threshold.

this plot all 16 filters were tuned to the same center frequency (i.e. V_{f1}=V_{f2} and V_{L1}=V_{L2}). The same data is plotted in Fig. 4.26 as a function of frequency. At a frequency of $100kHz$, the cochlear filter bank consumes $2\mu A$ ($10\mu W$), this is an upper bound on the power supply current in the normal operation (where filters are tuned to different frequencies). It can be shown (see Appendix D) that the power consumption P_T of the cochlear filterbank (LeqC5) is given by

$$P_T = V_{dd} I_o (3e^{\beta V_{L1}} + e^{\beta V_{f1}}) \frac{e^{\beta \frac{16}{15} \triangle} - 1}{e^{\beta \frac{1}{15} \triangle} - 1} \tag{4.27}$$

where $\triangle$=V_{f2}-V_{f1} (or V_{L2}-V_{L1}), $\beta = \frac{\kappa^2}{(1+\kappa)2V_T}$, and I_o is the zero bias current for the transistor. Fig. 4.27 shows the power supply current of the 128 cochlear neurons as a function of the comparator bias voltage V_b as measured from the cochlea chip LeqC7.

4.8 Conclusion

We have designed and fabricated an ultrasonic cochlea with 16 frequency channels and eight spiking neurons per channel that can span the frequency range from $20kHz$ to $100kHz$ with moderate quality factor Q. The eight cochlear neurons in each frequency channel can be set to have different thresholds.

In this chapter, we have presented tuning curves for the cochlear filters and investigated the mismatch in the quality factor Q. We have illustrated the spike rasters in response to a hyperbolic FM sweep. We showed the partial response of the cochlea chip to the recorded sounds from the artificial bat head and illustrated that the sound intensity is encoded in the spike latency of the cochlear neurons.

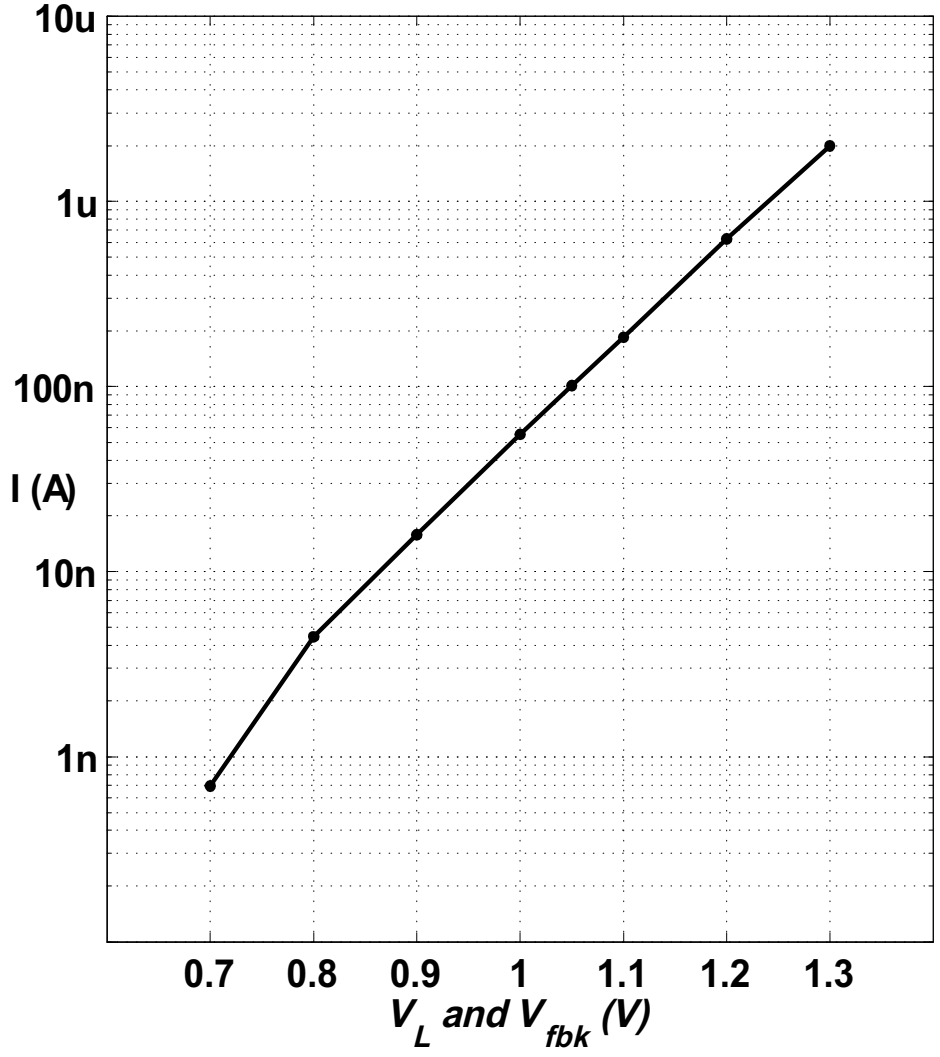

Figure 4.25: Power supply current of the cochlear filter bank as a function of the bias voltages V_L (or V_{fbk}) as measured from the cochlea chip LeqC5. In this plot all 16 filters were tuned to the same center frequency (i.e. V_{f1}=V_{f2} and V_{L1}=V_{L2}). The supply voltage (Vdd) was set at $5V$.

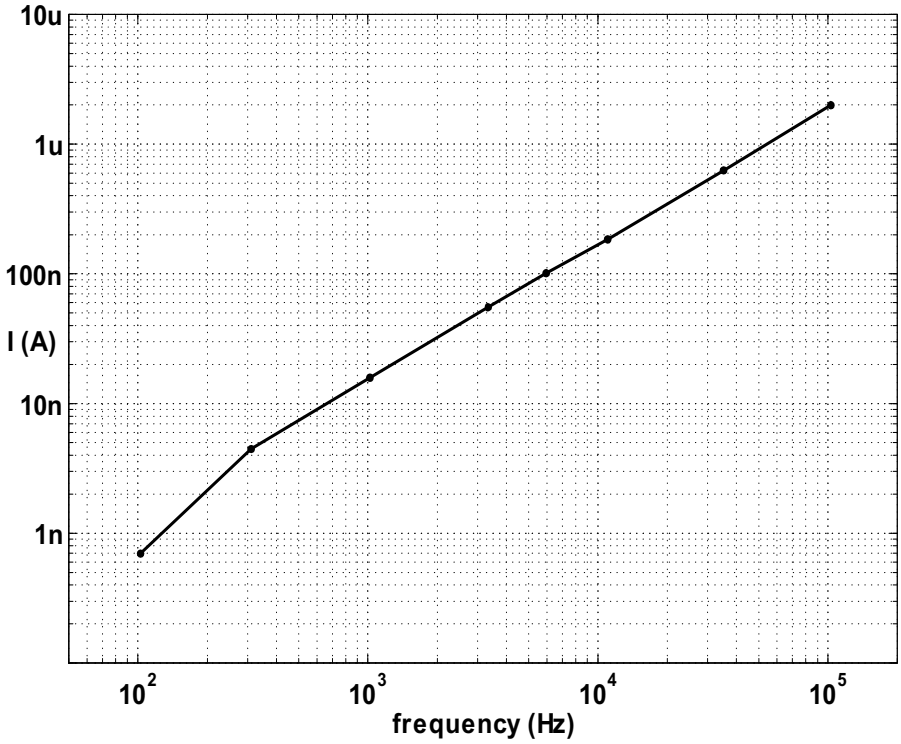

Figure 4.26: Power supply current of the cochlear filter bank as a function of frequency as measured from the cochlea chip LeqC5. In this plot all 16 filters were tuned to the same center frequency (i.e. V_{f1}=V_{f2} and V_{L1}=V_{L2}). Supply voltage Vdd=$5V$.

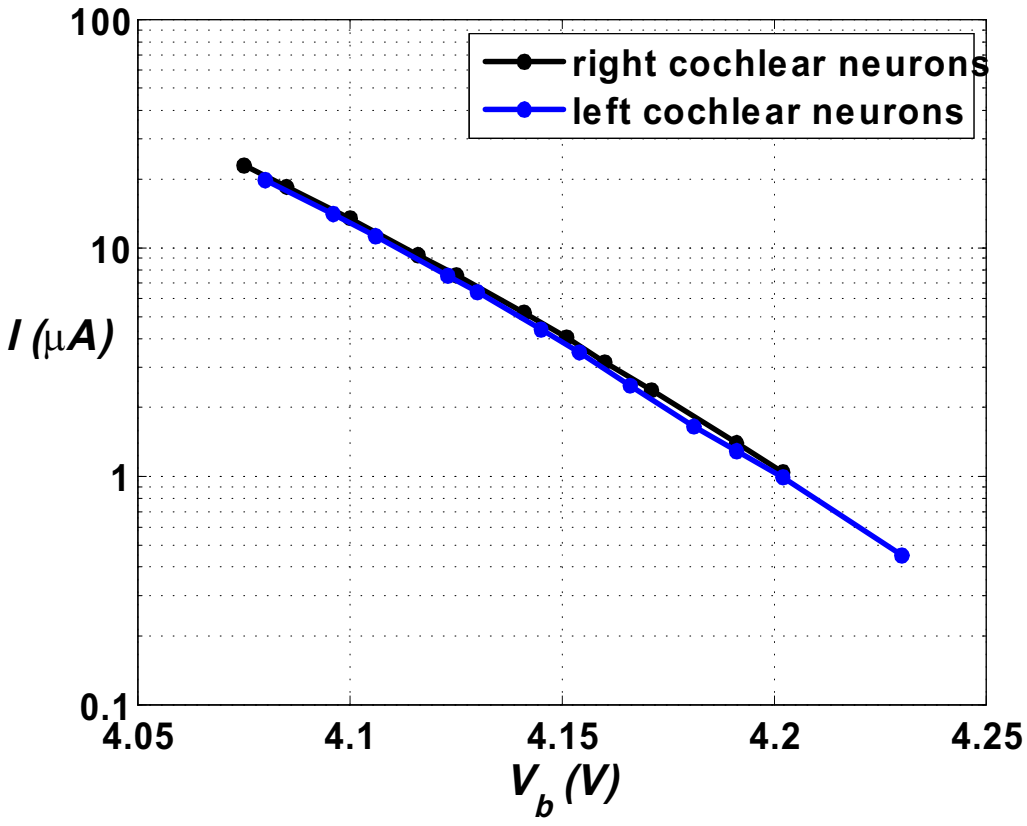

Figure 4.27: Measured power supply current of the 128 cochlear neurons as a function of the comparator bias voltage V_b as measured from the cochlea chip LeqC7.

The detailed response of the cochlea chip to the artificial bat head measurements will be deferred to Chapter 6.

Chapter 5

Feature Extraction Chips

5.1 Introduction

In Chapter 3, we presented an artificial bat head that can generate the necessary acoustic localization cues. In Chapter 4, we presented an ultrasonic cochlea with 16 frequency channels and eight spiking neurons per frequency channel that converts an analog signal into a spike-based representation. We also demonstrated that the intensity of the measured sounds from the artificial bat head is encoded in the latency of the cochlear spikes.

In this chapter we present two feature extraction chips: a monaural spectral difference (SD) chip and a binaural interaural level difference (ILD) chip, that can extract the localization cues from the cochlea's spiking output. Both feature extraction chips consist of a group of neurons. Each neuron receives excitation from a single cochlear filter and inhibition from another cochlear filter. In the binaural ILD chip, the excitation and inhibition are from different cochlea chips

(i.e. right and left cochlea chips) whereas in the spectral difference chip, both excitation and inhibition are from the same cochlea chip. Both feature extraction chips operate under the assumption that the intensity of the incoming sound is reflected in the latency of the cochlear spikes.

We also present a merge board which uses a microcontroller to combine (merge) the address-event spikes from two (right and left) cochlea chips into a single address bus and sends the spikes to the feature extraction chips. We begin, first, by describing the monaural spectral difference chip.

5.2 Spectral Difference Chip

The first of two feature extraction chips, the spectral difference (SD) chip is a monaural chip, it compares the activity of pairs of cochlear filters within the same cochlea. The chip compares the activity of neurons associated with one cochlear filter with the activity of neurons associated with another cochlear filter. In Fig. 5.1, the black dots represent the exponentially-increasing center frequencies of the 16 cochlear filters (from the cochlea chip). For any two cochlear filters i and j ($i,\ j$=$1, 2, ...16$) there are two possible neural configurations: neuron **H** receives excitation from the higher CF filter and inhibition from the lower CF filter, the second neuron **L** receives excitation from the lower CF filter and inhibition from the higher CF filter. Since there are 16 cochlear filters, there are 240 ($^{16}\mathrm{P}_2$) such neurons (permutations).

The block diagram of the spectral difference chip is shown in Fig. 5.2.

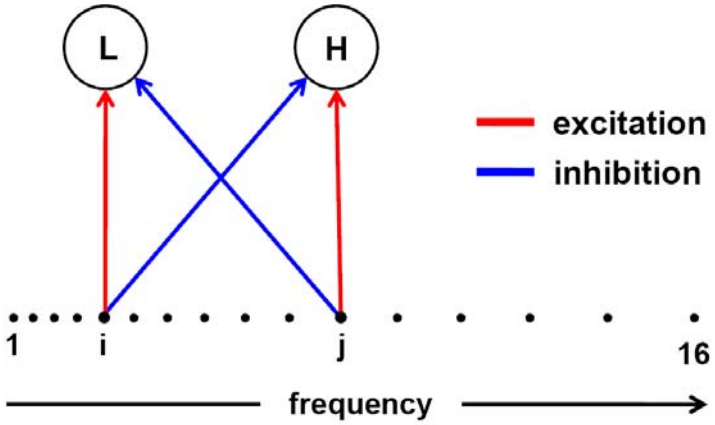

Figure 5.1: Spectral difference neurons. The black dots represent the exponentially-increasing center frequencies of the 16 cochlear filters. For any two cochlear filters i and j there are two possible neural configurations: neuron **H** receives excitation from the higher CF filter and inhibition from the lower CF filter, the other neuron **L** receives inhibition from the higher CF filter and excitation from the lower CF filter. Since there are 16 cochlear filters, there are 240 such neurons.

The chip has 240 neurons arranged in a 16 by 16 grid with no neurons on the diagonal. Each neuron has a single excitatory synapse and a single inhibitory synapse which are turned on and off via timing circuits. Sixteen timing blocks lie on the diagonal. Each timing block is connected to only one of the 16 cochlear filters through the address-event system. Each timing block has two identical timing circuits that are activated simultaneously; one timing circuit controls all excitatory synapses lying on the same row; the other timing circuit controls all inhibitory synapses lying in the same column as illustrated in Fig. 5.2, therefore, each row (column) receives excitation (inhibition) from a single cochlear filter.

Row (column) #1 receives excitation (inhibition) from the lowest CF cochlear filter while row (column) #16 receives excitation (inhibition) from the highest CF cochlear filter. A two-dimensional address-event arbiter is used to readout the neural activity.

Fig. 5.3 shows the schematic of the neuron circuit as well as the excitatory and inhibitory synapses. The neuron circuit is identical to that used in the cochlea chip. Transistors M23 through M30 form the voltage comparator that compares the voltage V_{mem} of the membrane capacitor C1 against the voltage V_{th}. Transistors M21 and M22 form the excitatory synapse whose current is controlled by the weight voltage V_{we}. Transistors M17 through M20 form the inhibitory synapse whose current is controlled by the weight voltage $V_{wi.}$. The excitatory and inhibitory synapses are voltage controlled current sources which can be switched on and off via the control signals tr_e and tr_i respectively. Because of the limited number of pins, all excitatory synapses have the same weight and their corresponding timing circuits have the same bias settings. Similarly, all inhibitory synapses have the same weight and their corresponding timing circuits have the same bias settings. V_{refr} controls the refractory period of the neuron. V_{Leak} controls the leakage current of the neuron and is the neuron's actual threshold as described in Chapter 4.

Shown in Fig. 5.4 is the schematic of the timing circuit along with the excitatory and inhibitory synapses. The timing circuit formed by transistors M1 through M7 along with capacitors C1 and C2 is placed along the diagonal of the spectral difference chip and controls the dynamics of the synapses. Transistors

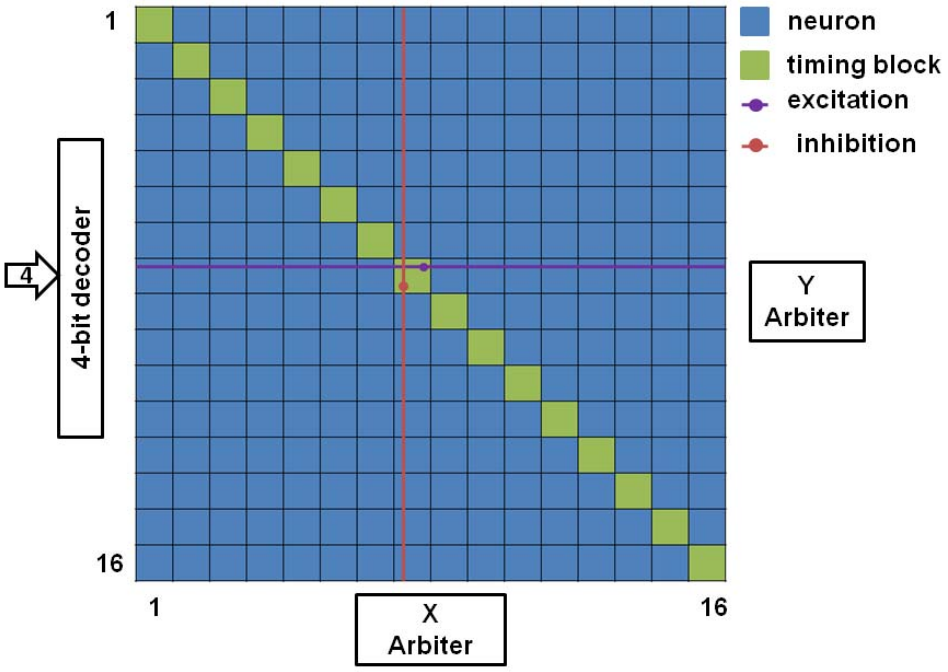

Figure 5.2: Block diagram of the spectral difference chip. The chip has 240 neurons (blue boxes) arranged in a 16 by 16 grid with no neurons on the diagonal. Each neuron has a single excitatory synapse and a single inhibitory synapse. The synapses are switched on and off via timing circuits. Sixteen timing blocks lie on the diagonal (green boxes) and are connected to the cochlea chip through the address-event system. Each timing block has two identical timing circuits; one timing circuit controls the timing of all excitatory synapses in the same row, and the other timing circuit controls the timing of all inhibitory synapses in the same column. A two-dimensional address-event arbiter is used to readout the activity of the neurons.

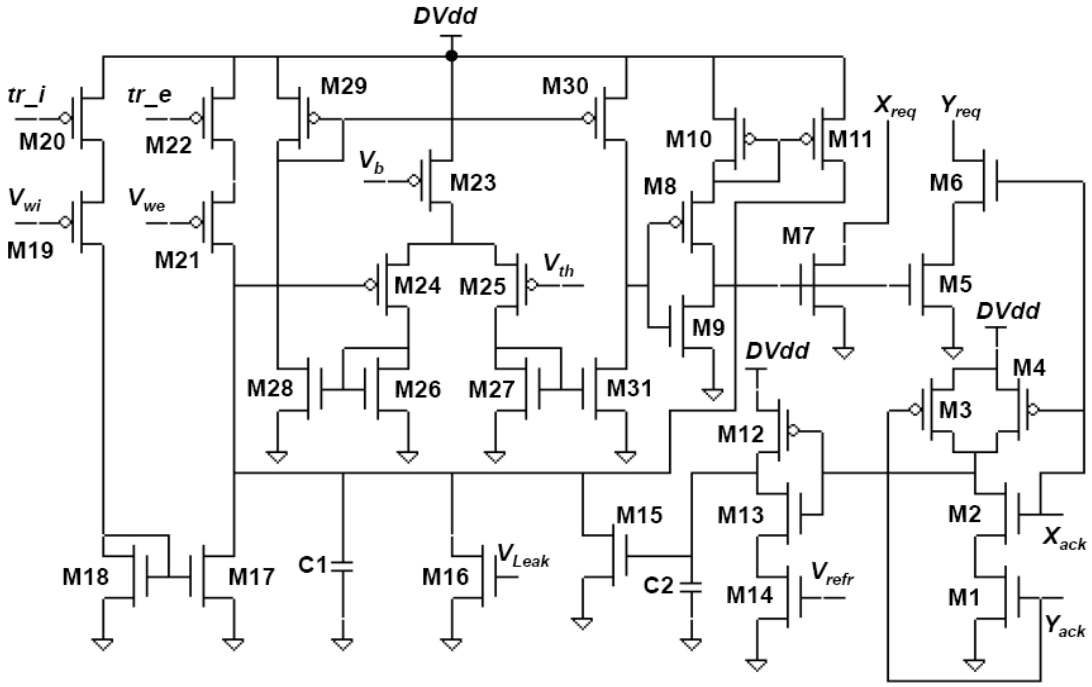

Figure 5.3: Schematic of the neuron circuit along with the excitatory and inhibitory synapses. The neuron circuit is identical to that used in the cochlea chip. Transistors M17 through M20 form the inhibitory synapse. The current of the inhibitory synapse is controlled by the weight voltage V_{wi} and is switched on and off via the control signal tr_i. Transistors M21 and M22 form the excitatory synapse whose current is controlled by the weight voltage V_{we}. The excitatory synapse is switched on and off via the control signal tr_e. Note that intersecting lines do not connect except where a black dot is shown.

M8 through M11 resemble the excitatory and inhibitory synapses shown in Fig. 5.3 and are not part of the timing circuit but have been included to illustrate how the timing circuit controls the synapses. The circuit consisting of transistors M8 and M9 is that of an excitatory synapse (M21 and M22 in Fig. 5.3) with V_w controlling the weight (current) of the synapse. Transistors M8 through M11 form an inhibitory synapse (M17 through M20 in Fig. 5.3) which is constructed out of an excitatory synapse (M8 and M9) by simply reversing the direction of current via a current mirror. When an active-low trigger pulse (*trg*) is applied to the input of the timing block circuit, it turns off M2 and turns on M3 allowing C1 to rapidly charge to *Vdd*. This in turn turns off M6 and turns on M5 allowing C2 to be discharged. V_{N1} controls how rapidly C2 is discharged. When V_{c2} is pulled-down below V_w, then the transistor pair M8 and M9 act as a constant current source whose current is controlled by the voltage V_w. When *trg* returns to its high state it turns off M3 and turns on M2 allowing C1 to be discharged through M1 and M2. V_{N1} controls the rate of discharge of C1.

The timing circuit has been operated in two modes: a single spike mode and a multiple spike mode. Fig. 5.5 shows the *simulated* response of the timing circuit to a $160ns$ wide active-low trigger pulse (*trg*) when applied to the inverter formed by transistors M2 and M3. The capacitor C1 is rapidly charged (middle plot) to *Vdd* and is discharged through M1 and M2 once *trg* returns to its high state. The rate of discharge is controlled by the voltage V_{N1}. In the single spike mode of operation the voltage V_{N2} is set such that the capacitor voltage V_{c2} can be discharged all the way to ground (bottom plot) during the interval in which

M5 is turned on. In this mode of operation, if a second spikes arrives before the capacitor C2 has charged up to *Vdd*, the voltage V_{c2} will be discharged to ground thus extending the time duration of the synaptic output by an amount equal to the interspike interval as shown in Fig. 5.6.

In the multiple spike mode, the voltage V_{N2} is set such that for a single input spike (*trg*) the capacitor voltage V_{c2} is not discharged all the way to ground as shown in Fig. 5.7. In this mode of operation, the effect of any additional spike is similar to that of the first spike (unless V_{C2} is close to $0V$) thus extending the time duration of the synaptic output by an amount proportional to the number of input spikes as shown in Fig. 5.8.

The layout of the spectral difference chip is shown in Fig. 5.9. The chip has a 4-line to 16-line input decoder that connects the spiking output of the cochlea chip to the timing blocks and a two-dimensional arbiter to readout the spike activity of the 240 neurons. The voltages V_{N1}, V_{N2}, and V_{P2} are the same for all inhibitory timing circuits. The membrane voltages for two neurons near the perimeter are connected through followers to chip pads for testing and monitoring. The voltage V_{c2} for an excitatory and an inhibitory timing circuit were also monitored through a follower. The weights for all excitatory synapses are identical (V_{wexc}) and the weights for all inhibitory synapses are identical (V_{winh}).

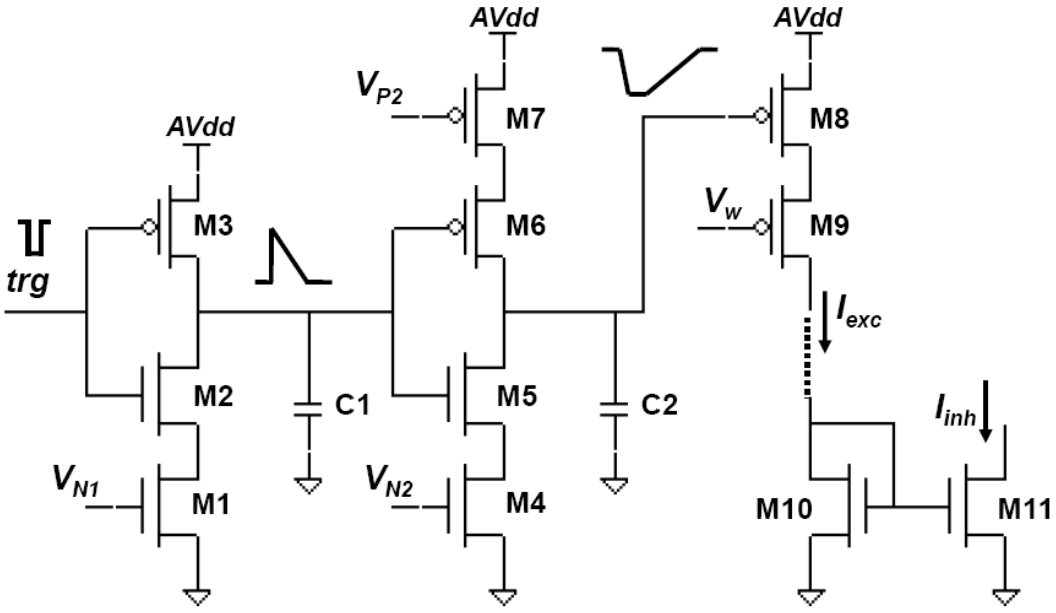

Figure 5.4: Schematic of the timing circuit along with the excitatory and inhibitory synapses. The timing circuit formed by transistors M1 through M7 along with capacitors C1 and C2 is placed along the diagonal of the spectral difference chip and controls the dynamics of the synapses. The circuit consisting of transistors M8 and M9 is the schematic of an excitatory synapse with V_w controlling the weight of the synapse. The circuit consisting of transistors M8 through M11 is the schematic of an inhibitory synapse with V_w controlling the weight of the synapse.

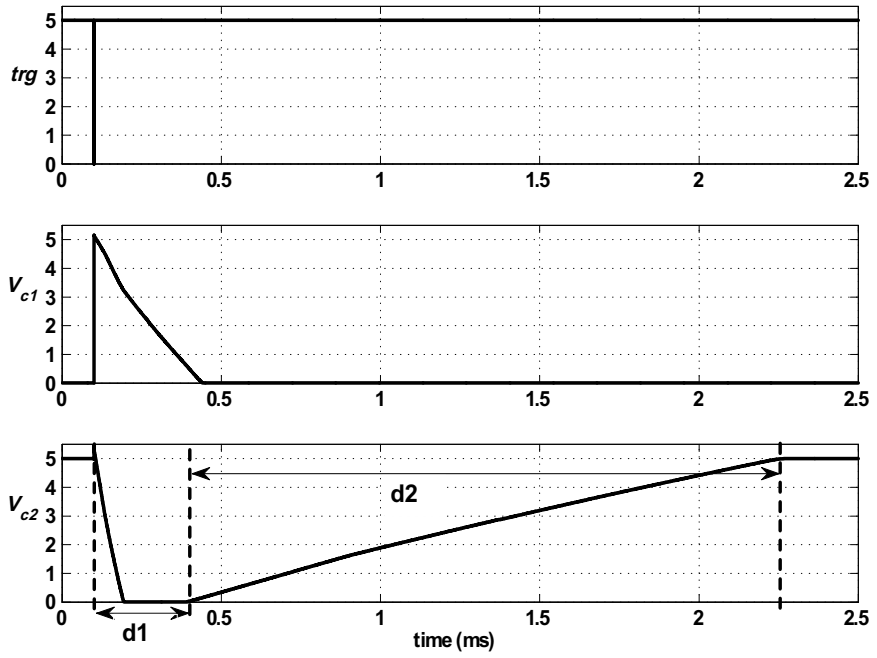

Figure 5.5: Timing circuit in the single spike mode. Simulated response of the timing circuit in response to a single input spike (*trg*). Upper plot: A $160ns$ wide active-low trigger pulse (*trg*) is applied to the input inverter formed by transistors M2 and M3. Middle plot: The capacitor C1 is rapidly charged to Vdd and is discharged through M1 and M2 once *trg* returns to its high state. The rate of discharge is controlled by the voltage V_{N1}. Bottom plot: In this mode of operation the voltage V_{N2} is set such that the capacitor voltage V_{c2} can be discharged all the way to $0V$ during the interval in which M5 is turned on.

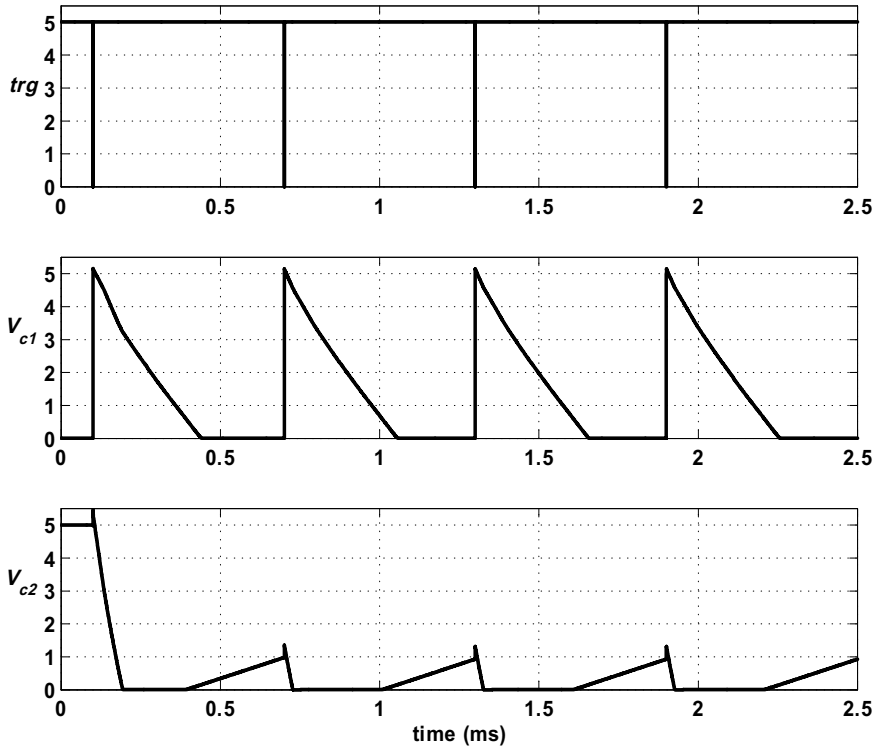

Figure 5.6: Timing circuit in the single spike mode. Simulated response of the synapse circuit in response to multiple spikes. In this mode of operation, if a second spikes arrives before the capacitor C2 has charged up to Vdd, the voltage V_{c2} will be discharged to ground thus extending the time duration of the synaptic output only by an amount equal to the interspike interval. The small glitches observed in V_{c2} coincide with the rapid increase of V_{c1} and are attributed to capacitive coupling.

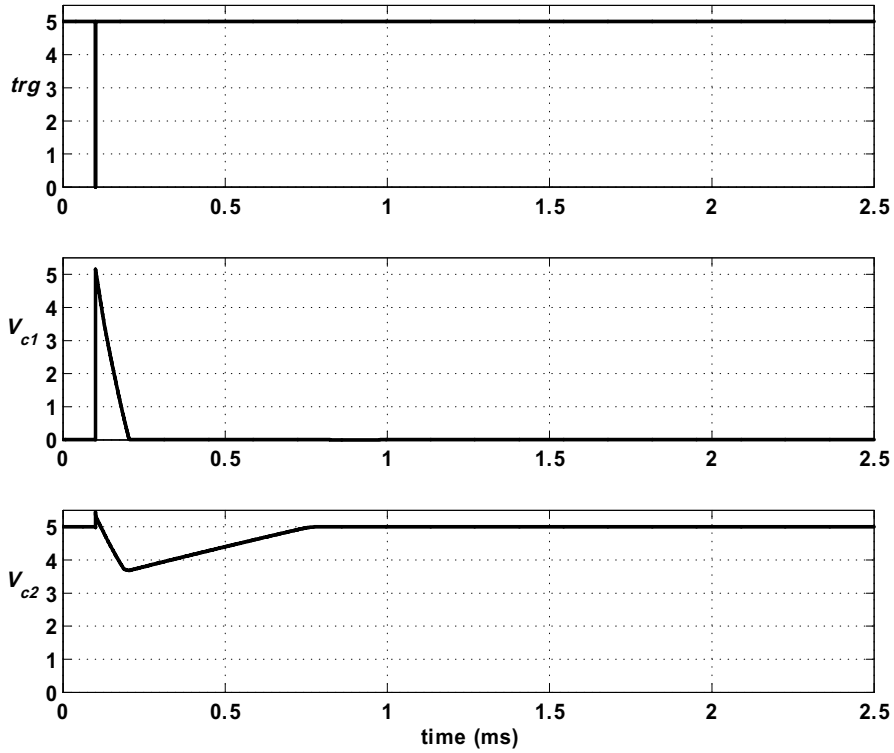

Figure 5.7: Timing circuit in the multiple spike mode. Simulated response of the timing circuit in response to a single input spike. In this mode of operation, the voltage V_{N2} is set such that for a single input spike (trg) the capacitor voltage V_{c2} is not discharged all the way to ground.

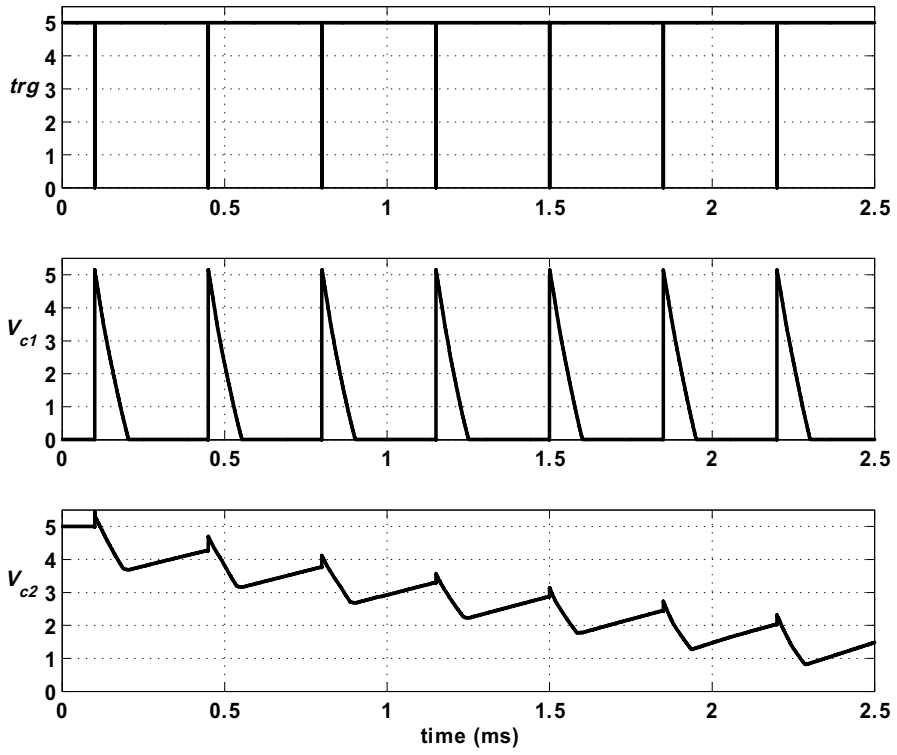

Figure 5.8: Timing circuit in the multiple spike mode. Simulated response of the synapse circuit in response to multiple spikes. In this mode of operation, the effect of any additional spike is similar to that of the first spike (unless V_{C2} is close to $0V$) thus extending the time duration of the synaptic output by an amount proportional to the number of input spikes.

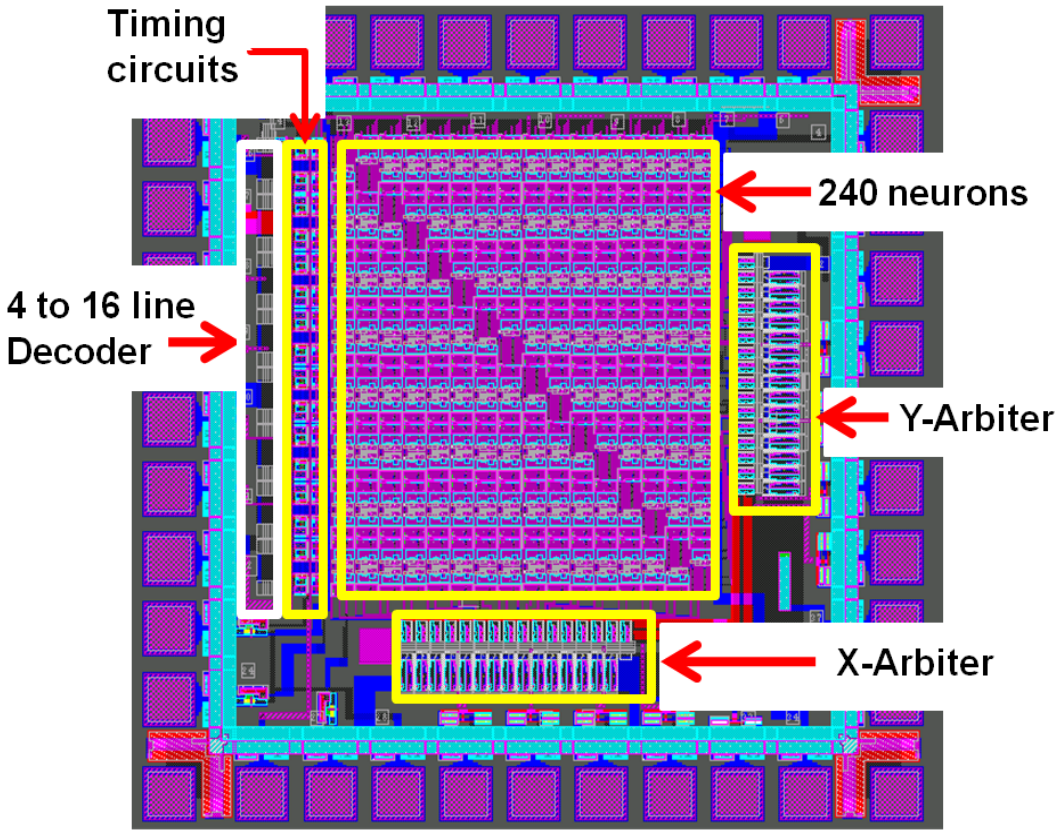

Figure 5.9: Layout of the spectral difference chip. The chip has 240 neurons arranged in a 16 by 16 grid with no neurons on the diagonal (lower right to upper left). Each neuron has an excitatory synapse and an inhibitory synapse. The timing of the synapses are controlled via timing circuits. The sixteen timing blocks shown lying on the diagonal in Fig. 5.2 were placed external to the neuron grid. A two-dimensional address-event arbiter is used to transmit the spikes off of the chip.

5.3 Interaural Level Difference (ILD) Chip

The second of the two feature extracting chips, the ILD chip is a binaural chip; it compares the activity of pairs of frequency-matched cochlear filters, one from the right cochlea chip and one from the left cochlea chip. Consider Fig. 5.10, the two sets of black dots represent the (exponentially-spaced) center frequencies of the 16 cochlear filters from both the right and left cochlea . For any two cochlear filters i (i=1, 2, ...16) with identical center frequencies there are two possible neural configurations: neuron **R-L** receives excitation from the right cochlear filter and inhibition from the left cochlear filter, while neuron **L-R** receives excitation from the left cochlear filter and inhibition from the right cochlear filter. Since there are 16 cochlear filters in each ear, there are 32 such neurons.Fig. 5.11 illustrates a detailed neural model of the ILD computation. The figure shows the output of the cochlear filter being half-wave rectified and synapsing onto the eight neurons (circles) associated with a particular cochlear filter as described in Chapter 4. The ILD neurons receive excitation from one cochlea and inhibition from the other.

Fig. 5.12 shows how the ILD computation is organized on the chip. The computation is broken down into ILD blocks. The ILD chip has 16 ILD blocks, one for each cochlear filter. Each ILD block has two neurons, one neuron (**R-L**) receives excitation from the right ear and inhibition from the left ear, while the other neuron (**L-R**) receives excitation from the left ear and inhibition from the right ear. Each neuron has eight excitatory synapses and eight inhibitory synapses corresponding to the eight cochlear neurons in each of the right and

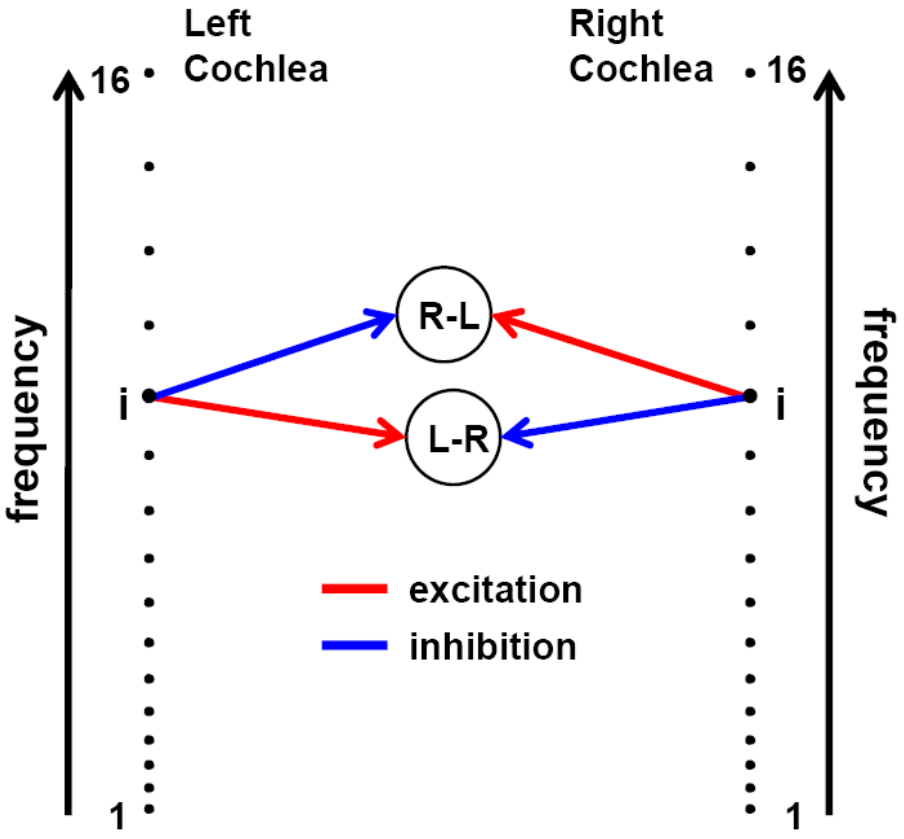

Figure 5.10: ILD neurons. The two sets of black dots represent the exponentially-increasing center frequencies of the 16 cochlear filters from both the right and left cochlea . For any two cochlear filters i (i=1, 2, ...16) with identical center frequencies there are two possible neural configurations: neuron **R-L** receives excitation from the right cochlear filter and inhibition from the left cochlear filter, while neuron **L-R** receives excitation from the left cochlear filter and inhibition from the right cochlear filter. Since there are 16 cochlear filters in each cochlea chip, there are 32 such neurons.

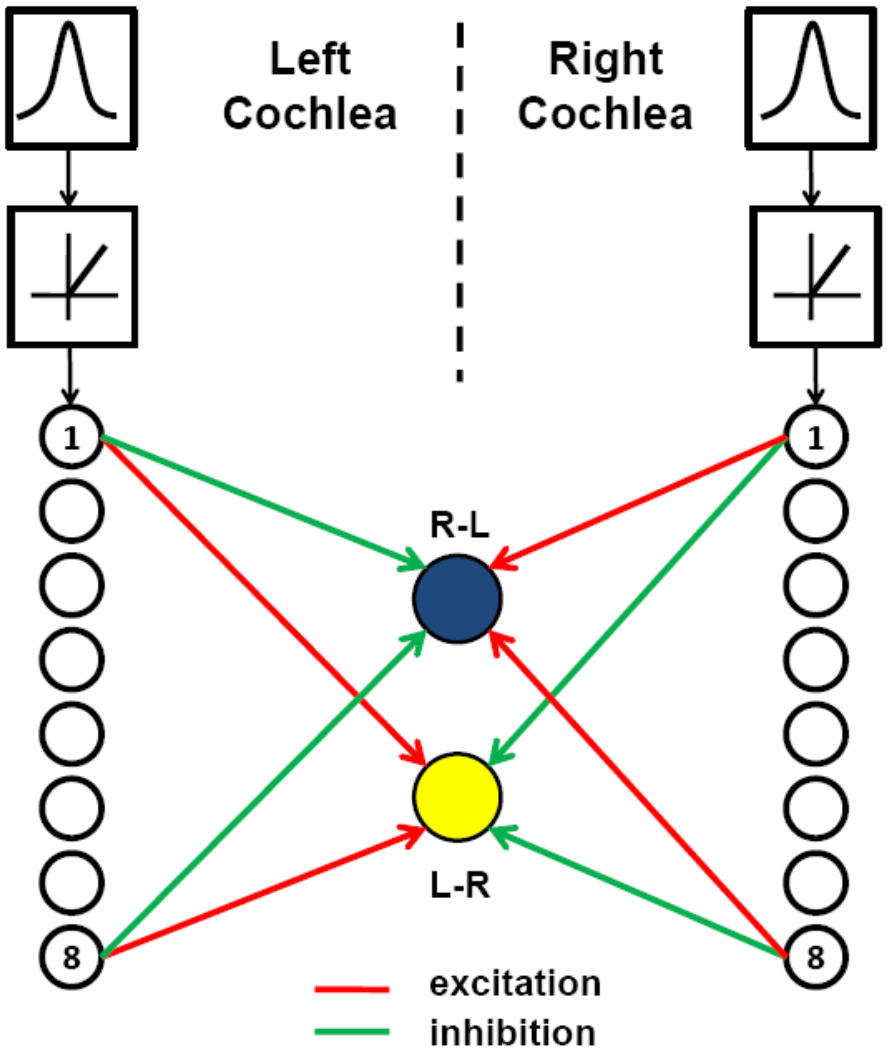

Figure 5.11: Neural model of the ILD computation. The ILD neurons receive excitation from one cochlea and inhibition from the other. Each circle represents one of the eight neurons receiving inputs from a particular cochlear filter as described in Chapter 4.

left cochlea chips. As illustrated in Fig. 5.11. each cochlear neuron makes two synaptic connections, one excitatory to an **R-L** neuron and one inhibitory to a **L-R** neuron. This is reflected in the ILD block in Fig. 5.12 where each row in the block has a single excitatory and inhibitory synapse and is connected to a single cochlear neuron. There is a one-to-one mapping from the 256 cochlear neurons (128 neurons/cochlea chip x 2 cochlea chips) to the 256 rows of the ILD blocks (16 ILD blocks x 16 rows/block). There are a total of 32 ILD neurons; the **R-L** have an even number (address) (0, 2, ...30) while the **L-R** neurons have an odd number (1, 3, ...31).

Shown in Fig. 5.13 is the schematic of the ILD neuron. The core of the ILD neuron is identical to that used in the cochlea and spectral difference chips, however, a one-dimensional address-event system was employed in the ILD chip requiring fewer transistors within each neuron. Transistors M1 through M9 form the voltage comparator that compares the voltage V_{mem} of the membrane capacitor C_{mem} against the voltage V_{th}. V_{refr} controls the refractory period of the neuron. V_{Leak} controls the leakage current of the neuron and is the neuron's actual threshold as described in Chapter 4.

The layout of the ILD chip is shown in Fig. 5.14. The chip has 16 ILD blocks, each block connects (through the address-event system) to one of the 16 cochlear filters. The chip has a 8-bit address-event input, seven of the eight bits represent the address of neurons coming off of the cochlea chip, the eighth bit (D bit) determines whether the incoming spike was sent from the right or left cochlea chip. The 8-bit input address is split into two 4-bit addresses, each with its own

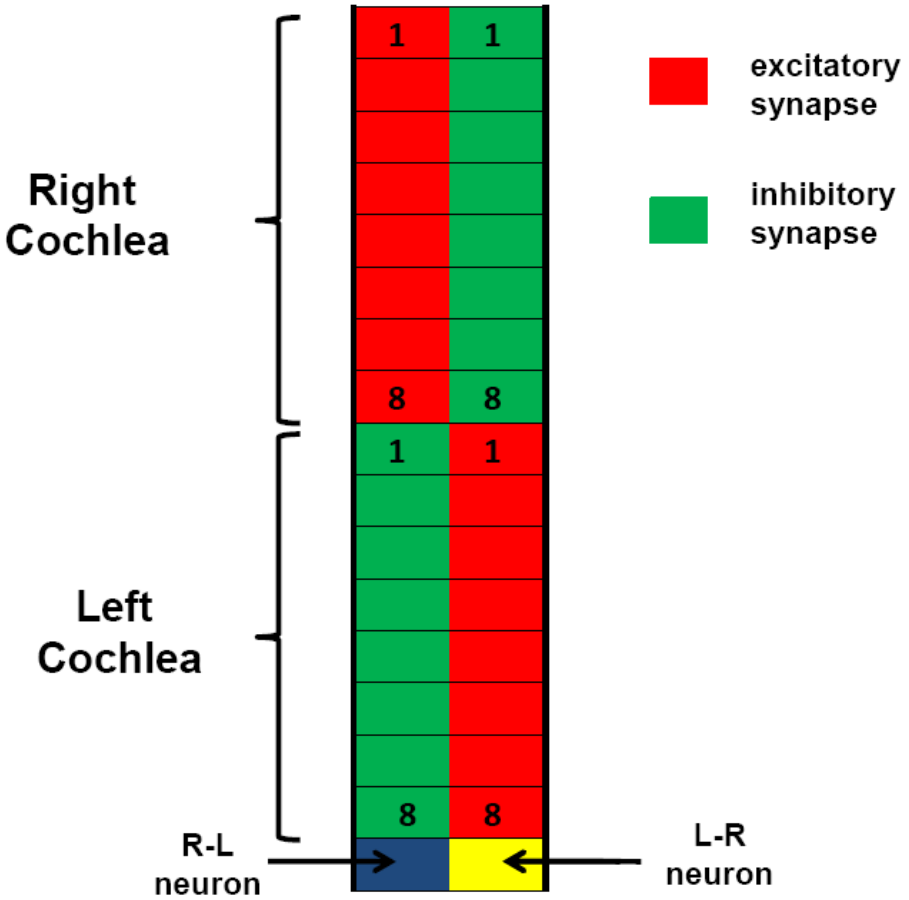

Figure 5.12: ILD Block. The ILD chip has 16 ILD blocks, one for each cochlear filter. Each ILD block has two neurons, one neuron (R-L) receives excitation from the right ear and inhibition from the left ear, while the other neuron (L-R) receives excitation from the left ear and inhibition from the right ear. Each neuron has eight excitatory synapses and eight inhibitory synapses. There is a one-to-mapping from the 256 cochlear neurons (128 neurons/cochlea chip x 2 cochlea chips) to the 256 rows of the ILD blocks (16 ILD blocks x 16 rows/block).

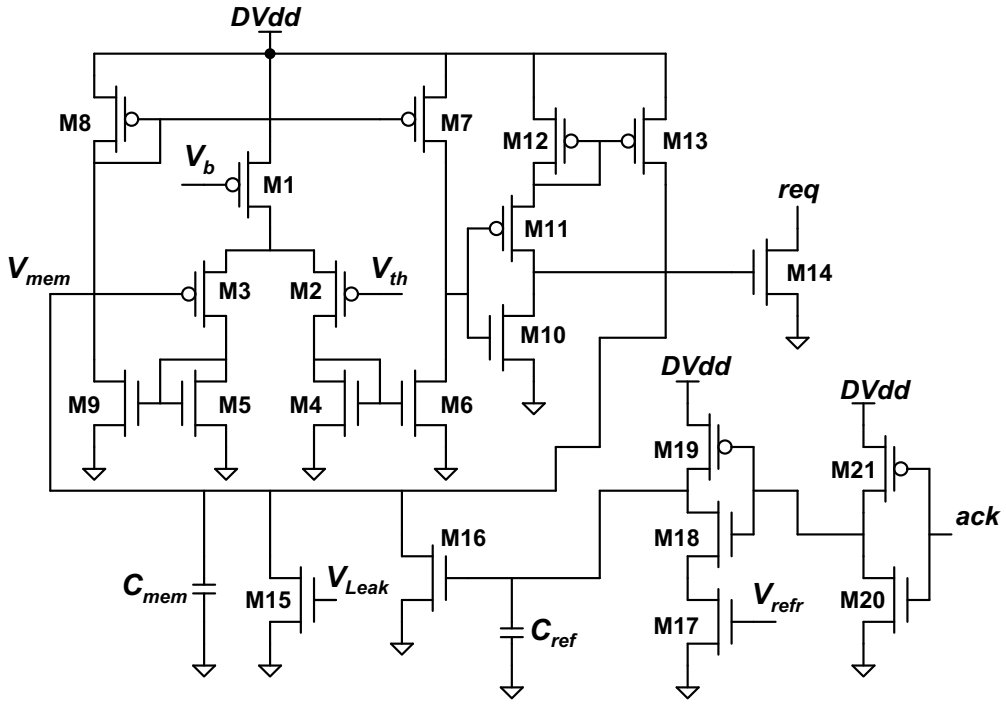

Figure 5.13: ILD neuron. The core of the ILD neuron is essentially the same as that used in the cochlea and spectral difference chip, however, a one-dimensional address-event system was employed in the ILD chip requiring fewer transistors within each neuron.

4-bit decoder. The horizontal decoder seen at the top of the chip decodes the X part (four bits) of the cochlea chip address (see Chapter 4) and selects one of the 16 ILD blocks to receive the incoming spike. The vertical decoder decodes the Y part (three bits) of the cochlea chip address along with the D bit. If D=0 the decoder selects the upper half of the ILD block indicating that the received spike is from the right cochlea and if D=1 it selects the lower half of the ILD block indicating that the received spike is from the left cochlea. The Y address selects which of the eight rows in the selected half of the ILD block to trigger. A single spike triggers both an excitatory and an inhibitory synapse simultaneously. All excitatory synapses receiving inputs from cochlear neurons with the same threshold have the same weight. All inhibitory synapses receiving inputs from cochlear neurons with the same threshold have the same weight. The weights are set via the two resistor ladders shown on the right of the chip.

5.4 Mismatch

The time duration (T) it takes for a capacitor (C) to (linearly) change by $\triangle V$ volts using a constant current I is given by:

$$T = \frac{C \triangle V}{I} \tag{5.1}$$

Let the current of an NMOS transistor be written as

$$I = I_o \exp(\frac{\kappa V_g}{V_T} + X) \tag{5.2}$$

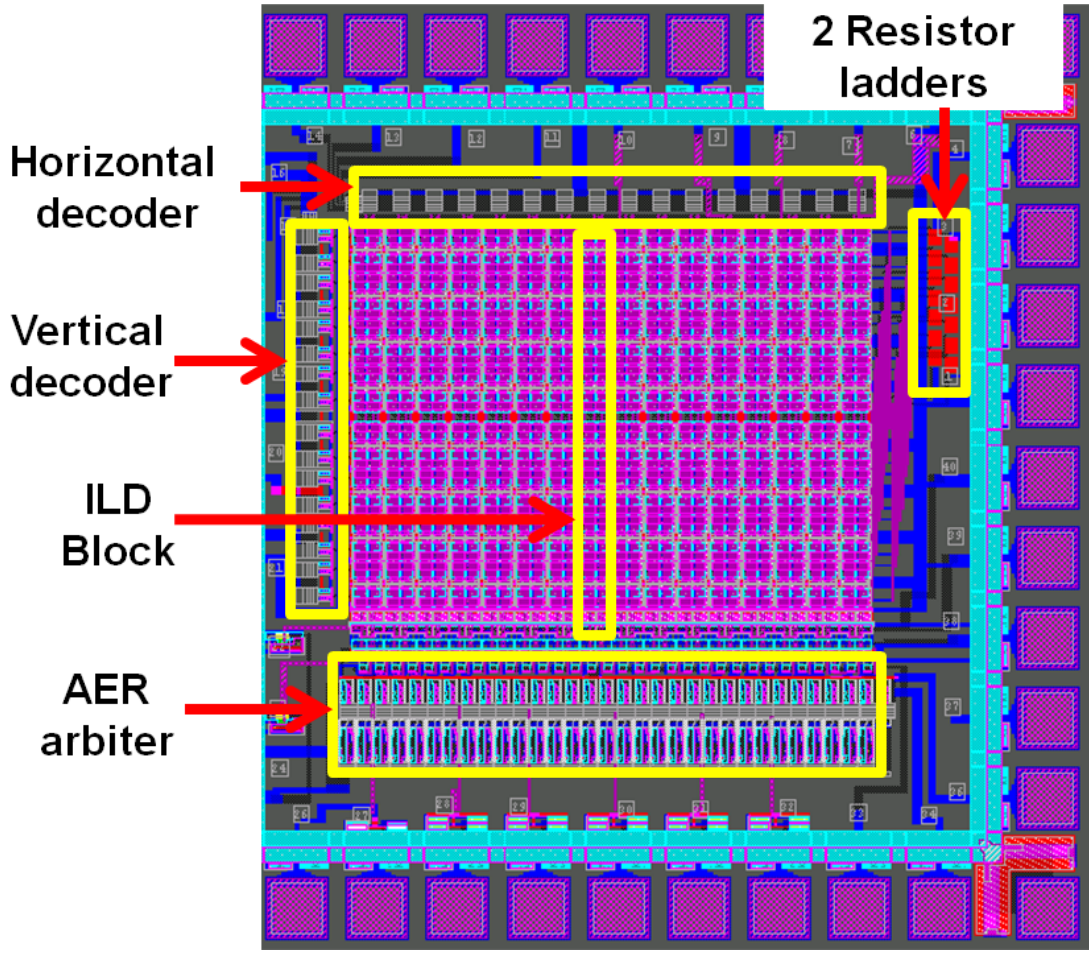

Figure 5.14: Layout of the ILD chip. The chip has 16 ILD blocks and an 8-bit address-event input. The 8-bit input address is split into two 4-bit addresses, each with its own 4-bit decoder. The horizontal decoder decodes the X part of the cochlea chip address and selects which of the 16 ILD blocks is to receive the incoming spike. The vertical decoder decodes the Y part of the cochlea chip address and selects which of the 16 rows of the ILD block to trigger.

where V_g is the gate bias of the transistor, V_T is the thermal voltage, I_o is the zero bias current, κ is the body effect [103], and X is a normal random variable ($\mu_X = 0$, variance=σ_X^2) that models all the sources of mismatch in a transistor as described in Chapter 4. By substituting equation 5.2 into 5.1 the time duration T may be written as

$$\begin{aligned} T &= \frac{C\triangle V}{I_o \exp(\frac{\kappa V_g}{V_T} + X)} \\ &= T_o \exp(-X) \\ &= T_o Y \end{aligned} \tag{5.3}$$

where

$$T_o = \frac{C\triangle V}{I_o \exp(\frac{\kappa V_g}{V_T})} \tag{5.4}$$

and

$$Y = \exp(-X) \tag{5.5}$$

Therefore, T will have a lognormal distribution as described in Chapter 4. For a fixed V_g, T_o is constant and the variance σ_T^2 of T may be written as

$$\sigma_T^2 = (T_o)^2 \sigma_Y^2 \tag{5.6}$$

and since

$$\sigma_Y = e^{(\mu_X + \frac{\sigma_X^2}{2})} \sqrt{(e^{\sigma_X^2} - 1)} \tag{5.7}$$

therefore, by substituting equation 5.7 in equation 5.6 we get

$$\sigma_T = T_o e^{(\frac{\sigma_X^2}{2})} \sqrt{(e^{\sigma_X^2} - 1)}. \tag{5.8}$$

Therefore, for a fixed mismatch X, increasing the current I (by increasing V_g) results in decreasing T_o and in turn a reduction in the mismatch in T. This is

in contrast to the cochlea chip where increasing the bias current resulted in the increase in mismatch of the quality factor Q.

5.5 Test Results

All chips were tested at a supply voltage of $5V$. For testing purposes, a dsPIC30F4011 microcontroller was programmed to send address-event spikes to the chips.

5.5.1 Spectral Difference Chip

In the spectral difference chip, because of the limited number of pins, only four analog voltages could be monitored; the membrane voltages (V_{mem}) for two neurons as well as two capacitor voltages: V_{C2} for an excitatory synapse(V_{C2e}) and V_{C2} for an inhibitory synapse (V_{C2i}). The two capacitor voltages are located in the same timing block (bottom row), one of the two membrane voltages was also located in the bottom row to able to observe the effect of the synapses.

Fig. 5.15 shows the scope trace (averaged 40 times) of the timing circuit voltage V_{C2e} and a neuron's membrane voltage V_{mem} in response to an input spike. In this plot the timing circuit was operated in the single spike mode, as illustrated in Fig. 5.5. The neuron was set to have an integration time of $1ms$. After the neuron fired a spike at $t = 3ms$ it went through a refractory period after which the membrane voltage starting charging up, however, the control voltage V_{C2e} caused the excitatory current to stop before the neuron could fire a second spike and the membrane voltage was then discharged through the leak transistor

(M15). The duration of d1 is controlled by V_{N1} as shown in Fig. 5.16. In the single spike mode d1 is set to be as small as possible, and d2 controls the duration of the excitatory of the neuron. Fig. 5.17 shows how V_{P2}. can be used to set the duration of d2. We will refer to d_{mem} (in Fig. 5.15) as the integration time of the neuron, defined as the time it takes the neuron to charge from zero to its firing threshold in the absence of any inhibitory current.

Fig. 5.18 shows the response of the spectral difference chip when its rows were stimulated in succession starting with the topmost row (no inhibitory currents). The time interval between stimulating the rows was set at $534\mu s$. The bias for the excitatory current was varied, the excitatory current in the upper plot is greater than that in the lower plots. In the upper plot, it is observed that all neurons in the same row fire in synchrony (same d_{mem}) while in the lower plot neurons within the same row are less synchronized. This is attributed to the increase in mismatch as the excitatory current decreases (see section 5.4).

Fig. 5.19 shows d_{mem} (blue line) as well as the deviation $\triangle\mathrm{d}_{mem}$ in d_{mem} (black dots) as a function of the excitatory weight voltage V_{wexc} (gate voltage of a PMOS transistor). For each individual row, $\triangle\mathrm{d}_{mem}$ was calculated as absolute time difference between first and last spike. As V_{wexc} increases the excitatory current decreases resulting in the increase of both d_{mem} and the mismatch in d_{mem} ($\triangle\mathrm{d}_{mem}$). Fig. 5.20 shows the standard deviation in d_{mem} ($\sigma_{d_{mem}}$) as a function of d_{mem}. First the mean firing time was calculated for each row, then the difference ($\triangle 2d_{mem}$) between the actual firing time and the mean firing time was computed, the standard deviation $\sigma_{d_{mem}}$ is that of $\triangle 2d_{mem}$. The linear relation

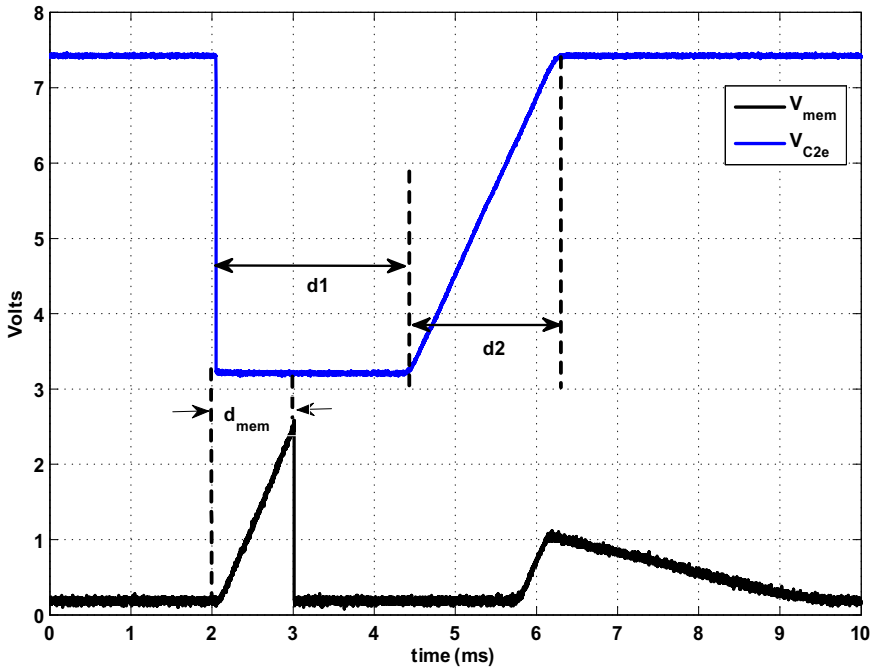

Figure 5.15: Spectral difference chip operated in the single spike mode. In response to an input spike the capacitor voltage V_{C2e} (also tr_e) was discharged to ground turning on the excitatory synapse. The membrane voltage increased to threshold, firing a spike, then went through a refractory period and then started integrating again, however, V_{C2e} was charged up back to Vdd turning off the synapse before the neuron could fire a second spike. The duration of d1 and d2 is controlled by V_{N1} and V_{P2} respectively. The measured response was averaged 40 times (by the oscilloscope).

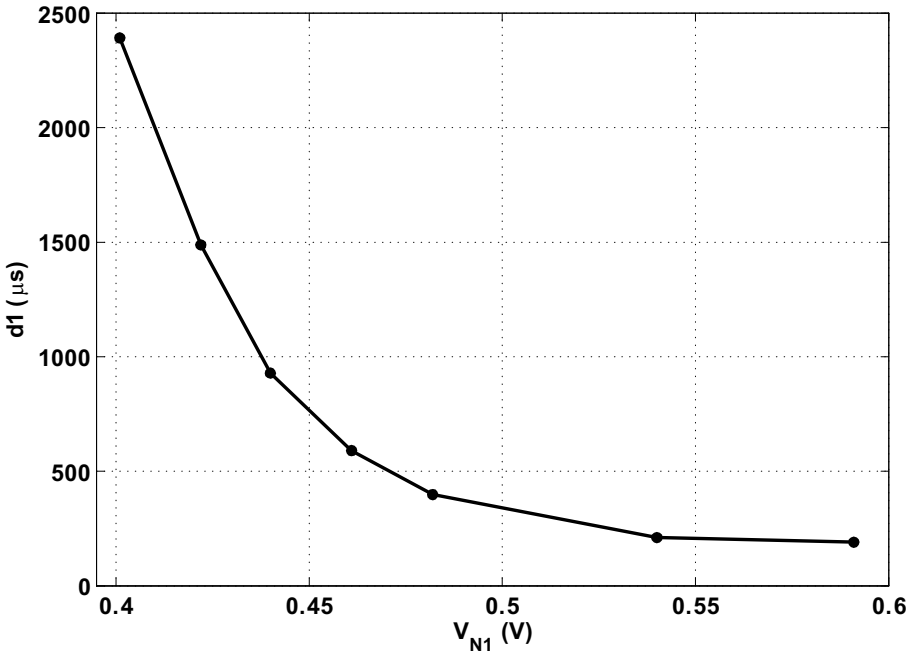

Figure 5.16: Spectral difference chip: Setting the timing circuit. The plot shows d1 (in Fig. 5.15) as a function of V_{N1}. This data is based on measuring V_{C2e} from a single timing block.

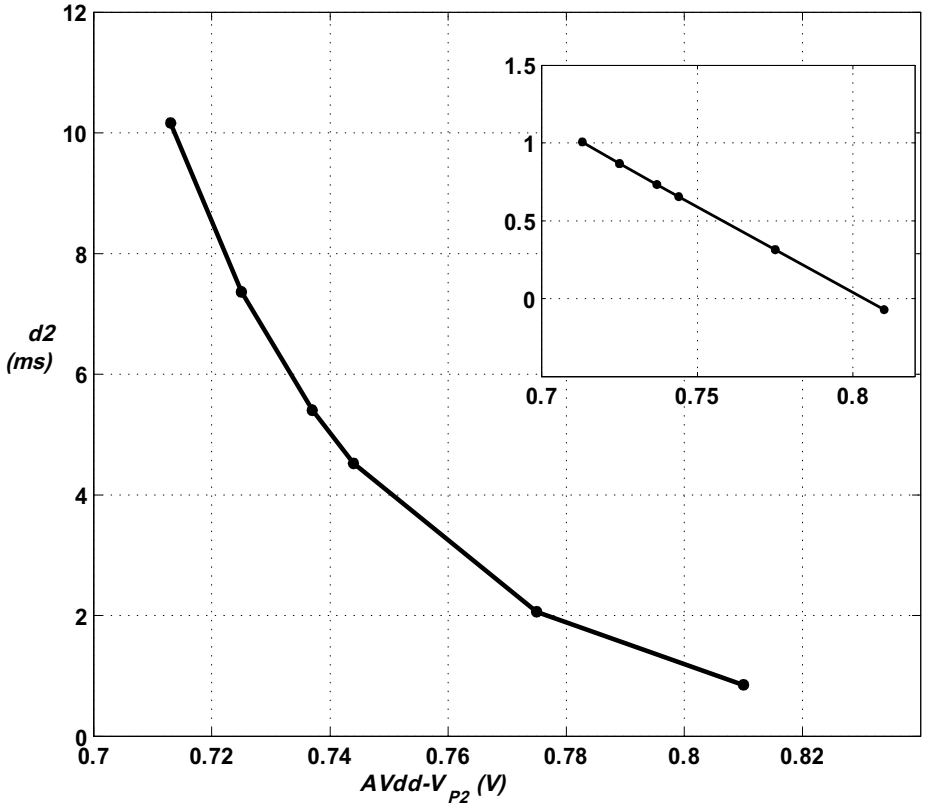

Figure 5.17: Spectral difference chip: Setting the timing circuit. The plot shows d2 (in Fig. 5.15) as function of V_{P2}. This data is based on measuring V_{C2e} from a single timing circuit. The inset shows the same data on a log scale (log(d2)), the linear relation is due to the exponential dependence of d2 on V_{C2e}.

between $\sigma_{d_{mem}}$ and d_{mem} is in agreement with equation 5.8.

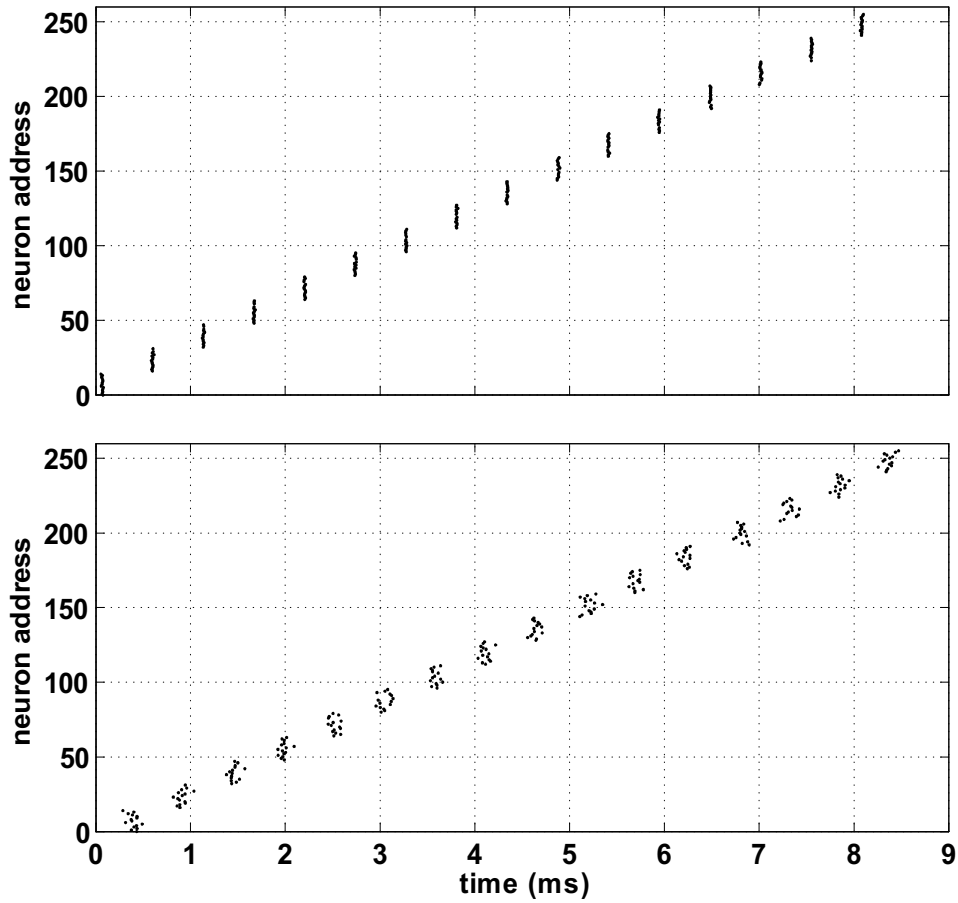

Figure 5.18: Spectral difference chip: Response of the chip to stimulating rows in succession. The difference between the upper and lower plots is that the excitatory current in the upper plot is greater.

The 16 by 16 grid of neuron responses in Fig. 5.21 mirrors the layout of the spectral difference chip. The presence of a black dot indicates that a neuron fired a spike as a response to the presented stimulus. There are no neurons along the diagonal (red asterisks). In this plot, the 16 rows of the spectral difference chip were stimulated in succession starting with the topmost row (row 0). The time

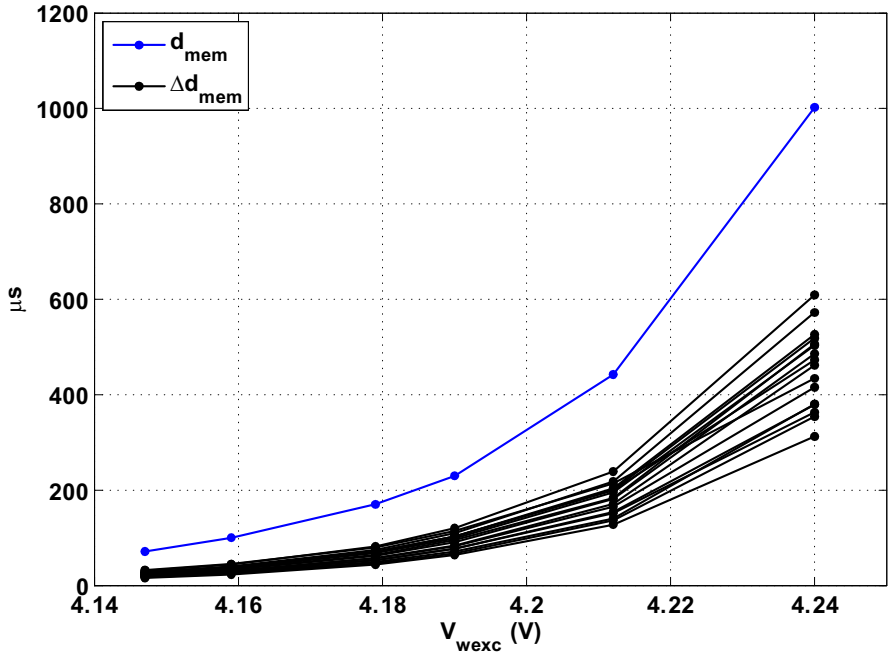

Figure 5.19: d_{mem} and $\triangle d_{mem}$ as a function of V_{wexc}. The plot shows d_{mem} (blue dots) as well as the deviation $\triangle d_{mem}$ in d_{mem} (black dots) as a function of the excitatory weight voltage V_{wexc} (gate voltage of a PMOS transistor). As V_{wexc} increases, the excitatory current decreases resulting in the increase of both d_{mem} and the mismatch in d_{mem} ($\triangle d_{mem}$).

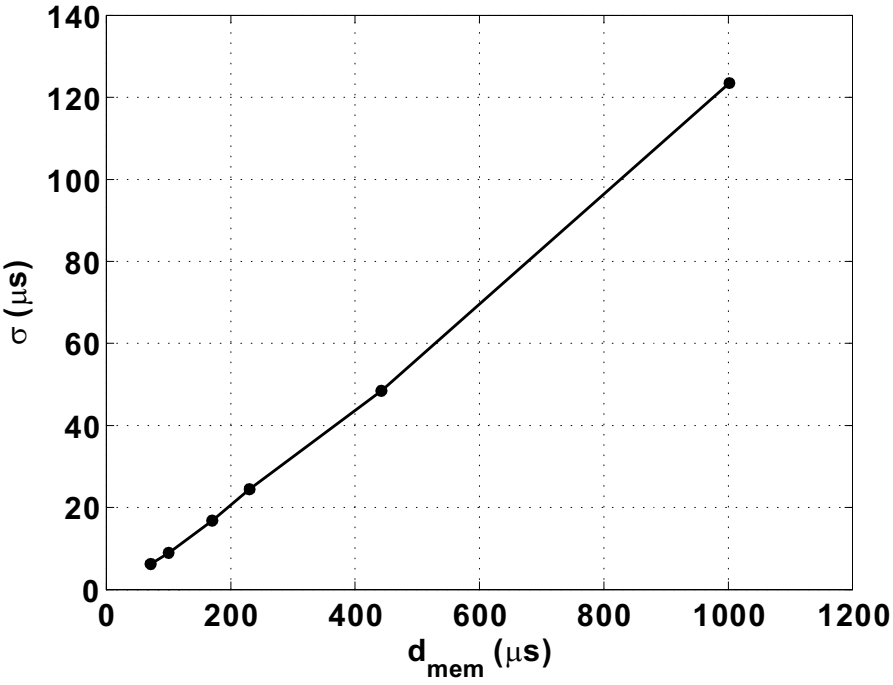

Figure 5.20: $\sigma_{d_{mem}}$ as a function of d_{mem}. The linear relationship between $\sigma_{d_{mem}}$ and d_{mem} is in agreement with equation 5.8.

interval (T_r) between stimulating successive rows was approximately $400\mu s$, the integration time d_{mem} for a neuron was approximately $148\mu s$, the duration of the excitatory current was approximately $1.68ms$, and the duration of the inhibitory current was approximately $4.7ms$. The direction of stimulating the rows was from top to bottom, therefore, all neurons above the diagonal received excitation before inhibition, and since $d_{mem}<T_r$, they all fired a spike. Neurons located below the diagonal and did not fire a spike received inhibitory current that outlasted their excitatory current and did not fire a spike. Neurons further away from the diagonal that did fire a spike either received excitation after the inhibition had ended or received excitation close to the end of the inhibitory current and thus their excitatory current outlasted the inhibition.

Fig. 5.22 shows the response of the spectral difference chip, the simulation pattern on similar to that in Fig. 5.21, stimulating the rows in succession starting with the topmost row. The time interval (T_r) between stimulating successive rows was approximately $267\mu s$, however, the integration time d_{mem} for a neuron was approximately $1ms$, the duration of the excitatory current was approximately $4.25ms$, and the duration of the inhibitory current was approximately $7.74ms$. As a result of the increased d_{mem}, some of the neurons above the diagonal received inhibition during their integration time before the neuron could fire a spike, therefore, the inhibition encroached into the region above the diagonal.

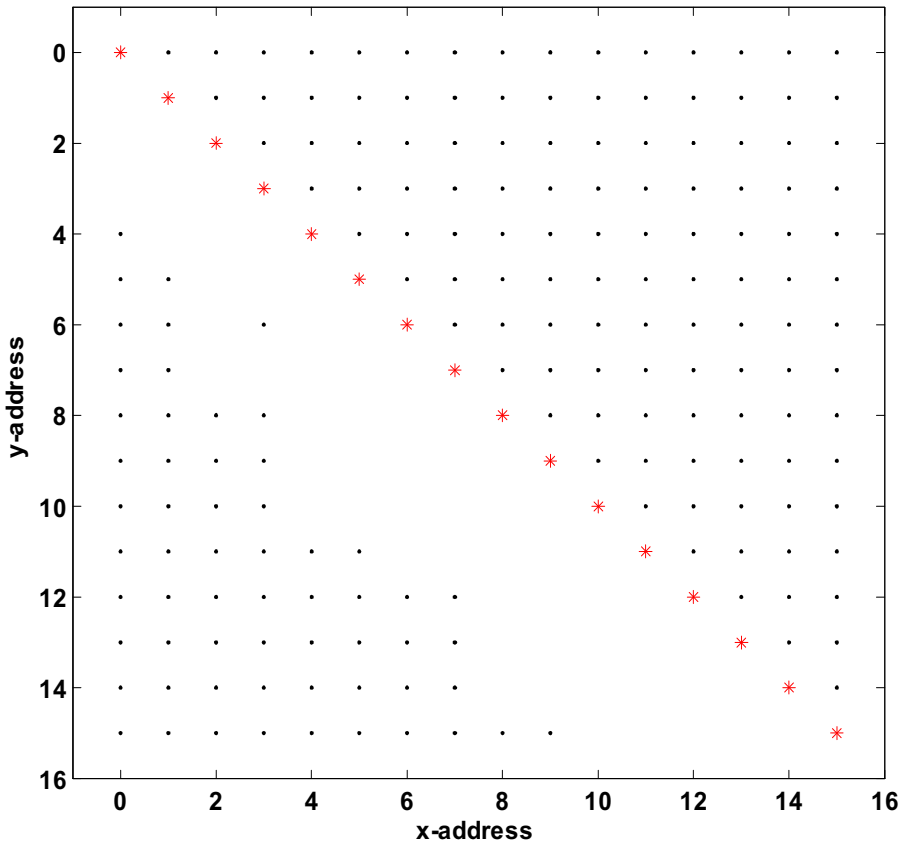

Figure 5.21: Response of the spectral difference chip. The 16 rows of the spectral difference chip were stimulated in succession starting with the topmost row. The time interval (T_r) between stimulating successive rows was approximately 400μs, the integration time d_{mem} for a neuron was approximately 148μs, the duration of the excitatory current was approximately 1.68ms, and the duration of the inhibitory current was approximately 4.7ms. All neurons above the diagonal received excitation before inhibition, and since $d_{mem} < T_r$, they all fired a spike.

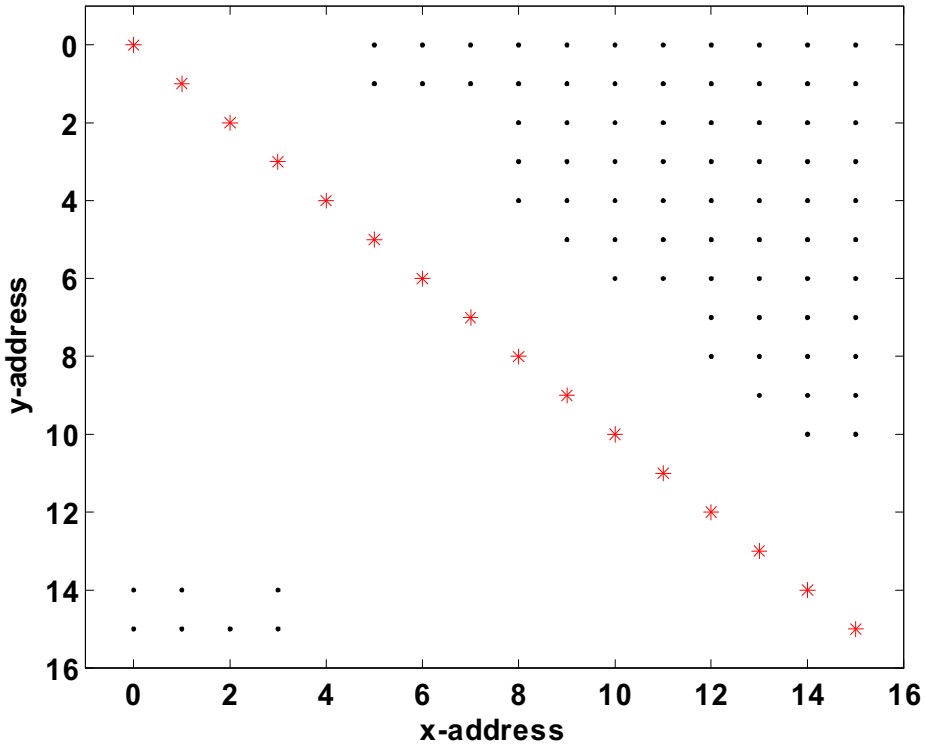

Figure 5.22: Response of the spectral difference chip. The rows were stimulated in succession starting with the topmost row (row 0). The time interval (T_r) between stimulating successive rows was approximately 267μs. The neurons had an integration time of 1ms. The inhibition lasted for a duration of 7.74ms and excitation lasted for 4.25ms.

5.5.2 ILD Chip

Two dsPIC30F4011 microcontrollers were programmed to emulate the right and left cochlea chips and send address-event spikes to the ILD chip for testing. The ILD chip has 32 neurons, 16 are **R-L** neurons receiving excitation from the right ear and inhibition from the left ear, while the other 16 are **L-R** neurons receiving excitation from the left ear and inhibition from the right ear. The **R-L** neurons are even numbered (0, 2, ...30), while the **L-R** neurons are odd numbered (1, 3, ...31).

Fig. 5.23 shows the ILD chip being testing by stimulating all 256 excitatory synapses. First, all 128 right excitatory synapses were triggered in succession resulting in all even-numbered neurons firing a spike. All 128 left excitatory synapses were then triggered in succession resulting in all odd-numbered neurons firing a spike. Recall that triggering an excitatory synapse in an ILD block is accompanied by triggering an inhibitory synapse, however, in this plot all inhibition was turned off. This plot demonstrates that all 256 excitatory synapses are functional.

Fig. 5.24 shows the ILD chip being tested by a similar pattern. In this plot, however, an excitatory right ear synapse was triggered followed by an excitatory left ear synapse. It is observed that all 32 ILD neurons fire in succession (again, all inhibition was turned off). Recall that the eight rows of inhibitory neurons can be set to have different weights. Fig. 5.25 shows the response of the ILD chip to the same stimulus in Fig. 5.24, however, in this plot a gradient was created in the inhibitory weights. It is observed that the inhibition was sufficient to inhibit some of the ILD neurons. As the inhibitory weights were increased, the number

of neurons that were inhibited also increased as shown in Fig. 5.26. These plots demonstrate the basic functionality of the ILD chip; in the following chapter we will present the response of the ILD chip to the measured responses from the artificial bat head.

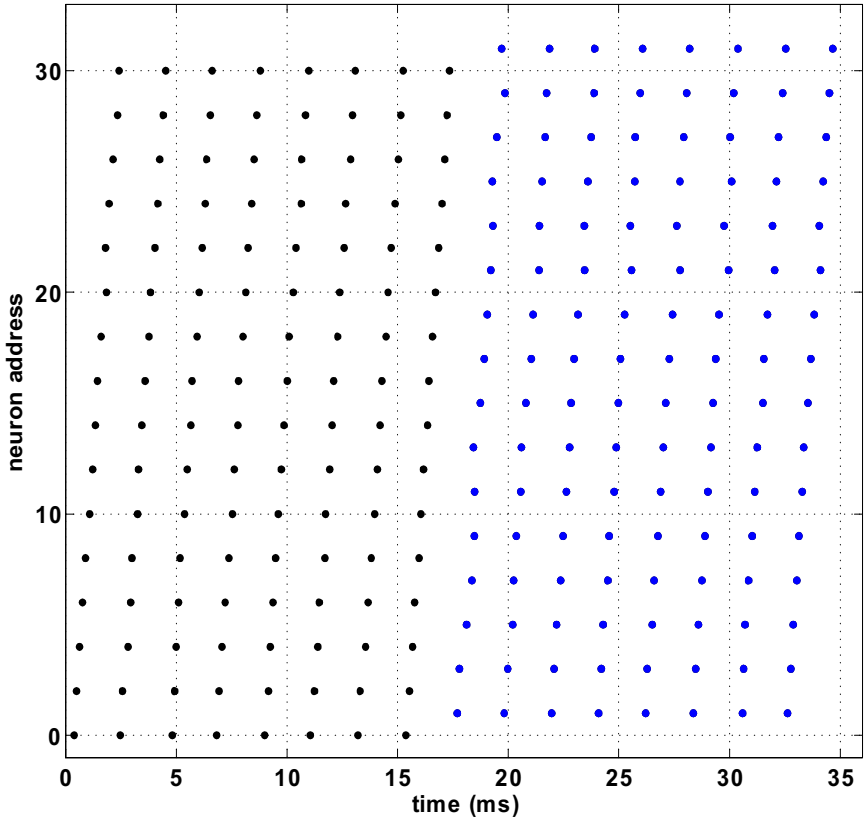

Figure 5.23: ILD chip. Testing the synapses. All 256 excitatory synapses were stimulated. First, all 128 right excitatory synapses were triggered in succession resulting in all even-numbered neurons firing a spike. All 128 left excitatory synapses were then triggered in succession resulting in all odd-numbered neurons firing a spike.

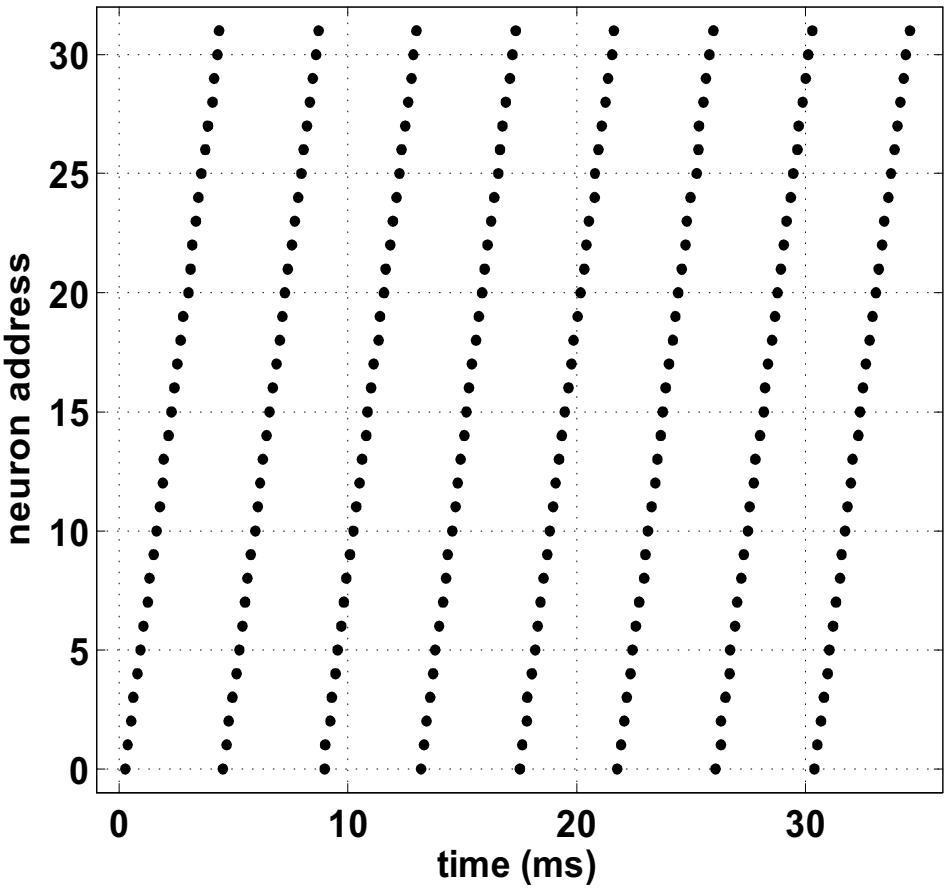

Figure 5.24: ILD chip. In this plot, an excitatory right ear synapse was triggered followed immediately by an excitatory left ear synapse. It is observed that all 32 ILD neurons fire in succession (all inhibition was turned off).

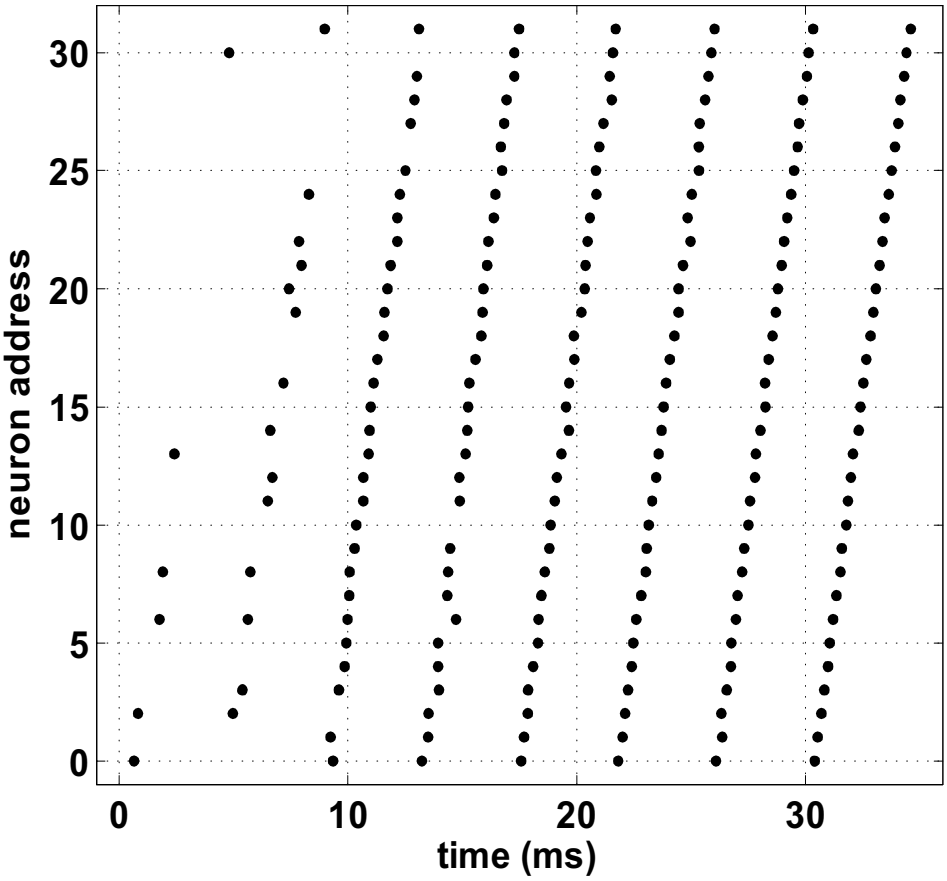

Figure 5.25: ILD chip. The ILD chip was stimulated with the same stimulus as in Fig. 5.24, however, the inhibitory weights (V_{winh}) were increased for some of the synapses causing some of the ILD neurons to be inhibited.

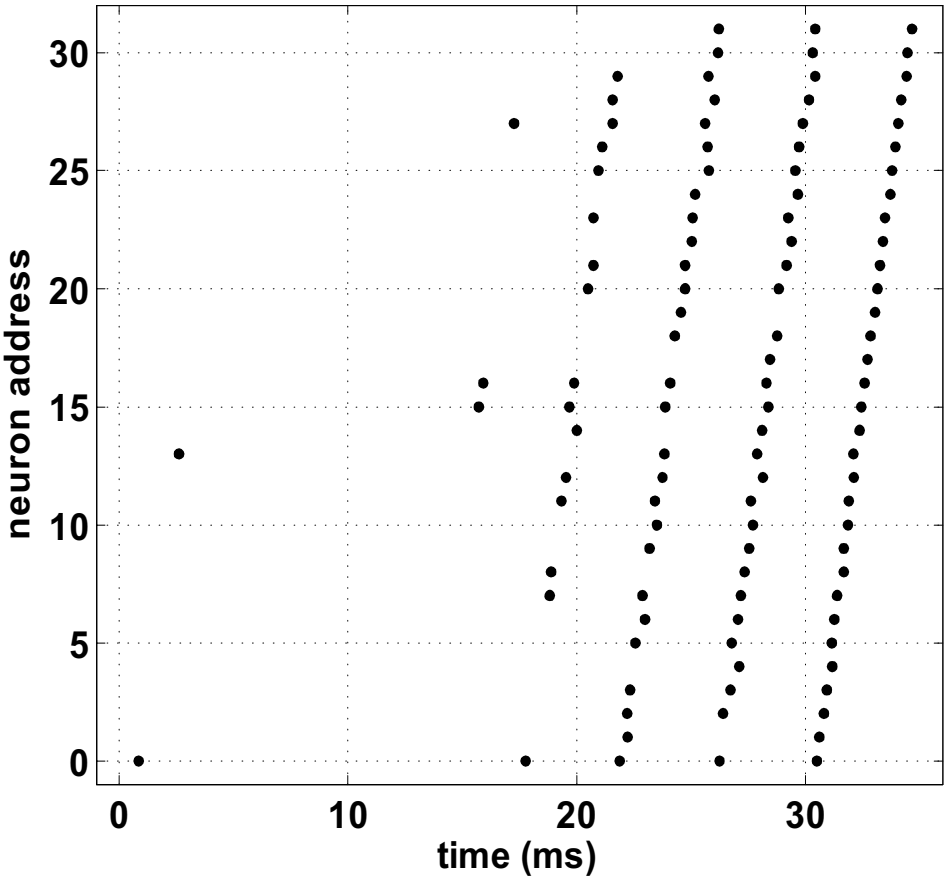

Figure 5.26: ILD chip. The ILD chip was stimulated with the same stimulus as in Figs. 5.24 and 5.25, however, the inhibitory weights (V_{winh}) were increased compared to those in Fig. 5.25 causing more ILD neurons to be inhibited.

5.5.3 Power Consumption

Fig. 5.27 shows the power supply current of the ILD and spectral difference chips (along with the cochlear neuron array) as a function of the bias voltage V_b. Even though the spectral difference chip has nearly twice as many neurons as the cochlear neuron array, the measured current is almost 25% less than that of the cochlea chip. The ILD chip has 1/4 as many neurons as that of he cochlear neuron array, however, its current is almost an order of magnitude smaller.

5.6 Merge Board

The localization task requires the extraction of both binaural and monaural cues from the spiking cochlear outputs. The spectral difference chip is a monaural chip and thus can receive its input directly from the cochlea chip. On the other hand, the interaural level difference (ILD) chip is a binaural processing stage and because of the limited number of available chip pins, the chip was designed so that its input address-event spikes from both the right and left cochlea share the same chip pins. This requires that the address bus from both cochleae be merged into a single address bus.

Each cochlea chip has 128 neurons which require a 7-bit address bus in addition to an active-low request line and an active-high acknowledge line. To distinguish between the right and left cochlea, an eighth address bit was added (externally) to the output of each cochlea chip, this additional line was set to the logic state "0 " for the right cochlea and to the logic state "1" for the left cochlea.

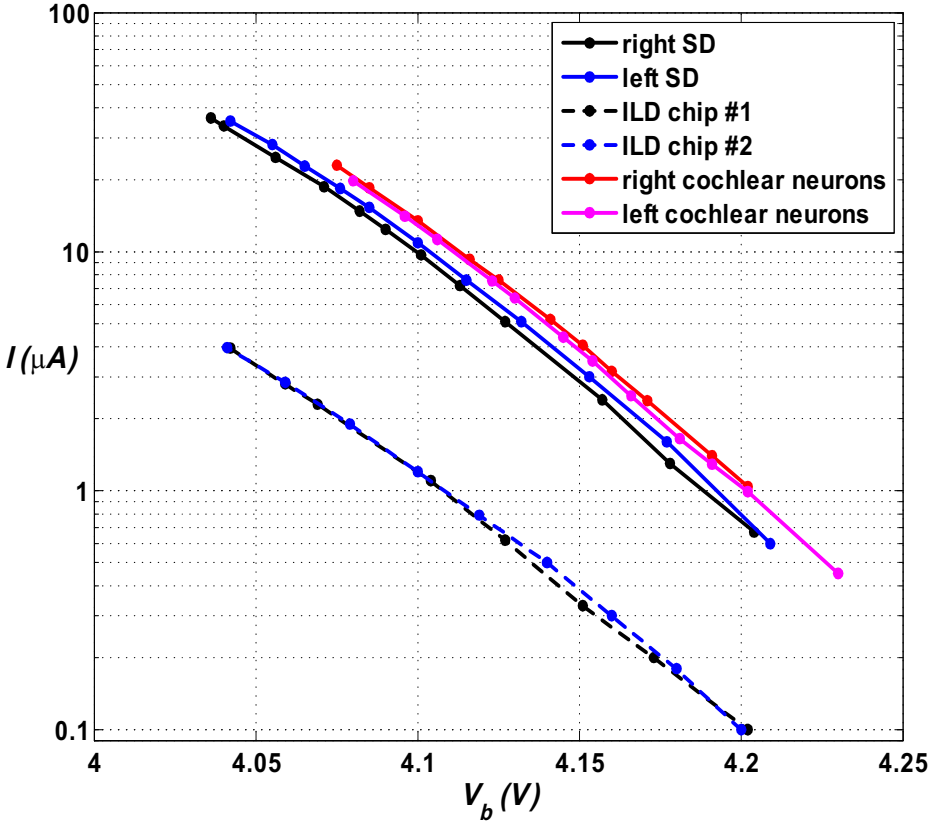

Figure 5.27: Power supply current measurement of the ILD and spectral difference chips (as well as the cochlear neuron array) as a function of the bias voltage V_b as measured from the cochlea chip LeqC7. Even though the spectral difference chip has nearly twice as many neurons as the cochlea chip, the measured current is almost 25% less than that of the cochlea chip. The ILD chip has 1/4 as many neurons as that of the cochlea chip, however, its current is almost an order of magnitude smaller.

A dsPIC30F411 microcontroller (Microchip^{TM}) was programmed to control the merge process. The microcontroller operates at a clock frequency of approximately 118MHz(external clock of 7.3728MHz, internally multiplied by 16). The clock cycle has a duration of $8.47ns$ and each instruction takes four clock cycles ($33.9ns$). The block diagram of the hardware localization system is shown in Fig. 5.28. The merge board consists of the microcontroller and the two 74F541 tri-state buffers. The 74F541 tri-state buffer has an active-low enable. Because the buffer has a TTL output, pull-up resistors (not shown) were used to pull up the address bus to $5V$.

Initially the microcontroller sets the two (active low) enable signals, en1 and en2, to the logic state high resulting in the outputs of the two buffers to be in their floating state and the merged address bus is pulled-up to $5V$ via the pull-up resistors. The two acknowledge signals, ack1 and ack2, as well as the three trigger signals, trg (trigger), trg_R (trigger_Right), and trg_L (trigger_Left) are also set to the logic state high.

After initialization, the microcontroller polls the request lines to check if a request has been asserted. If, for example, the right cochlea puts out a request (i.e. req1=0), low, the microcontroller first enables the right buffer by setting en1=0 thus connecting the output of the right cochlea chip to the common address bus. A delay (delay1) is provided to allow the buffer to turn on and for the cochlea chip's output to settle on the bus. The receiving chips are then triggered by pulling the trigger lines low (trigger=0 and trigger_right=0). A delay (delay2) was provided to control the width of the trigger pulse after which the trigger signals were set

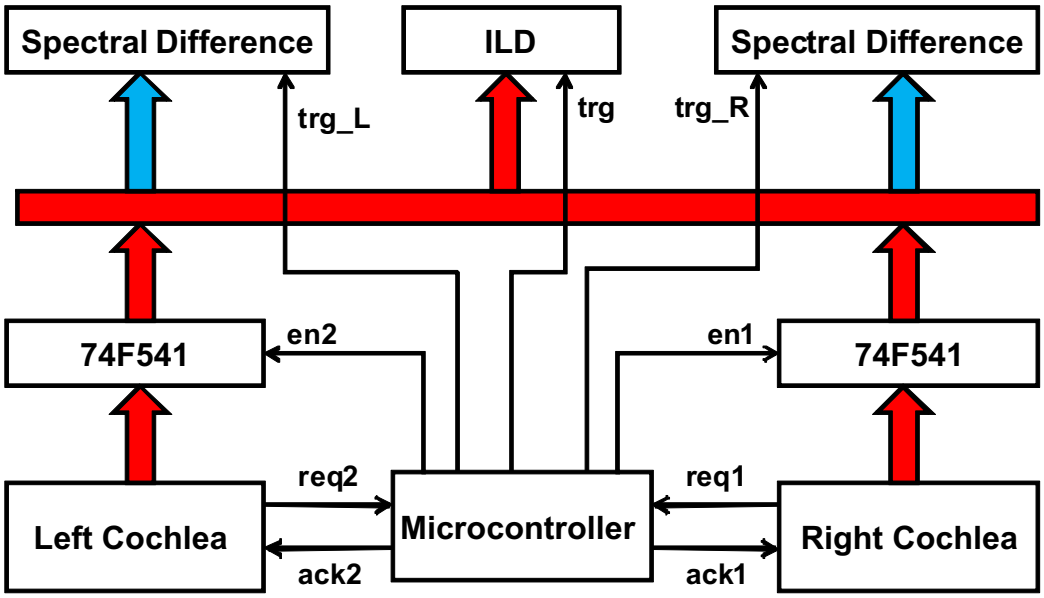

Figure 5.28: Block diagram of the hardware localization system. The merge board consists of the microcontroller and the two 74F541 tri-state buffers. A dsPIC30F4011 microcontroller was programmed to merge the address-event spikes of the two cochlea chips into a single address-event bus using the 74F541 tri-state buffer. The red address bus is eight bits wide while the light blue address bus is four bits wide.

back to their high states. The right buffer is then disabled (en1=1) before the acknowledge signal (ack1=1) is sent back to the cochlea chip to complete the handshaking process. The flowchart of the program in shown in Fig. 5.29.

The merge board was tested by programming two dsPIC30F4011 microcontrollers to simulate the address-event output of the right and left cochlea chips. The merge board merges the address-event spikes from the two (simulated) cochlea chips onto a single address bus. Fig. 5.30 shows the output of the merge board.

The black dots represent spikes from the right cochlea chip while the blue dots represent spikes from the left cochlea chip. The output shows 6878 spikes merged in $6ms$ ($873ns$ interval between spikes).

5.7 Conclusion

In this chapter we have presented two feature extraction chips: a monaural spectral difference chip with 240 neurons (each neuron compares the activity of pairs of cochlear filters within the same ear) and a binaural interaural level difference (ILD) chip with 32 neurons (each neuron compares the activity of pairs of frequency-matched cochlear filters, one from each ears). In this chapter we tested the chips with spikes generated from a dsPIC30F4011 microcontroller. In the next chapter we will present the response of the feature extraction chips to the FM sweeps measured from the artificial bat. We have also presented a merge board which uses a microcontroller to combine (merge) the address-event spikes from two (right and left) cochlea chips into a single address bus and sends the spikes to the feature extraction chips.

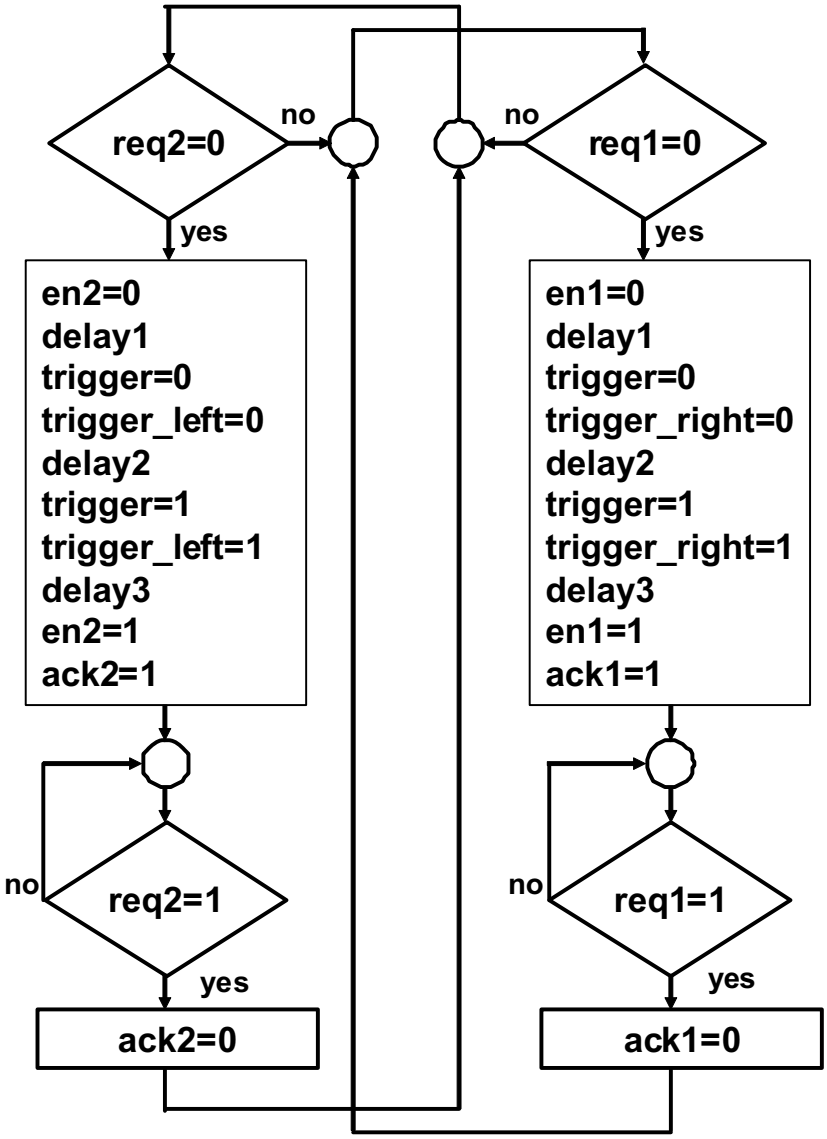

Figure 5.29: Flowchart of the merge process. After an initialization routine the microcontroller polls the request lines from the two cochlea chips; if a request has been asserted, it enables the corresponding buffer, sends a trigger pulse to both the ILD chip and the corresponding spectral difference chip and then completes the handshaking process with the requesting cochlea chip.

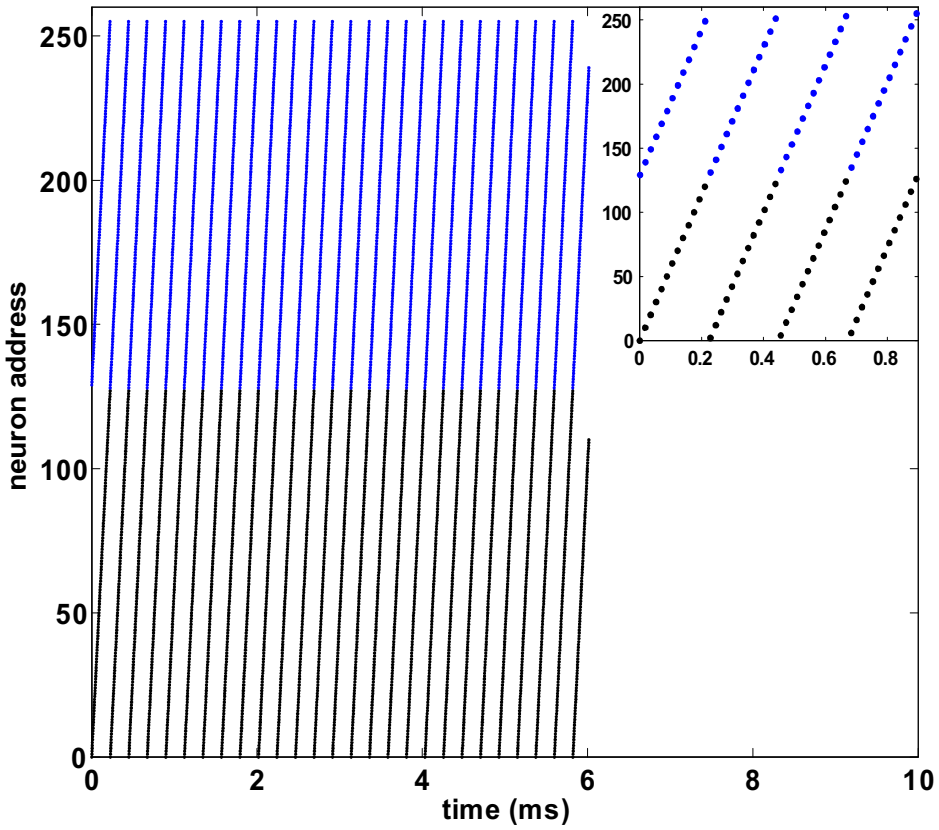

Figure 5.30: Output of the merge board. The merge board merges the address-event spikes from two (simulated) cochlea chips onto a single address bus. The output of each cochlea was simulated by programming a dsPIC30F4011 microcontroller. The black dots represent spikes from the right cochlea chip while the blue dots represent spikes from the left cochlea chip. The inset shows the same data sampled by showing every tenth spike and only the first 900μs. In this plot, the two "cochleae" put out address-event spikes at the same rate and thus the alternation between the right and left spikes on the merged address bus.

Chapter 6

Localization

6.1 Introduction

In Chapter 3, we presented an artificial bat head and measured its response to FM hyperbolic sweeps from 703 directions. In Chapter 4, we presented an electronic ultrasonic cochlea that can transform the microphone signals into a spike-based cochlear representation. In Chapter 5, we presented two acoustic feature extraction chips: a monaural spectral difference chip and a binaural ILD chip designed to extract the localization cues from the spike-based representation.

To estimate the elevation and azimuth of the sound source, we first demonstrate in *software* that the measured responses from the artificial bat head contain enough information to be used for localization. Next, we show that the *hardware* system (consisting of the cochlea chips, spectral difference chips and ILD chips) can extract the relevant localization cues as shown in the software analysis and that the spatiotemporal pattern of spiking outputs from the feature extraction

chips can be decoded to simultaneously estimate the elevation and azimuth of the sound source.

6.2 Localization Procedure

The frequency response of the artificial bat head to FM hyperbolic sweeps was measured from 703 directions. Let $\theta(i)$, $\phi(i)$, and $\phi_m(i)$ denote the elevation, azimuth, and modified azimuth (recall the definition in section 3.7) from the i^{th} direction respectively ($i = 1, 2, ...703$). The localization algorithm that has been adopted is as follows: for each of the 703 measured directions, the spectrum from both ears is transformed into a binary code word (code for short) c_i of length L where i is the index representing the i^{th} direction. Each bit in the code is represented using ± 1; this ensures that all codes have the same autocorrelation (equal to L, the length of the code); mathematically,

$$c_i c_i^T = L \quad \forall i \ (i = 1, 2, ...703) \tag{6.1}$$

where c_i is a row vector. The code c_i is the "*fingerprint*" of the sound arriving from the i^{th} direction. It is, therefore, important to investigate the relationship between the different codes. When cross-correlating codes, codes that are identical will result in a cross-correlation equal to L. For every one bit difference between the two codes the cross-correlation will decrease by two; mathematically,

$$c_i c_j^T = L - 2d_{ij} \tag{6.2}$$

where d_{ij} is the *Hamming distance* (number of different bits) between codes i and j. The matrix of codes for all 703 directions will be referred to as the code matrix C.

The code matrix C has 703 rows and L columns. Each row in the matrix corresponds to a code from a unique direction and each column in the matrix corresponds to a single bit (dimension) in the code. Several metrics will be used to evaluate the performance of the code matrix; one such metric is the number of unique codes (NUC). Ideally the number of unique codes should be equal to the number of (unique) directions; if not, there will be some ambiguity in determining the direction of sound for at least one of the codes. Another metric is the code distance. Let D be the vector representing the minimum distance between every code in the matrix and all other codes; mathematically,

$$D(i) = \min(c_i c_j^T) \qquad \forall j \neq i,\ i, j = 1, 2, ...703 \tag{6.3}$$

The relationship between the number of unique codes and D is that:

$$NUC = 703 \Longleftrightarrow D(i) \geq 1 \quad \forall i \tag{6.4}$$

i.e. for the number of unique codes to be equal to 703, the minimum distance $D(i)$ for every code word c_i must be greater than zero (or vice versa). Given the binary code word x of an unknown direction (under the assumption that x is one of the 703 possible code words), the direction of the sound source can be estimated by cross-correlating x with all 703 code words and selecting the direction that results in a cross-correlation equal to L (see equation 6.1); mathematically,

$$y(i) = c_i x^T, \quad i = 1, 2, ...703 \tag{6.5}$$

Using this notation, finding the direction of the sound source amounts to finding i such that:

$$y(i) = L. \tag{6.6}$$

and the estimated elevation θ^{est}, azimuth ϕ^{est}, and modified azimuth ϕ_m^{est} are given by:

$$\begin{aligned} \theta^{est} &= \theta(i) \\ \phi^{est} &= \phi(i) \\ \phi_m^{est} &= \phi_m(i). \end{aligned} \tag{6.7}$$

If the number of unique codes (NUC) is equal to 703 and x comes from one of the 703 locations, then the direction of the sound source can be estimated with zero error. Since the 703 code words are the rows of the code matrix C, the cross-correlation in equation 6.5 can be written as a matrix multiplication:

$$y = Cx^T \tag{6.8}$$

where x is a row vector (similar to c_i) and y is a column vector. Each entry $y(i)$ is the cross-correlation of the code word c_i from the i^{th} direction with x.

If the number of unique codes (NUC) is less than 703, then for some code word x, there will be more than one entry (i.e. direction) in y that is equal to L. In that case, the estimated direction is taken as the mean of all directions (i) which satisfy the condition $y(i) = L$.

It was assumed above that the code word x of the unknown direction was one of the 703 possible binary code words, i.e. zero bits of error. If we allow x

to have n bits of error then the estimated direction is taken as the mean of all directions (i) which satisfy the condition:

$$y(i) \geqslant L - 2n. \tag{6.9}$$

In the extreme case that $n = L$, i.e. allowing errors in all L bits, then all directions will satisfy equation 6.9 and it can be shown that the estimated direction will always be:

$$\begin{aligned} \theta^{est} &= 0 \\ \phi^{est} &= 0 \\ \phi_m^{est} &= 0 \end{aligned} \tag{6.10}$$

and that the standard deviation of error for estimating θ, ϕ, and ϕ_m, is σ_θ=41.1°, σ_ϕ=53.4°, and σ_{ϕ_m}=38.2° respectively. The reason that the maximum standard deviation for θ and ϕ_m is less than that of ϕ is that θ was limited (in our artificial bat head measurement) to ±67.5° and ϕ_m is not uniformly distributed (see Fig. 3.18). It is noted that the minimum number of bits required to uniquely encode all 703 directions is 10 bits ($\lceil \log_2 703 \rceil$).

Fig. 6.1 is an overview of the localization schemes presented in this chapter. We first demonstrate, in *software*, that the measured responses from the artificial bat head contain enough information to be used for localization. In the *threshold* code scheme, the measured responses will be converted directly into a binary code without any additional processing, whereas in the *difference* code scheme, the measured responses will first be processed to extract the binaural ILD cues and the monaural spectral cues before being converted into a binary code.

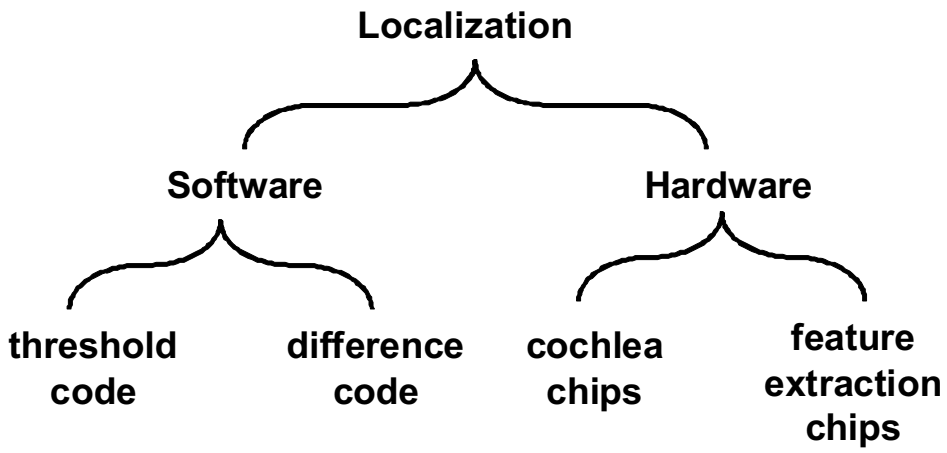

Figure 6.1: Overview of the localization schemes presented in this chapter.

We then show the equivalent processing in the *hardware* system. The cochlea chips convert the recorded microphone sounds into their spike-based cochlear representation (which is a binary output). The feature extraction chips then process the spiking output of the cochlea chips to extract the binaural and monaural localization cues. Therefore, localization using the output of the cochlea chips resembles the software-based threshold code, whereas localization using the output of the feature extraction chips resembles that of the software-based difference code.

6.3 Localization in Software

In this section we demonstrate in software (MATLAB^{TM}) that the measured responses from the artificial bat head can be used for localization. We present two different codes: a threshold code and a difference code. In the threshold code,

the binary code c_i for each direction is computed by simply comparing the magnitude spectrum of the sound at each frequency bin against a threshold. In the difference code, the binary code is computed by comparing binaural and monaural differences in the magnitude spectrum against a threshold.

6.3.1 Threshold Code (L=1312)

The threshold code is inspired by the fact that some auditory neurons (auditory nerve fibers) have thresholds below which they do not fire any action potentials (spikes). The threshold code assumes that for every frequency bin there is an artificial neuron (with a threshold) that is responsive only to that frequency bin. Fig. 6.2 illustrates how the threshold code is obtained; for each of the 703 directions the magnitude spectrum at every frequency bin in the FFT (from both ears) is compared against a threshold. The code bit corresponding to that frequency bin is either set to 1 (i.e. spike fired) if the magnitude spectrum is greater than the threshold or to -1 (i.e. no spike) otherwise. The length (L) of the code is 1312 bits (656 bits from the right spectrum and 656 bits from the left spectrum), each spanning the frequency range from $20kHz$-$100kHz$. The spacing between the adjacent frequency bins is $122Hz$.

Fig. 6.3 shows the binary receptive fields of four dimensions (columns) of the code matrix C. The left column shows the magnitude spectrum at four different frequencies, while the right column shows the corresponding receptive field obtained by setting a threshold of 5dB. The black curve is the decision

boundary separating the directions for which the code bit =+1 (green region) from the directions for which the code bit =−1 (white region). Notice that the notch seen at $70kHz$ and $90kHz$ splits the +1 region into two non-connected regions.

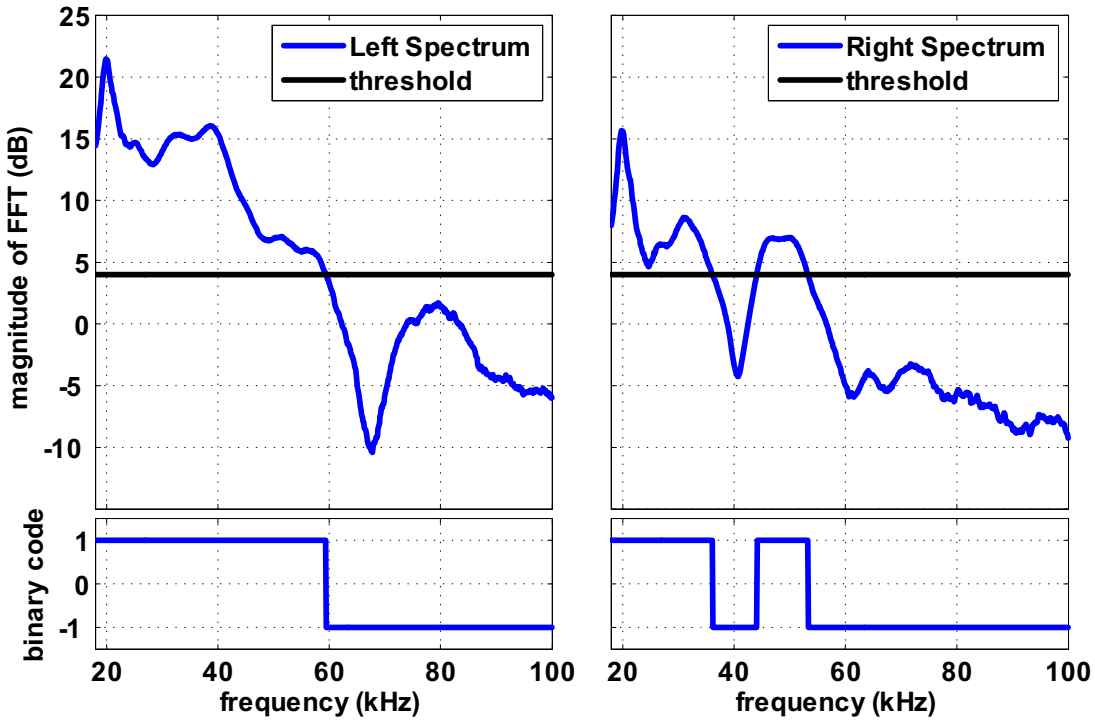

Figure 6.2: Constructing the threshold code. The magnitude spectrum at every frequency bin is compared against a threshold, the code bit corresponding to that frequency bin is set to 1 if the magnitude spectrum is greater than the threshold and is set to −1 otherwise.

To investigate how the threshold level affects the code matrix, the threshold was changed from −20dB to 30dB in steps of 1dB. For every threshold value a code matrix was computed. Fig. 6.4 (a) shows the number of unique codes

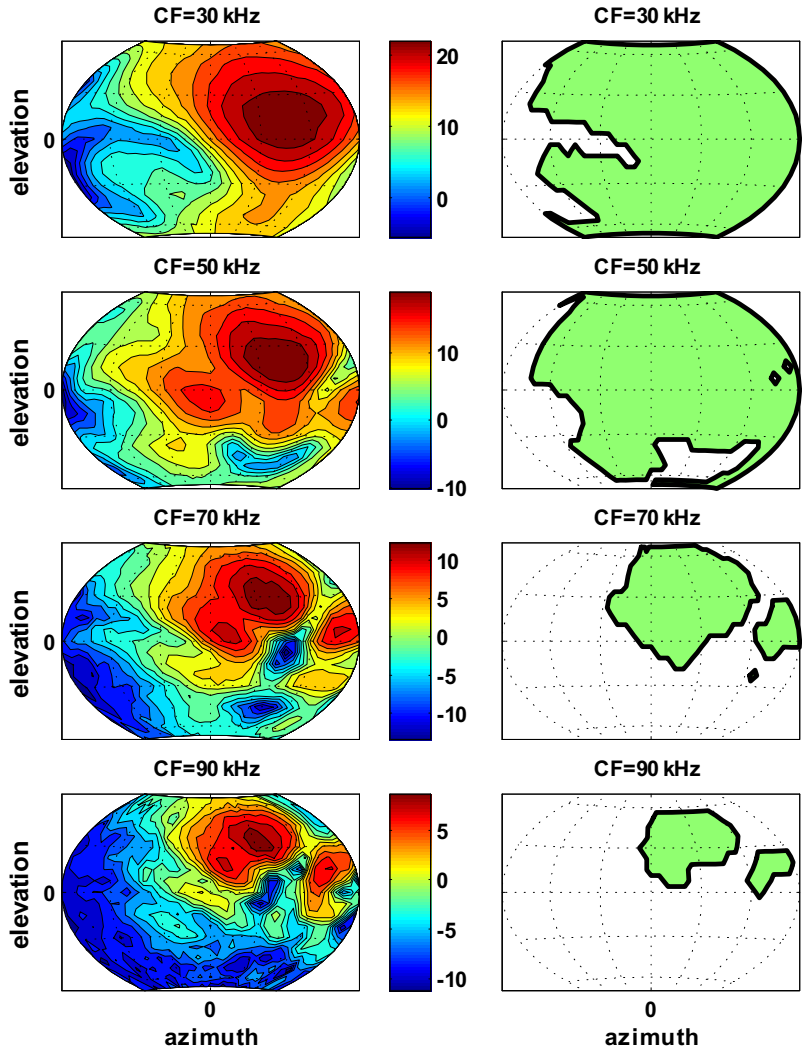

Figure 6.3: Binary receptive fields of four dimensions (columns). The left column shows the magnitude spectrum at four different frequencies (listed at the top of each subplot) while the right column shows the corresponding binary receptive field obtained by setting a threshold of 5dB. The black curve is the decision boundary separating the directions for which the code bit $=1$ (green region) from the directions for which the code bit $=-1$ (white region).

(NUC) for the code matrix C as a function of threshold. If the threshold is too low, all bits are in C are equal to 1 and there is only one unique code. If the threshold is too high, all bits in C are equal to -1 and, again, there is only one unique code. Fig. 6.4 (b) shows the activity ($\mathbf{P}$) of the code matrix C (defined as the percentage of C that is equal to 1) as a function of threshold. Fig. 6.4 (c) shows the median and mean of the code distance vector D for a particular code matrix C as a function of threshold, as well as the standard deviation of error in estimating the elevation θ and the modified azimuth ϕ_m. Observe that at the threshold extremes, the number of unique codes (NUC) is equal to 1 which results in σ_θ=41.1°, and σ_{ϕ_m}=38.2° as discussed in the previous section. Also notice that as NUC approaches 703, the error drops to zero.

This shows that the measured responses from the artificial bat head contain enough information to be used in localization and that given a code word x of an unknown direction, the direction of the sound source can be estimated with zero error (with the assumption that x comes from one of the 703 measured directions as described in the previous section).

Fig. 6.5 shows the correlation between the code matrices calculated at the different thresholds. The black contour lines enclose the region where the correlation is 90% or higher. It is observed that only for a small range of thresholds does the code matrix not change significantly.

We have chosen the code matrix calculated at a threshold of 5dB to further examine the properties of the code matrix . We will refer to this code matrix as C_{5dB}. It has the greatest mean distance (31.4) and the number of unique codes

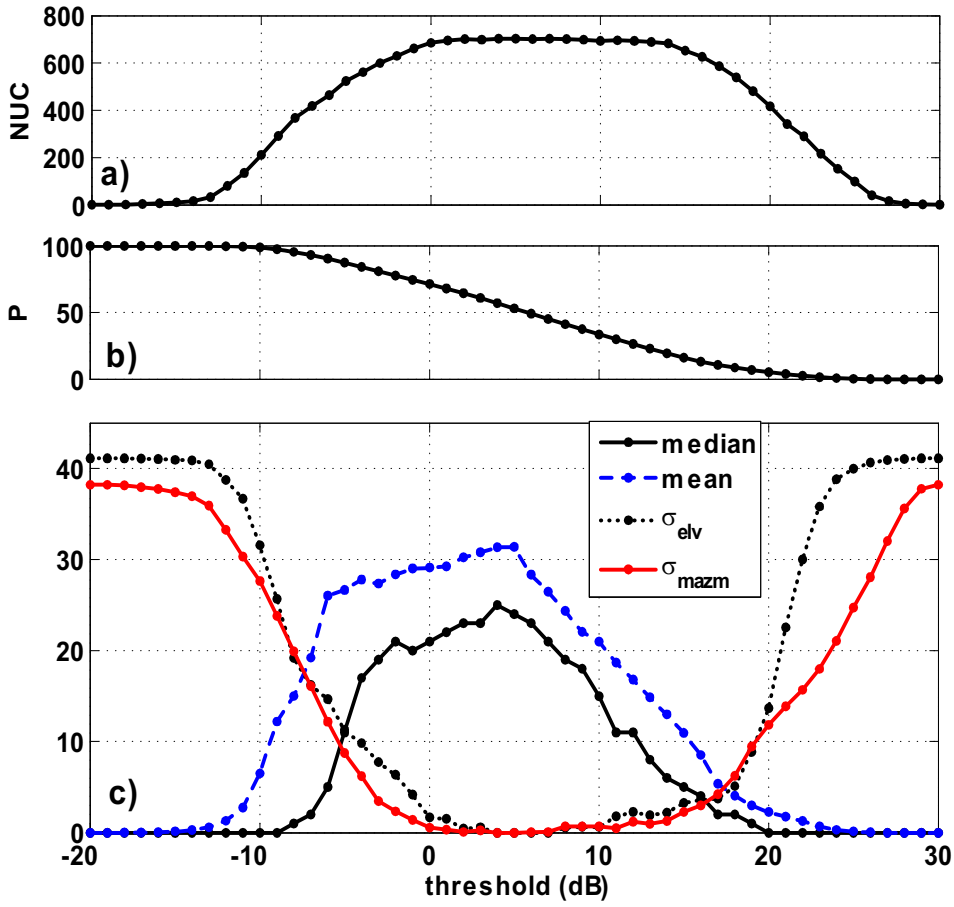

Figure 6.4: Threshold Code ($L = 1312$ bits). a) Number of unique codes (NUC) as a function of threshold. b) Activity (**P**) of the code matrix as a function of threshold. c) Median and mean of the code distance vector D as a function of threshold as well as the standard deviation of error in estimating the elevation σ_{elv} and the modified azimuth ($\sigma_{mazm} \equiv \sigma_{\phi_m}$) assuming no bit errors.

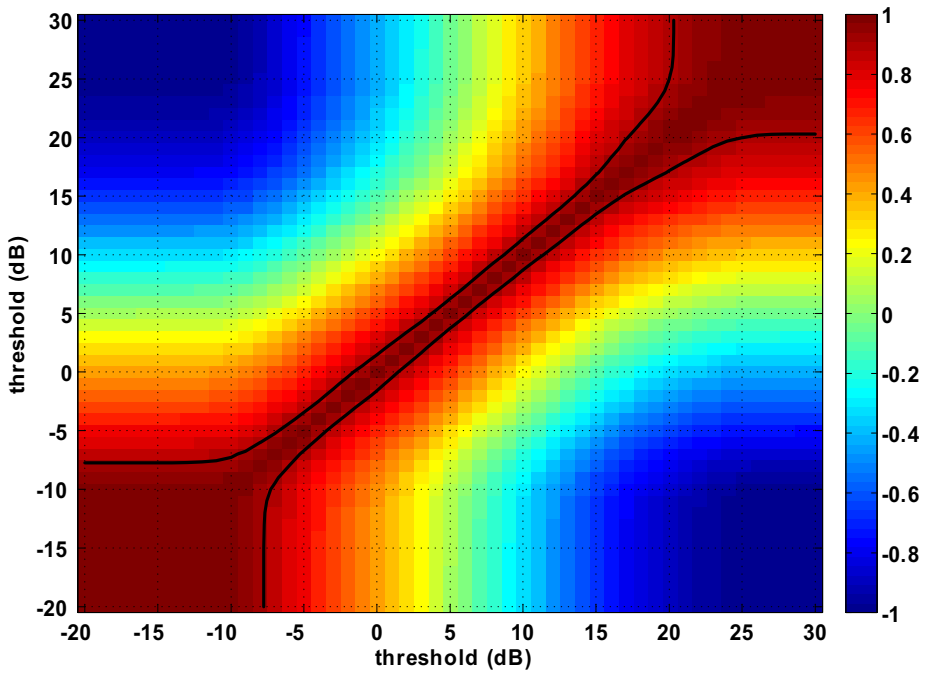

Figure 6.5: Correlation between the code matrices calculated at different thresholds. The black contour lines enclose the region where the correlation is 90% or higher. The narrowband of high correlation along the diagonal shows that over most of its range, changes in the threshold level produce significant changes in the code matrix.

(NUC) for this matrix is 703. This code matrix has 1312 columns, each of which corresponds to a single bit (dimension) in the code. Given that the frequency spacing between adjacent bits is only $122Hz$, one might question how closely correlated the 1312 columns (dimensions) of C_{5dB} are? Fig. 6.6 shows the cross-correlation of the columns of C_{5dB}. The black contour lines enclose the region where the correlation is 90% or higher. Columns 1-656 represent the portion of the code derived from the left spectrum while columns 657-1312 represent the portion of the code derived from the right spectrum.

As expected, because of the $122Hz$ spacing between the adjacent frequency bins, the adjacent columns of C_{5dB} are highly correlated and thus provide similar information. It is also observed that there is a weak correlation between matching frequencies of the two cochlea. It is notable that the low frequencies are more correlated between the two cochlea.

Consider the two code matrices $C1$ and $C_{sampled}$ where:

$$\begin{aligned} C1 &= [a\ a\ a\ b\ b\ b\ c\ c\ c] \\ C_{sampled} &= [a\ b\ c] \end{aligned} \tag{6.11}$$

where a, b, and c are column vectors. It is clear that the code matrix $C1$ is a mere repetition of the smaller code matrix $C_{sampled}$. If the matrix $C_{sampled}$ has a length L_s and a mean distance D_s, then the code matrix $C1$ has a length $3L_s$ and mean distance $3D_s$. The code matrix $C_{sampled}$ can be obtained from the larger matrix $C1$ by picking every third column (i.e. sampling by three). Similarly, by sampling the 1312 columns of C_{5dB}, i.e. by selecting every n^{th} column (where n

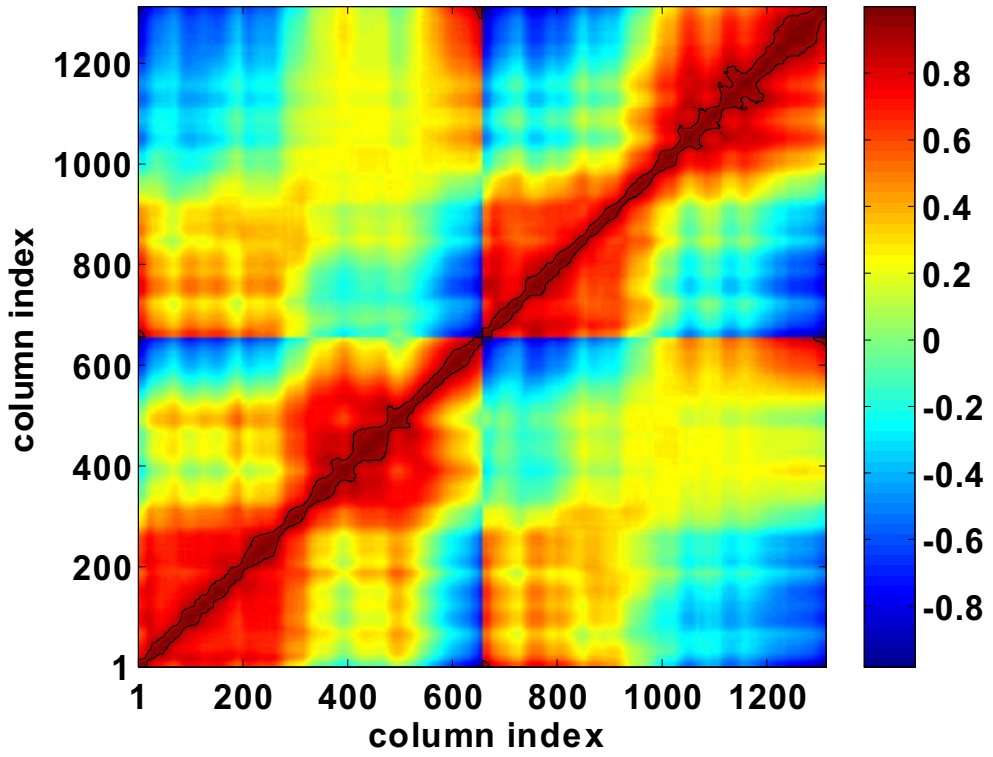

Figure 6.6: Cross-correlation of the 1312 columns of the code matrix C_{5dB}. Each column in the matrix corresponds to a single bit (dimension) in the code. Columns 1-656 represent the portion of the code derived from the left spectrum while columns 657-1312 represent the portion of the code derived from the right spectrum. The black contour lines enclose the region where the correlation is 90% or higher. Because of the $122Hz$ separation between the adjacent frequency bins, the adjacent columns are highly correlated.

is the sampling period), a smaller code matrix can be generated. The effect of sampling on the mean distance of the new code matrices is shown in Fig. 6.7 where $n = 1$ corresponds to the original code matrix C_{5dB} with a mean code distance of 31.4. For example, when $n = 2$ (i.e. selecting every other column), the size of new code matrix is half that of C_{5dB} (new code length=$L/2$) and the mean code distance is 15.6. When $n = 10$, the size of the new code matrix is 1/10 that of C_{5dB} (new code length=$L/10$), and the mean code distance is 2.9. The observation that the mean of the sampled code matrix can be predicted by the formula mean(1)/n, where mean(1)=mean of C_{5dB}, suggests that, similar to $C1$, the original code matrix C_{5dB} is, to a large extent, composed of a repetition of a smaller code matrix. For a given value of n, there are n different sampled code matrices that can be constructed (since there are n possible starting points from which to begin sampling). By computing and comparing the n different code matrices it was found that for $n \leq 14$, the correlation between matrices of the same code length was $> 90\%$.

The previous analysis suggests that for a fair comparison between different localization schemes, the compared code matrices should have the same (or at least comparable) code length. A software program was written to process the code matrix C_{5dB} and remove dimensions (columns) such that the dimensions of the new (reduced) code matrix $C_{reduced}$ do not have a cross-correlation that exceeds a threshold $xcorr$. Table 6.1 shows the effect of limiting the maximum cross-correlation of the dimensions of $C_{reduced}$. The columns in the table represent (in order) maximum allowable cross-correlation $xcorr$, length of the reduced code

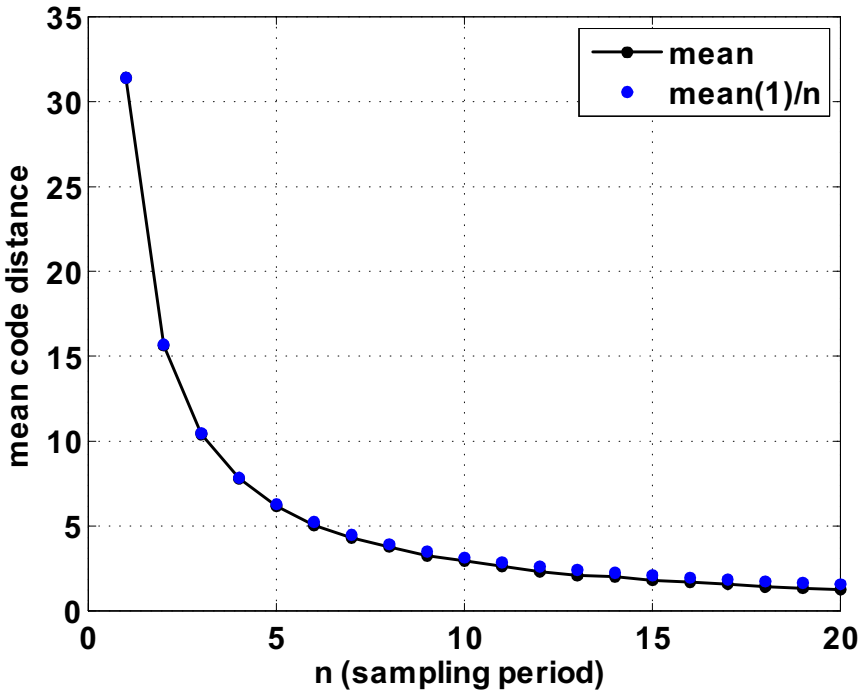

Figure 6.7: The effect of sampling the 1312 columns of the code matrix C_{5dB} on the mean of the new code matrix. The black curve is defined only at the integer points. mean(1) is the mean distance of the original code matrix C_{5dB} where L=1312.

(L), activity (P), number of unique codes (NUC), mean distance, median distance, minimum distance, and maximum distance of the reduced code matrix $C_{reduced}$. The last two columns, σ_θ, and σ_{ϕ_m}, are the standard deviation of error (in degrees) in estimating the elevation (θ) and the modified azimuth (ϕ_m) of the sound source respectively (assuming no errors in the code bits).

The first row corresponds to the original code matrix C_{5dB}; however, five of the 1312 dimensions had a fixed value of -1 and were, therefore, removed (thus L=1307). Because the number of unique codes (NUC) in the first row is 703, there is no error in estimating the elevation θ or the modified azimuth ϕ_m of the sound source; however, for the rest of the entries in the table, the number of unique code (NUC) is less than 703, and as NUC decreases, the error in estimating θ and ϕ_m increases.

The table illustrates that just reducing the maximum allowable cross correlation to 99% resulted in the number of unique codes (NUC) to drop below 703 which means that there are at least two directions that have the same code.

Consider the code matrix $C_{reduced}$ that has no cross correlation greater than 95% (6^{th} row in table 6.1). The length (L) of the codes of this matrix is 145 and the NUC is 664. We will refer to this matrix as $C^{145}_{reduced}$. Fig. 6.8 shows the correlation between the columns of $C^{145}_{reduced}$, which resembles that of C_{5dB} in Fig. 6.6. Fig. 6.9 shows the binary receptive fields of the matrix $C^{145}_{reduced}$ (only the first 144 dimensions). The black regions indicate the directions for which the code is $+1$, while the white regions indicate the directions for which the code $=-1$.

Fig. 6.10 shows the summation of all 145 receptive fields. The maximum

Table 6.1: Reducing cross-correlation in the code matrix C_{5dB}.

$xcorr$	L	P	NUC	mean(D)	median(D)	min(D)	max(D)	σ_θ	σ_{ϕ_m}
100	1307	53	703	31	24	1	162	0	0
99	578	56	698	15	11	0	73	0.7	0.2
98	316	57	693	8	6	0	42	1.2	0.2
97	235	57	691	6	5	0	32	1.6	0.4
96	176	57	681	4	3	0	24	2.0	0.7
95	145	58	664	4	3	0	20	2.3	1.0
90	70	58	568	1	1	0	10	5.4	1.9
85	46	58	479	1	1	0	5	7.1	2.8
80	30	59	330	0	0	0	4	9.8	3.8
75	22	58	233	0	0	0	2	16.2	5.8
70	18	56	161	0	0	0	2	17.2	6.6
65	14	56	113	0	0	0	1	17.0	7.8
60	12	53	80	0	0	0	1	18.5	9.5
55	9	52	43	0	0	0	1	22.9	12.6
50	8	51	26	0	0	0	1	24.6	15.9

and minimum possible value are 145 and -145 respectively. Regions where the summation is equal to zero indicate equal "time" spent on either side (+1 and −1) of the decision boundary, while regions where the summation is close to 145 (or −145) indicate that the directions are almost always on the same side of the decision boundary. Fig. 6.11 shows the distance vector D for $C^{145}_{reduced}$ as a function of azimuth and elevation. Notice the similarity between the pattern of directions that have a large distance and those that have a summation close to zero.

Shown in Fig. 6.12 is the spatial correlation plot for 16 different directions (listed at the top of each subplot where azm=azimuth and elv=elevation). The spatial correlation is calculated by cross-correlating a code from a single direction with codes from all (703) directions (see equation 6.8). The maximum correlation is L for directions with identical codes and decreases by two for every one bit difference. The codes were selected from the code matrix $C^{145}_{reduced}$. The colorbar shows the different contour levels that were plotted. The highest contour level, 139, shows all the directions whose codes are within three bits of error to the code of the selected direction, while the smallest contour level, 85, shows all the directions whose codes are within 30 bits of error to the code of the selected direction. The white regions show the directions whose codes have a distance greater than 30 bits to the code of the selected direction.

Fig. 6.13 shows the true and estimated directions for the code matrix $C^{145}_{reduced}$ assuming no bit errors. The standard deviation of error in estimating elevation is σ_θ=2.3° and the standard deviation of error in estimating the modified azimuth is σ_{ϕ_m}=1°.

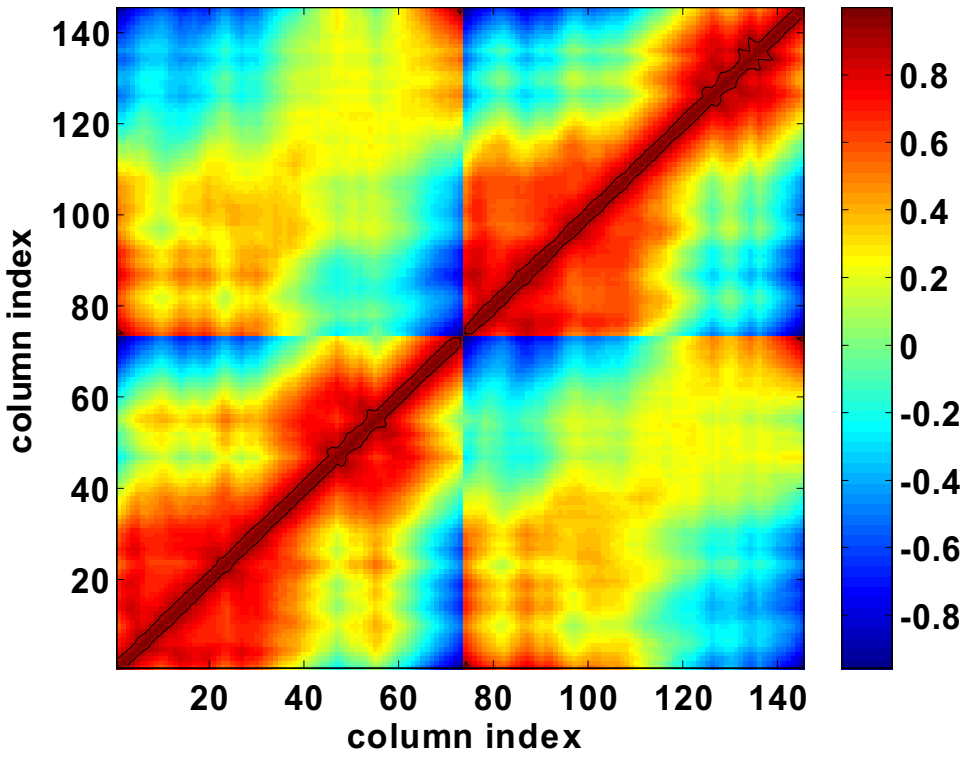

Figure 6.8: Cross-correlation of the 145 columns of the threshold code matrix $C^{145}_{reduced}$. Each column in the matrix corresponds to a single dimension in the code. The black contour lines enclose the region where the correlation is 90% or higher.

Fig. 6.14 shows the true and estimated directions for the code matrix $C^{145}_{reduced}$ when eight bits of error is randomly imposed on the code. The standard deviation of error in estimating elevation is σ_θ=13.5° and the standard deviation of error in estimating the modified azimuth is σ_{ϕ_m}=5.2°. The localization performance of the code matrices introduced in the rest of this chapter will be compared against that of the code matrix $C^{145}_{reduced}$ (as well as the rest of the entries of Table 6.1).

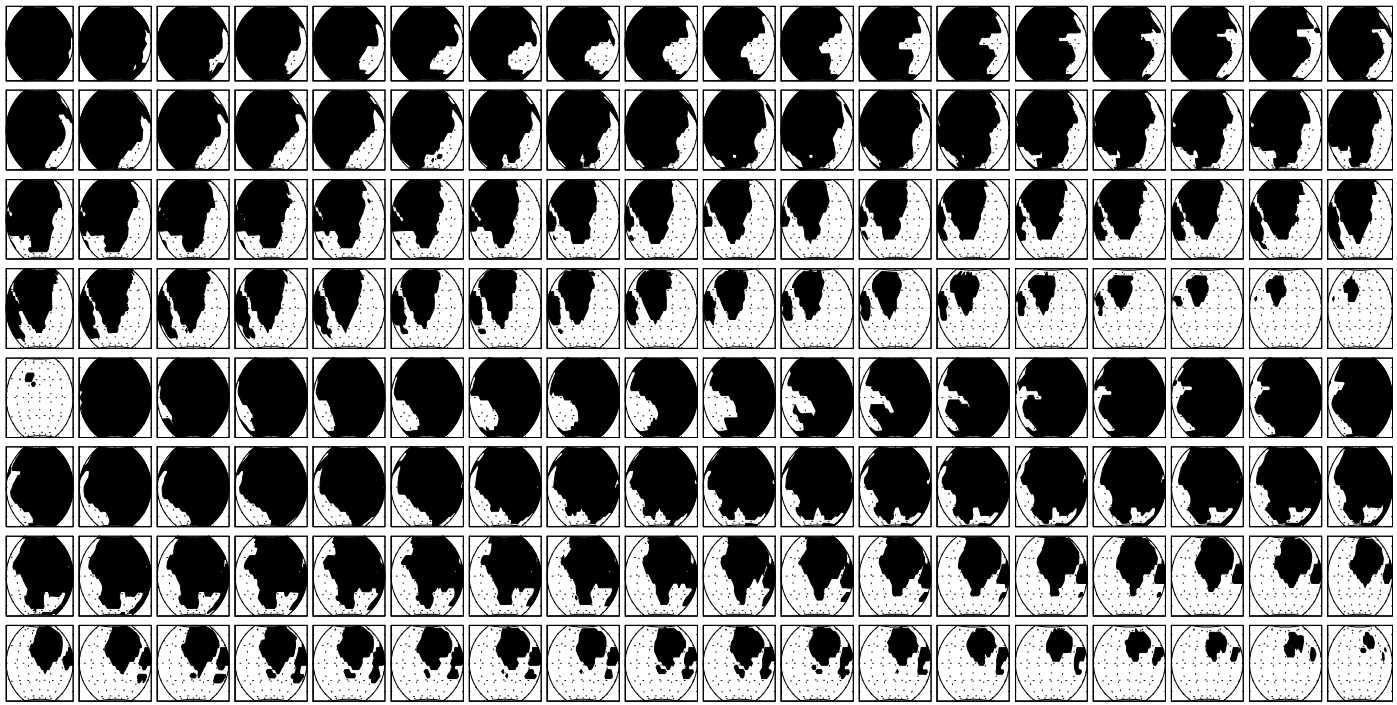

Figure 6.9: Binary receptive fields of the (first 144) dimensions of the threshold code matrix $C_{reduced}$ (L=145). The black regions indicate the directions where the code is equal to +1 while the white regions indicate the directions where the code is equal to -1.

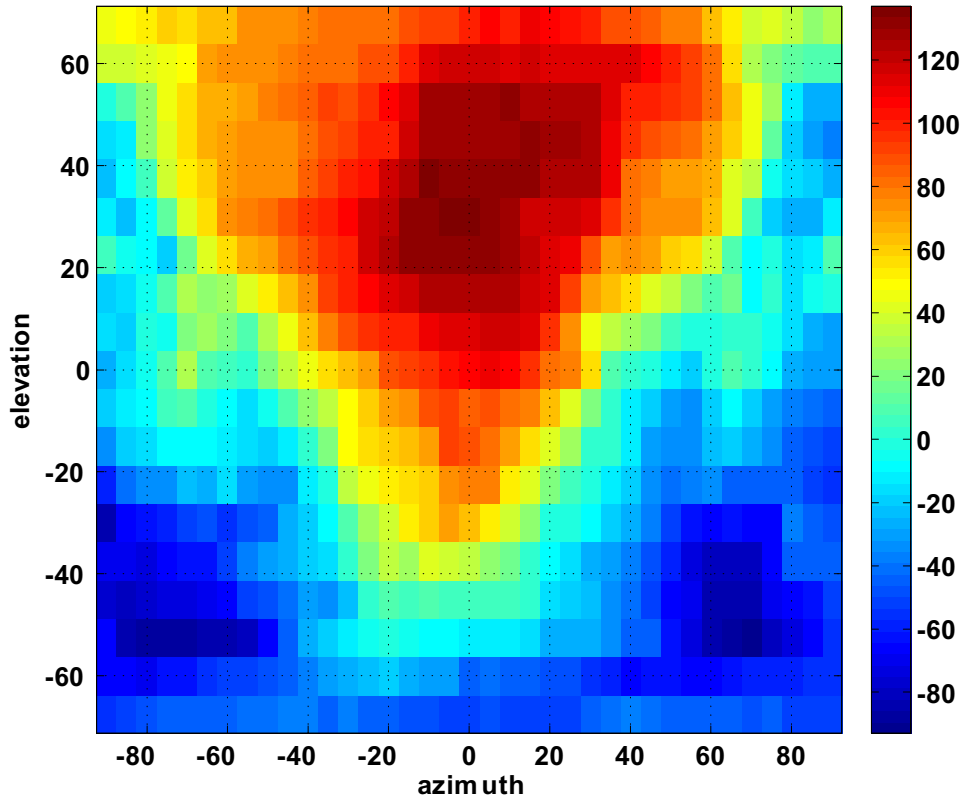

Figure 6.10: Summation of all 145 binary receptive fields of the threshold code matrix $C^{145}_{reduced}$. The maximum possible value is 145 and the minimum possible value is -145. Where the the summation is equal to zero indicates equal "time" spent on either side ($+1$ and -1) of the decision boundaries, while regions where the summation is close to 145 (or -145) indicate that the directions are almost always on the same side of the decision boundaries.

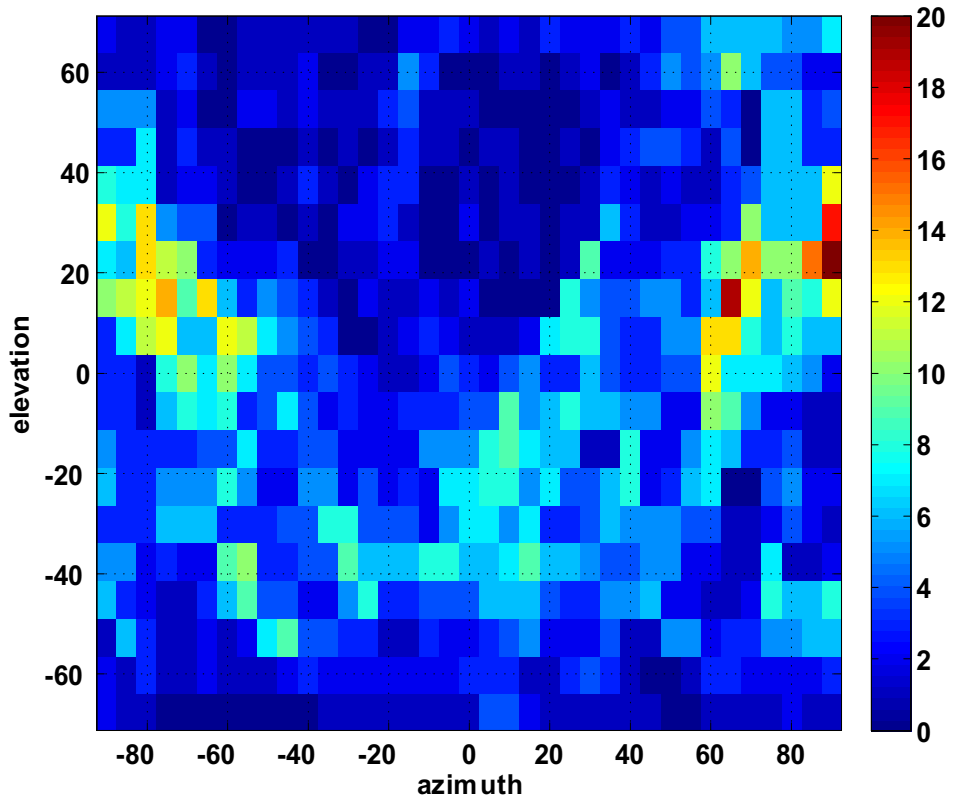

Figure 6.11: Distance vector (D) for the threshold code matrix $C^{145}_{reduced}$ as a function of azimuth and elevation. Notice the similarity between the pattern of directions that have a large distance and those that have a summation close to zero.

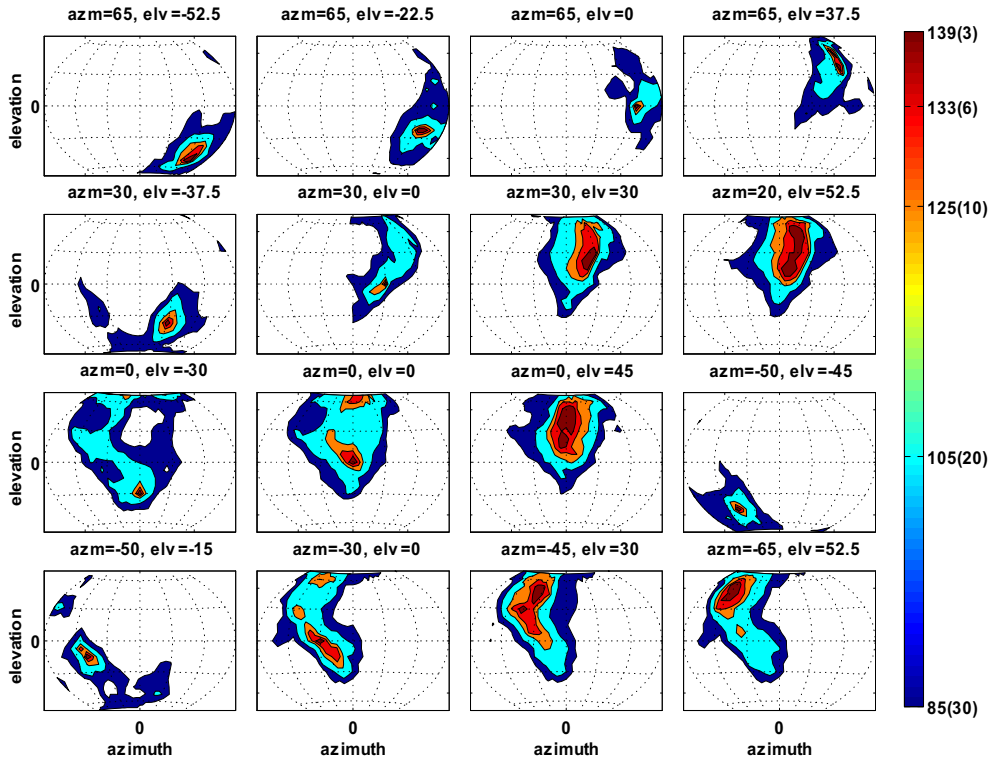

Figure 6.12: Spatial correlation plots for the threshold code matrix $C^{145}_{reduced}$. Shown is the spatial correlation plot for 16 different directions (listed at the top of each subplot where azm=azimuth and elv=elevation). The spatial correlation plot is calculated by cross-correlating a code from a single direction with codes from all 703 directions. The maximum correlation is equal to L (L=145) and decreases by two for every one bit difference. The codes were selected from the threshold code matrix $C^{145}_{reduced}$. The smallest contour level, 85, shows all the directions that are within 30 bits of error to the selected direction, while the highest contour level, 139, shows all the directions that are within three bits of error to the selected direction.

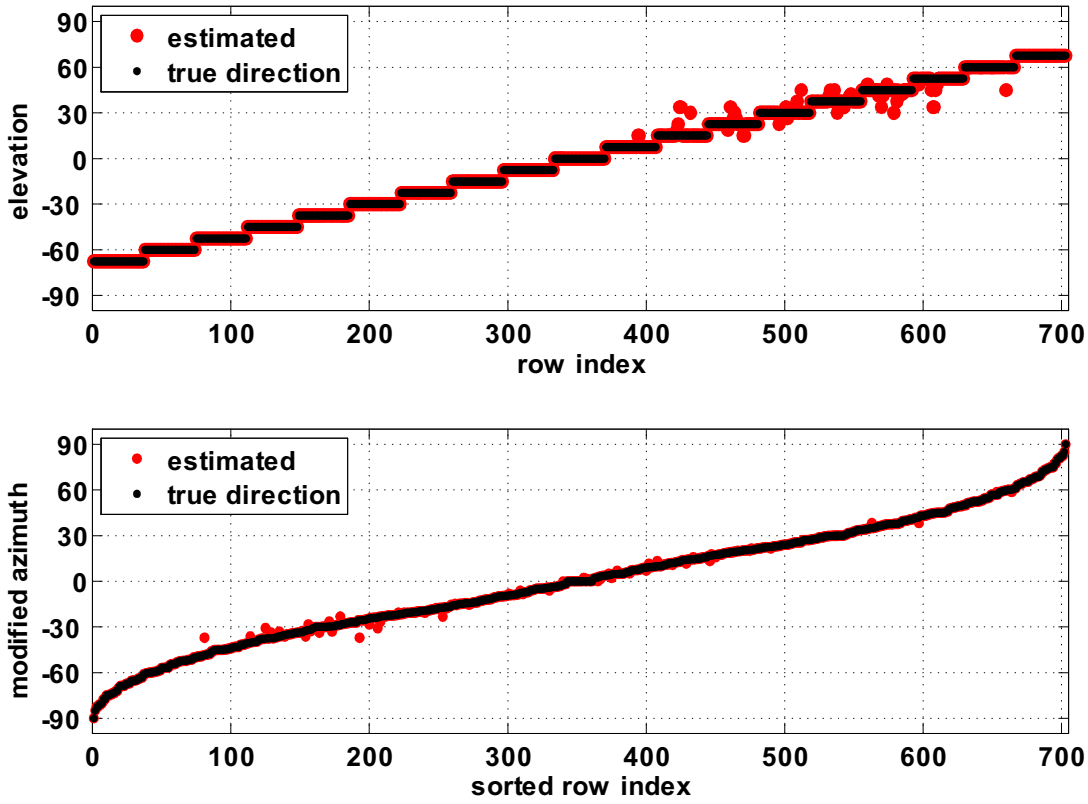

Figure 6.13: True and estimated directions for the threshold code matrix $C^{145}_{reduced}$ assuming no bit errors. The standard deviation of error in estimating elevation is 2.3° and the standard deviation of error in estimating the modified azimuth is 1°.

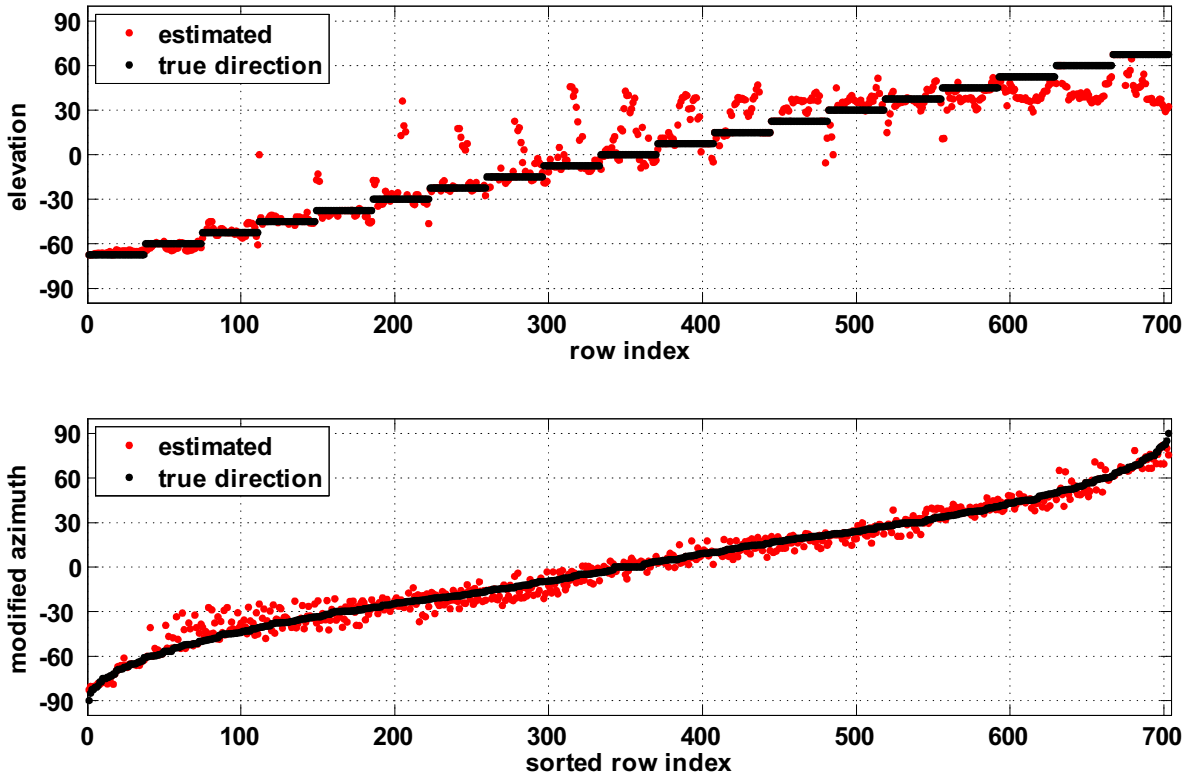

Figure 6.14: True and estimated directions for the threshold code matrix $C^{145}_{reduced}$ when eight bits of error is randomly imposed on the code. The standard deviation of error in estimating elevation is 13.5° and the standard deviation of error in estimating the modified azimuth is 5.2°.

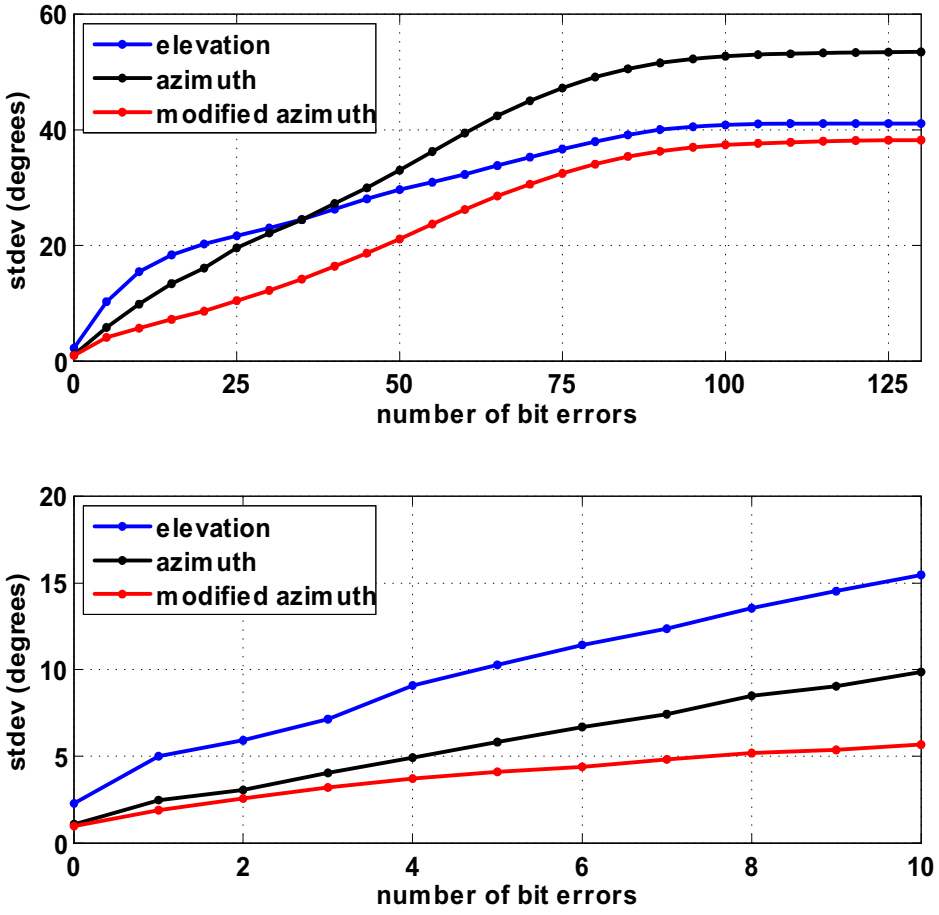

Figure 6.15: Standard deviation of error (in degrees) in estimating elevation, azimuth, and modified azimuth as a function of the number of bit errors for the threshold code matrix $C^{145}_{reduced}$. The lower plot is similar to the upper plot but focuses on the initial part of the curve.

6.3.2 Threshold Code (L=32)

For a direct comparison with the output of the two cochlea chips (which have 16 cochlear filters each), we generate a threshold code that compares the magnitude spectrum at 16 frequency bins (in each ear). First, the frequency axis is sampled at 16 frequency channels (bins), however, rather than uniformly sampling the frequency axis, the frequency bins will be chosen according to the formula

$$\mathrm{CF}(i) = 20kHz * 2^{\alpha(i-1)} \qquad i = 1, 2, ...16 \tag{6.12}$$

where $\alpha = 0.1548$ and i is the channel index. This formula generates frequencies that are exponentially-spaced (linearly spaced on the logarithmic axis) similar to that of the mammalian cochlea as described in Chapter 2. The channel index $i = 1$ corresponds to the lowest frequency channel ($20kHz$) and $i = 16$ corresponds to the highest frequency channel ($90kHz$). The frequencies generated by this formula are approximately:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
CF(kHz)	20	22	24	27	30	33	36	40	45	49	55	60	67	74	82	90

The codes have a length $L = 32$ (16 frequency bins x 2 cochlea). The code matrix for this 32 bit threshold code will be referred to as C_{th}^{32}. Figs. 6.16a and 6.16b show respectively the number of unique codes (NUC) and the activity (**P**) for the code matrix C_{th}^{32} as a function of threshold. Fig. 6.16c shows the median and mean of the code distance vector D for a particular code matrix C_{th}^{32} as a function of threshold as well as the standard deviation of error in estimating the

Table 6.2: Localization performance of the 32-bit threshold code (threshold=7dB).

$xcorr$	L	P	NUC	mean(D)	median(D)	min(D)	max(D)	σ_θ	σ_{ϕ_m}
100	32	60	325	0	0	0	3	6.5	3.9

elevation θ and the modified azimuth ϕ_m. It is observed that the maximum value of NUC is 336 (< 703) and that σ_θ and σ_{ϕ_m} do not drop to zero. The minimum values for σ_θ and σ_{ϕ_m} are 6.4° and 3.7°, respectively (and occur at a threshold of 8dB and 6dB respectively). The localization performance of the 32-bit threshold code, for a threshold of 7dB, is summarized in Table 6.2. This will be compared against the localization performance of the hardware system, namely, localization using the spiking output of the two cochlea chips.

6.3.3 Difference Code

One of the drawbacks of the threshold code is that it is inherently sensitive to the choice of the threshold level, or equivalently, to changes in the absolute intensity of the received signal. A code that relies on the relative differences in intensity rather than the absolute value has the potential of being more robust to changes in intensity. The are two types of differences: *binaural differences* where the spectrum from one ear is compared against the spectrum from the other ear, and *monaural differences* where comparisons are made between different frequencies within the same ear. The difference code presented in this section is also a binary code and consists of two parts: a binaural difference code which is the interaural

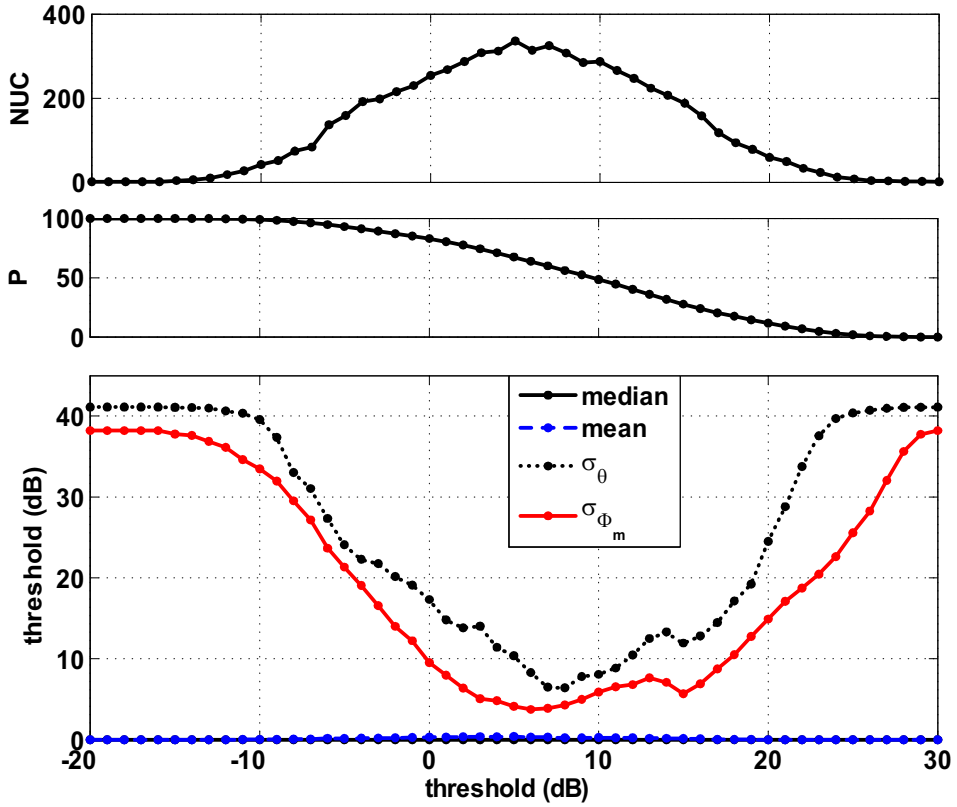

Figure 6.16: Threshold Code (L=32 bits). a) Number of unique codes (NUC) as a function of threshold. b) Activity (**P**) of the threshold code matrix C_{th}^{32} as a function of threshold. c) Median and mean of the code distance vector D as a function of threshold as well as the standard deviation of error in estimating the elevation (σ_θ) and the modified azimuth (σ_{Φ_m}) of the sound source assuming no bit errors.

level difference (ILD), and a monaural difference code, which is the monaural spectral difference (cue). This code is similar to that which will be implemented in hardware. In this code we will use the same 16 frequency channels (bins) given by equation 6.12.

Binaural Difference Code

Let $\mathrm{HR}_k(i)$ and $\mathrm{HL}_k(i)$ be the magnitude spectrum (in dB) of the right and left ear respectively at the i^{th} center frequency and from the k^{th} direction (more accurate is $\mathrm{HR}_k(\mathrm{CF}(i))$ and $\mathrm{HL}_k(\mathrm{CF}(i))$, but to simplify the notation, CF has been dropped).

First, for each of the 16 frequency channels, the binaural interaural level difference (ILD) is computed by subtracting the left spectrum at the i^{th} frequency from the right spectrum at the *same frequency*; mathematically,

$$ILD_{RL}(k,i) = \mathrm{HR}_k(i) - \mathrm{HL}_k(i), \quad i = 1,2,...16, \quad k = 1,2,...703 \tag{6.13}$$

where the subscript RL denotes the order of the computation (Right minus Left). Similarly, the ILD can also be computed by subtracting the right spectrum from the left spectrum:

$$ILD_{LR}(k,i) = \mathrm{HL}_k(i) - \mathrm{HR}_k(i), \quad i = 1,2...16, \quad k = 1,2,...703 \tag{6.14}$$

where the subscript LR denotes (Left minus Right). The equations show that the ILD does not depend on the absolute value of HR and HL, but rather on the difference between the two.

The binaural part of the difference code matrix is then computed by comparing both ILD_{RL} and ILD_{LR} against an ILD threshold (ILD_{th}) for all 16

frequencies and for all 703 directions, mathematically,

$$C_{RL}(k,i) = \begin{Bmatrix} +1 & ILD_{RL}(k,i) \geqslant ILD_{th} \\ -1 & ILD_{RL}(k,i) < ILD_{th} \end{Bmatrix} \quad i = 1,2,...16,\ k = 1,2...703$$

and

$$C_{LR}(k,i) = \begin{Bmatrix} +1 & ILD_{LR}(k,i) \geqslant ILD_{th} \\ -1 & ILD_{LR}(k,i) < ILD_{th} \end{Bmatrix} \quad i = 1,2,...16,\ k = 1,2...703 \tag{6.15}$$

where k is the direction index and C_{RL} and C_{LR} are the two binary ILD matrices (each has the dimensions 703 by 16). Although C_{LR} might appear redundant at first (since $ILD_{LR}(k,i) = -ILD_{RL}(k,i)$), for a non-zero threshold, it is not. The difference $\mathrm{HR}_k(i)$-$\mathrm{HL}_k(i)$ has three possible outcomes:

$$\begin{aligned} \mathrm{HR}_k(i) &> \mathrm{HL}_k(i) \Longrightarrow C_{RL}(k,i) = +1 \textit{ and } C_{LR}(k,i) = -1 \\ \mathrm{HR}_k(i) &< \mathrm{HL}_k(i) \Longrightarrow C_{RL}(k,i) = -1 \textit{ and } C_{LR}(k,i) = +1 \\ \mathrm{HR}_k(i) &\approx \mathrm{HL}_k(i) \Longrightarrow C_{RL}(k,i) = -1 \textit{ and } C_{LR}(k,i) = -1 \end{aligned} \tag{6.16}$$

The first outcome, $\mathrm{HR}_k(i)$>$\mathrm{HL}_k(i)$, can, in general, be expected to occur if the sound source is located in the right hemisphere, while the second outcome, $\mathrm{HR}_k(i)$ <$\mathrm{HL}_k(i)$, can be expected to occur if the sound source is located in the left hemisphere. The third outcome, $\mathrm{HR}_k(i) \approx \mathrm{HL}_k(i)$, can be expected to occur if the sound source is located near the plane of symmetry (mid-sagittal plane). Therefore, the combination of C_{RL} and C_{LR} divides the entire space into three regions. This is demonstrated in Fig. 6.17, which shows the ILD binary receptive fields computed at two different center frequencies (CF= $20kHz$ for the upper row and $40kHz$ for the lower row). The middle column shows the ILD in dB

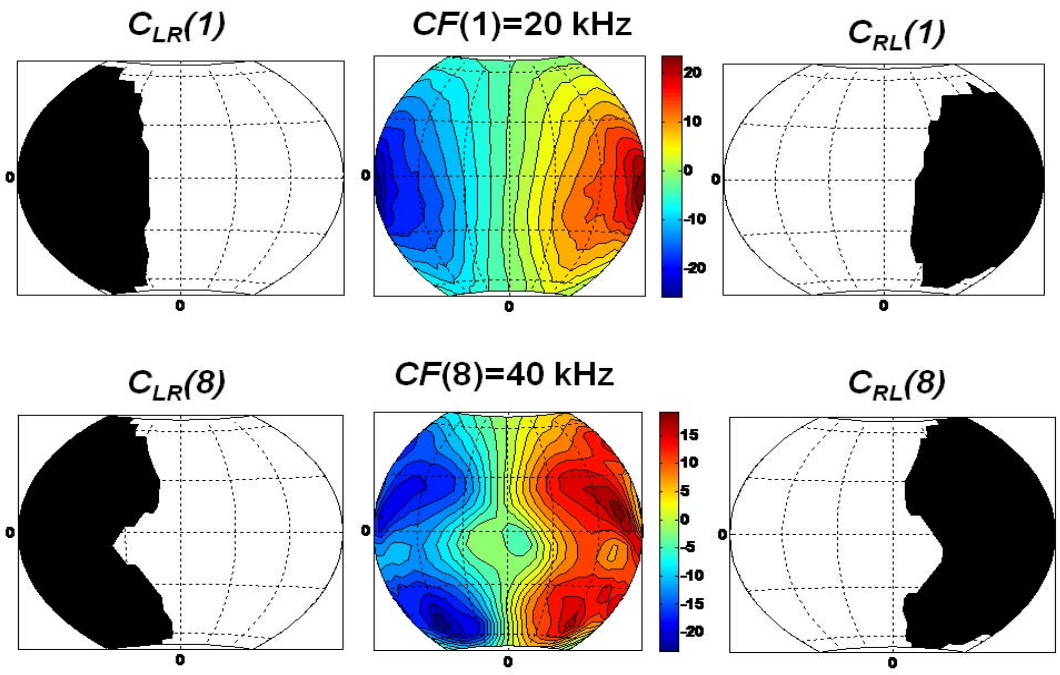

Figure 6.17: ILD binary receptive fields computed at two different center frequencies (CF=$20kHz$ for the upper row and $40kHz$ for the lower row). The middle column shows the ILD in dB computed at the two frequencies while the right (left) column shows the corresponding binary receptive field C_{RL} (C_{LR}) computed using an ILD threshold (ILD_{th}) of 5dB. The black regions indicate the directions where the code is equal to +1 while the white regions indicate the directions where the code is equal to -1.

computed at the two center frequencies while the right (left) column shows the corresponding binary receptive field C_{RL} (C_{LR}) computed using an ILD threshold (ILD_{th}) of 5dB.

Fig. 6.18 shows the 32 binary ILD receptive fields computed from ILD_{RL} and ILD_{LR}. The top two rows correspond to the 16 ILD receptive fields computed

from ILD_{RL}, while the bottom two rows correspond to the 16 ILD receptive fields computed from ILD_{LR}.

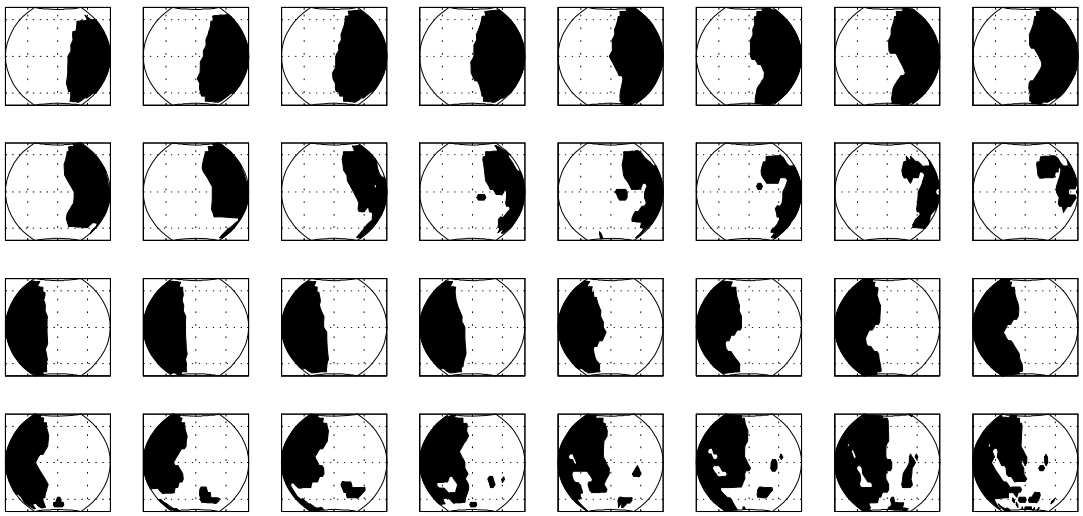

Figure 6.18: ILD binary receptive fields. The top two rows correspond to the 16 ILD binary receptive fields C_{RL} derived from ILD_{RL}, while the bottom two rows correspond to the 16 ILD binary receptive fields C_{LR} derived from ILD_{LR}. The black regions indicate the directions where the code is equal to +1 while the white regions indicate the directions where the code is equal to -1.

Monaural Difference Code

The monaural spectral difference (SD) is computed by comparing the spectrum at different frequencies but within the same ear. The frequency axis has been sampled at 16 frequency channels. For the monaural difference part of our code we will compare the spectrum at one frequency channel with the other 15 fre-

quency channels. There are 240 ($^{16}\mathrm{P}_2$) possible comparisons in each ear. First, the spectral difference (SD) is computed at the right ear by subtracting

$$SD_{RR}(k,i,j) = \mathrm{HR}_k(i) - \mathrm{HR}_k(j), \quad i,j = 1,2...16,\ i \neq j,\ k = 1,2...703 \quad (6.17)$$

where the subscript RR denotes the comparison being made at the right ear and k is the direction index. Similar to the ILD computation, the equation shows that spectral difference does not depend on the absolute value of HR but rather on the relative difference.

Similarly for the left ear

$$SD_{LL}(k,i,j) = \mathrm{HL}_k(i) - \mathrm{HL}_k(j), \quad i,j = 1,2...16, i \neq j,\ k = 1,2...703 \quad (6.18)$$

where the subscript LL denotes the comparison being made at the left ear. The monaural part of the difference code matrix is then computed by comparing both $SD_{RR}(k,i,j)$ and $SD_{LL}(k,i,j)$ against a spectral difference threshold (SD_{th}) for all 16 frequencies and for all 703 directions; mathematically,

$$C_{RR}(k,i,j) = \begin{cases} +1 & SD_{RR}(k,i,j) \geqslant SD_{th} \\ -1 & SD_{RR}(k,i,j) < SD_{th} \end{cases} \quad i,j = 1,2,...16,\ i \neq j,\ k = 1,2...703 \quad (6.19)$$

$$C_{LL}(k,i,j) = \begin{cases} +1 & SD_{LL}(k,i,j) \geqslant SD_{th} \\ -1 & SD_{LL}(k,i,j) < SD_{th} \end{cases} \quad i,j = 1,2,...16,\ i \neq j,\ k = 1,2...703 \quad (6.20)$$

It should be pointed out that in the actual implementation, whether it be in the real biological system or even in the electronic hardware system, the ILD and spectral difference processing takes place at a level higher than the cochlea, i.e. after the sound has been converted into its spike-based representation. The

implication is that, for a condition $SD_{RR}(k,i,j) \geqslant SD_{th}$ (or $SD_{LL}(k,i,j) \geqslant SD_{th}$) to result in a spike requires that at least that the louder of the two spectra be capable of eliciting a spike in the cochlea.

In Fig. 6.19, the subplot in the upper left corner shows the spatial contour plot of the magnitude spectrum (in dB) of the sound measured at the right ear at $20kHz$ while the other 15 subplots show the spectral difference between the magnitude spectrum at $20kHz$ and that of the other 15 center frequencies listed at the top of each subplot.

The corresponding binary spectral difference (SD_{th}=7dB) receptive fields are shown in Fig. 6.20 under the assumption that the magnitude spectrum at $20kHz$ is greater than 4dB (the subplot in the upper left corner is the binary threshold receptive field at $20kHz$ using a firing threshold of 4dB and not a spectral difference receptive field). The binary spectral difference receptive fields for frequencies $55kHz$ and greater are essentially the same as the threshold receptive field at $20kHz$ (as shown in the upper left corner).

The entire code matrix C_{diff} is formed by joining the four matrices C_{RR}, C_{LL}, C_{RL}, and C_{LR}. The dimensions of C_{diff} is 703 (rows) by 512 (columns). The length L (number of columns of C_{diff}) for the codes in this matrix is 512 (240 Right SD + 240 Left SD + 32 ILD).

Using an ILD threshold (ILD_{th}) of 5dB, and a spectral difference threshold (SD_{th}) of 7dB (as well as a minimum firing threshold of 4dB), the number of unique codes for the difference code matrix C_{diff} is 700. The matrix has a mean distance of 10 and a median distance of 9. Of the 512 columns (dimensions), 203

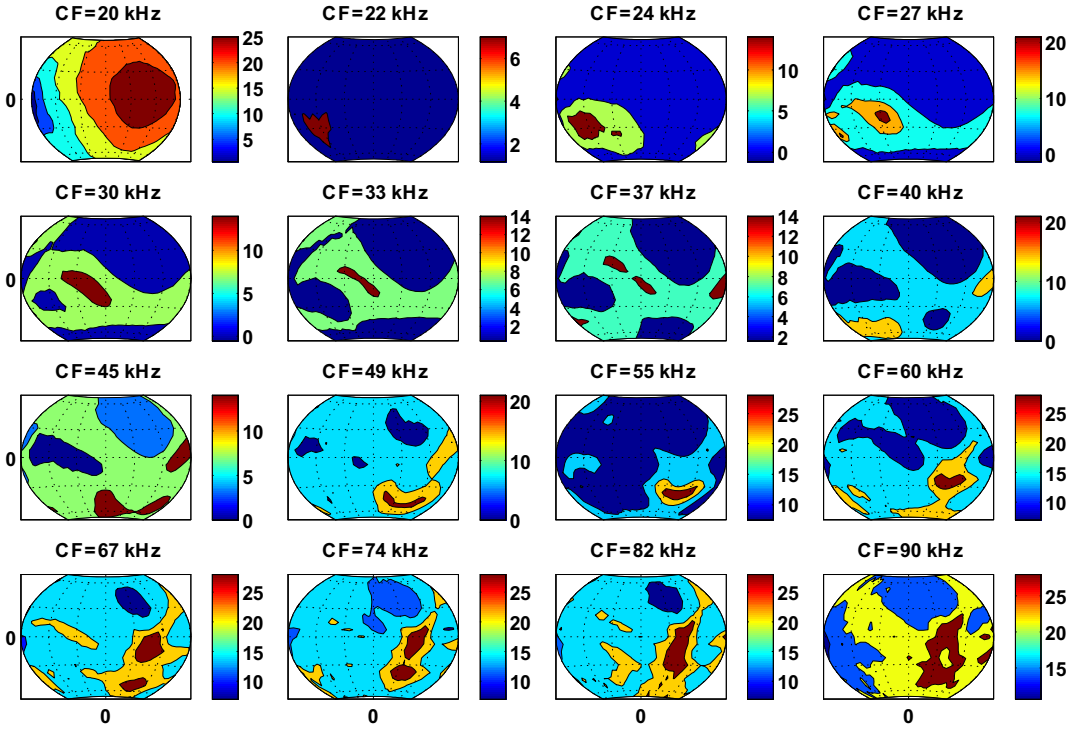

Figure 6.19: Spectral difference in the right ear. The subplot in the upper left corner shows the spatial contour plot of the magnitude spectrum(in dB) of the sound measured at the right ear at $20kHz$ while the other 15 subplots show the spectral difference between the magnitude spectrum at $20kHz$ and that of the other 15 center frequencies listed at the top of each subplot.

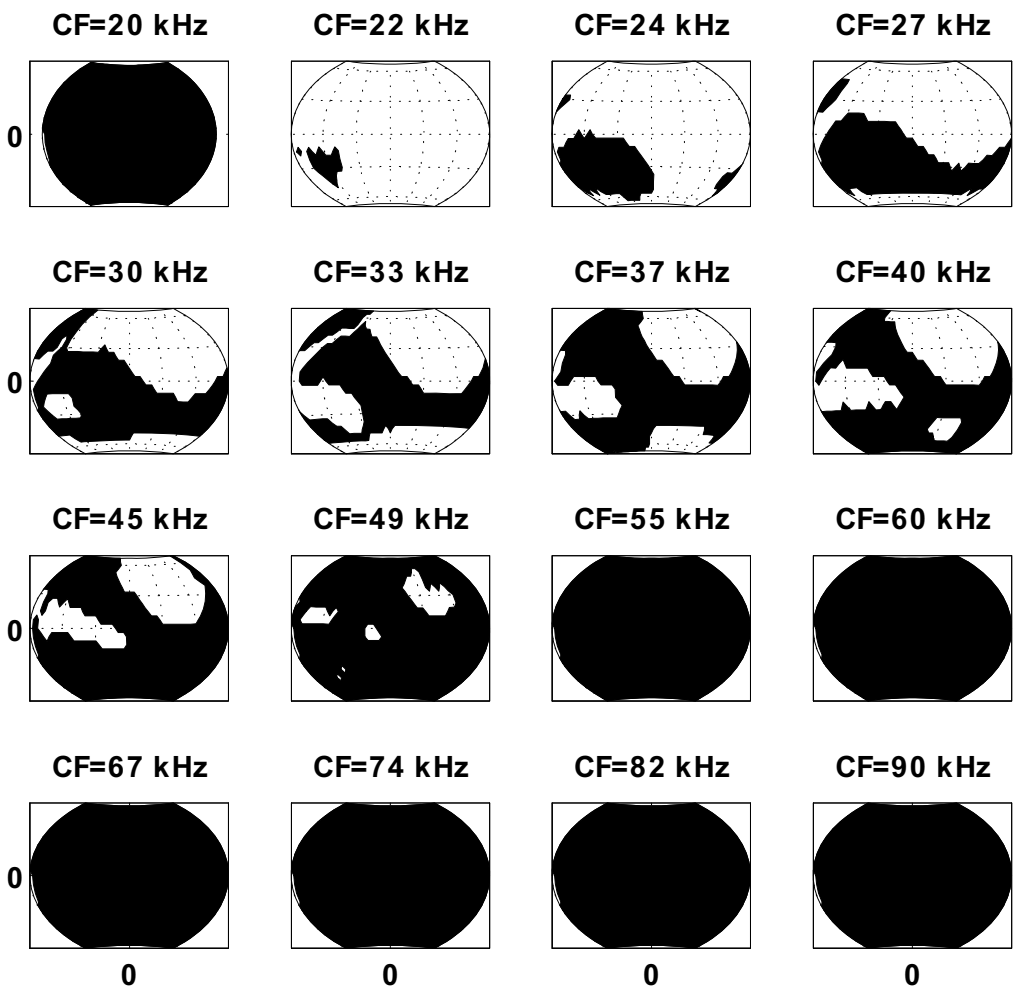

Figure 6.20: Binary receptive fields. The subplot in the upper left corner shows the threshold receptive field at $20kHz$ (threshold =7dB). The other 15 subplots show the binary spectral difference (SD_{th}=7dB) receptive fields corresponding to the 15 spectral differences shown in Fig. 6.19 under the assumption that the magnitude spectrum at $20kHz$ is greater than 4dB. Notice that the binary spectral difference receptive fields for frequencies $55kHz$ and greater are essentially the same as the threshold receptive field at $20kHz$ (as shown in the upper left corner). The black regions indicate the directions where the code is equal to +1 while the white regions indicate the directions where the code is equal to -1.

are always equal to -1 and thus provide no information.

Table 6.3 shows the effect of limiting the maximum cross correlation of the dimensions of C_{diff}. The first row corresponds to the original 512 code matrix after removing the 203 non-contributing dimensions. Notice that even after removing dimensions such that the maximum cross-correlation is $\leqslant 80\%$, the number of unique codes (NUC) remains high at 696, the mean distance is six bits and the length of the code is 124. Recall that for the threshold code matrix $C^{145}_{reduced}$, L=145, NUC=664, mean(D)=4 bits, and the maximum correlation=95%). Also observe that, in the last row, after limiting the maximum cross-correlation to 50%, the length (L) of the code was reduced to 31 while still outperforming the 32-bit threshold as shown in Table 6.2.

Fig. 6.21 shows the spatial correlation plot for 16 different directions (listed at the top of each subplot where azm=azimuth and elv=elevation). The maximum correlation is equal to L (L=124) and decreases by two for every one bit difference. The codes were selected from the code matrix C^{124}_{diff}. The lowest contour level, 64, shows all the directions whose codes are within 30 bits of error to the selected direction, while the highest contour level, 118, shows all the directions whose codes are within three bits of error to the selected direction. The white regions show the directions whose codes have a distance greater than 30 bits to the code of the selected direction. Compared to the spatial correlation plot for the code matrix $C^{145}_{reduced}$ (shown in Fig. 6.12), the spatial extent of the correlation appears to be smaller.

Fig. 6.22 shows the binary receptive fields of the 124 dimensions of the code

Table 6.3: Reducing cross-correlation in the code matrix C_{diff}.

$xcorr$	L	P	NUC	mean(D)	median(D)	min(D)	max(D)	σ_θ	σ_{ϕ_m}
100	341	32	700	10	9	0	37	0	0.1
99	295	32	700	10	9	0	35	0	0.1
98	273	33	700	9	9	0	34	0	0.1
97	258	35	700	9	8	0	34	0	0.1
96	240	36	700	9	8	0	33	0	0.1
95	233	37	700	9	8	0	31	0	0.1
90	191	39	699	8	7	0	27	0	0.1
85	152	43	698	7	6	0	22	0	0.1
80	124	45	696	6	5	0	20	0	0.2
75	97	45	687	5	4	0	16	0	0.3
70	75	46	683	3	3	0	14	0	0.3
65	61	48	673	3	2	0	11	0.4	0.5
60	49	47	645	2	2	0	9	0.7	1.0
55	38	50	613	1	1	0	7	0.8	0.9
50	31	50	561	1	1	0	6	2.3	2.0

matrix C^{124}_{diff}, the black regions indicate the directions for which the code is $+1$ while the white regions indicate the directions for which the code $=-1$. The last 15 subplots in the bottom row correspond to the ILD binary receptive fields.

Fig. 6.23 shows the summation of all 124 binary receptive fields of the code matrix C^{124}_{diff}. Where the the summation is equal to zero indicates equal "time" spent on either side ($+1$ and -1) of the decision boundaries.

Fig. 6.24 shows the distance vector (D) for the code matrix C^{124}_{diff} as a function of azimuth and elevation. Compared to the distance plot of the threshold code matrix $C^{145}_{reduced}$ shown in Fig. 6.11, the region of maximum distance covers the central direction (azimuth=0, elevation=0) and the surrounding directions.

Fig. 6.25 shows the cross-correlation of the 124 columns (dimensions) of the code matrix C^{124}_{diff}. Compared to the cross-correlation of the code matrix $C^{145}_{reduced}$ shown in Fig. 6.8, the dimensions of C^{124}_{diff} are less correlated.

Fig. 6.26 shows the true and estimated directions for the code matrix C^{124}_{diff} when 15 bits of error is randomly imposed on the code. The standard deviation of error in estimating elevation and the modified azimuth is σ_θ=5° and σ_{ϕ_m}=3.9° respectively, compared to the σ_θ=13.5° and σ_{ϕ_m}=5.2° for the threshold code matrix $C^{145}_{reduced}$ allowing eight bits of error.

Fig. 6.27 shows σ_θ, σ_ϕ, and σ_{ϕ_m} (in degrees) as a function of the number of bit errors for the code matrix C^{124}_{diff}. The lower plot is similar to the upper plot but focuses on the initial part of the curve. Also shown in the lower plot are the curves for the threshold code matrix $C^{145}_{reduced}$ (as shown in Fig. 6.15). It is clear that the difference code outperforms the threshold code of comparable length. It

will be shown in the next section that the same applies to the hardware system.

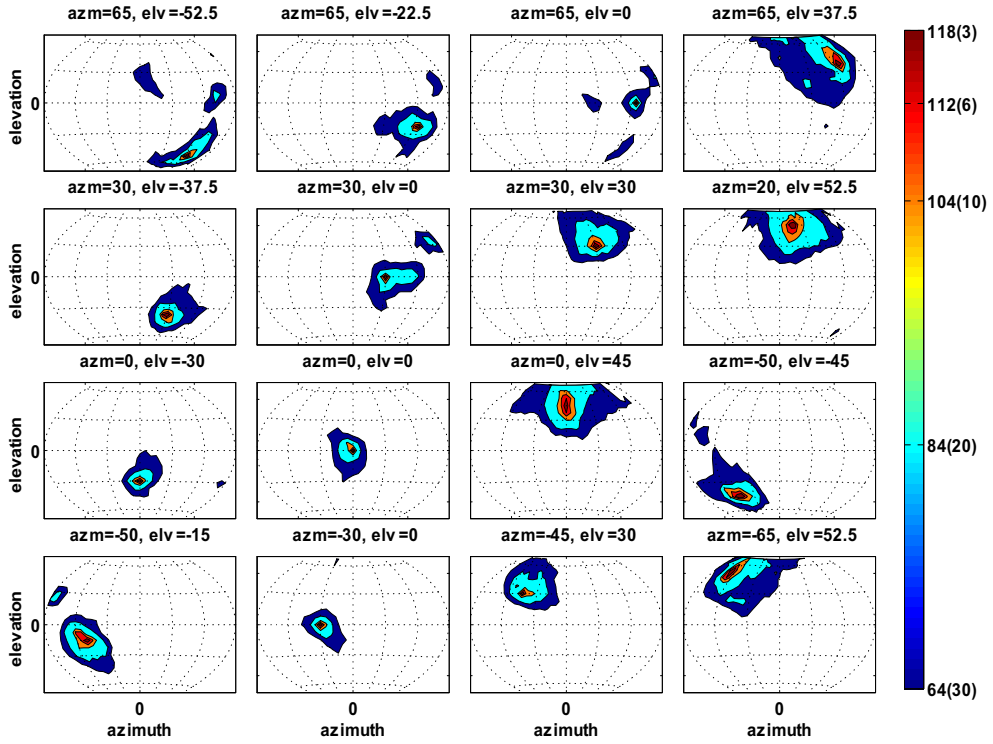

Figure 6.21: Spatial correlation plots for the difference code matrix C^{124}_{diff}. Shown is the spatial correlation plot for 16 different directions (listed at the top of each subplot where azm=azimuth and elv=elevation). The spatial correlation plot is calculated by cross-correlating a code from a single direction with all 703 codes. The maximum correlation is equal to L (L=124) and decreases by two for every one bit difference. The codes were selected from the code matrix C^{124}_{diff}. The lowest contour level, 64, shows all the directions that are within 30 bits of error to the selected direction, while the highest contour level, 118, shows all the directions that are within three bits of error to the selected direction.

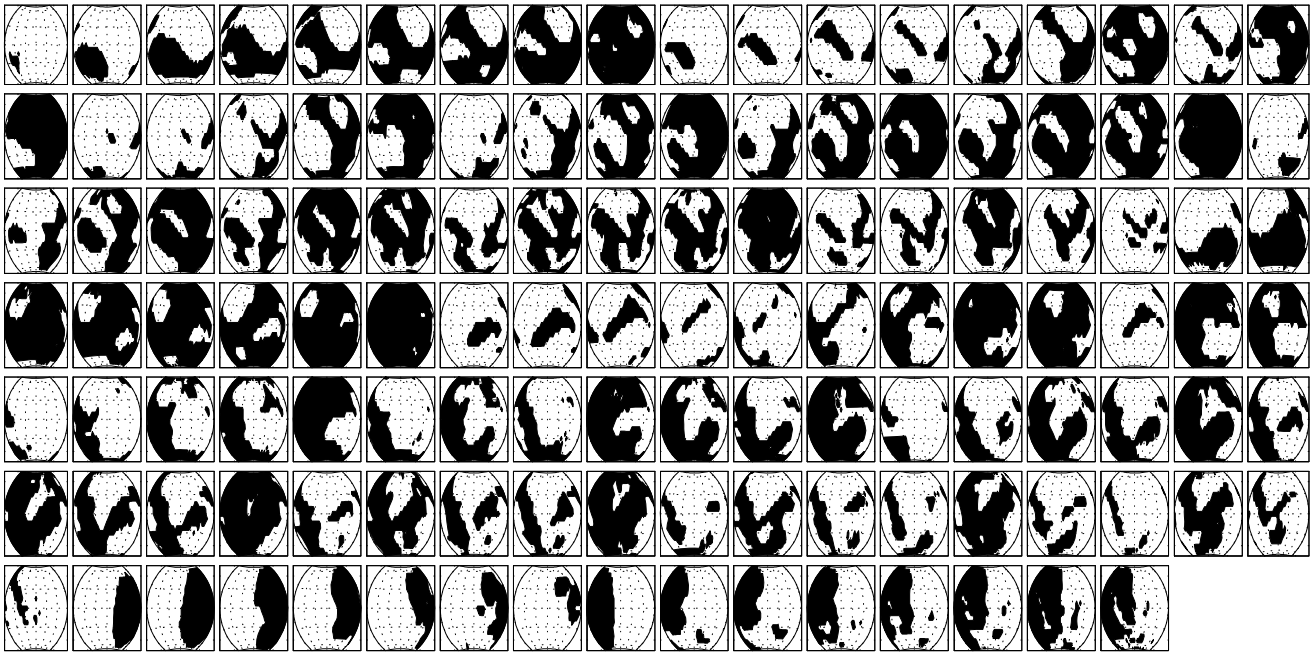

Figure 6.22: Binary receptive fields of the dimensions of the code matrix C_{diff} (L=124). The black regions indicate the directions where the code is equal to +1 while the white regions indicate the directions where the code is equal to -1.

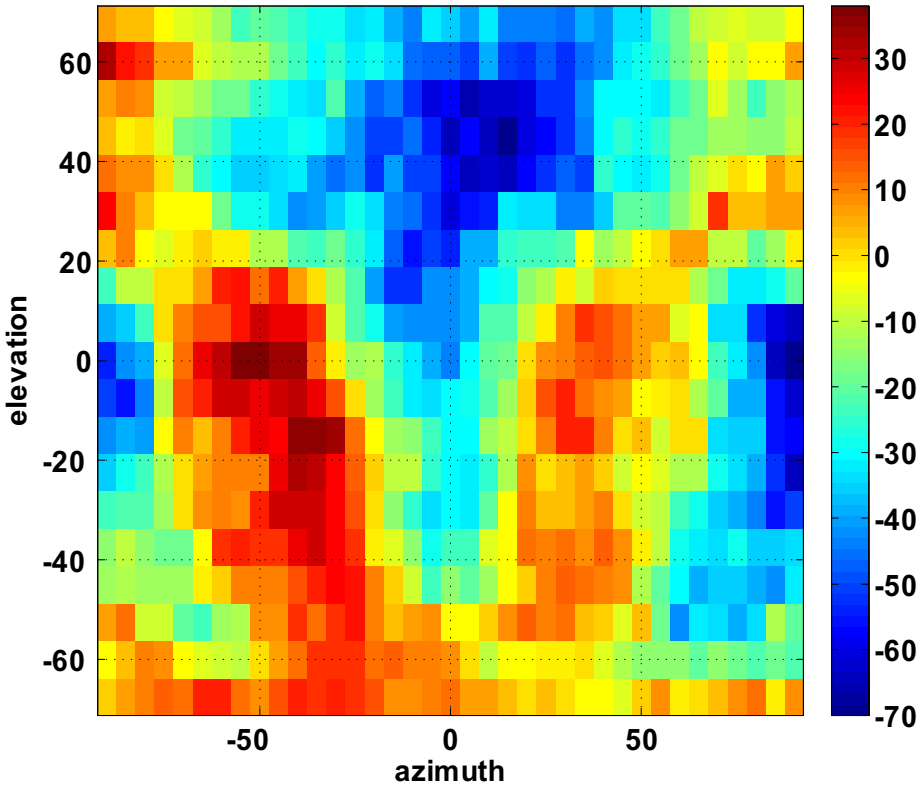

Figure 6.23: Summation of all 124 binary receptive fields of the code matrix C_{diff}^{124}. The regions where the summation is equal to zero indicates equal "time" spent on either side (+1 and −1) of the decision boundaries.

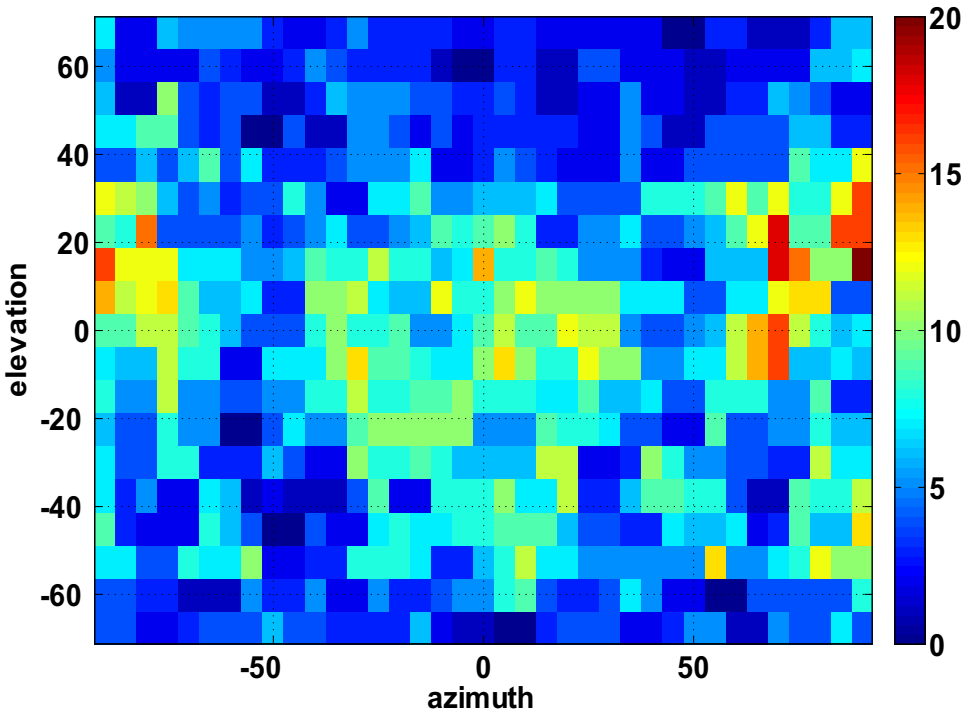

Figure 6.24: Distance vector (D) for the code matrix C_{diff}^{124} as a function of azimuth and elevation.

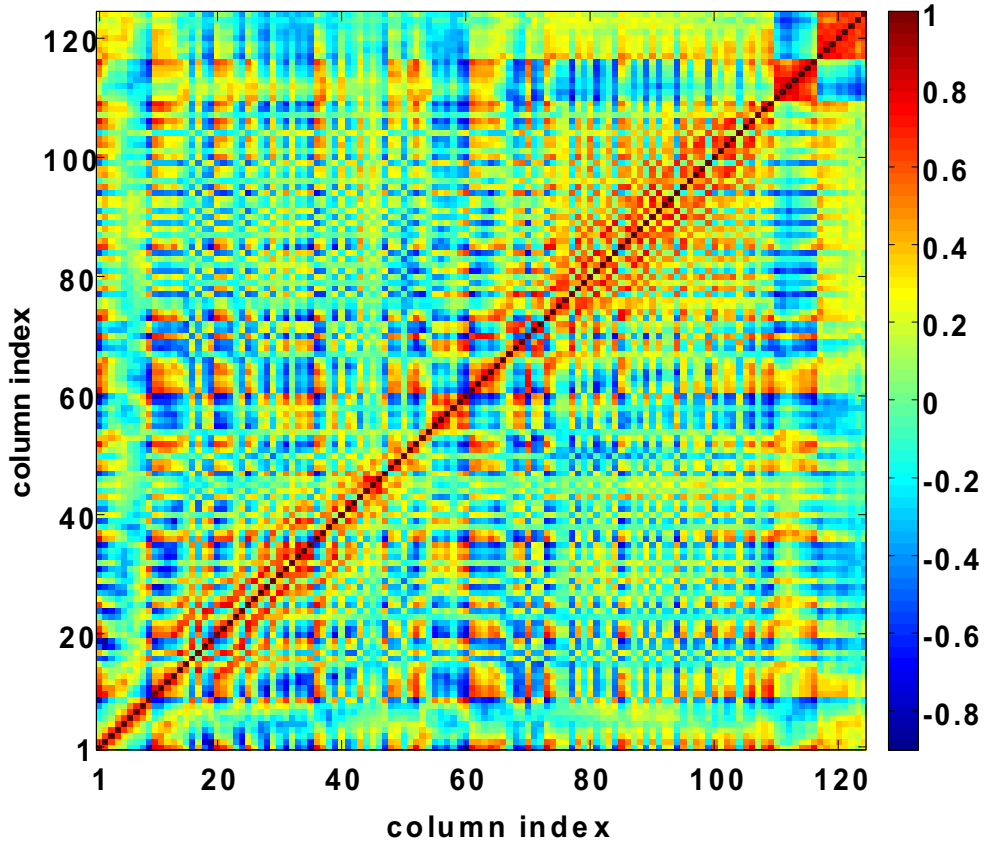

Figure 6.25: Cross-correlation of the 124 columns (dimensions) of the code matrix C^{124}_{diff}. Compared to the cross-correlation plot of the threshold code matrix $C^{145}_{reduced}$ shown in Fig. 6.8, the columns of the difference code matrix are less correlated.

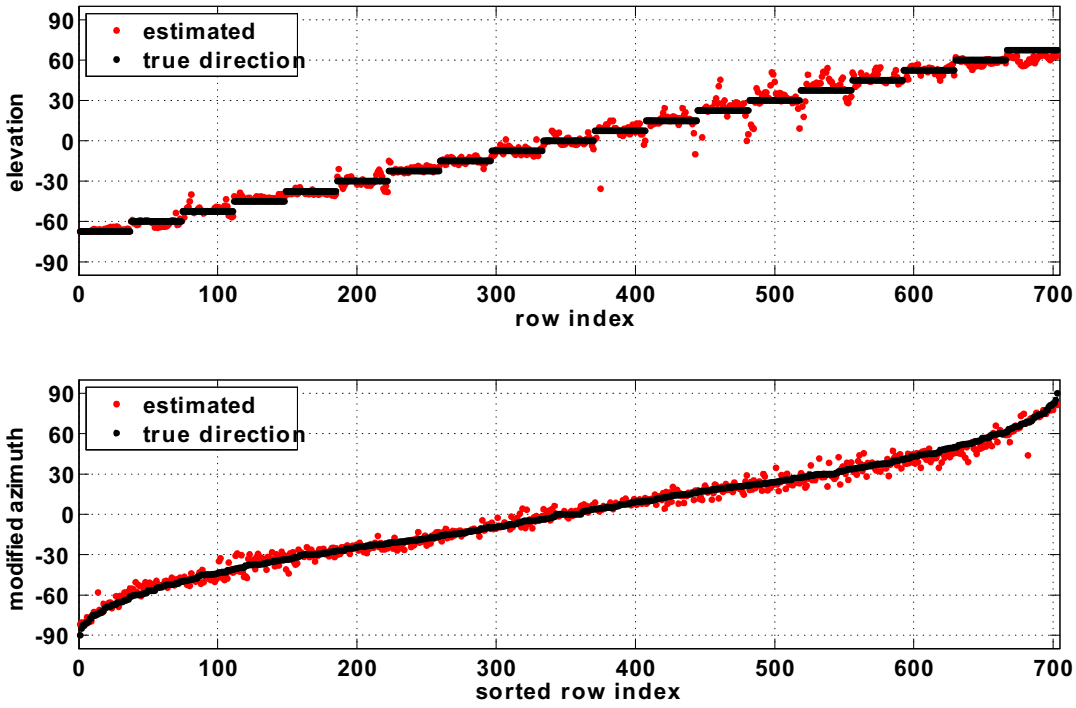

Figure 6.26: True and estimated directions for the code matrix C_{diff}^{124} when 15 bits of error were randomly imposed on the code. The standard deviation of error in estimating elevation is 5° and the standard deviation of error in estimating the modified azimuth is 3.9°. In the lower plot, the row index was sorted so that the modified azimuth appears to change monotonically.

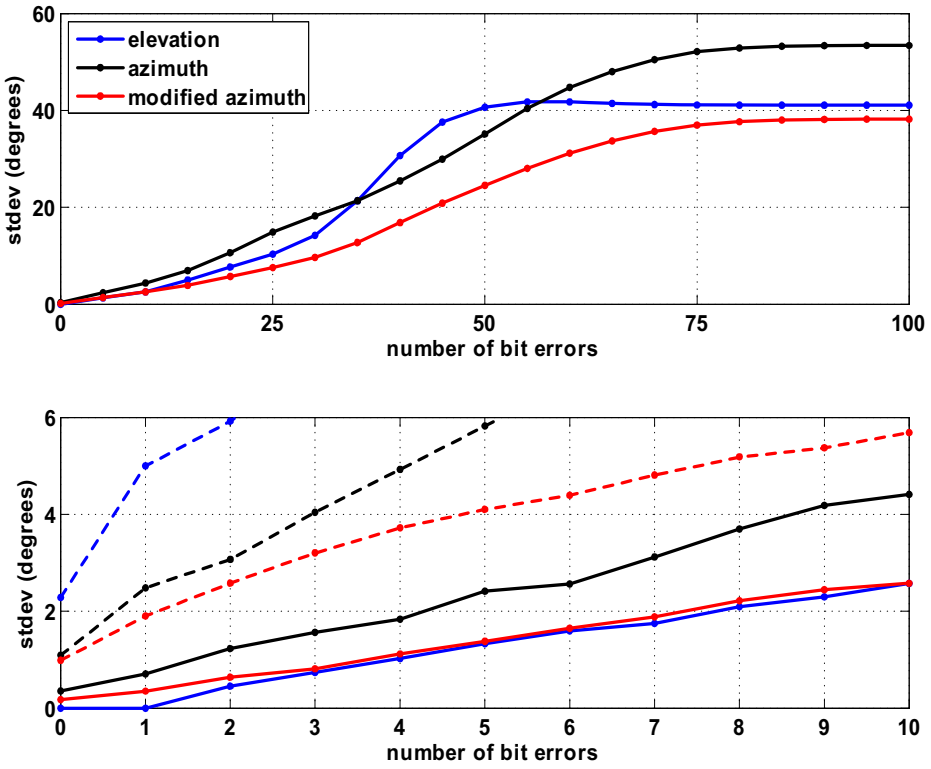

Figure 6.27: Standard deviation of error (in degrees) in estimating elevation, azimuth, and modified azimuth as a function of the number of bit errors for the code matrix C_{diff}^{124}. The lower plot is the similar to the upper plot but focuses on the initial part of the curve. The dashed lines in the lower plot are those for the threshold code matrix $C_{reduced}^{145}$. It is clear that the difference code outperforms the threshold code of comparable length.

6.4 Localization in Hardware

Fig. 6.28 shows a picture of the localization system. Shown are two (right and left) ultrasonic cochlea chips, two (right and left) spectral difference chips, two ILD chips, and the merge board. The FM sweeps recorded from the right and left ears (microphones) of the artificial bat head are uploaded to two Agilent arbitrary waveform generators (not shown) and played back to the two cochlea chips. The address-event output spikes of the cochlea chips connect to the merge board, where a dsPIC30F4011 microcontroller handles the address-event communication and merges their address-event spikes onto a single address bus. The microcontroller also sends the trigger signals to the ILD and spectral difference chips. The merged address bus connects to the spectral difference chips and the ILD chips. The merge board was designed to connect to two spectral difference chips and up to three ILD chips.

A MATLAB^{TM} program running on a personal computer (PC not shown) was used to control the data acquisition. The bat head measurements were uploaded from the PC to the two (right and left) arbitrary waveform generators. A third waveform generator (not shown) was used to simultaneously trigger (for synchronization) a logic analyzer (not shown) and the other two waveform generators. The logic analyzer ($80ns$ sampling period) was used to capture the different control signals as well as the address-event spikes from both the feature extraction chips and the merged address bus. The acquired data is then transferred from the logic analyzer to the PC. The localization system was operated in two modes: the

single spike mode and the *multiple spikes mode*. We begin first by describing the single spike mode.

6.4.1 Single Spike Latency Code

Recall that the cochlea chip has 16 cochlear filters and eight neurons for each cochlear filter. The 128 neurons are arranged in eight rows (16 neurons per row), and each row of neurons can be set to have a different threshold. In the single spike mode, only one row (threshold) of cochlear neurons was turned on (i.e. only one neuron per cochlear filter), and the neurons (for all chips) were allowed to fire a single spike. In this mode of operation, the terms cochlear filter and cochlear neuron will be used interchangeably. As a result, the only way to encode intensity information in a given cochlear filter is through the latency of the spike, and thus we refer to this code as the single spike latency code. For the ultrasonic cochlea chips, the measured center frequencies (CF in kHz) of the 16 ($k = 1, 2, ...16$) right (**R**) and left (**L**) cochlear filters are:

CF$\backslash k$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
R	20	22	24	28	31	35	39	43	50	53	60	65	73	81	93	101
L	21	23	25	30	32	37	41	45	52	58	62	72	79	85	89	106

The right cochlea has a mean quality factor (Q) of 8.8 with a standard deviation of 1.7 while the left cochlea has a mean quality factor (Q) of 10.8 with a standard deviation of 4.1.

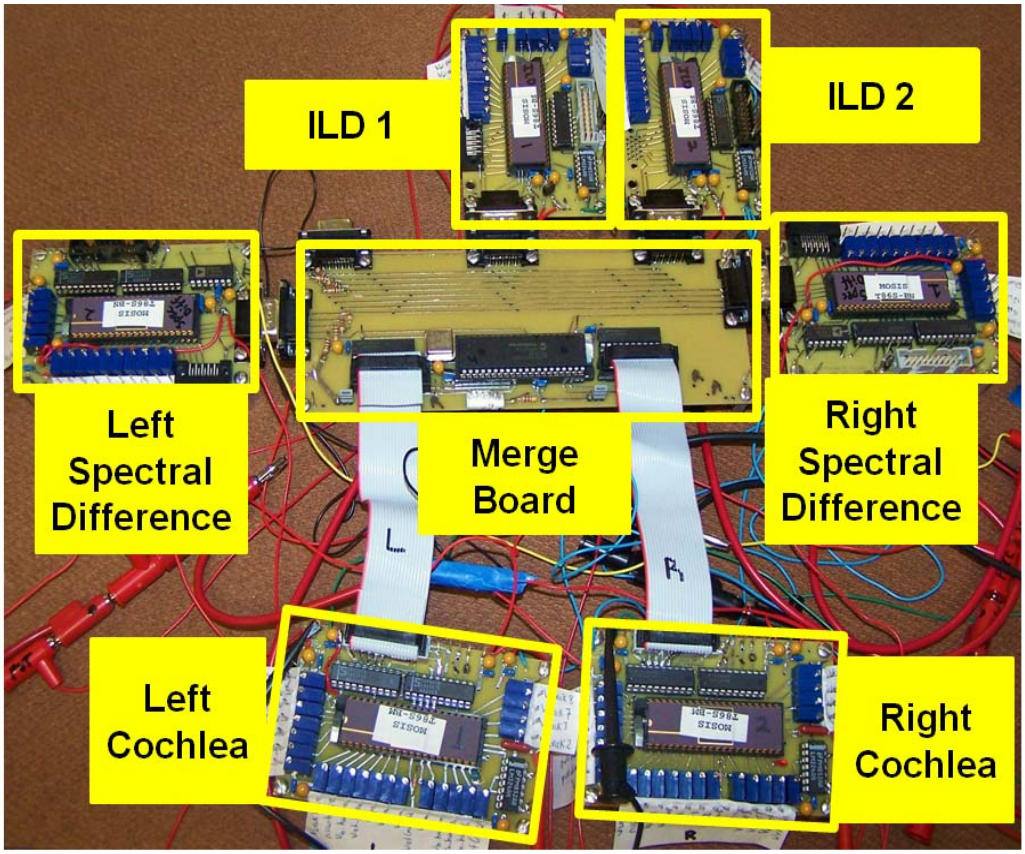

Figure 6.28: Photograph of the localization system. A MATLAB^{TM} program uploads the FM sweeps recorded from the right and left ears of the artificial bat head to two arbitrary waveform generators (not shown). The sweeps are played back to the right and left cochlea chips. The cochlea chips connect to the merge board where a microcontroller merges the address-event spikes onto a single address bus. The bus then connects to the right and left spectral difference chips as well as the ILD chips. A logic analyzer (not shown) is used to capture all the address-event spikes and control signals.

Spike Rasters and Spike Count

Fig. 6.29 shows the response of the two cochlea chips to sounds from four different directions (listed at the top of the plot where elv denotes elevation and azm denotes azimuth). For the sound emitted from the direction located on the plane of symmetry (i.e. elv=0, azm=0), there is a mismatch in latency and the majority of the left cochlear neurons have a shorter latency. The mean mismatch in latency is $94\mu s$ and the standard deviation is $168\mu s$. Fig. 6.30 shows the total number of spikes from the right and left cochlea as a function of direction. The curves saturate at 16 since there are only 16 neurons that are turned on in each cochlea.

Spike Latency

Figs. 6.31 and 6.32 show the spike latencies for all 16 cochlear filters as a function of elevation and azimuth in the left and right cochlea chips respectively. The white regions indicate that no spikes were fired for those directions. The colorbars represent the spike latency. Because of the direction of the sweep (from $120kHz$ down to $18kHz$), the higher the CF of the filter, the smaller the latency.

Fig. 6.33 shows the spike latency for the 16 cochlear neurons (filters) as a function of the peak gain (in dB) at the CF of the cochlear filter. The independent variable (the peak gain) is plotted on the y-axis. Black dots represent spikes from the right cochlea, blue dots represent spikes from the left cochlea, and the horizontal axis represents time in ms. It is clear that the spike latency decreases as the peak gain increases and that the change is monotonic. This plot, however,

does not take into account the fact that a cochlear filter responds to a wide range of frequencies and not just its center frequency. Fig. 6.34 shows the same spike latency data as a function of the mean magnitude spectrum of sound (in dB) at the CF of the cochlear filter. The mean spectrum was computed by summing the inner product of the magnitude spectrum of the sound at each direction with the transfer function of the cochlea and then normalizing with respect to the number of frequency bins. The left cochlear neurons (generally) have a shorter latency than the right cochlear neurons. In this plot, the latency appears to be linear with the mean spectrum. It should be noted, however, that the above method of computing the mean does not take into account the effect of the neuron's leak current or the fact that the portion of the sweep after the spike has been generated does not contribute to the firing of the spike.

ILD Coding

The top row in Fig. 6.35 shows the spatial receptive fields (at $20kHz$) of the left (a) and right (b) ears in response to the FM sweep as measured from the artificial bat head along with the corresponding ILD plot (c), where the ILD was computed by subtracting the left contour plot from the right contour plot. The second row shows the spatial contour plot of the spike latency from the left (d) and right (e) cochlea respectively (where the white regions indicate that no spikes were fired for the corresponding directions) as well as the difference (f) in the spike latency between the spikes fired from the right and left cochlear filters at $\simeq 20kHz$. The white regions in f) correspond to the white regions in d) and e), indicating that

the difference in spike latency is not defined for those directions (because one of the two cochlear neurons did not fire). The spike latencies in d) and e) are correlated to the spatial receptive fields in a) and b); the higher the intensity (in a and b) the shorter the latency (in d and e). Observe the resemblance between the ILD computed from the magnitude spectrum (c) and that computed from the difference in spike latencies (f).

Another example is shown in Fig. 6.36 but for a center frequency of $60kHz$. Both plots illustrate that the measured ILD has been successfully encoded in the difference in spike latencies between the right and left ears.

Fig. 6.37 shows the ILD neuron spike latency for the four lowest center frequencies (CF) versus the difference in cochlear spike latency. The blue dots represent spikes from the **L-R** ILD neurons while the black dots represent spikes from the **R-L** ILD neurons. No ILD spikes are generated if the absolute difference in cochlear spike latency is less than $200\mu s$.

Fig. 6.38 shows the receptive fields for all 32 ILD neurons. The plot is divided into 16 pairs of neurons, each pair corresponds to one of the 16 ILD blocks (see Chapter 5). The neuron to the right of the index corresponds to the **R-L** neuron while the neuron to the left of the index corresponds to the **L-R** neuron. The white regions indicate that no spikes were generated from those directions. The binary receptive field for the ILD neurons is merely the black and white version of this plot. The asymmetry between the **R-L** and **L-R** neurons is attributed to the mismatch in latencies as described earlier.

Fig. 6.39 shows the difference in spike latency at $20kHz$ as a function of

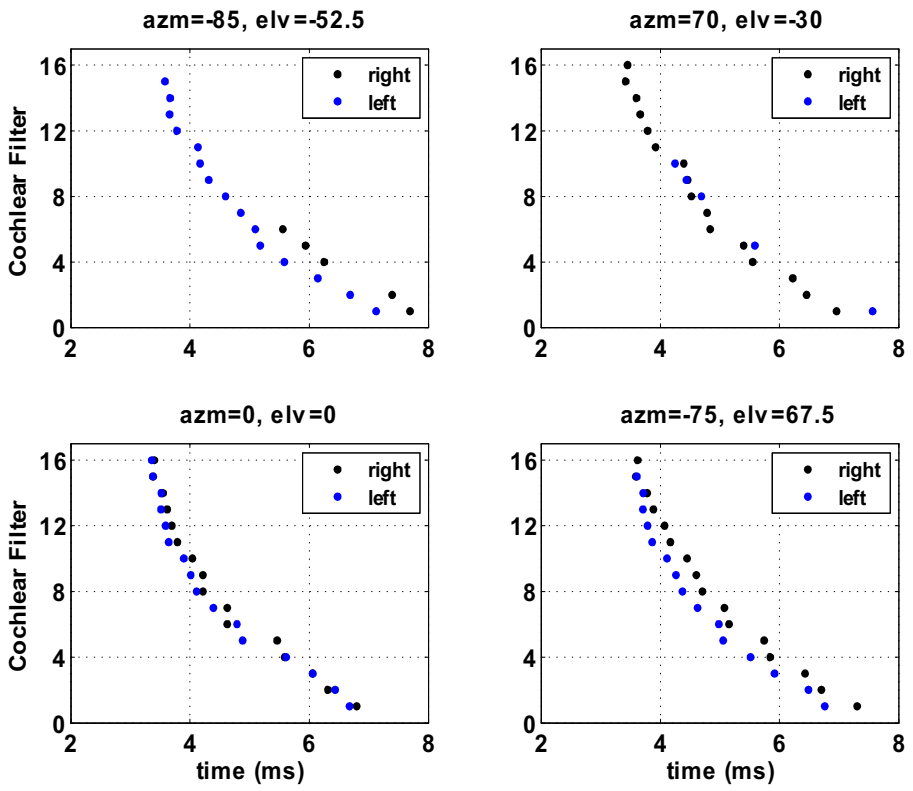

Figure 6.29: Response of the right and left cochlea chips to sounds from four different directions (listed at the top of the plot where elv denotes elevation and azm denotes azimuth). Notice that for the sound emitted from the direction elv=0, azm=0 (i.e. located on the plane of symmetry) there is a mismatch in latency and that the majority of the left cochlear neurons have a shorter latency.

azimuth ϕ (top plot) as well as the modified azimuth ϕ_m (bottom plot). This plot, similar to Figs. 3.27 and 3.28, illustrates that the difference in spike latency (i.e. ILD) correlates better with ϕ_m.

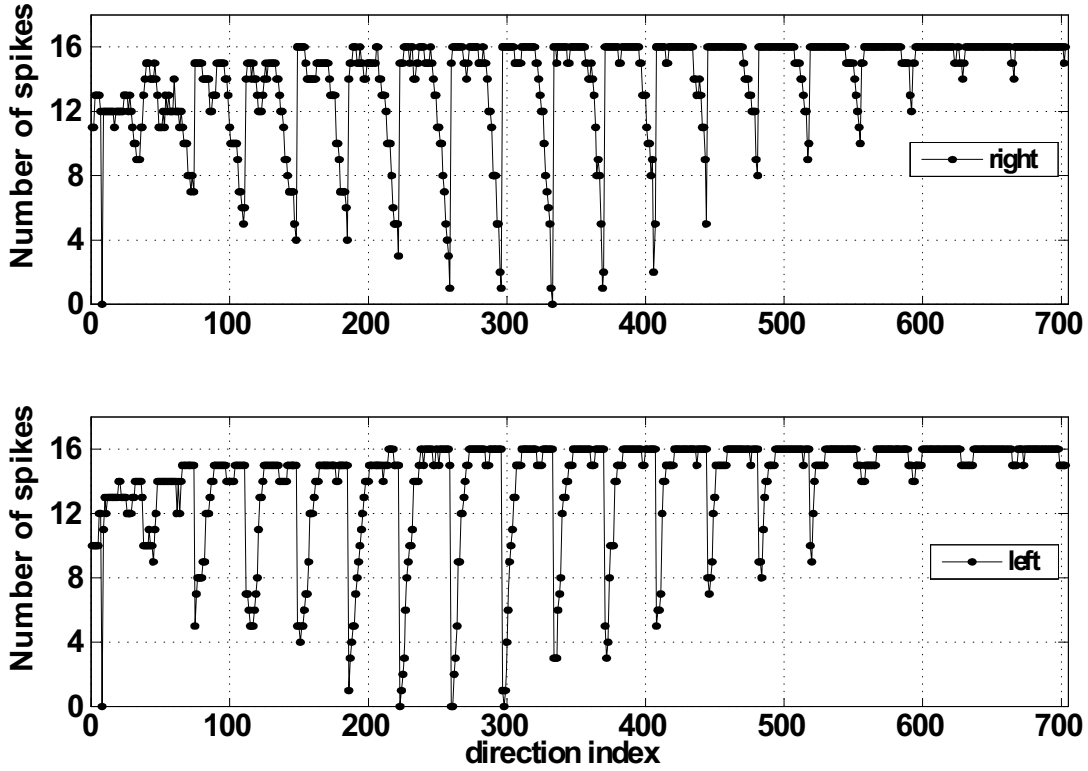

Figure 6.30: Total number of spikes from the right cochlea (upper plot) and left cochlea (lower plot) as a function of direction. The curves saturate at 16 since there are only 16 neurons that are turned on in each cochlea. Recall that the artificial bat head measurement starts at a fixed elevation of -67.5° (and stops at 67.5°) sweeping the azimuth from 90° to -90° in steps of 5 degrees . The elevation is then increased by 7.5 degrees and the process is repeated. Direction indices 1-37 correspond to an elevation of -67.5° while direction indices 667-703 correspond to an elevation of 67.5°.

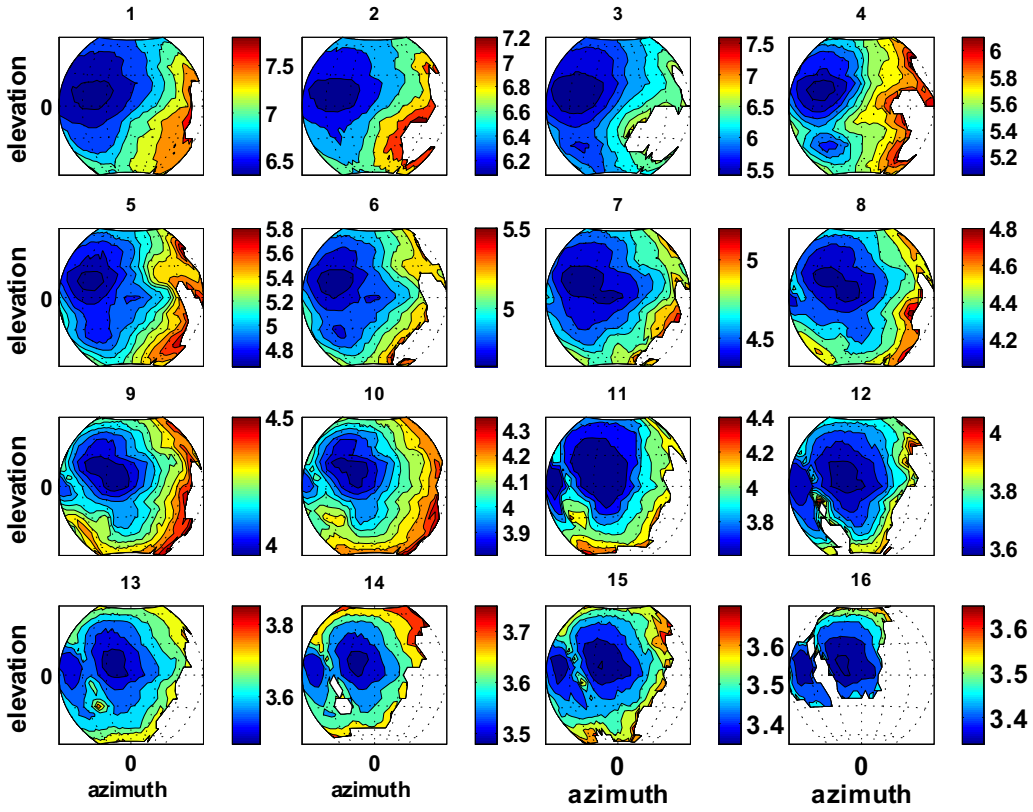

Figure 6.31: Latency of the left cochlea spikes. Shown are the spike latencies for the 16 cochlear filters (only one neuron per cochlear filter). The white regions indicate that no spikes were fired for those directions. Subplot number 1 corresponds to the lowest center frequency while subplot number 16 corresponds to the highest center frequency. Grid lines are separated by 30 degrees.

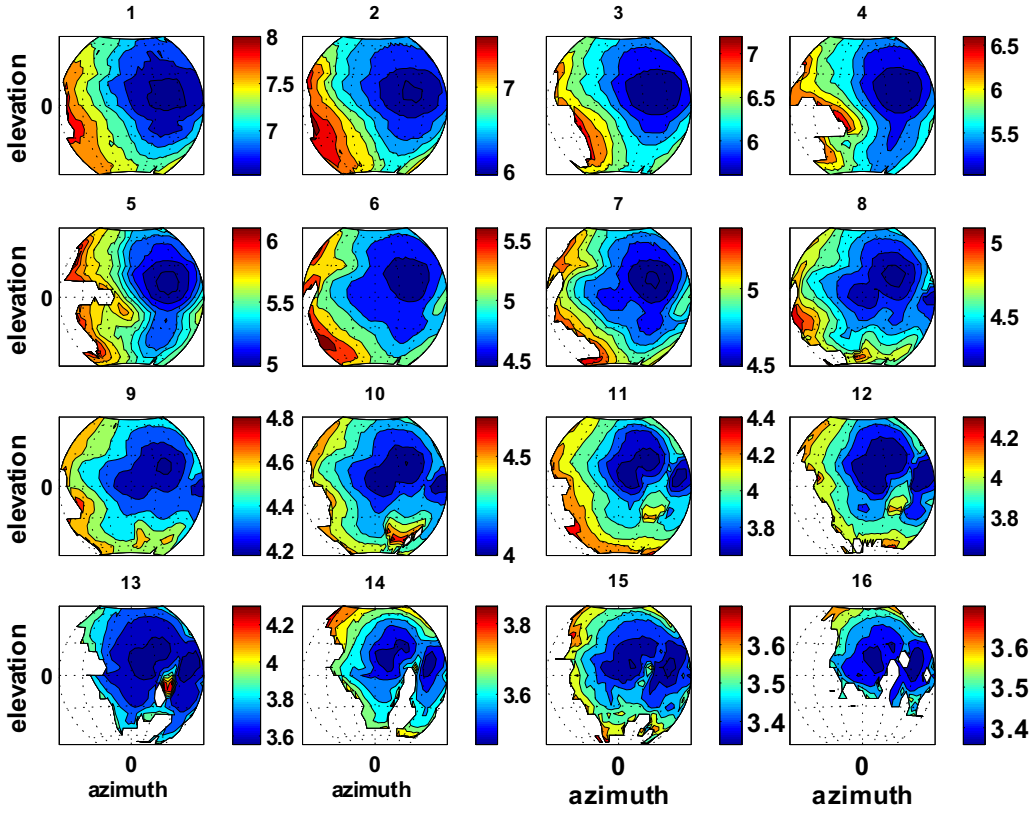

Figure 6.32: Latency of the right cochlea spikes. Shown are the spike latencies for the 16 cochlear filters (only one neuron per cochlear filter). The white regions indicate that no spikes were fired for those directions. Subplot number 1 corresponds to the lowest center frequency while subplot number 16 corresponds to the highest center frequency. Grid lines are separated by 30 degrees.

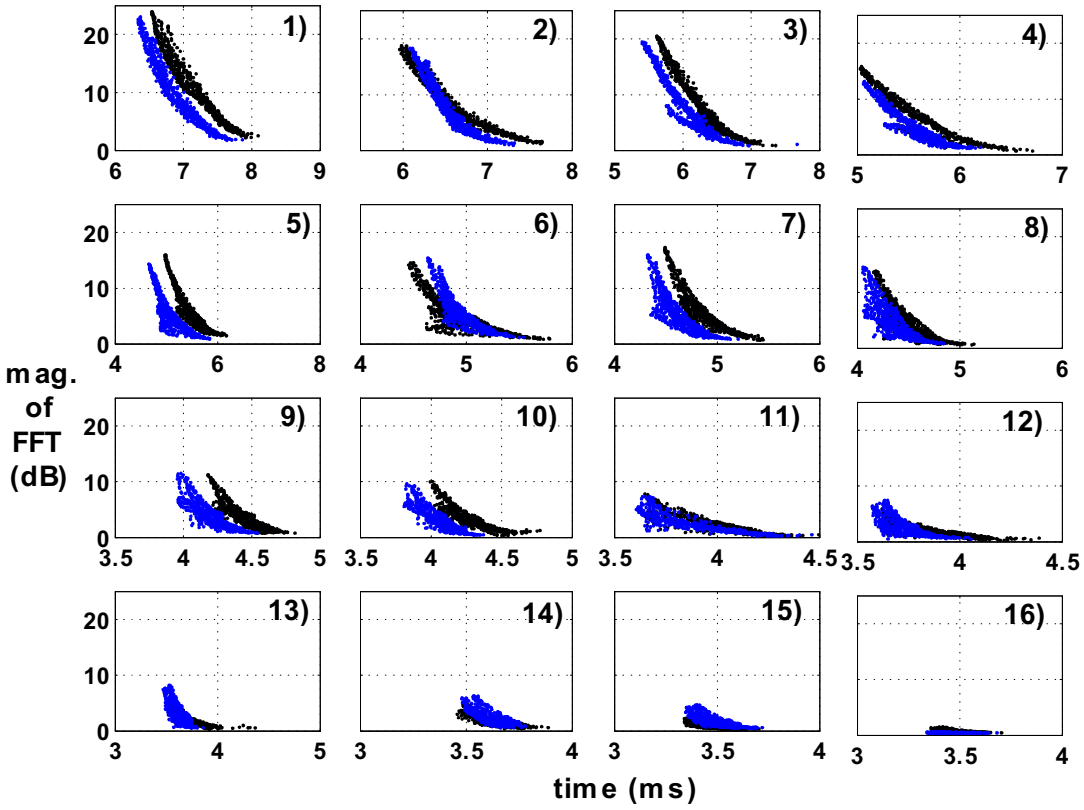

Figure 6.33: Spike latency for the 16 cochlear filters (neurons) as a function of the magnitude spectrum (FFT in dB) of sound at the CF of the cochlear filter (notice the independent variable is plotted on the y-axis). Black dots represent spikes from the right cochlea while blue dots represent spikes from the left cochlea. The horizontal axis represents time in *ms*.

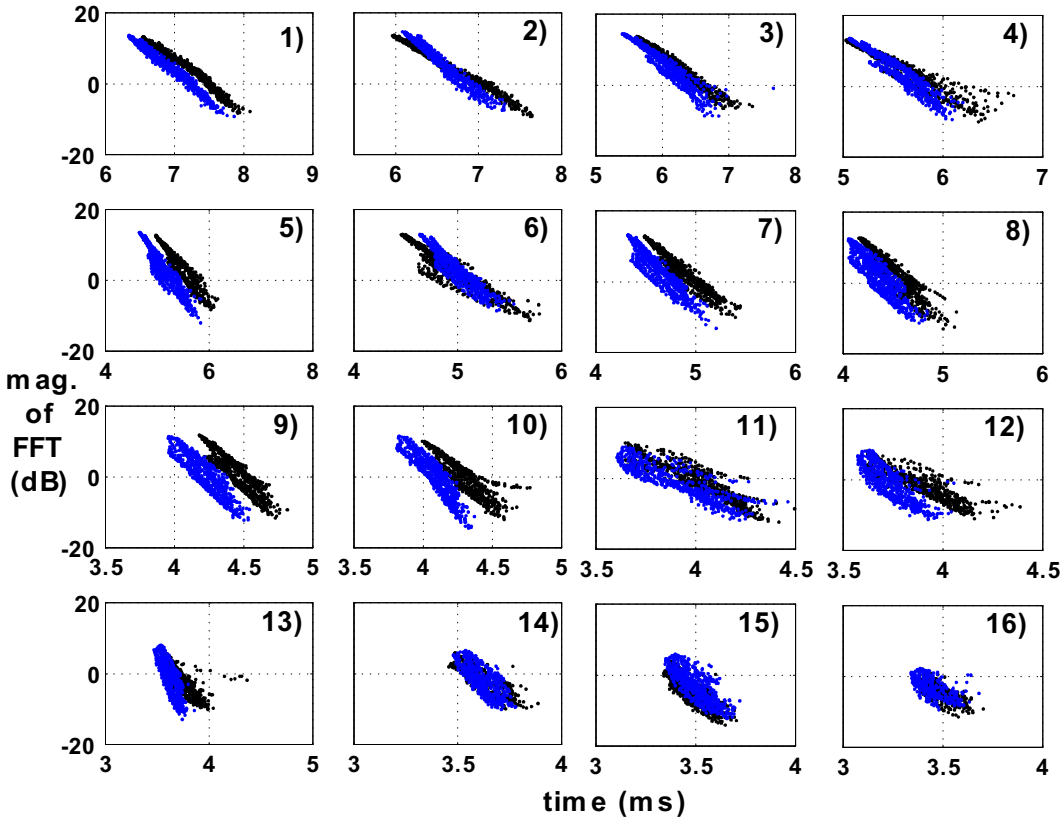

Figure 6.34: Spike latency for the 16 cochlear filters (neurons) as a function of the mean magnitude spectrum (FFT in dB) of sound at the CF of the cochlear filter (notice the independent variable is plotted on the y-axis). Black dots represent spikes from the right cochlea while blue dots represent spikes from the left cochlea, the horizontal axis represents time in *ms*.

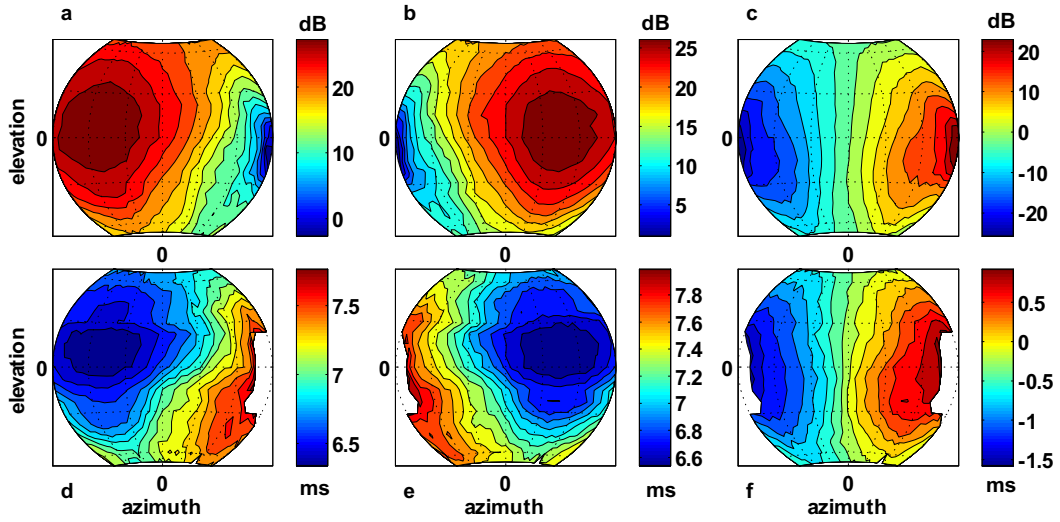

Figure 6.35: ILD at $20kHz$. a,b) Spatial contour plots of the magnitude spectrum at $20kHz$ as measured from the left and right ears respectively. c) ILD computed by subtracting the left contour plot from the right contour plot. d,e) Spatial contour plot of the spike latency from the left and right cochlea respectively (the white regions indicate that no spikes were fired for the corresponding directions. f) Difference in the spike latency between the spikes fired from the left and right cochlear filters at $\simeq 20kHz$. The white regions in f) correspond to the white regions in d) and e) indicating that the difference in spike latency is not defined for those directions. Grid lines are separated by 30 degrees.

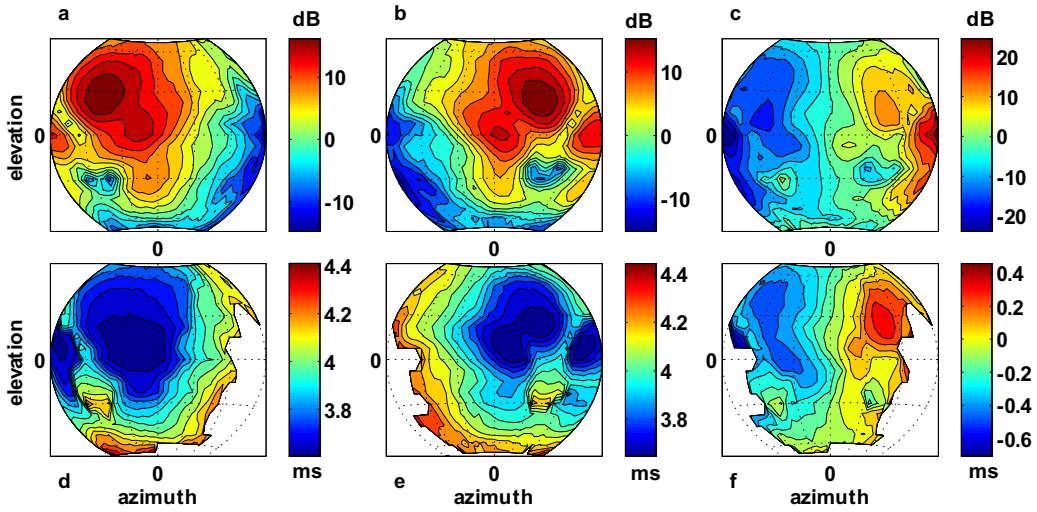

Figure 6.36: ILD at $60kHz$. a,b) Spatial contour plots of the magnitude spectrum at $60kHz$ as measured from the left and right ears respectively. c) ILD computed by subtracting the left contour plot from the right contour plot. d,e) Spatial contour plot of the spike latency from the left and right cochlea respectively (the white regions indicate that no spikes were fired for the corresponding directions. f) Difference in the spike latency between the spikes fired from the left and right cochlear filters at $\simeq 60kHz$. The white regions in f) correspond to the white regions in d) and e) indicating that the difference in spike latency is not defined for those directions. Grid lines are separated by 30 degrees.

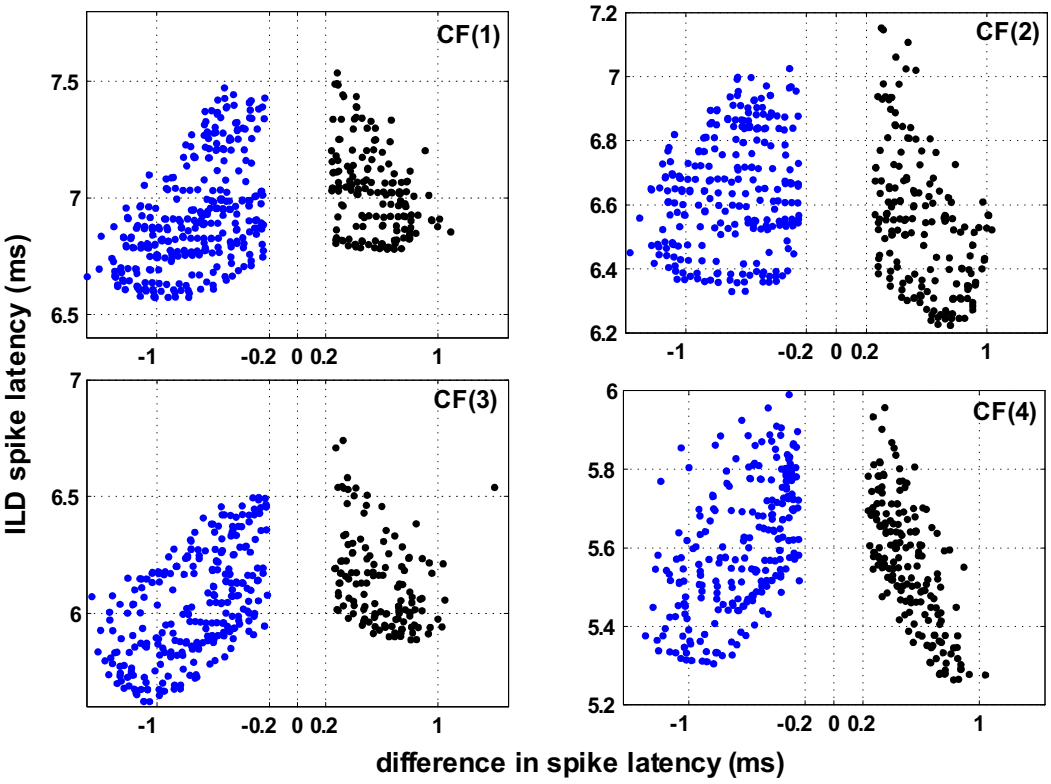

Figure 6.37: ILD spike latency for the ILD neurons corresponding to the four cochlear filters with the lowest center frequency (CF) versus the difference in cochlear spike latency. The blue dots represent spikes from the **L-R** ILD neurons while the black dots represent spikes from the **R-L** ILD neurons. No ILD spikes are generated if the absolute difference in cochlear spike latency is less than $200\mu s$.

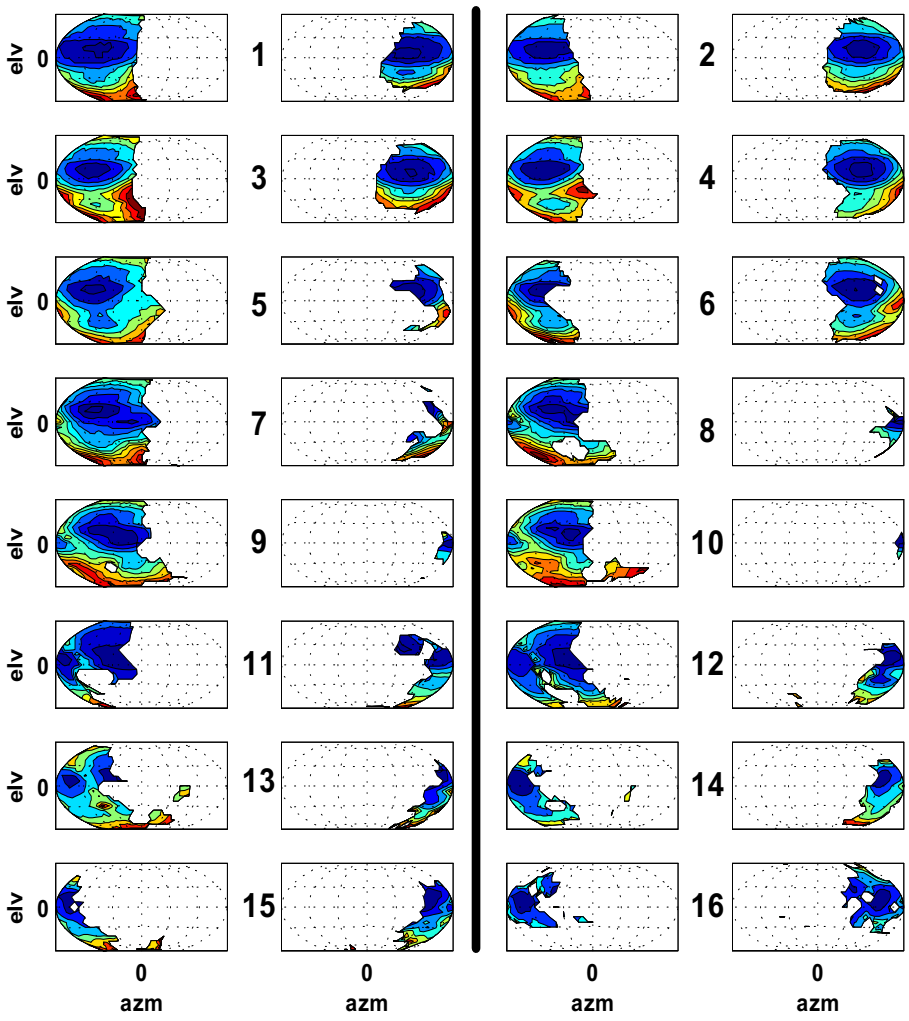

Figure 6.38: ILD receptive fields. The plot shows the receptive fields for all 32 ILD neurons. The plot is divided into 16 pairs of neurons, each of which corresponds to one of the 16 ILD blocks. The neuron to the right of the index corresponds to the **R-L** neuron while the neuron to the left of the index corresponds to the **L-R** neuron. The white regions indicate that no spikes were generated from those directions. The binary receptive field for the ILD neurons is merely the black and white version of this plot.

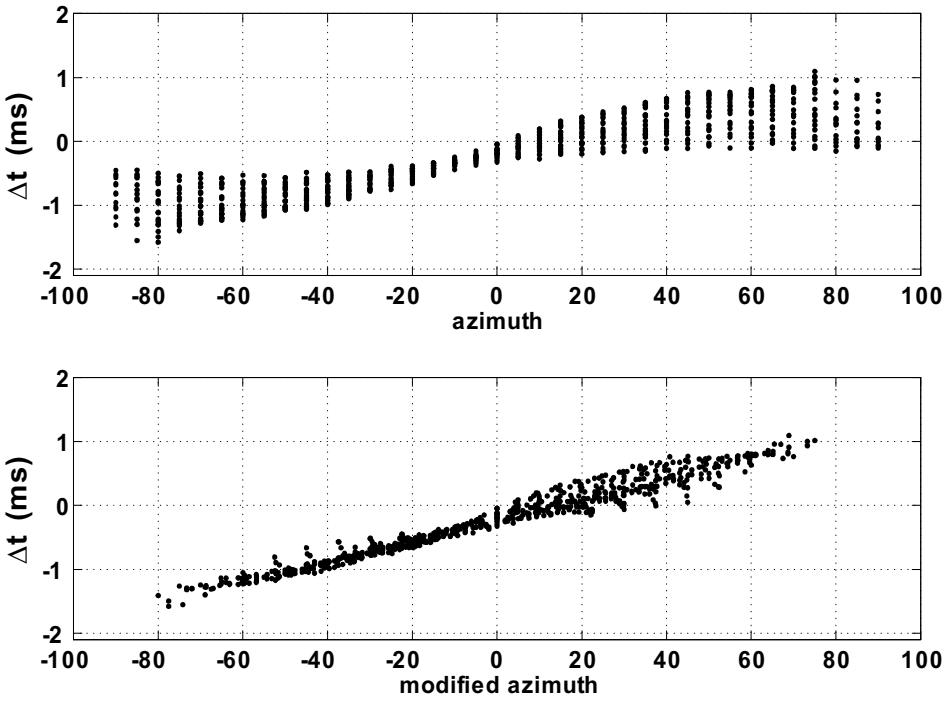

Figure 6.39: Spike latency difference at $20kHz$. The plot shows the spike latency difference at $20kHz$ as a function of azimuth ϕ (top plot) as well as the modified azimuth ϕ_m (bottom plot). The plot (similar to Figs. 3.27 and 3.28) illustrates that the spike latency difference correlates better with ϕ_m.

Spectral Difference Coding

Fig. 6.41 shows the response of the right spectral difference (SD) chip. Recall that the spectral difference chip has 240 neurons arranged in a 16 by 16 array with no neurons along the diagonal (gray boxes). Each row in the array receives excitation from a single cochlear filter while each column in the array receives inhibition from a single cochlear filter. The index i (or j)=1 corresponds to the cochlear filter with the lowest center frequency (CF), while the index i (or j)=16 corresponds to the cochlear filter with the highest CF. For example, row 1 (column 1) receives excitation (inhibition) from the cochlear filter with lowest CF while row 16 (column 16) receives excitation (inhibition) from the cochlear filter with highest CF. The dark regions indicate the directions for which a neuron fired a spike. In this measurement, the integration time (d_{mem}, see Fig. 5.15) for the right and left spectral difference neurons was $202\mu s$ and $222\mu s$ respectively. Because the FM sweep is a downward sweep ($120kHz$ to $18kHz$ in $5ms$), all neurons below the diagonal receive excitation before inhibition.

To explain the operation of the spectral difference chip in the single spike latency mode of operation, let us consider the subset of neurons shown in Fig. 6.40. The figure shows the interaction between the five cochlear filters with the lowest center frequencies. The leftmost column (c=1, j=$1, 2, ...5$) represents the binary cochlear receptive fields of the five cochlear neurons (filters). The 5 by 5 array (i, j=$1, 2...5$), to the right of the solid black line, represents the receptive fields of the corresponding 20 spectral difference neurons (recall that there are no

neurons along the diagonal).

In this mode of operation, the spectral difference chip was set such that an inhibitory spike results in a strong inhibition that outlasts the duration of the entire sweep. As a result, if an inhibitory cochlear spike precedes an excitatory cochlear spike, the spectral difference neuron will not fire a spike.

Since all neurons *below* the diagonal receive excitation before inhibition and since the SD neurons have a short integration time, their receptive fields are nearly identical to the corresponding excitatory cochlear receptive field (c=1, j) with the exception of one neuron with coordinates (i=4, j=5). The behavior of this neuron can be explained by observing that in Fig. 6.29 the spike generated from the 5^{th} right cochlear filter has a relatively long latency (because of mismatch), bringing it closer to the inhibitory spike from the 4^{th} right cochlear filter.

All SD neurons *above* the diagonal receive inhibition before excitation; since the duration of inhibition, if present, lasts for the entire duration of the sweep, the SD neurons can fire a spike *only* if the inhibitory cochlear neuron did not fire a spike. Observe that in the first row (j=1) of Fig. 6.40, the excitatory cochlear neuron (c=1, j=1) has a receptive field that covers almost the entire space; however, the SD neuron (i=2, j=1) did not fire any spikes because its inhibitory receptive field (c=1, j=2) is the same as the excitatory receptive field (c=1, j=1). On the other hand, the SD neuron (j=1, i=3) has an inhibitory receptive (c=1, j=3) that does not cover all directions allowing the SD neuron to fire spikes for the directions that do not provide any inhibition. The same applies to the rest of the neurons above the diagonal.

Consider now the entire response of the right and left spectral difference chips shown in Figs. 6.41 and 6.42 respectively. Observe that *below* the diagonal, neurons in the same *row* have similar receptive fields and fire for most of the space. Their receptive fields reflect that of their excitatory cochlear filters.

For neurons *above* the diagonal, it is observed that neurons in the same column have similar receptive fields. This is attributed to the fact that neurons in the same *column* receive the same inhibition. Consider, for example, neurons in the 14^{th} row and column. The white regions in the receptive fields of the 14^{th} row SD neurons indicate that no spikes were fired for those directions. This indicates no excitatory input for those directions and, in turn, no inhibition, which is reflected in the receptive fields of the 14^{th} column SD neurons.

Localization

Table 6.4 shows the localization performance of the hardware system operated in the single spike latency mode (with the integration time (d_{mem}) for all neurons set at approximately $200\mu s$) when using the spiking output of the 32 cochlear neurons ($L = 32$).

Recall that the localization acuity for the software-based, 32-bit threshold code matrix C_{th}^{32} (see Table 6.2) was σ_θ=6.5°, σ_{ϕ_m}=3.9°, and P=60%. Table 6.4 shows that for the code matrix corresponding to the spiking cochlea output, P=87%, which indicates that the cochlear neurons fire spikes for most directions as observed in Figs. 6.31 and 6.32. As a result, the localization accuracy is worse than the software-based 32-bit threshold code.

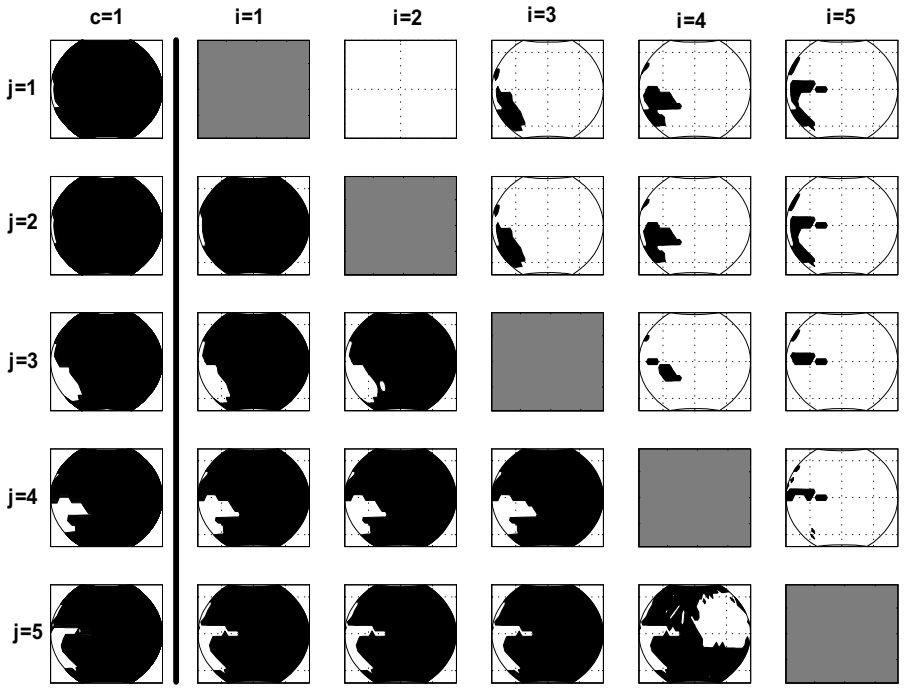

Figure 6.40: Right spectral difference chip. The leftmost column (c=1, j=1, 2, ...5) represents the binary cochlear receptive fields of the five cochlear neurons (filters). The 5 by 5 array (i, j=1, 2...5) to the right of the solid black line represents the receptive fields of the corresponding 20 spectral difference neurons (recall that there are no neurons along the diagonal).

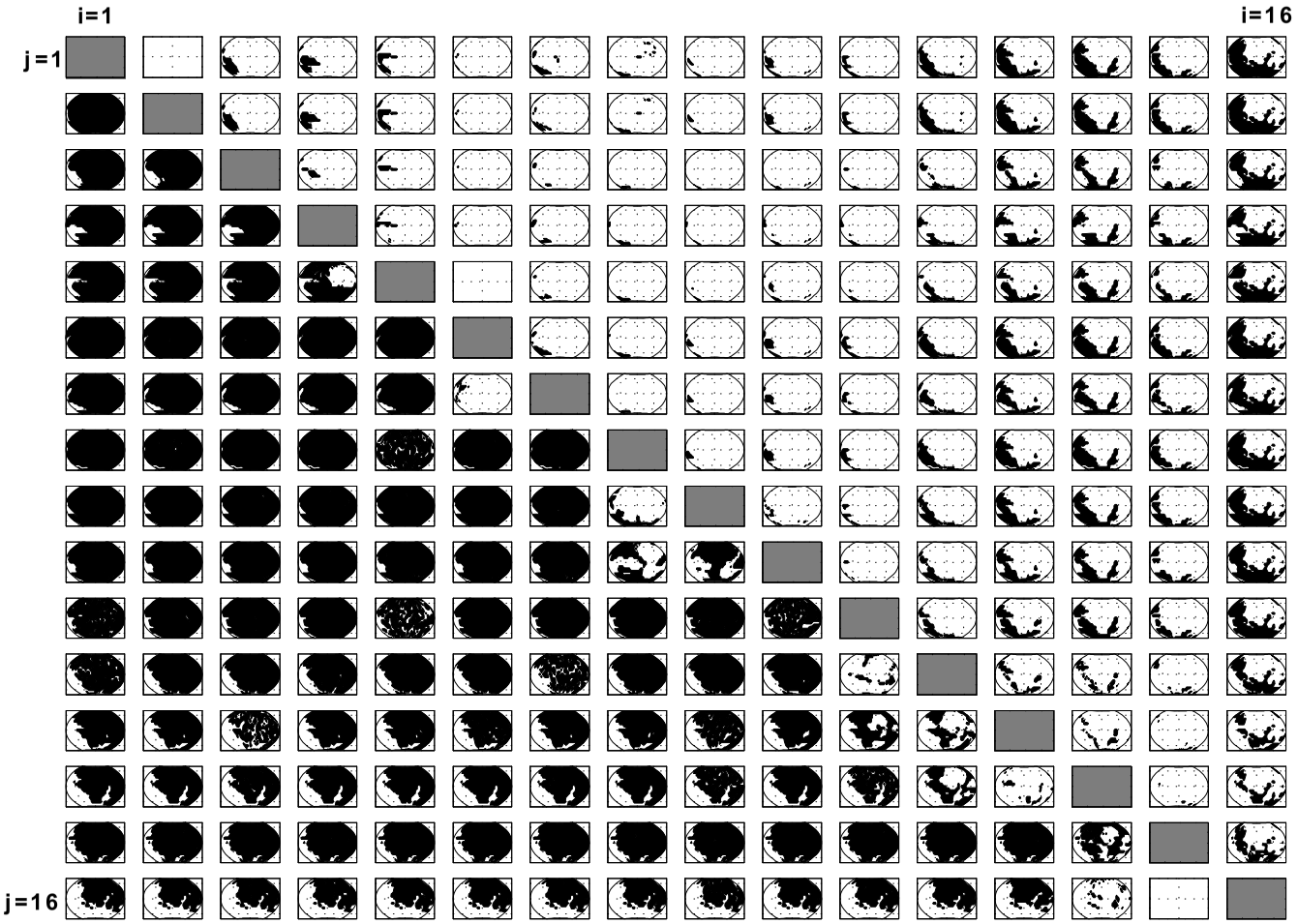

Figure 6.41: Binary receptive fields of the right spectral difference chip operated in the single spike mode. The chip has 240 neurons with no neurons along the diagonal (gray boxes). Each row (column) receives excitation (inhibition) from a single cochlear filter.

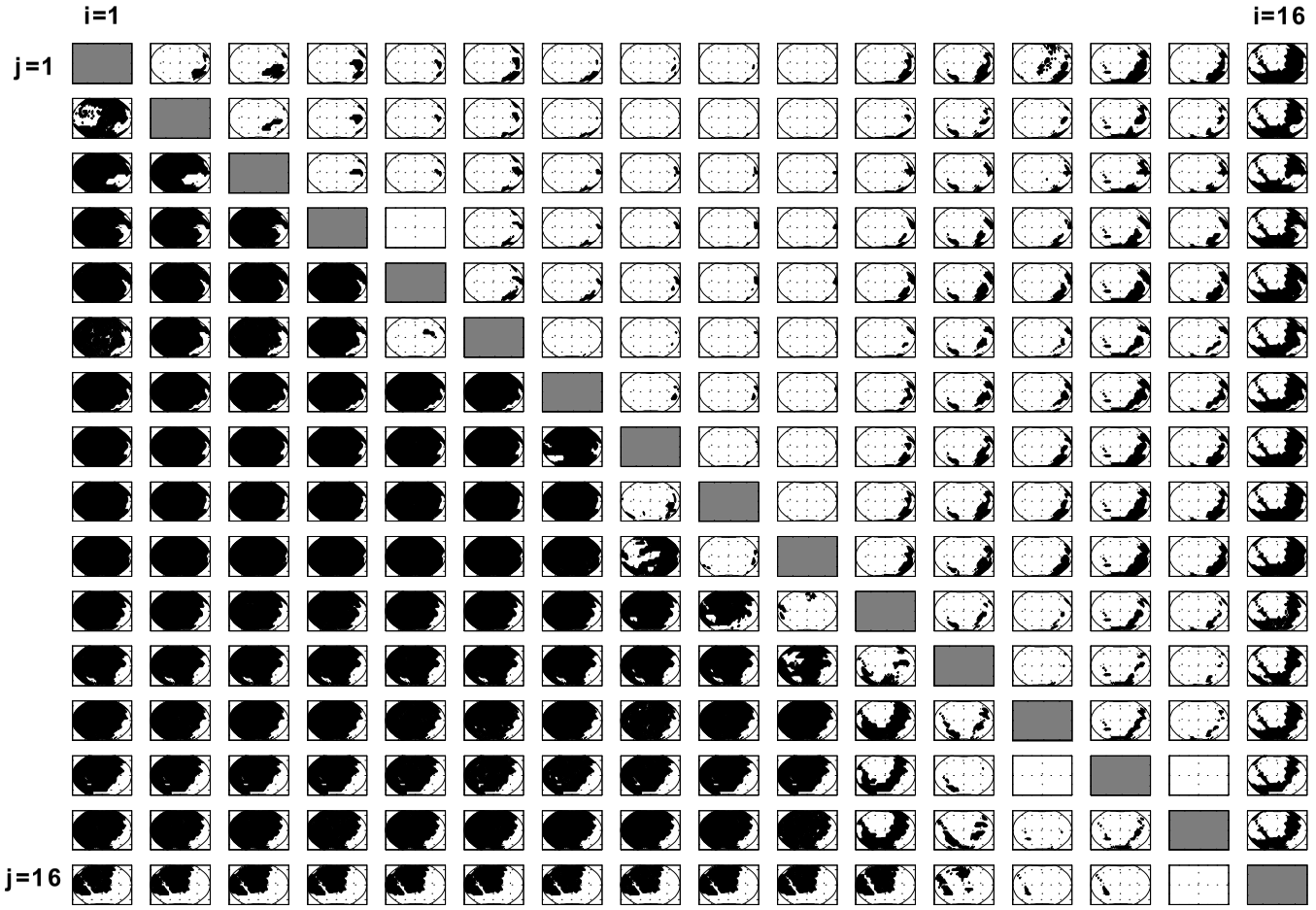

Figure 6.42: Response of the left spectral difference chip operated in the single spike mode. The chip has 240 neurons with no neurons along the diagonal (gray boxes). Each row (column) receives excitation (inhibition) from a single cochlear filter.

The code matrix formed by the spiking output of the two (right and left) spectral difference chips and the ILD chip will be referred to as C_{h1} (subscript h denotes hardware), which has 512 columns.

The integration time (d_{mem}) of the neurons of the right and left spectral difference chips was increased to $600\mu s$ and $580\mu s$ respectively. The response of the right and left spectral difference chips is shown in Figs. 6.43 and 6.44 respectively. The code matrix for this setting will be referred to as C_{h2}.

The integration time (d_{mem}) of the neurons of the right and left spectral difference chips was also increased to $1080\mu s$ and $976\mu s$ respectively. The response of the right and left spectral difference chips shown is in Figs. 6.45 and 6.46 respectively. The code matrix for this setting will be referred to as C_{h3}.

Recall that in Chapter 5 it was demonstrated that inhibition mostly occurred close to the diagonal where CF(i) $\simeq$CF(j) and the inhibitory spikes fire in close proximity to the excitatory spikes and that the amount of encroachment of inhibition below the diagonal (above the diagonal in Chapter 5 because of the opposite direction of the sweep) is related to the integration time of the neuron.

Similarly, it is observed here that as the integration time increases, the encroachment of inhibition below the diagonal increases. It is also observed that the encroachment into the higher frequency region (bottom of the plot) is greater than in the lower frequency region. This is attributed to the shape of the hyperbolic sweep where the first $80kHz$ of the FM sweep occurs within the first $2ms$.

Table 6.4: Localization performance of the hardware system operated in the single spike mode using the cochlea output. Only one cochlear neuron was turned on for each cochlear filter for a total number of 16 cochlear neurons per cochlea chip.

$xcorr$	L	P	NUC	mean(D)	median(D)	min(D)	max(D)	σ_θ	σ_{ϕ_m}
100	32	87	233	0	0	0	15	19.2	12.7

The localization performance for all three code matrices is shown in Table 6.5. As described earlier, neurons which do not contribute to localization (i.e. whose output does not change in response to any of the 703 measured directions) were not included and thus $L < 512$ in Table 6.5. All three matrices have a good localization performance (individually). The table shows that the feature extraction chips have successfully extracted the localization cues from the spiking cochlear output.

It should be noted that although localization using the spiking output of the cochlea chips resulted in a poor localization performance, the spiking cochlear output still contains information (in the latency of the spikes) about the direction of the sound, which was then extracted by the feature extraction chips.

The correlation between the three code matrices C_{h1}, C_{h2}, and C_{h3} is shown in Table 6.6. The correlation between the 32 columns corresponding to the ILD neurons (whose chip settings were not changed) have a correlation $\geqslant 97\%$.

Table 6.5: Localization performance of the code matrices C_{h1}, C_{h2}, and C_{h3} for the hardware system operated in the single spike mode.

	xcorr	L	P	NUC	mean(D)	median(D)	min(D)	max(D)	σ_θ	σ_{ϕ_m}
C_{h1}	100	505	43	701	11	7	0	119	0.4	0.4
C_{h2}	100	496	37	693	9	5	0	51	0.5	0.4
C_{h3}	100	476	31	675	6	3	0	60	0.7	0.5

The high correlation, for example between C_{h1} and C_{h2}, does not imply that the two matrices are interchangeable. We tested the localization accuracy by selecting a code word x from the code matrix C_{h2} and used C_{h1} to estimate the unknown direction. The mean error for estimating the modified azimuth was 1.8° with a standard deviation of 8.1° while the mean error for estimating the elevation was 2° with a standard deviation of 20.7°.

Table 6.7 shows the localization performance of the difference code matrix C_{h1}. The three feature extraction chips have a total of 512 neurons. The first row corresponds to the original C_{h1} matrix (after removing seven non-contributing dimensions). The remaining rows show the effect of limiting the cross-correlation as described before. The eighth row shows that the localization performance of the hardware-based difference code matrix with L=112, mean(D)=5, σ_θ=0.5°, and σ_{ϕ_m}=0.5°, is comparable to that of the software-based difference code C_{diff}^{124} (L=124, mean(D)=6, σ_θ=0°, and σ_{ϕ_m}=0.2°). We will refer to the code matrix corresponding to the eighth row as C_{h1}^{112}.

The last row in Table 6.7 shows that even when the maximum cross-correlation was limited to 50%, the code matrix (L=27 bits) still had a reasonably good localization performance (σ_θ=2.9° and σ_{ϕ_m}=2.0°) compared to the poor localization using the spiking output of the cochlea (see Table 6.4) where the code matrix (L=32) had σ_θ=19.2°, and σ_{ϕ_m}=12.7°. This shows that the feature extraction chips have successfully extracted additional information from the spiking cochlear output.

Fig. 6.47 shows the distance vector (D) of the code matrix C_{h1}^{112} as a function of elevation and azimuth. Although C_{h1}^{112} is a hardware-based difference code matrix, its distance plot resembles that of the threshold code matrix $C_{reduced}^{145}$ shown in Fig. 6.8.

Fig. 6.48 shows the spatial correlation plot for the code matrix C_{h1}^{112} for 16 different directions (listed at the top of each subplot where azm=azimuth and elv=elevation). Recall that the spatial correlation plot is calculated by cross-correlating a code from a single direction with codes from all 703 directions. The maximum correlation is equal to L (L=112) and decreases by two for every one bit difference. The smallest contour level, 52, shows all the directions that are within 30 bits of error to the selected direction, while the highest contour level, 106, shows all the directions that are within three bits of error to to the selected direction.

Fig. 6.49 shows the cross-correlation of the 112 columns of the code matrix C_{h1}^{112}. Similar to the software-based difference code matrix C_{diff}^{124}, the columns of C_{h1}^{112} are less correlated than those of the threshold code matrix $C_{reduced}^{145}$.

Table 6.6: Correlation between the code matrices C_{h1}, C_{h2}, and C_{h3}.

code matrix	C_{h1}	C_{h2}	C_{h3}
C_{h1}	100	83.2	69.6
C_{h2}	83.2	100	83
C_{h3}	69.6	83	100

6.4.2 Multiple Spike Mode

In this mode of operation, each cochlear filter feeds its output into more than one cochlear neuron (up to a maximum of eight). The cochlear neurons are set to have different thresholds. Recall that in the multiple spike mode of operation (see section 5.2) the duration of inhibition in the feature extraction chips does not extend for the entire duration of the sweep and that additional spikes extend the duration of inhibition. Figs. 6.50 and 6.51 show the binary receptive fields of the spectral difference chips being operated in the multiple spike mode with three cochlear neurons turned on for each cochlear filter. The inhibitory weight was set be stronger than the excitatory weight by approximately $30mV$. The duration of the inhibition was set to be longer than excitation (approximately $1ms$ duration in response to a single spike). Observe that above the diagonal, the further the neuron is from the diagonal, the less it is affected by inhibition. This can be clearly seen the rightmost column (column number 16) in Fig. 6.50 where the neuron

Table 6.7: Localization performance of the code matrix C_{h1} obtained from the hardware system operated in the single spike mode. The average integration time for all neurons is 0.2ms.

$xcorr$	L	P	NUC	mean(D)	median(D)	min(D)	max(D)	σ_θ	σ_{ϕ_m}
100	505	43	701	11	7	0	119	0.4	0.4
99	359	33	701	9	6	0	60	0.4	0.4
98	315	32	700	9	6	0	53	0.4	0.5
97	290	32	699	9	6	0	48	0.5	0.5
96	253	33	699	8	6	0	45	0.5	0.5
95	236	33	699	8	6	0	42	0.5	0.5
90	156	38	693	6	5	0	30	0.5	0.5
85	112	40	695	5	4	0	25	0.5	0.5
80	82	43	688	4	4	0	20	0.6	0.6
75	68	42	686	4	3	0	13	0.9	0.8
70	56	45	686	3	3	0	13	0.9	0.7
65	37	48	654	2	2	0	6	2.0	1.1
60	34	51	657	2	2	0	6	1.3	1.3
55	31	53	654	2	2	0	6	1.5	1.4
50	27	51	596	1	1	0	5	2.9	2.0

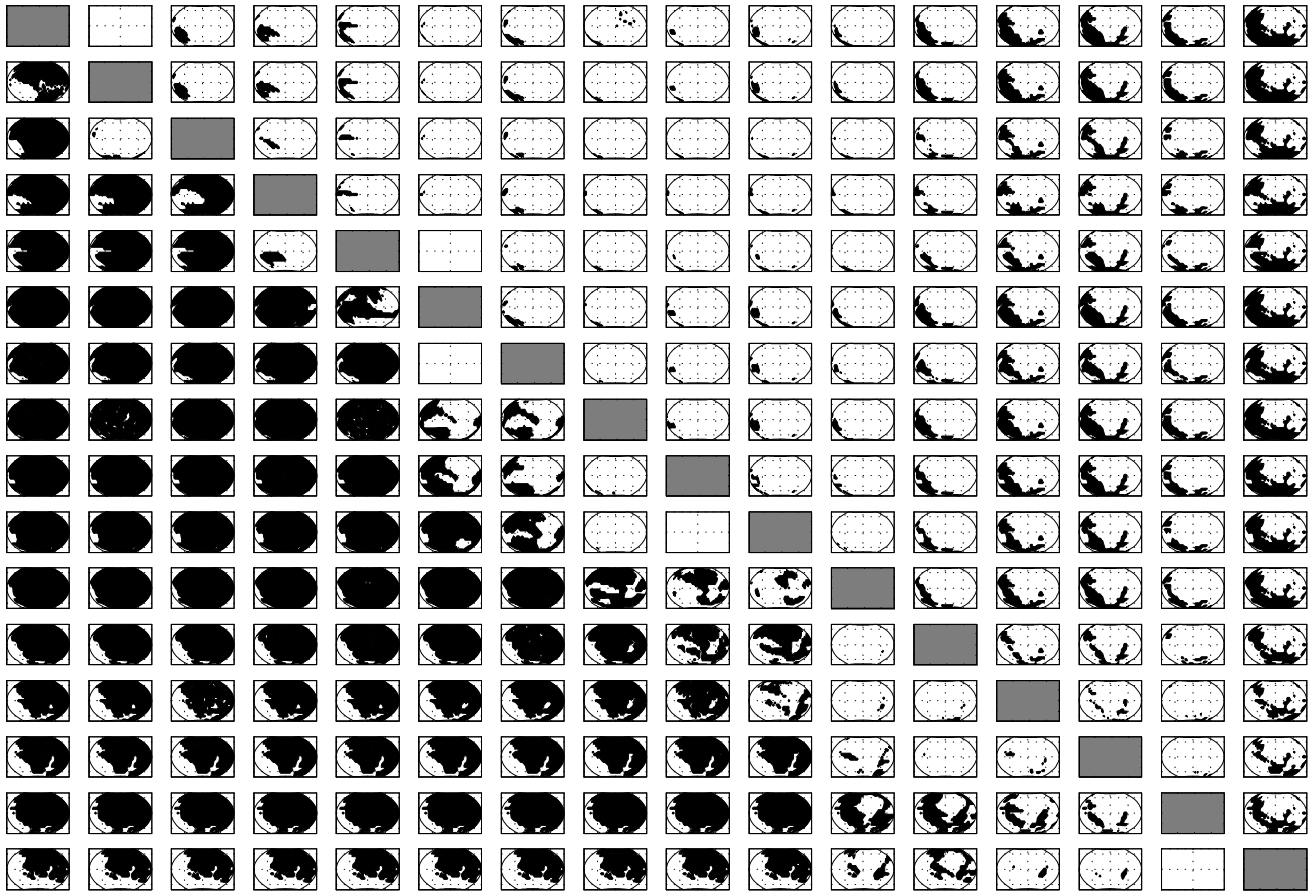

Figure 6.43: Response of the right spectral difference chip operated in the single spike mode. The neurons were set to have an integration time of approximatley 0.6ms. The chip has 240 neurons with no neurons along the diagonal (gray boxes). Each row (column) receives excitation (inhibition) from a single cochlear filter.

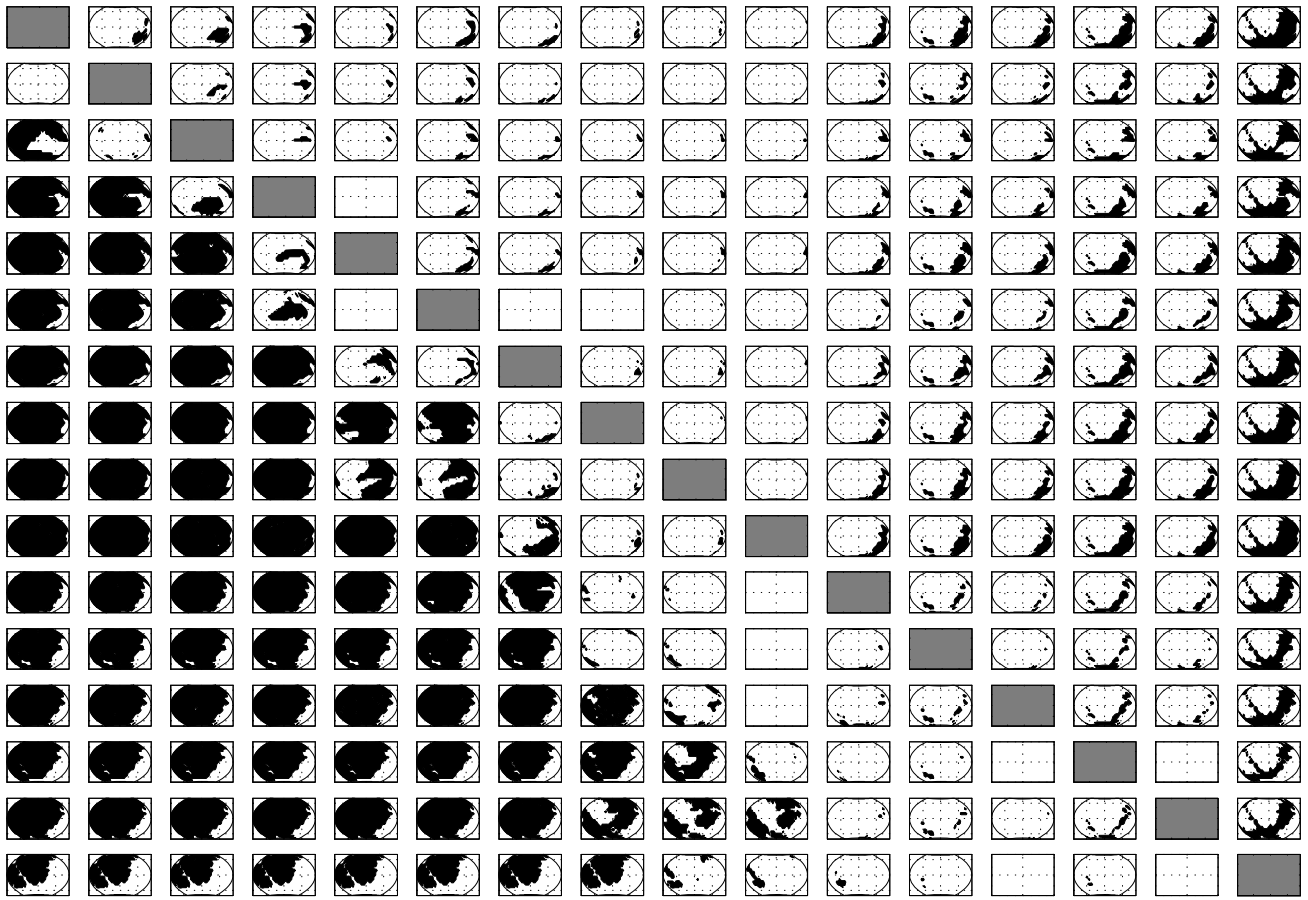

Figure 6.44: Response of the left spectral difference chip operated in the single spike mode. The neurons were set to have an integration time of approximatley 0.58ms. The chip has 240 neurons with no neurons along the diagonal (gray boxes). Each row (column) receives excitation (inhibition) from a single cochlear filter.

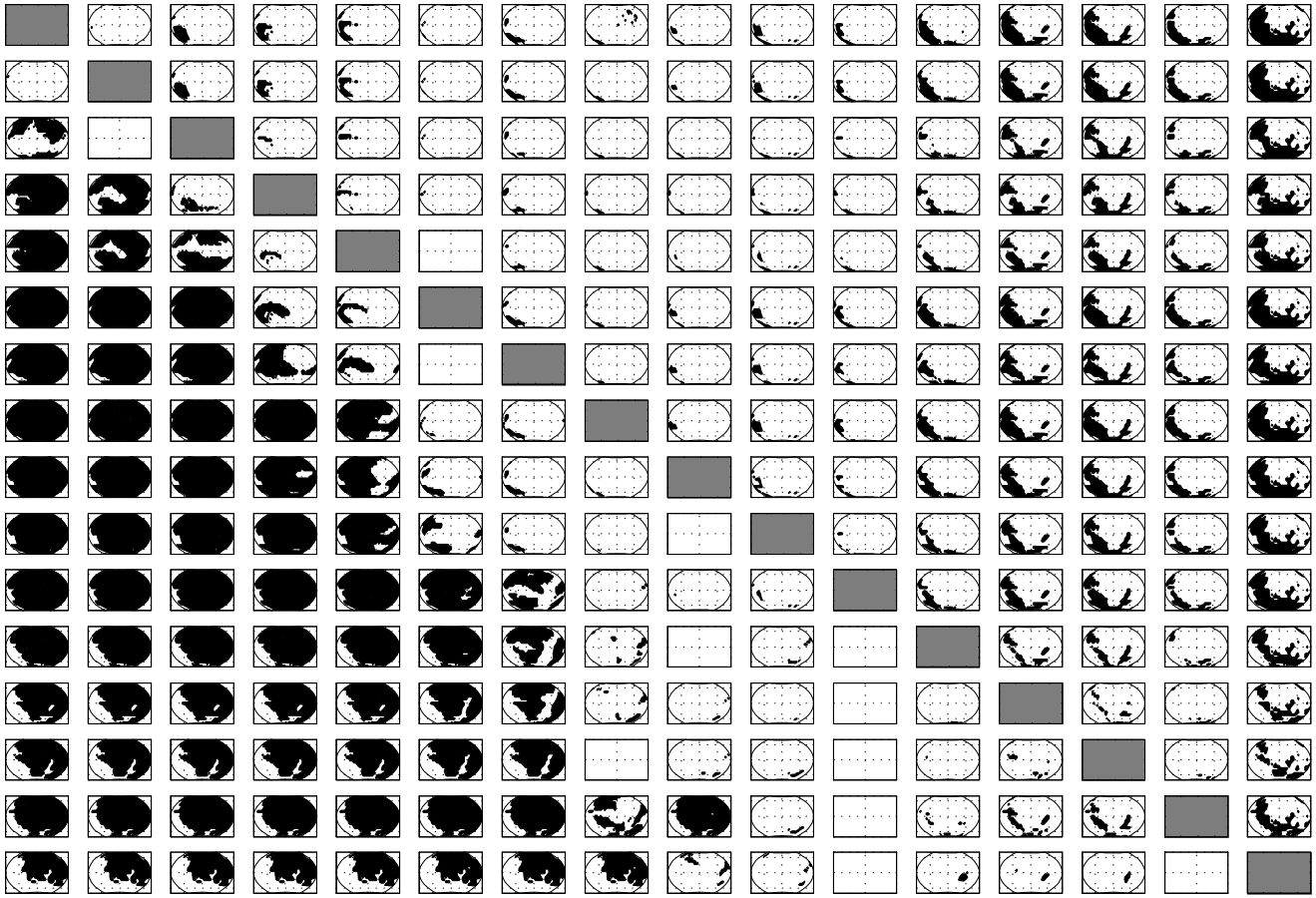

Figure 6.45: Response of the right spectral difference chip operated in the single spike mode. The neurons were set to have an integration time of approximatley 1ms. The chip has 240 neurons with no neurons along the diagonal (gray boxes). Each row (column) receives excitation (inhibition) from a single cochlear filter.

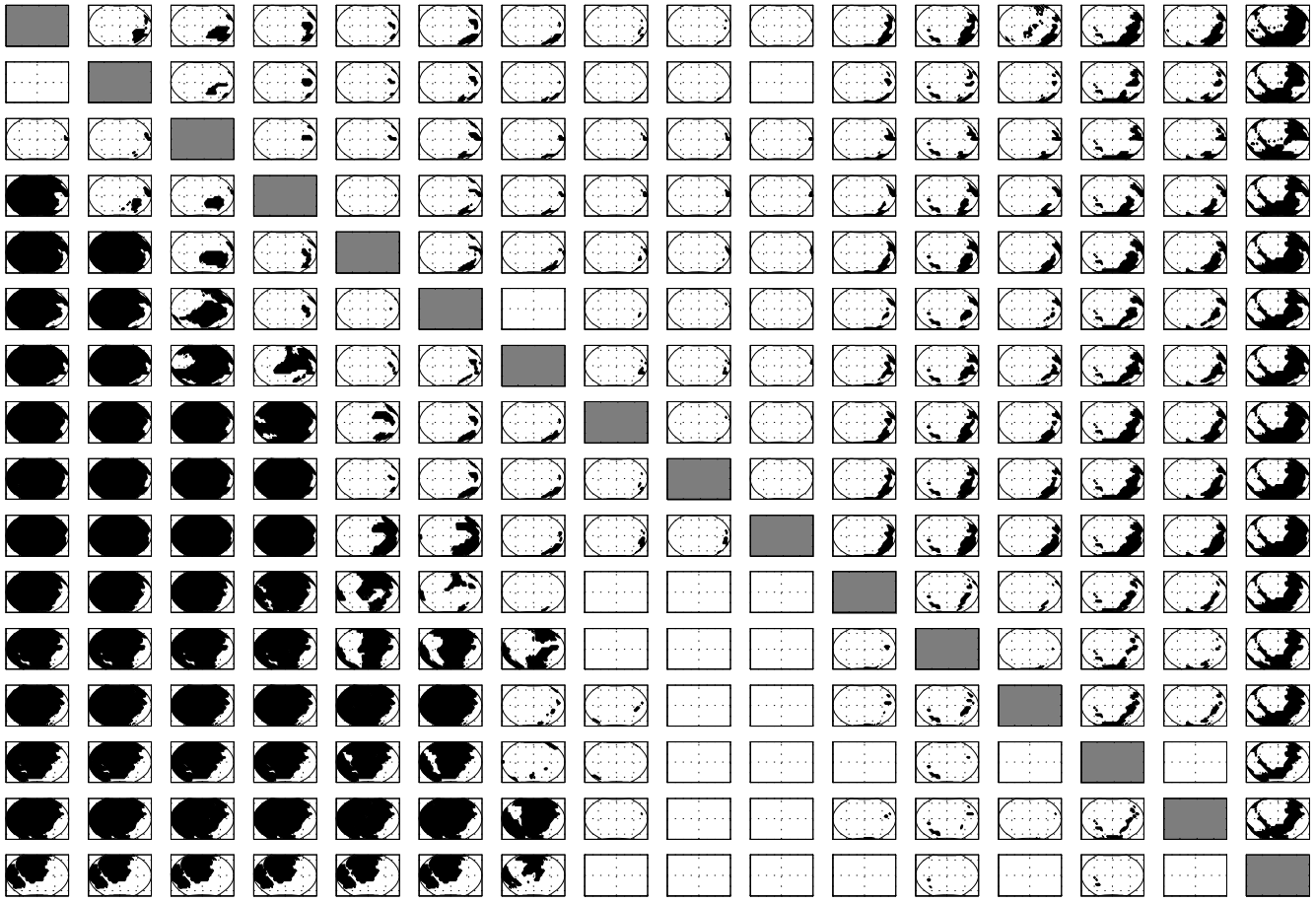

Figure 6.46: Response of the left spectral difference chip operated in the single spike mode. The neurons were set to have an integration time of approximatley 1ms. The chip has 240 neurons with no neurons along the diagonal (gray boxes). Each row (column) receives excitation (inhibition) from a single cochlear filter.

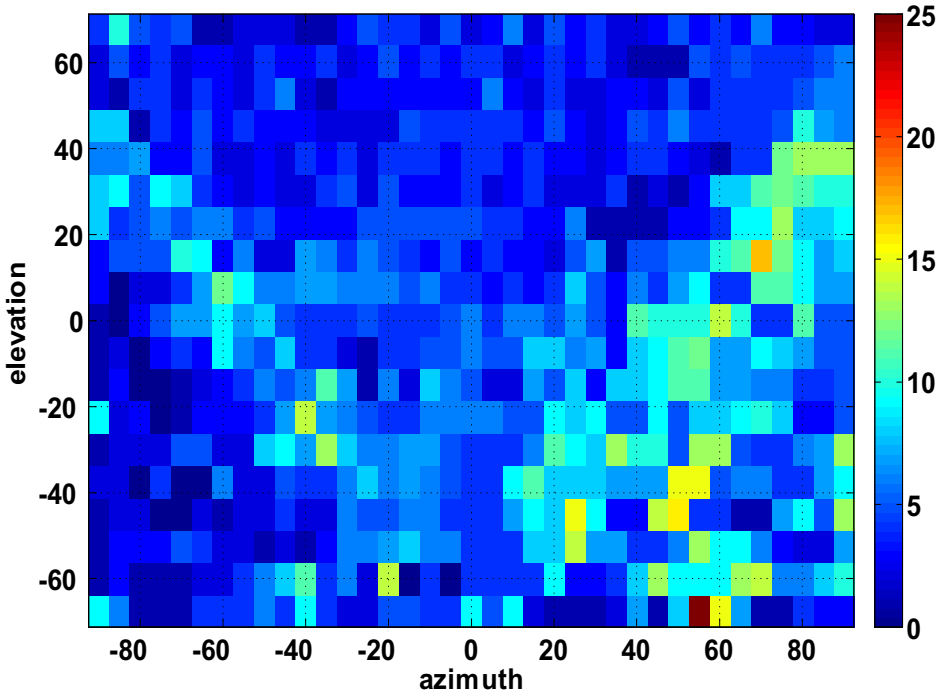

Figure 6.47: Distance vector (D) for the hardware-based difference code matrix C_{h1}^{112} as a function of azimuth and elevation. Although C_{h1}^{112} is a hardware-based difference code matrix, its distance plot resembles that of the threshold code matrix $C_{reduced}^{145}$ shown in Fig. 6.11.

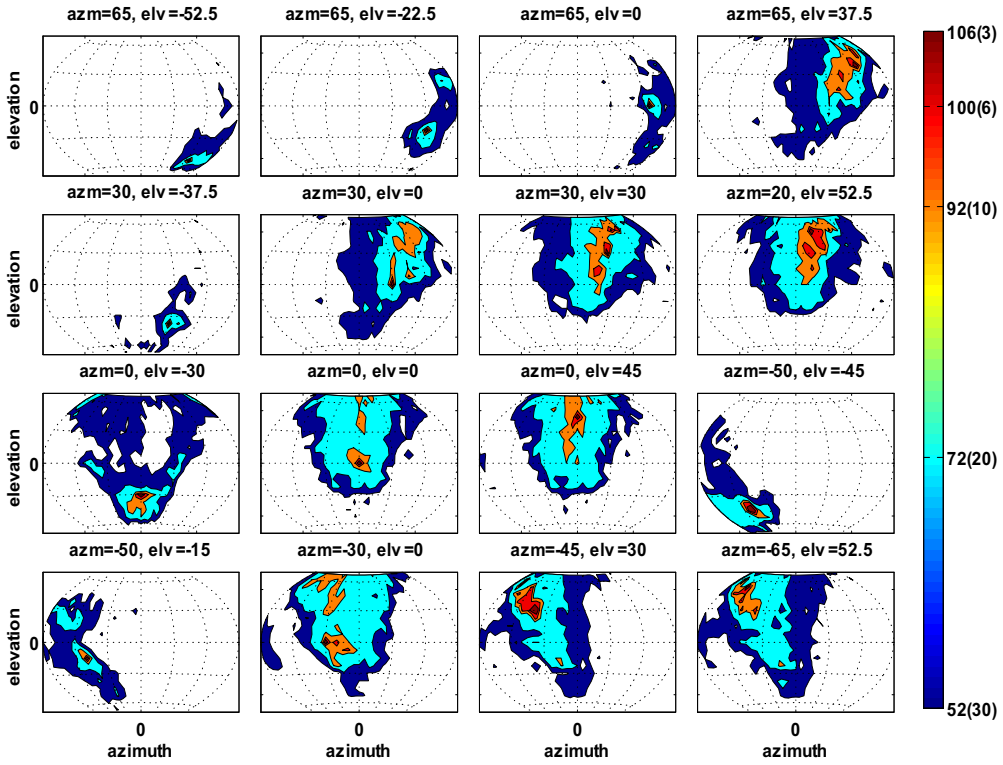

Figure 6.48: Spatial correlation plots for the hardware-based difference code matrix C_{h1}^{112}. Shown is the spatial correlation plot for 16 different directions (listed at the top of each subplot where azm=azimuth and elv=elevation). The spatial correlation plot is calculated by cross-correlating a code from a single direction with codes from all 703 directions. The maximum correlation is equal to L (L=112) and decreases by two for every one bit difference. The codes were selected from the threshold code matrix C_{h1}^{112}. The smallest contour level, 52, shows all the directions that are within 30 bits of error to the selected direction, while the highest contour level, 106, shows all the directions that are within three bits of error to the selected direction.

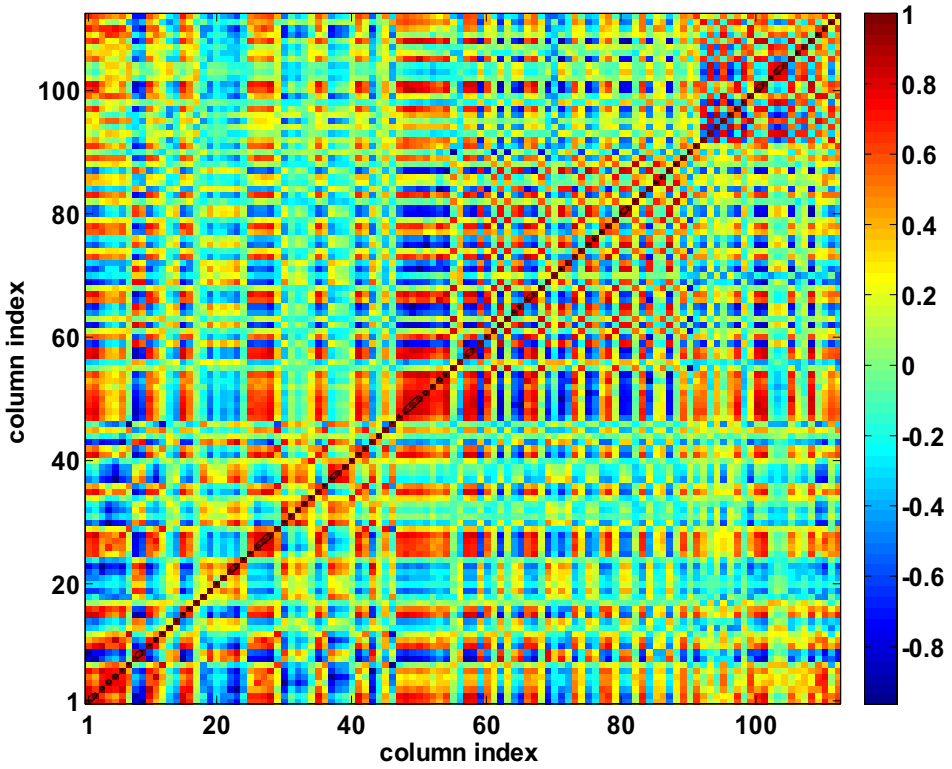

Figure 6.49: Cross-correlation of the 112 columns (dimensions) of the hardware-based difference code matrix C_{h1}^{112}. Similar to the software-based difference code matrix C_{diff}^{124}, the columns of C_{h1}^{112} are less correlated than those of the threshold code matrix $C_{reduced}^{145}$.

Table 6.8: Localization performance of the hardware system operated in the multiple spike mode using the cochlea output. Three cochlear neurons were turned on for each cochlear filter for a total number of 48 cochlear neurons per cochlea chip.

$xcorr$	L	P	NUC	mean(D)	median(D)	min(D)	max(D)	σ_θ	σ_{ϕ_m}
100	96	70	564	2	1	0	9	3.2	1.9

closest to the gray box is almost entirely inhibited and the inhibition decreases as we move up in the column. Fig. 6.52 shows the binary receptive fields of the 32 ILD neurons.

Table 6.8 shows the localization performance of the hardware system operated in the multiple spike mode using the output of the cochlea chip. Since there are three neurons that are turned on for each cochlear filter, there are a total of 48 neurons (that are turned on) in each cochlea chip. The table shows that the localization performance of the hardware system using the spiking output of the cochlea chips operated in the multiple spike mode is much better than the corresponding performance in the single spike mode. This increase in localization acuity is attributed to the increase in the number of neurons (with different thresholds).

The localization performance of the 512-column code matrix is shown in Table 6.9. The first row shows that the mean distance of the code matrix operated in the multiple spike mode (30) is greater than that operated in the sin-

gle spike mode (11). The tenth row shows that the localization performance of the hardware-based difference-code matrix with $L = 131$ (referred to as C_h^{131}), mean(D)=13, σ_θ=0°, and σ_{ϕ_m}=0.1° is comparable to that of the software-based difference code C_{diff}^{124} (L=124, mean(D)=6, σ_θ=0°, and σ_{ϕ_m}=0.2°).

It is not clear why some of the neural responses appear to be "noisy" as observed in the first row of the right spectral difference response. Nevertheless, the measurement was repeated a month later with nearly identical chip settings and the two code matrices had a correlation of 84%. We tested the localization accuracy by selecting a code word x from one code matrix and used the other matrix to estimate the unknown direction. The mean error for estimating the modified azimuth was -0.3° with a standard deviation of 5.6°, while the mean error for estimating the elevation was -2.3° with a standard deviation of 9.7°.

Fig. 6.54 shows the distance vector (D) for the threshold code matrix C_h^{131} as a function of azimuth and elevation. Similar to the software-based difference code C_{diff}^{124}, the the region of maximum code distance covers the central direction (azimuth=0, elevation=0) and the surrounding directions.

Fig. 6.55 shows the spatial correlation plots for the hardware-based difference code matrix C_h^{131}. Shown is the spatial correlation plot for 16 different directions (listed at the top of each subplot where azm=azimuth and elv=elevation). The maximum correlation is equal to L (L=131) and decreases by two for every one bit difference. The smallest contour level, 71, shows all the directions that are within 30 bits of error to the selected direction, while the highest contour level, 125, shows all the directions that are within three bits of error to to the selected

direction.

Fig. 6.56 shows the cross-correlation between the columns of the code matrix C_h^{131}. Similar to the software-based difference code C_{diff}^{124}, the columns of the matrix are less correlated compared to those of the threshold code matrix $C_{reduced}^{145}$.

Table 6.10 shows the localization performance of the hardware system operated in the multiple spike mode using the spiking output of the two cochlea chips. As described earlier, three neurons were turned on for each of the 16 cochlear filters for a total of 48 cochlear neurons per cochlea. The localization performance is better than that of single spike mode (see Table 6.4) when only one neuron per cochlear filter was turned on.

We extended the duration of the inhibition for the spectral difference chips operated in the multiple spike mode (approximately $2ms$ in response to a single spike). The response of the right spectral difference chip is shown in Fig. 6.53. The extent of inhibition (above the diagonal) extends to the neuron furthest from the diagonal. The code matrix corresponding to this chip setting has a mean distance of 23 and the number of unique codes (NUC) is equal to 703 (thus zero localization error).

Other Frequency Sweeps

In addition to measuring the response of the artificial bat head to a hyperbolic sweep, three other frequency sweeps were also used, namely, a linear sweep, a

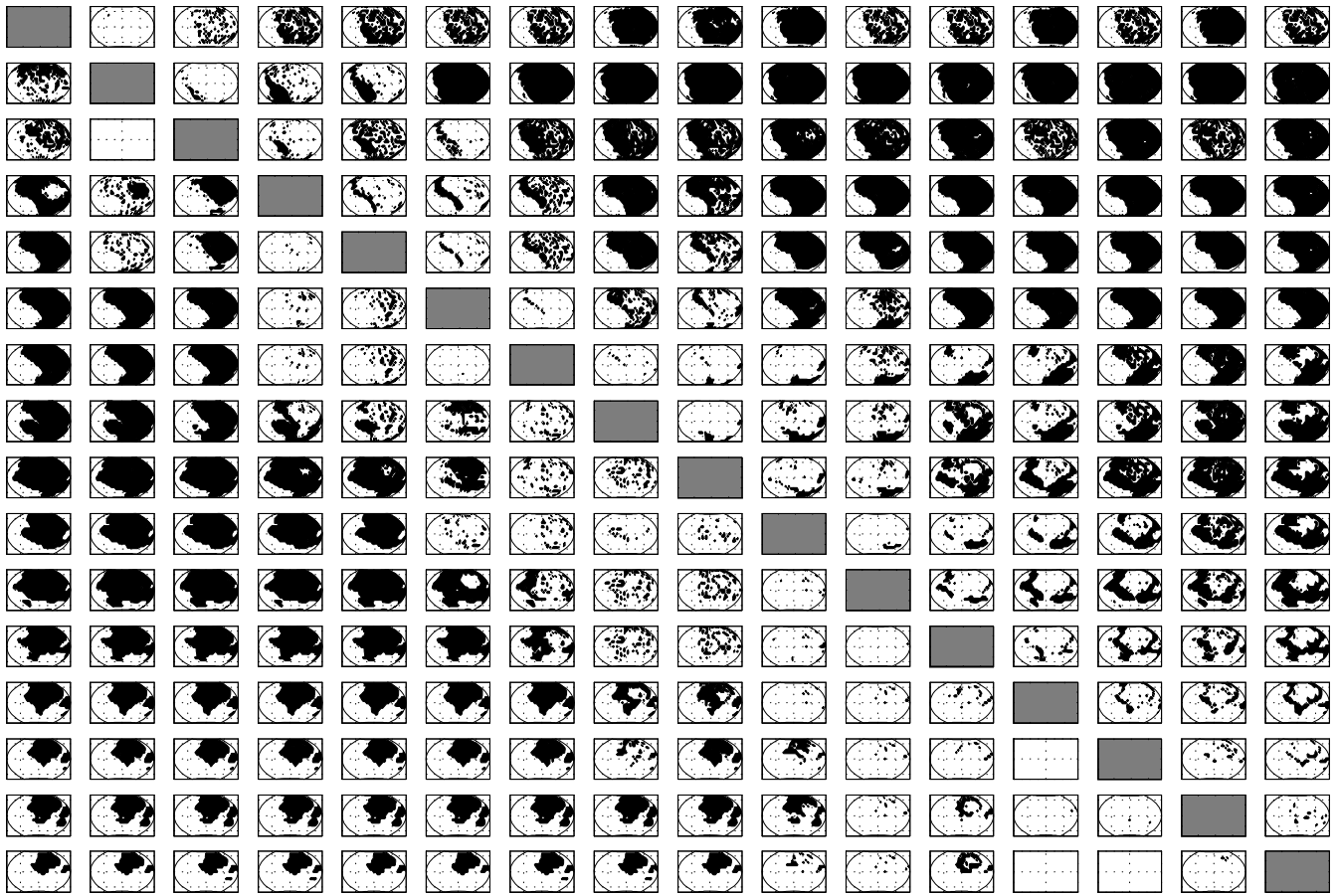

Figure 6.50: Response of the right spectral difference chip operated in the multiple spike mode. The chip has 240 neurons with no neurons along the diagonal (gray boxes). Each row (column) receives excitation (inhibition) from a single cochlear filter.

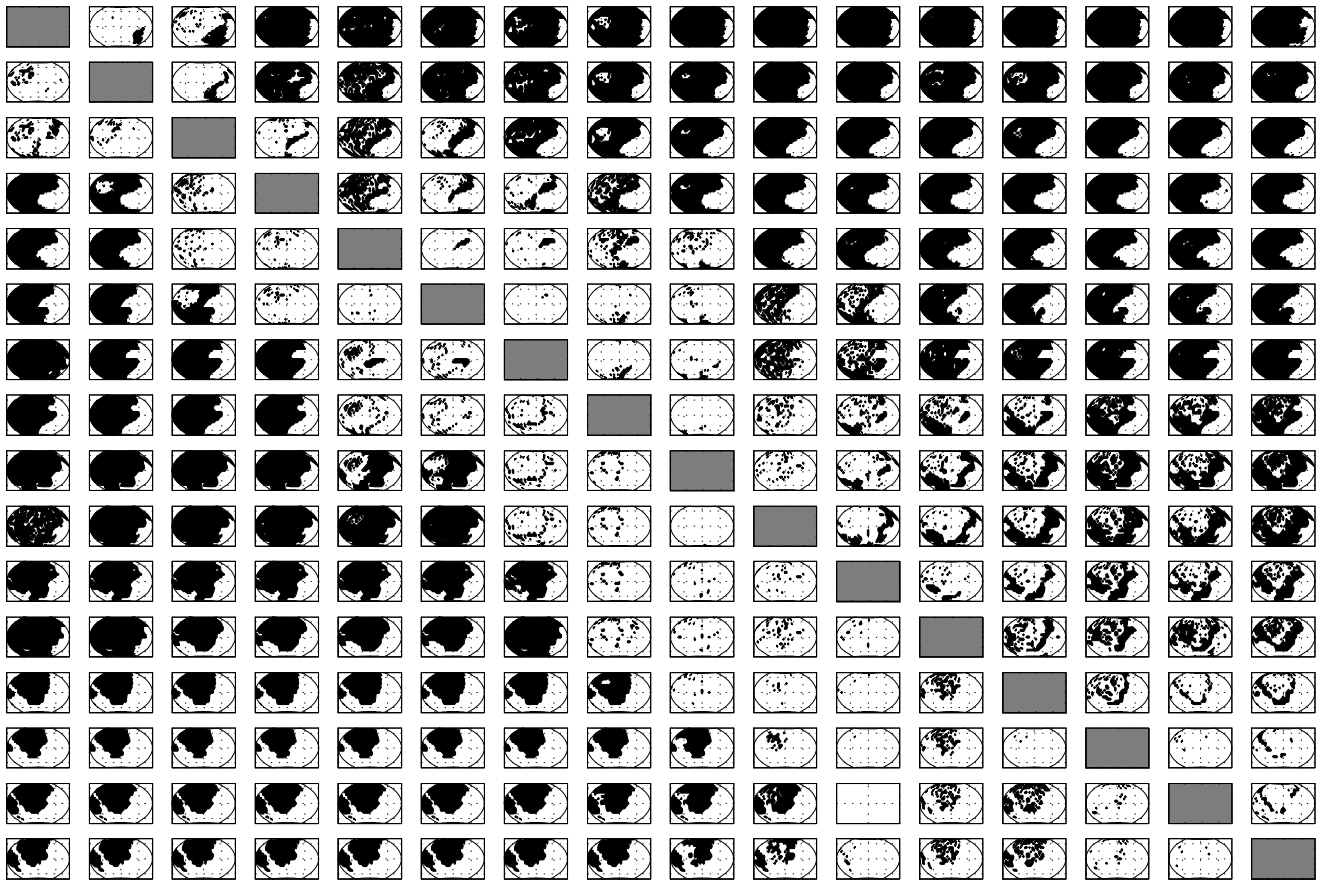

Figure 6.51: Response of the left spectral difference chip operated in the multiple spike mode. The chip has 240 neurons with no neurons along the diagonal (gray boxes). Each row (column) receives excitation (inhibition) from a single cochlear filter.

Table 6.9: Localization performance of the hardware system operated in the multiple spike mode.

$xcorr$	L	P	NUC	mean(D)	median(D)	min(D)	max(D)	σ_θ	σ_{ϕ_m}
100	507	41	703	30	30	4	75	0	0
99	378	37	703	27	28	4	63	0	0
98	346	36	703	26	26	3	59	0	0
97	333	36	703	26	26	3	58	0	0
96	314	36	703	25	25	2	55	0	0
95	297	36	703	24	24	2	55	0	0
90	238	36	703	20	20	2	45	0	0
85	196	38	703	18	18	2	40	0	0
80	166	38	703	16	15	1	38	0	0
75	131	40	702	13	13	0	30	0	0.1
70	101	40	701	10	10	0	24	0.2	0.1
65	85	41	701	9	9	0	20	0.3	0.2
60	72	42	697	8	8	0	18	0.5	0.4
55	60	42	700	6	6	0	14	0.3	0.2
50	46	42	694	5	5	0	11	0.3	0.4

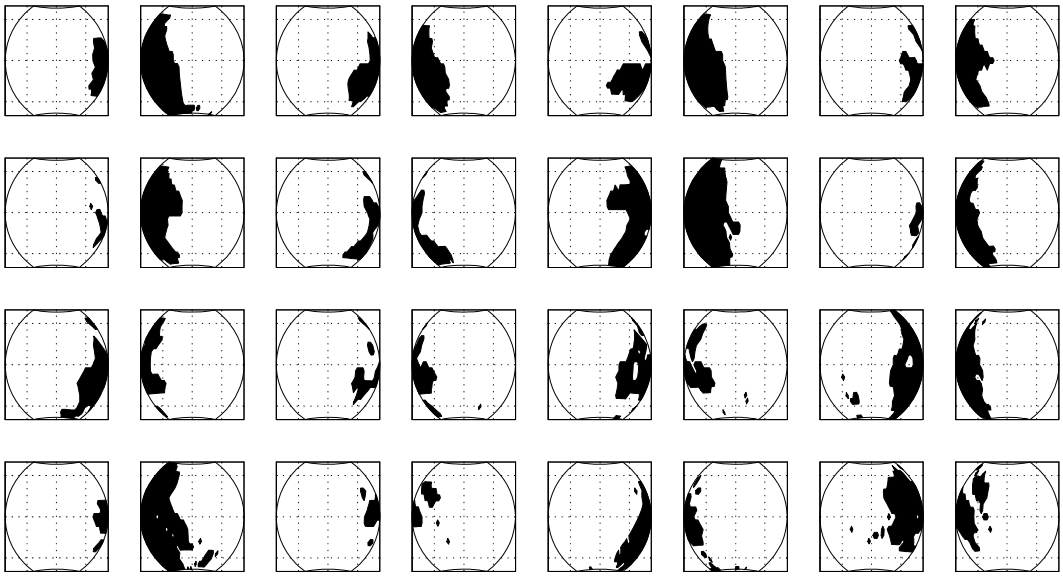

Figure 6.52: Binary receptive fields of the 32 ILD neurons.

Table 6.10: Localization performance of the hardware system operated in the multiple spike mode using the spiking output of the two cochlea chips. There are three cochlear neurons that were turned on for each of the 16 cochlear filters for a total of 48 cochlear neurons per cochlea chip.

$xcorr$	L	P	NUC	mean(D)	median(D)	min(D)	max(D)	σ_θ	σ_{ϕ_m}
100	96	70	564	2	1	0	9	3.2	1.9
95	80	67	552	1	1	0	7	3.3	2.0
90	58	64	502	1	1	0	5	4.1	2.3
85	43	62	440	1	0	0	3	5.1	3.1
80	30	61	330	0	0	0	3	6.6	4.2

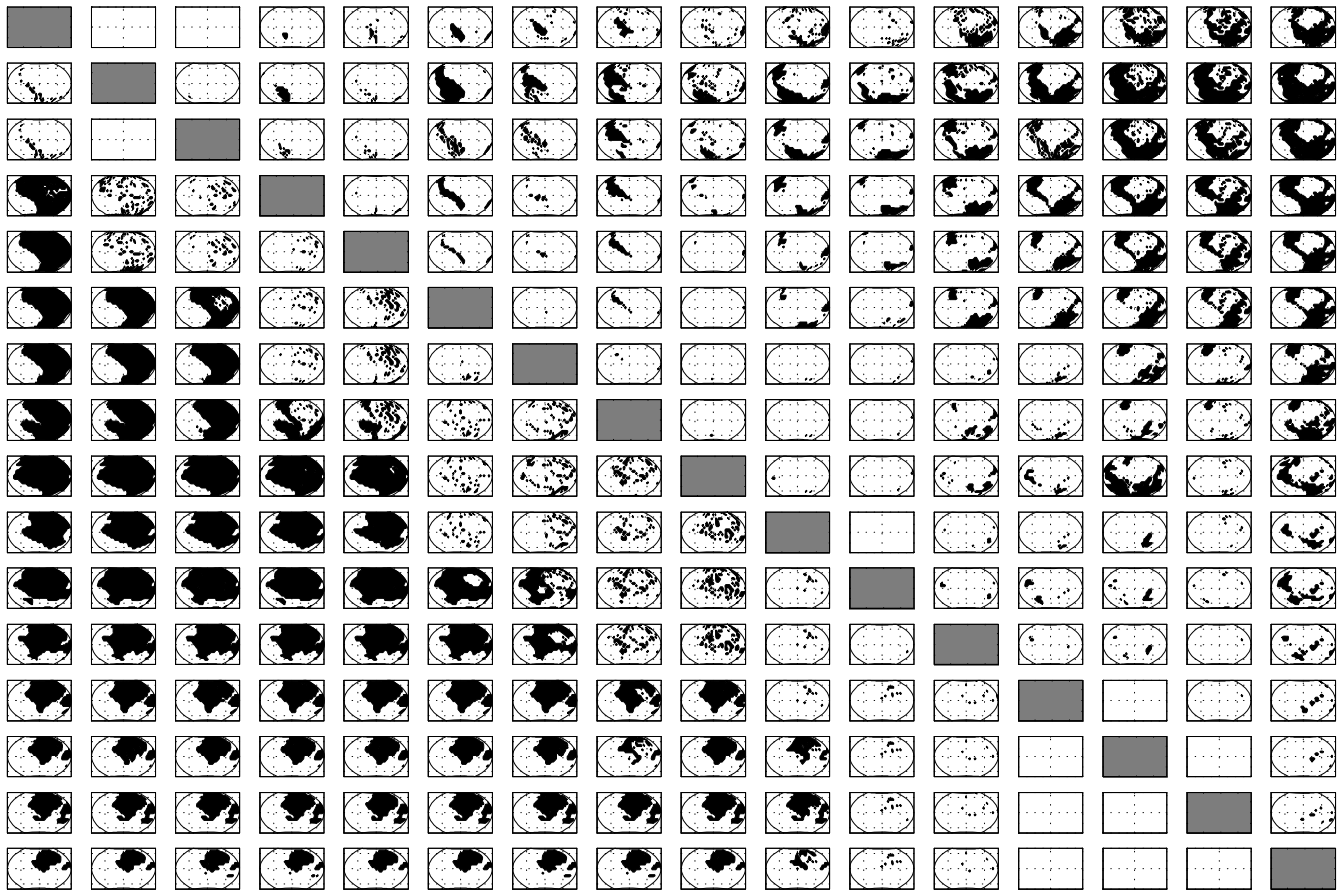

Figure 6.53: Response of the right spectral difference chip operated in the multiple spike mode after extending the duration of the inhibition. The chip has 240 neurons with no neurons along the diagonal (gray boxes). Each row (column) receives excitation (inhibition) from a single cochlear filter.

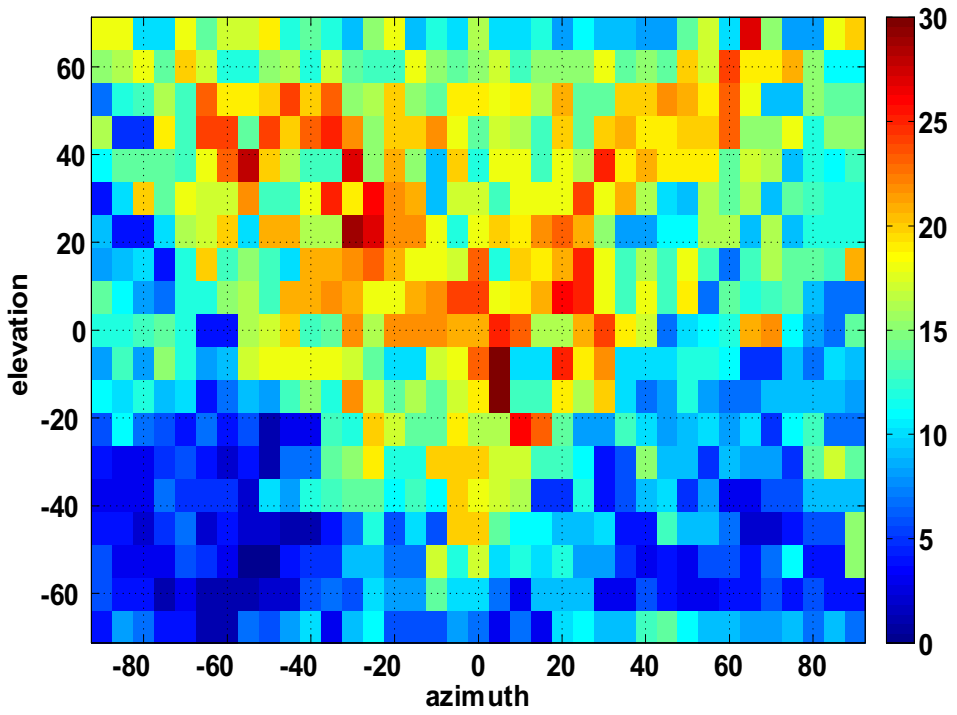

Figure 6.54: Distance vector (D) for the threshold code matrix C_h^{131} as a function of azimuth and elevation.

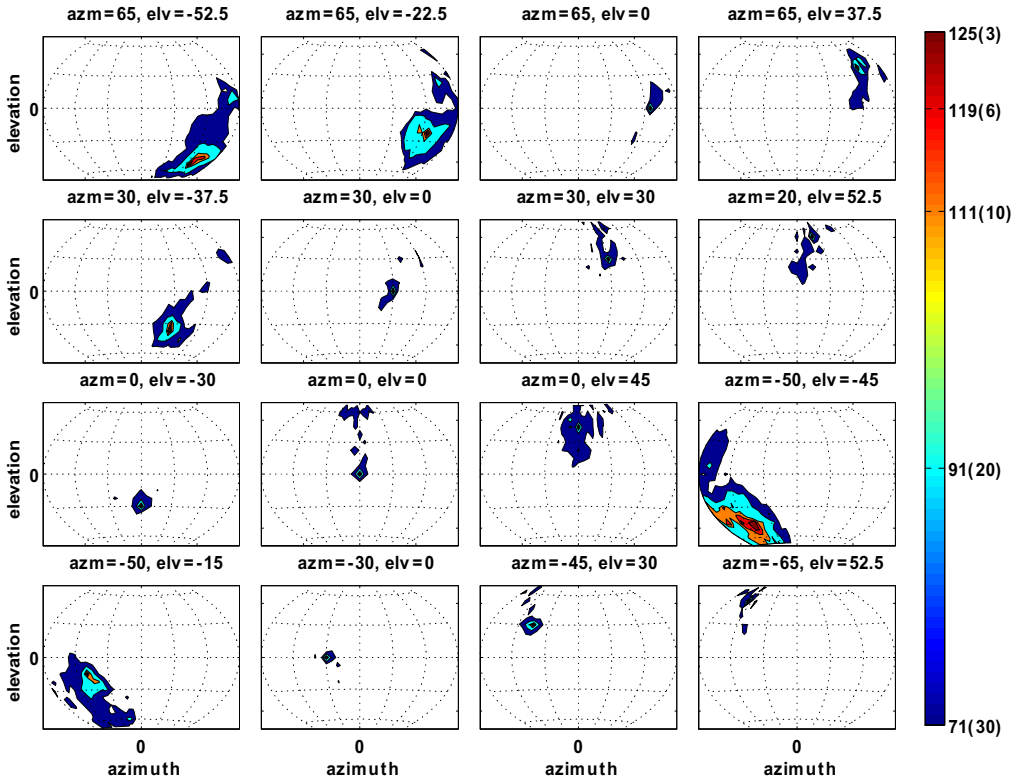

Figure 6.55: Spatial correlation plots for the hardware-based difference code matrix C_h^{131}. Shown is the spatial correlation plot for 16 different directions (listed at the top of each subplot where azm=azimuth and elv=elevation). The maximum correlation is equal to L (L=131) and decreases by two for every one bit difference. The smallest contour level, 71, shows all the directions that are within 30 bits of error to the selected direction, while the highest contour level, 125, shows all the directions that are within three bits of error to the selected direction.

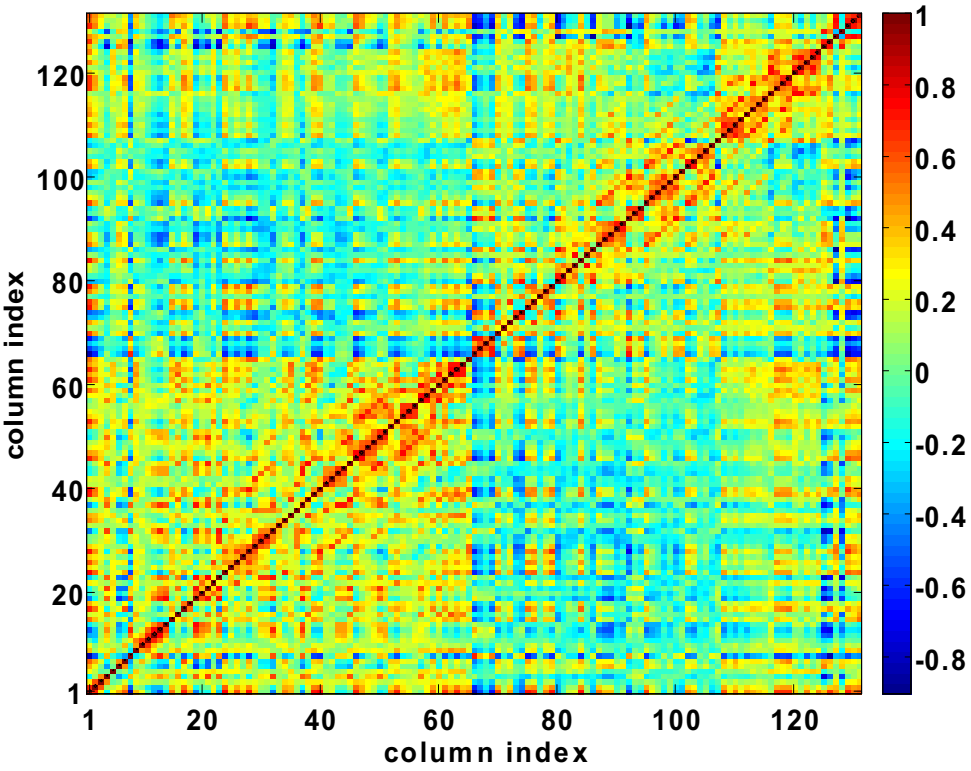

Figure 6.56: Cross-correlation of the 131 columns (dimensions) of the code matrix C_h^{131}.

logarithmic sweep, and a logarithmic sweep along with its second harmonic. Fig. 6.57 shows the instantaneous frequency as a function of time for all four sweeps which are $5ms$ in duration. The frequency of the hyperbolic sweep changes from $120kHz$ to $18kHz$. The frequency of the linear sweep changes from $110kHz$ to $15kHz$. The frequency of the logarithmic sweep changes from $120kHz$ to $16kHz$. In the logarithmic plus harmonic sweep, the fundamental sweeps from $55kHz$ to $15kHz$ (the second harmonic sweeps from $110kHz$ to $30kHz$). It is observed that the instantaneous frequency of the logarithmic sweep is the closest to the hyperbolic sweep.

The hardware localization systems was tested with all four frequency sweeps using the same chip settings. The hardware system was set to operate in the multiple spike mode with three cochlear neurons turned on for each cochlear filter. An additional ILD chip was connected to the system; the two ILD chips have different ILD thresholds. Fig. 6.58 shows the binary receptive fields for 64 ILD neurons taken from two ILD chips in response to the hyperbolic sweep. The top four rows correspond to one ILD chip while the bottom four rows correspond to the second ILD chip.

The response of the right spectral difference chip to the hyperbolic, linear, logarithmic, and logarithmic (plus harmonic) sweeps are shown in Figs. 6.59, 6.60, 6.61, and 6.62 respectively.

The localization performance of the four sweeps is shown in Table 6.11 where the code matrices $C1$, $C2$, $C3$, and $C4$ correspond to the hyperbolic, logarithmic, linear, and logarithmic (plus harmonic) respectively. The table shows a good

Table 6.11: Localization performance of the hardware system operated in the multiple spike mode in response to four different frequency sweeps. The code matrices C1, C2, C3, and C4 correspond to the hyperbolic, logarithmic, linear, and logarithmic plus harmonic sweeps respectively.

	xcorr	L	P	NUC	mean(D)	median(D)	min(D)	max(D)	σ_θ	σ_{ϕ_m}
$C1$	100	520	25	703	27	27	2	59	0	0
$C2$	100	523	26	703	26	26	3	68	0	0
$C3$	100	493	28	703	19	18	1	68	0	0
$C4$	100	476	25	703	27	25	2	74	0	0

localization performance for all four sweeps.

Table 6.12 shows the correlation between the four code matrices. The ILD part of the code had a correlation $\geqslant 82\%$ between all four codes. For $C1$ (hyperbolic sweep) and $C2$ (logarithmic sweep), the ILD part had a 97% correlation.

Both Tables 6.11 and 6.12 are in agreement with the observation that the logarithmic sweep is similar to the hyperbolic sweep. We tested the localization accuracy by selecting a code word (x) from the code matrix $C2$ (logarithmic sweep), and using $C1$ (hyperbolic sweep) to estimate the unknown direction; the mean error for estimating the modified azimuth was 0° with a standard deviation of 4.4° while the mean error for estimating the elevation was -0.4° with a standard deviation of 7°.

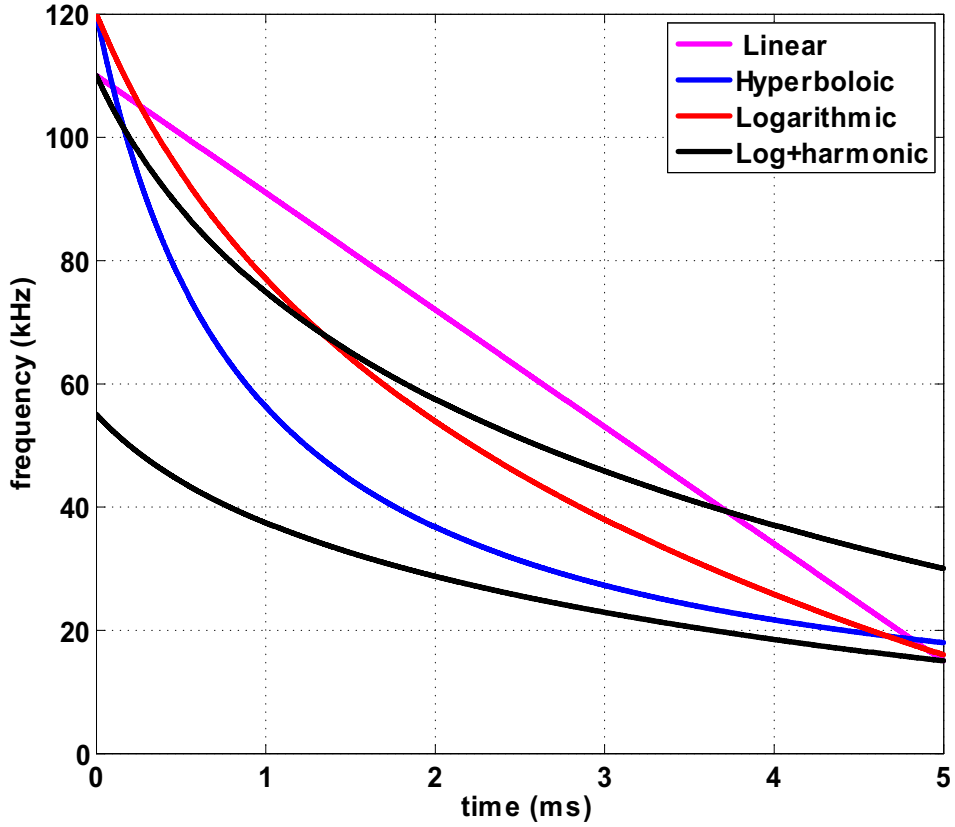

Figure 6.57: The artificial bat head was tested with four different frequency sweeps: a linear sweep, a hyperbolic sweep, a logarithmic sweep, and a logarithmic sweep in addition to its second harmonic.

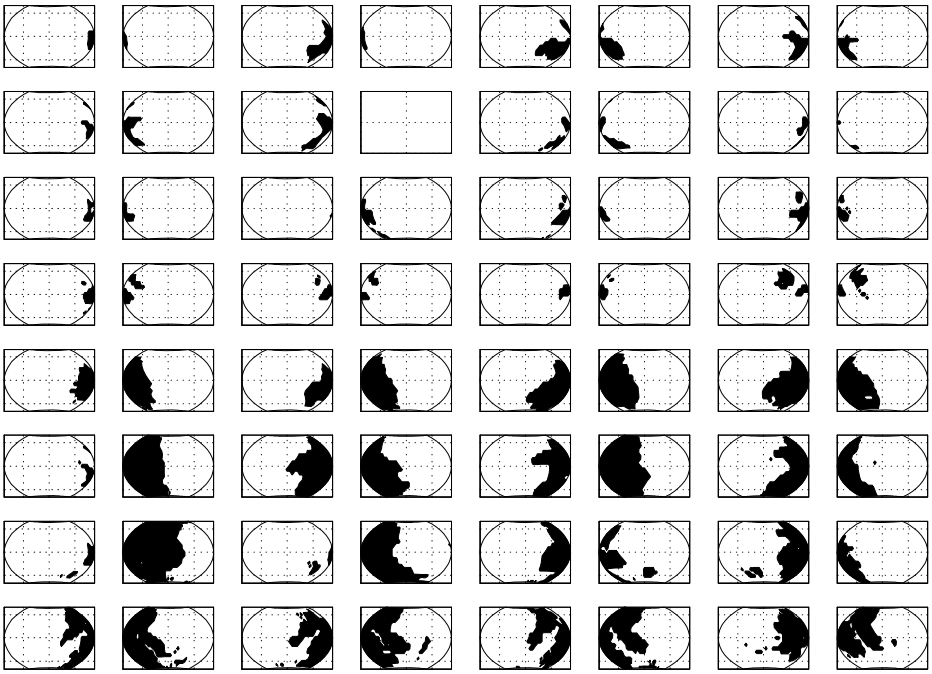

Figure 6.58: Binary receptive fields for 64 ILD neurons taken from two ILD chips. The two ILD chips have different ILD thresholds. The top four rows correspond to one ILD chip while the bottom four rows correspond to the second ILD chip.

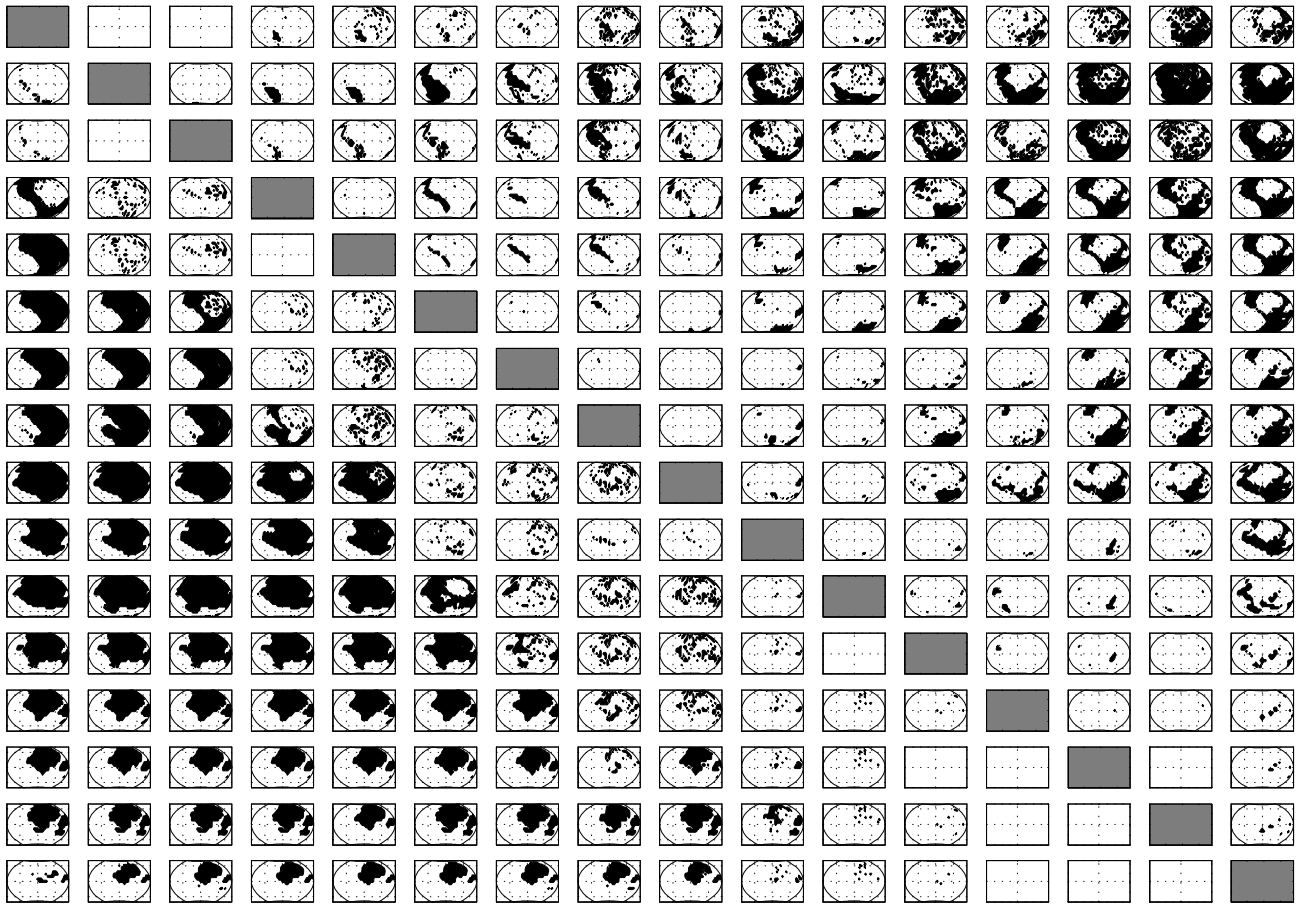

Figure 6.59: Response of the right spectral difference chip to a hyperbolic sweep. The chip has 240 neurons with no neurons along the diagonal (gray boxes). Each row (column) receives excitation (inhibition) from a single cochlear filter.

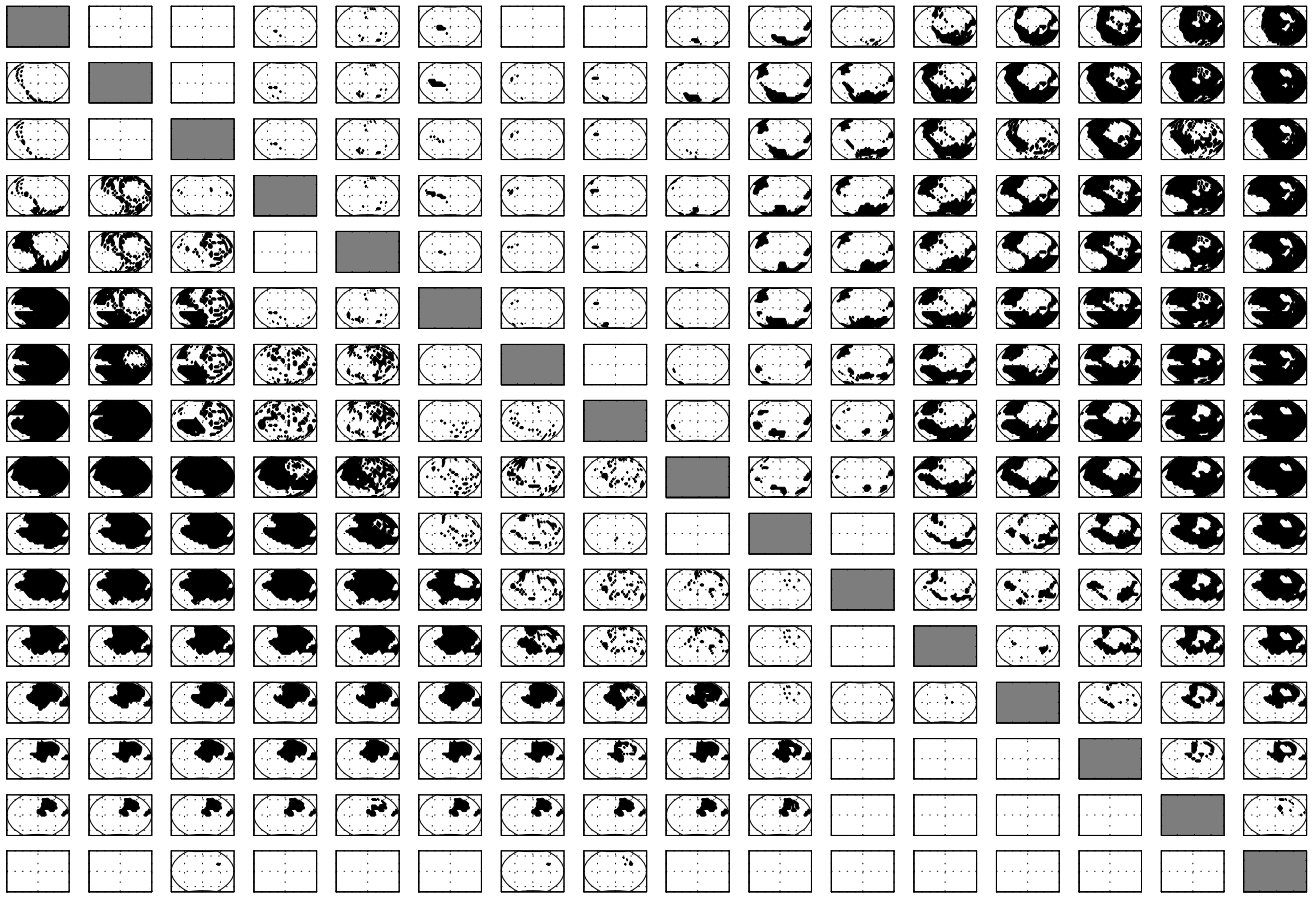

Figure 6.60: Response of the right spectral difference chip to a linear sweep. The chip has 240 neurons with no neurons along the diagonal (gray boxes). Each row (column) receives excitation (inhibition) from a single cochlear filter.

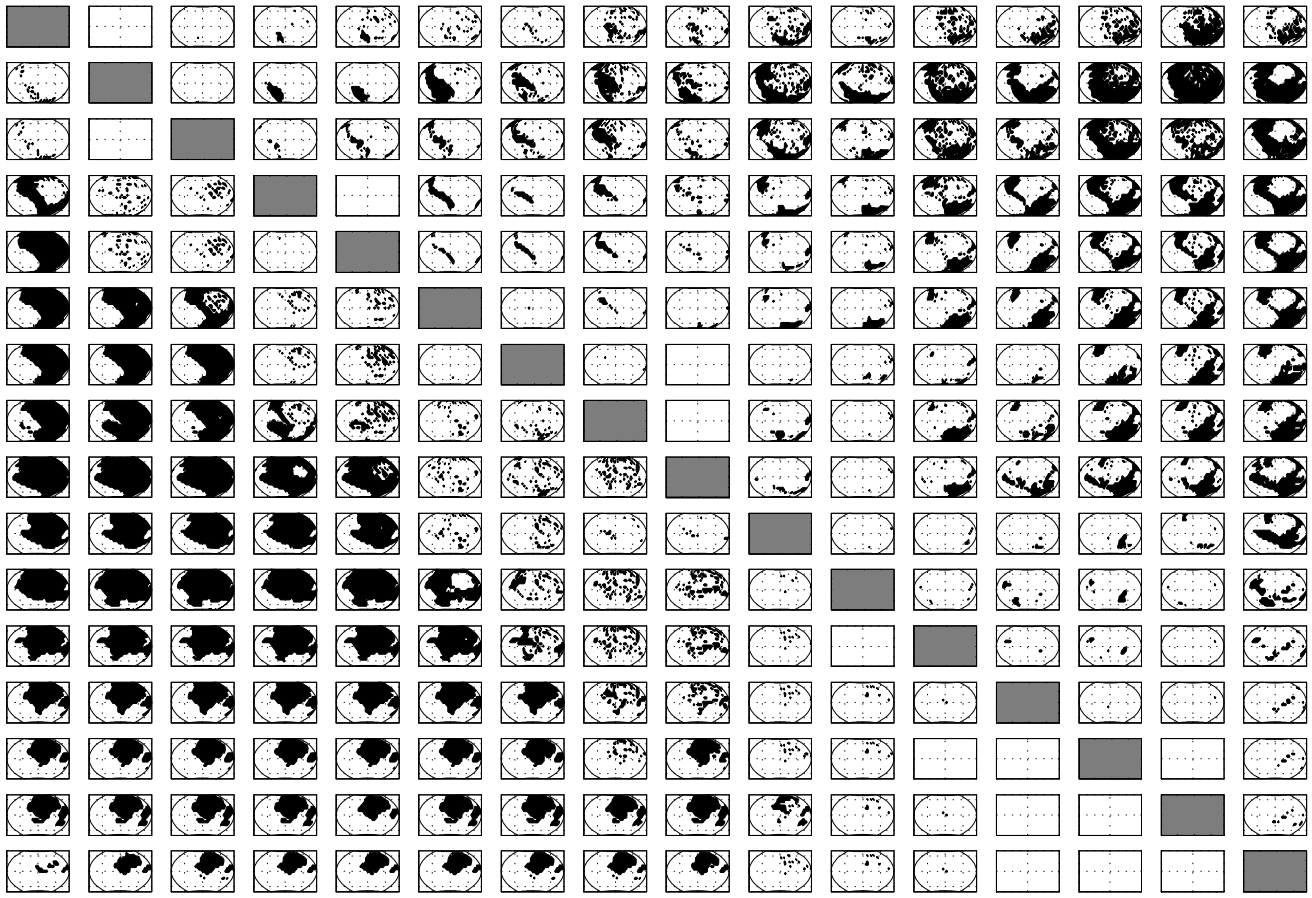

Figure 6.61: Response of the right spectral difference chip to a logarithmic sweep. The chip has 240 neurons with no neurons along the diagonal (gray boxes). Each row (column) receives excitation (inhibition) from a single cochlear filter.

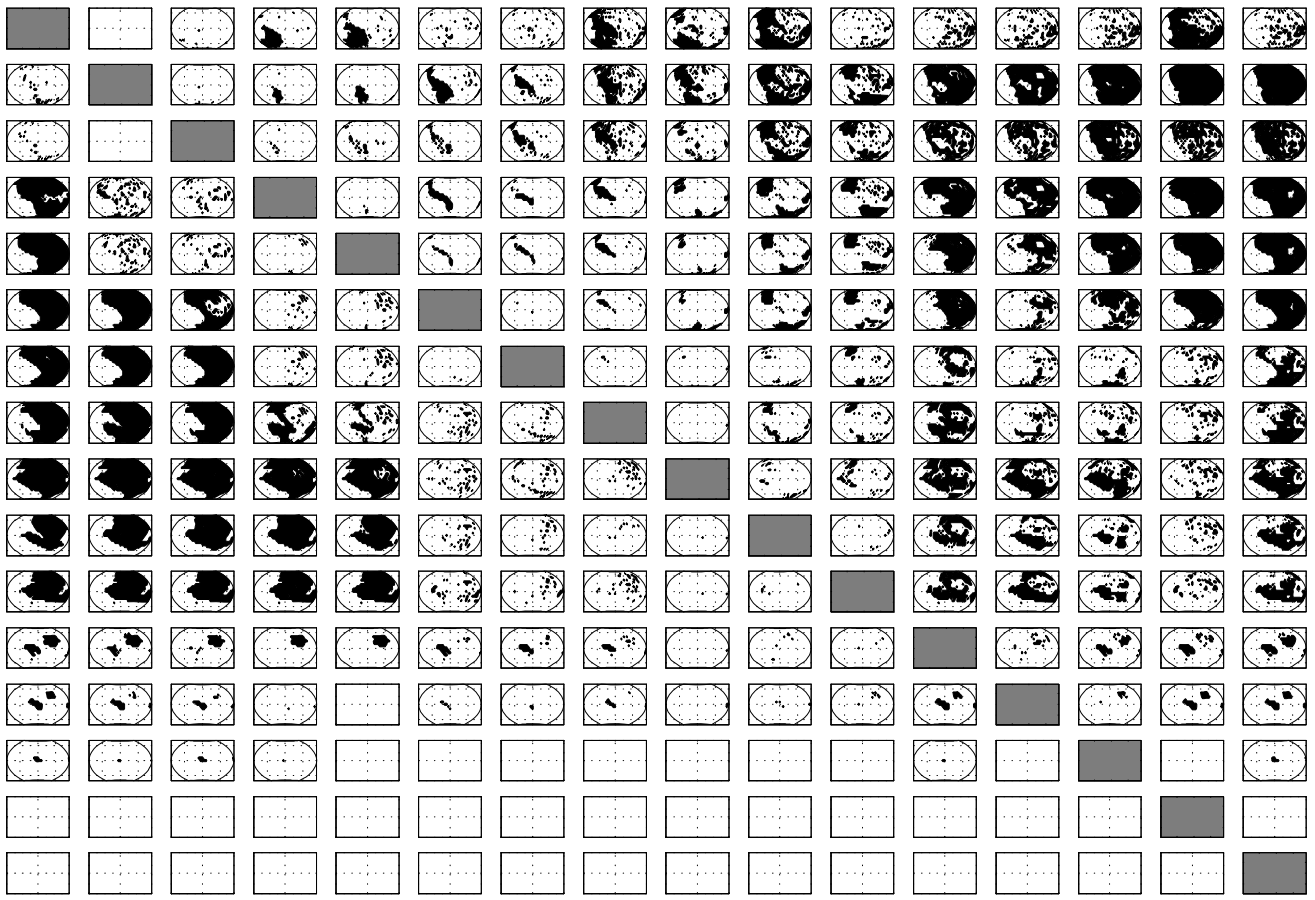

Figure 6.62: Response of the right spectral difference chip to a logarithmic sweep in addition to its second harmonic. The chip has 240 neurons with no neurons along the diagonal (gray boxes). Each row (column) receives excitation (inhibition) from a single cochlear filter.

Table 6.12: Correlation between the code matrices obtained from the four different frequency sweeps. The code matrices C1, C2, C3, and C4 correspond to the hyperbolic, logarithmic, linear, and logarithmic plus harmonic sweeps respectively.

	$C1$	$C2$	$C3$	$C4$
$C1$	100	89	67	62
$C2$	89	100	68	63
$C3$	67	68	100	63
$C4$	62	63	63	100

6.5 Conclusion

In this chapter we demonstrated that the measured responses from the artificial bat head contain enough information to be used for localization. We demonstrated that the intensity of the measured signals is encoded in the latency of the cochlear spikes and we demonstrated that the two feature extraction chips: the spectral difference chip and ILD chip can extract the localization cues from the spiking output of the cochlea. We showed that the spatiotemporal pattern of spiking outputs from the feature extraction chips can be decoded to simultaneously estimate the elevation and azimuth of the sound source. We also demonstrated that the hardware localization system was not only successful in localizing hyperbolic chirps, but also was capable of localizing other frequency sweeps.

Chapter 7

Periodicity Detection

7.1 Introduction

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 [104], 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 [105] can be used to detect unknown harmonic signals, a dedicated time domain approach would likely result in even lower power consumption. Goldberg et al. demonstrated a low power time-domain approach based on autocorrelation [106]. 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 [107][108][109]. By applying a peak-detector and measuring the period of the

envelope peaks, a sufficient number of consecutively matching periods is interpreted as the presence of a periodic signal of interest. By limiting the range of frequencies that are acceptable for the periodicity, detection specific to human voices or other periodic signals is possible. This algorithm rejects impulse events, wind noise, and can be implemented to be intensity-independent. Clearly, as this type of algorithm will trigger on pure tones, repetitive impulses, or non-voice or non-vehicular harmonic sounds in the right range, this is best used as a wake-up signal for more sophisticated algorithms or when false positives are acceptable. While the primary focus of this work is the design and testing of our VLSI circuit, we consider and discuss the effects of signal-to-noise ratio, bandwidth, and other parameters on this algorithm.

7.2 System

Fig. 7.1 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

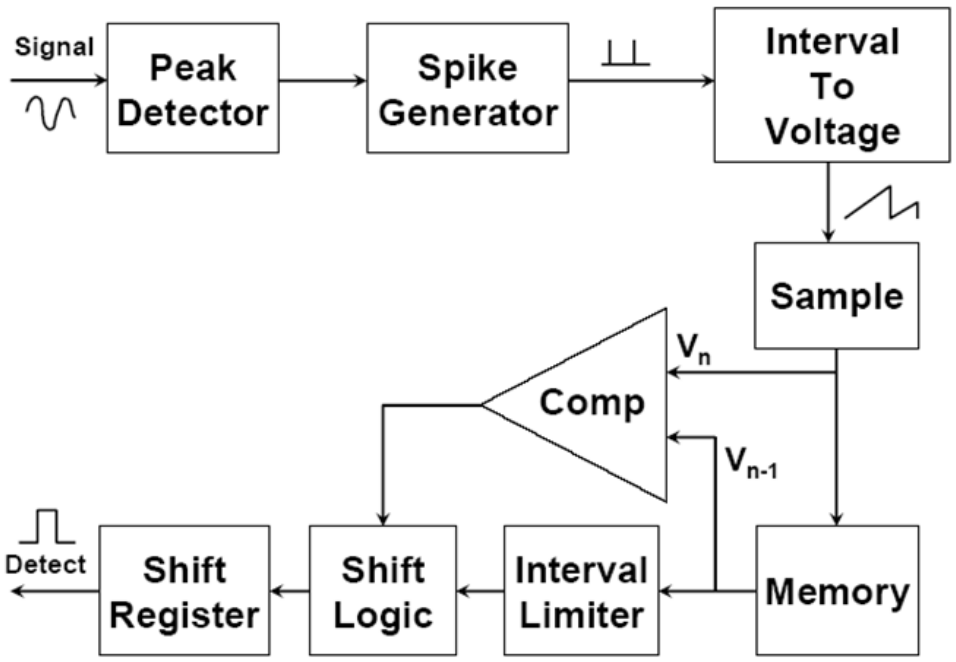

Figure 7.1: System block diagram of the peak-periodicity detection chip.

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. In the experimental data we show, the rate of three-consecutive hits or five-consecutive hits above a threshold rate is interpreted to be a detection event, however, a single occurrence of three-consecutive hits or five-consecutive hits can also be used. Not shown in the block diagram is the 'Timing Signals' stage which is responsible for generating the signals necessary to sequence the sampling, voltage conversion and decision-making processes following each peak in the signal. Fig. 7.2 shows the timing diagram for the first four stages.

We acknowledge that the falling edge of $o2$ is the obvious edge to use as an indication of the absolute time of the signal's peak; our goal, however, is to estimate the period of the envelope which depends on the relative timing of the spikes (interspike interval). Our SPICE simulations showed that during the time that the peak voltage tracked the input; it was possible to get oscillations on the digital pulse $o2$. We also note that at the onset of the tracking period, the input has its maximum slope (relative to the rest of the tracking period). Therefore, we designed M1 (Fig. 7.3) to have a large W/L and we chose to use the rising edge of $o2$ where the charging signal is strongest and $o2$ was the most reliable for producing a sharp transition. Although we are not truly measuring the interpeak interval, for clean periodic signals we reliably obtain the same period measurement.

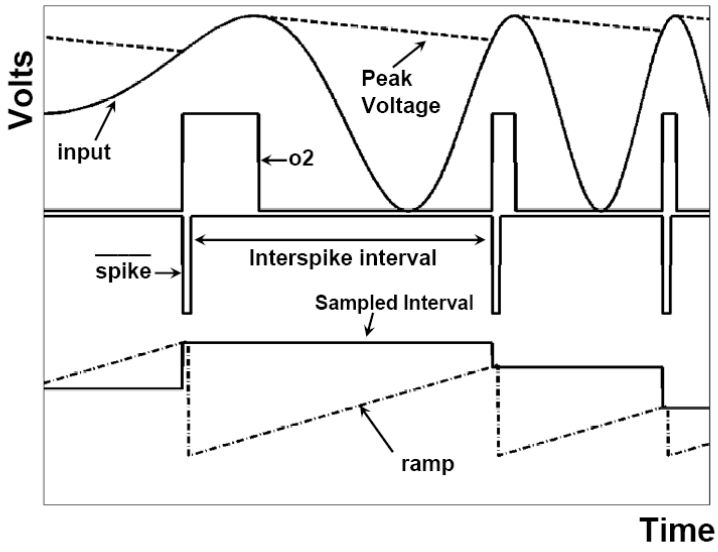

Figure 7.2: Timing diagram of the first four stages: The peak voltage (dotted line) follows the peaks of the signal. o2 is an internal signal of the peak detector. The rising edge of o2 marks the onset of a peak in the envelope and is used to generate a spike. The ramp (dotted line) converts the interspike interval into a voltage. Sampled interval is the voltage representation of the current interspike interval. ramp is then reset to encode a new interval.

7.3 Circuits

In an effort to reduce power and to match the longer time intervals of interest in the audio frequency range using integrated capacitors, most of the transistors are biased in the subthreshold region of operation where the saturation drain current equation is exponential and the drain-source saturation voltage is about $100mV$. The drain current I_d is given by:

$$I_d = I_o \exp(\frac{\kappa V_{gb} - V_{sb}}{V_T}) \tag{7.1}$$

where I_o is the saturation current, κ is the body effect, V_T is the thermal voltage, V_{gb} is the gate-to-bulk voltage, and V_{sb} is the source-to-bulk voltage [103]. All circuits described below were fabricated in a commercially available 2-poly $1.5\mu m$ CMOS process. The chip occupies an area of about 0.242 mm^2 (read-out circuitry not included) and was operated at a supply voltage (Vdd) of three volts and consumes less than $1.8\mu W$.

7.3.1 Peak Detector

The first stage of our circuit is the peak detector shown in Fig. 7.3. As the capacitor's voltage drops below the input voltage V_{in}, the wide-range transconductance amplifier (transamp) [110] turns M1 on and C1 is charged to V_{in}.

When the input drops below the capacitor's voltage, M1 is turned off and the capacitor is linearly discharged by the current source formed by transistors

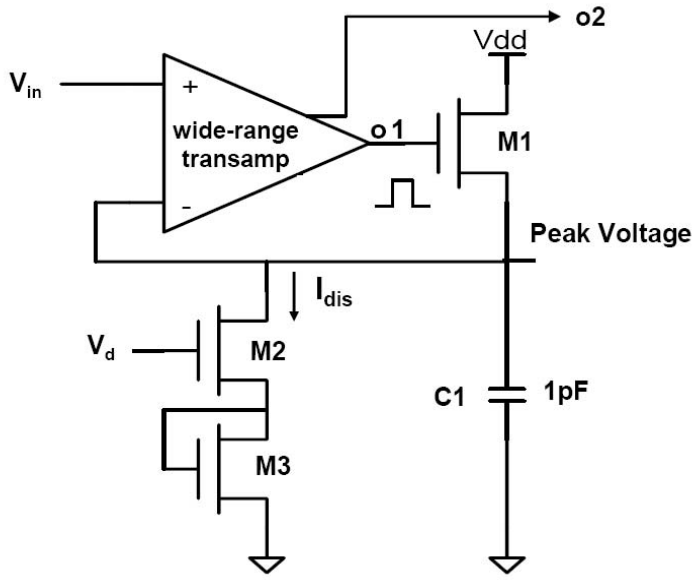

Figure 7.3: Peak Detector with constant current discharging controlled by V_d. The schematic of the wide-range transamp is shown in Fig. 7.4. The transamp has two similar outputs $o1$, and $o2$. The $o2$ output is used in the edge detector circuit in Fig. 7.5.

M2/M3. The discharging current I_{dis} is given by the following relation:

$$I_{dis} = I_o \exp(\frac{\kappa^2 V_d}{(1+\kappa)V_T}). \tag{7.2}$$

For proper operation of the peak detector, the discharge rate should be small enough to prevent the peak detector from tracking the input. The fixed linear discharge of C1 makes the chip's performance amplitude-dependent. The discharging current can, however, be made adaptive by externally controlling V_d. Fig. 7.4 shows the wide-range transconductance amplifier used in the peak detector. Since M1 is used as a source follower, the output $o1$ will not, in general, swing to Vdd because it will stop when the peak value is attained on the amplifier's input. The second output, $o2$, is designed to hit the rail to provide a digital signal when the peak detector is charging the capacitor. Transistors M6/M7 have a W/L ratio of $4\mu m/10.4\mu m$ while that for M4/M5 is $10.4\mu m/10.4\mu m$. The second output $o2$, therefore, has a larger transconductance and will swing from rail to rail effectively generating a voltage digital pulse. Fig. 7.2 illustrates the simulated response of the peak detector to a sinusoidal input, the transamp's outputs $o1$ and $o2$ will be high while the peak voltage tracks the input.

7.3.2 Spike Generator

The rising edge of the peak detector's output $o2$ marks the onset of the peak of the envelope. This edge will be used to generate a voltage spike. The edge detector circuit shown in Fig. 7.5 is used as our spike generator. On the rising edge of $o2$, the voltage V_a swings from ground to Vdd, the capacitor sends a pulse of current

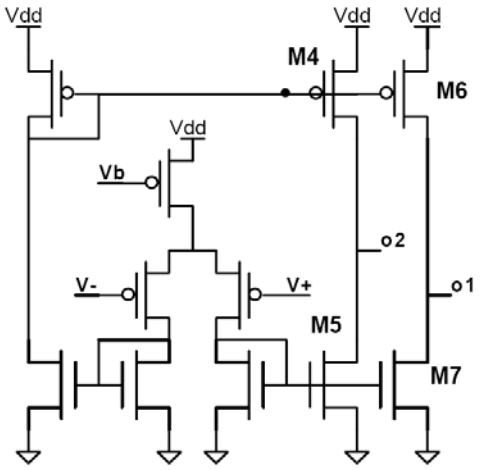

Figure 7.4: Dual-output, wide-range transconductance amplifier used to generate digital pulses at peaks in the signal's envelope. $o2$ has a larger transconductance than $o1$ because transistors M4/M5 have a larger W/L ratio than M6/M7.

down the current mirror pulling V_o low, thus generating a voltage spike. The width of the spike can be controlled by V_{b3} and was set to be approximately $270\mu s$. In order to generate a second spike, V_a must first return to its low state before which the mirrored capacitive current is too small to pull V_o low. Therefore, similar to biological neurons, the circuit has a refractory period T_{refr}; after a spike has been generated, another spike cannot be generated for a certain period of time [111]. In our circuit, after a spike has been generated, $V_{refractory}$ controls how quickly V_a returns to its low state thereby controlling the refractory period. We can think of the rising edges of $o2$ as a train of input spikes. If a second input spike arrives before V_a has returned to its low state, V_a will be pulled up to Vdd, resetting the "refractory timer". Therefore, after a spike has been generated, if

the inter-peak intervals are less than the refractory period, this circuit will remain in its "refractory state" and no longer generate spikes. We note that the width of the spike pulse itself ($\approxeq 0.3ms$) imposes a lower bound on the refractory period. The choice of a particular value for T_{refr} is application dependent and must be less than the smallest period (highest frequency) of interest. Fig. 7.6 shows the peak voltage and the spikes that correspond to the peaks of the envelope of the input signal.

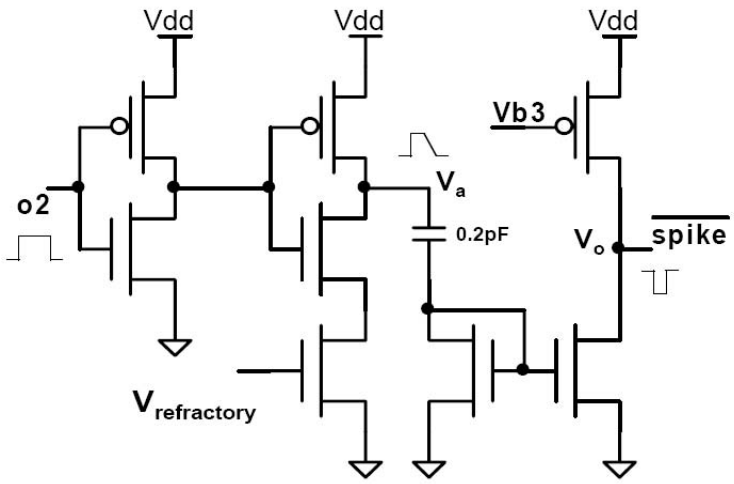

Figure 7.5: Edge Detector/Spike Generator for generating spikes on the rising edges of $o2$. V_a must return to its low state before another spike can be generated resulting in a refractory period.

7.3.3 Timing Signals

The proper sequencing and time duration of the ramp voltage sampling and reset events is a crucial task. The sampling pulse must be long enough to complete

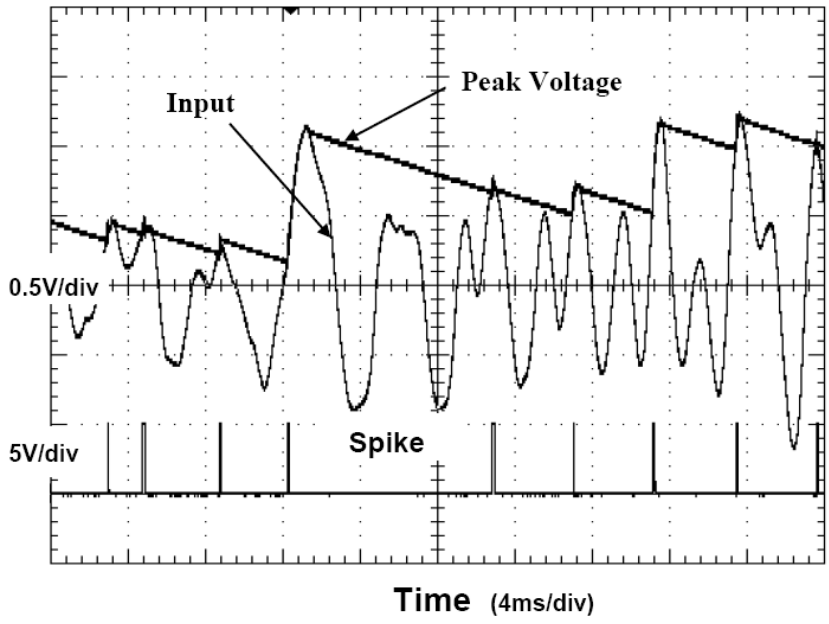

Figure 7.6: Oscilloscope trace shows the peak voltage and the spikes that correspond to the peaks of the envelope of the input signal.

the sampling process and must end before the onset of the reset pulse. The reset pulse must be long enough to allow the complete discharge of C3 (see Fig. 7.10). The comparison between the new interval voltage (V_n) and the previous interval voltage, (V_{n-1}) must occur (in the interval comparator circuit) after the sample pulse and before the onset of the reset pulse. Therefore, a guard interval that can be adjusted to meet these requirements has been included in the design of the circuit. Fig. 7.7 shows an oscilloscope trace of the timing signals that are generated with every new spike. The circuits that generate the sample and reset pulses are shown in Figs. 7.8 and 7.9 respectively. $\overline{spike}$ begins high, holding V_x low and "turn off" low.

The falling edge of $\overline{spike}$ starts the process of charging capacitors C6, C7 (Fig. 7.8), and C8 (Fig. 7.9). The charging rate is determined by the charging currents as well as the size of the capacitors. As each capacitor's voltage crosses the threshold of an inverter, a pulse is generated. The circuit in Fig. 7.8 is responsible for generating the sample pulse. V_y is pulled high abruptly on the falling edge of $\overline{spike}$ which pulls down sample and starts the sampling process. The sampling process continues until the voltage ramp V_x causes the turn off pulse to become high. C6 is then discharged which results in sample switching back to its high state. The rate at which C7 charges up can be controlled by the bias voltage V_{to} which in turn controls the width of the sampling pulse. The circuit in Fig. 7.9 generates the reset pulse. The voltage ramp developed across C8 will turn on the reset pulse. The bias voltage V_r controls how fast the ramp rises to threshold. This delayed reset turn-on is used to adjust the separation between

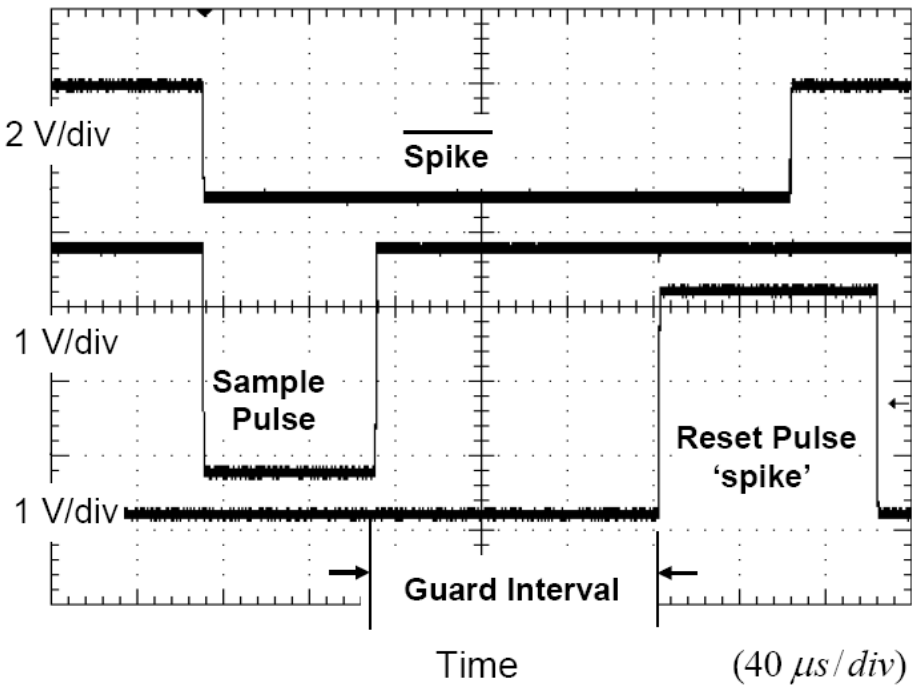

Figure 7.7: Oscilloscope trace of the timing signals used to sequence and control sampling, comparison and memory storage.

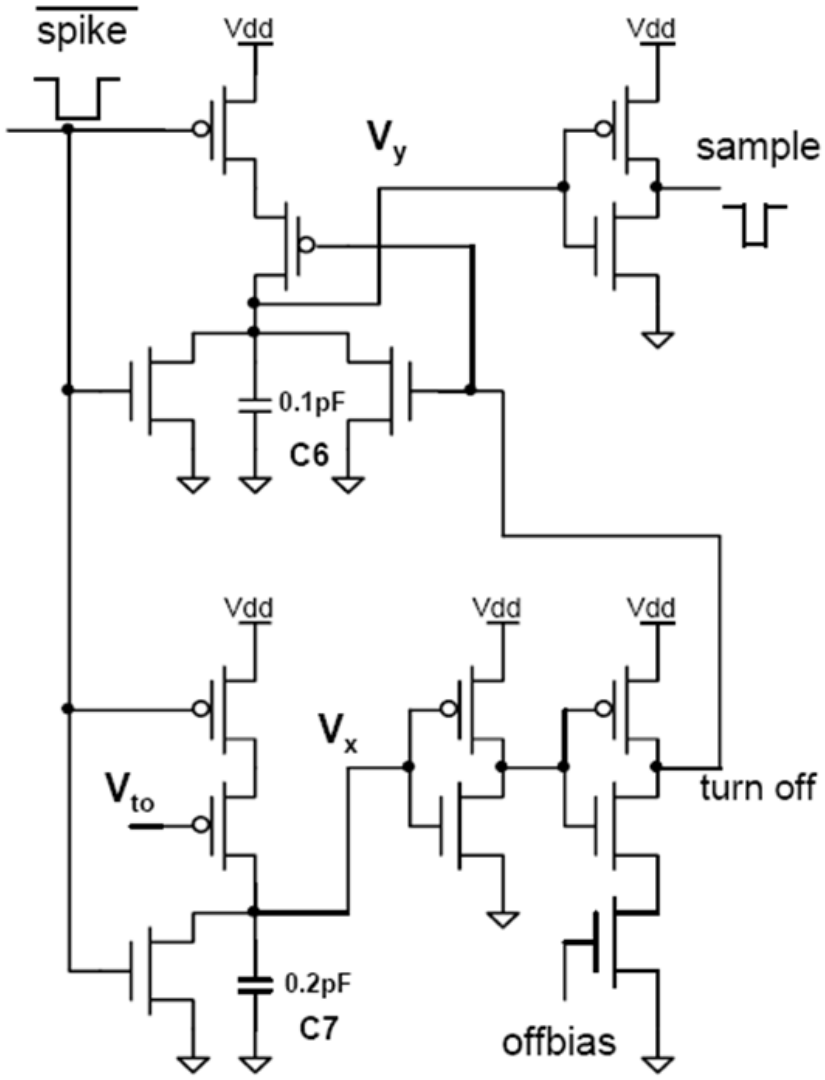

Figure 7.8: Timing signal generator: generating the sample pulse.

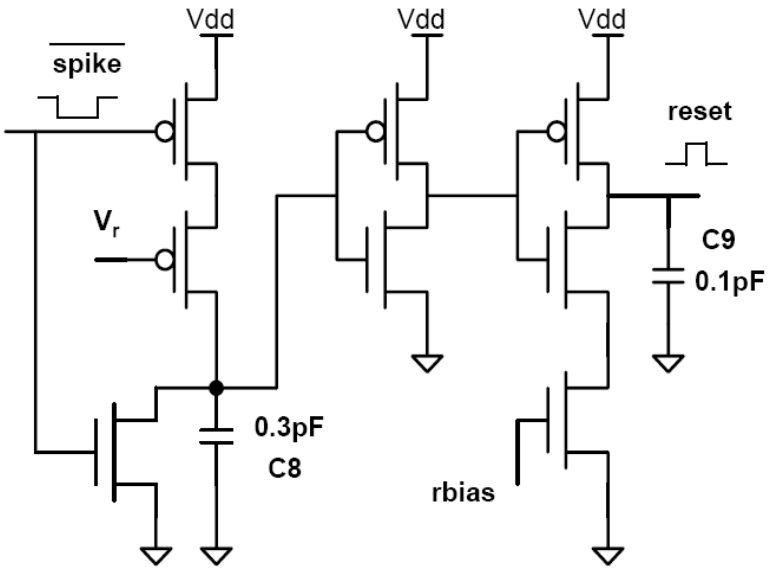

Figure 7.9: Timing signal generator: generating the reset pulse.

the sampling and reset pulses. The voltage rbias can be used to extend the width of the reset pulse beyond the rising edge of $\overline{spike}$.

7.3.4 Voltage Conversion, Sampling and Storage

Using a technique very similar to [112], we converted the interspike intervals into voltages by sampling (and holding) a fixed-slope voltage ramp. The interval is then stored for further processing. All three operations are shown in Fig. 7.10. The constant current source formed by transistor M1 and fixed bias V_{ramp} charges C3. The voltage ramp V_{c3} is used to convert an interspike interval to a voltage. As a spike arrives, the ramp is sampled and the capacitor C4 stores this new interspike interval V_n to be compared (in the interval comparator stage) with the preceding interval V_{n-1} stored on the capacitor C5. The sample pulse that is generated with

every new spike turns the follower amp1 on for a short period of time, sampling the linear ramp. The reset pulse that follows the sample pulse discharges C3 through M2 in preparation for a new interval to voltage conversion. The sampled voltage V_n is temporarily stored in C4. The follower amp2 is weakly biased so that its output slowly follows its input. In effect, this allows the sampled interval V_n to be slowly transferred to memory as V_{n-1} on C5. Since the current charging C3 is constant, the voltage on the capacitor increases linearly with time. Therefore an interspike interval is linearly transformed into a voltage. Mathematically:

$$V_n = mT_n \tag{7.3}$$

where m is the slope of the voltage ramp V_{c3}, is the interspike interval, is the voltage representation of the interspike interval and n is the time index. The slope of the ramp, m, is controlled by V_{ramp}. Because the dynamic range is constrained by the power rails, the ramp voltage V_{c3} saturates for long intervals. If V_{sat} is the saturation voltage, then the interval, T_{sat}, at which the transformation begins to saturate is given by:

$$T_{sat} = \frac{V_{sat}}{m} \tag{7.4}$$

Fig. 7.11 shows the implementation of the wide-range transconductance amplifiers amp1 and amp2. Because sample swings down to ground, the transistor with control voltage V_c (see Fig. 7.11(a)) was added to limit the current flow. Fig. 7.12 shows oscilloscope traces that demonstrate the voltage conversion process and the sampling process. As a spike comes in, the ramp voltage V_{c3} is first sampled then reset, V_{c3} then charges linearly with time encoding the new interspike interval.

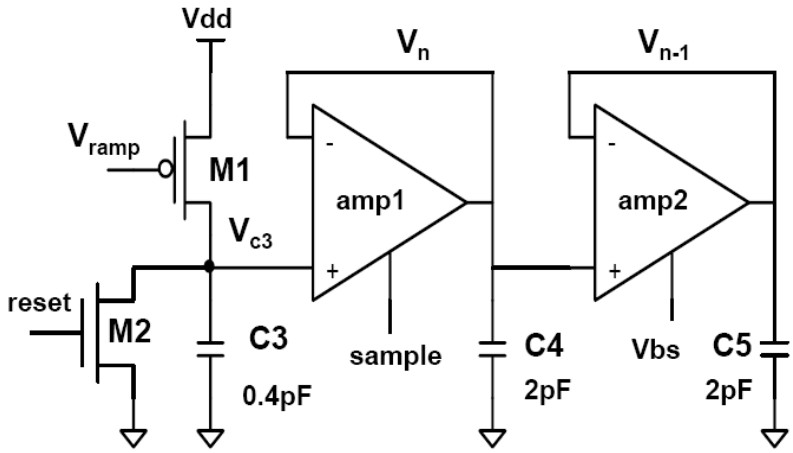

Figure 7.10: Interval-to-Voltage Converter, Sampling and Storing of intervals. The constant current source M1 charges C3 producing a voltage proportional to the interspike interval. This voltage is sampled and stored in C4 and is slowly copied and stored in C5.

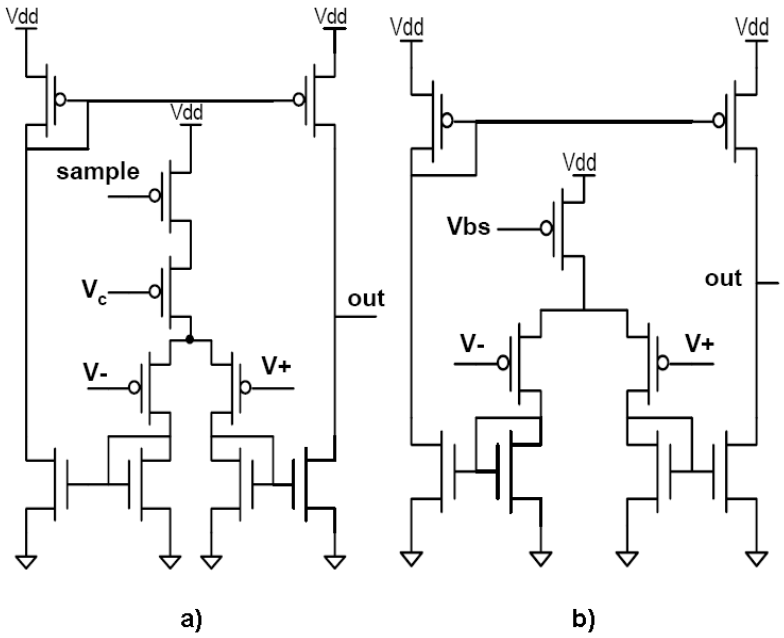

Figure 7.11: Wide-range transconductance amplifiers used in Sampling and Storage. (a) amp1 (b) amp2.

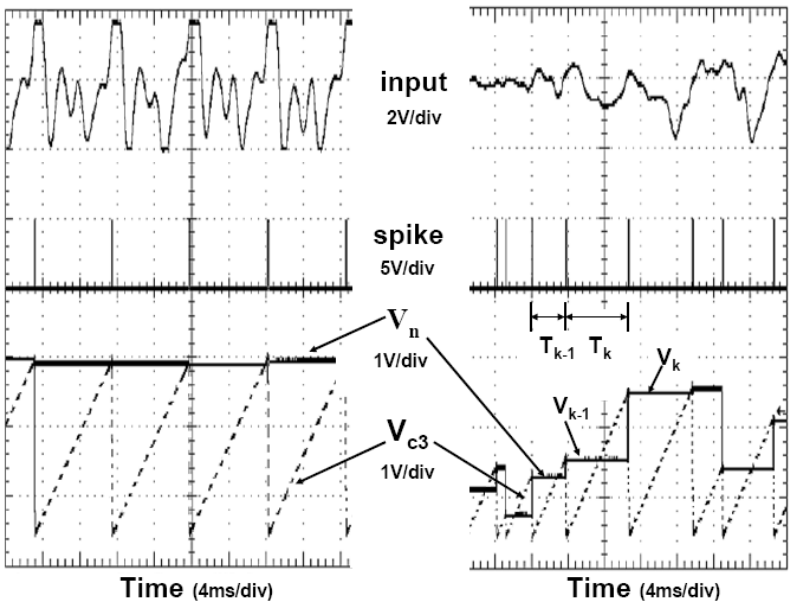

Figure 7.12: Two example stimuli: (left) harmonic complex and (right) irregular noise-like signal. Oscilloscope traces show the spike waveform, the interval to voltage conversion and the sampling processes. V_n remains constant throughout the interspike interval and gets updated with every new spike.

When another spike arrives, the encoded interval is sampled and the voltage V_{c3} is reset repeating the whole process. The sampled voltage interval V_n remains constant throughout the interspike interval and gets updated with every new spike via the sampling process.

7.3.5 Interval Limiter

In the case of a noisy input, it is possible to get very small interspike intervals that match causing false alarms. Saturation of the voltage transformation for long intervals is another concern and may also result in false alarms. It is therefore desirable to have a circuit that defines a range of "valid interspike intervals". This validation circuit is shown in Fig. 7.13. As long as the stored "interval" V_{n-1} satisfies the inequality $V_L < V_{n-1} < V_H$, where V_L and V_H are externally defined by the user, the output transistors M1 and M2 will be on. The output of the circuit will be low, indicating a "valid interval". Equivalently in the time (interval) domain, we find that the circuit evaluates the inequality $T_{min} < T_{n-1} < T_{max}$, where T_{min} and T_{max} are the minimum and maximum valid intervals respectively and are given by:

$$\begin{aligned} T_{\min} &= \frac{V_L}{m} \\ T_{\max} &= \frac{V_H}{m} \end{aligned} \tag{7.5}$$

Fig. 7.14 illustrates the linear interval-to-voltage transformation that saturates for intervals longer than T_{sat}. Since no interval T_n can be less than T_{refr}, T_{refr} must be chosen to be less than T_{min}.

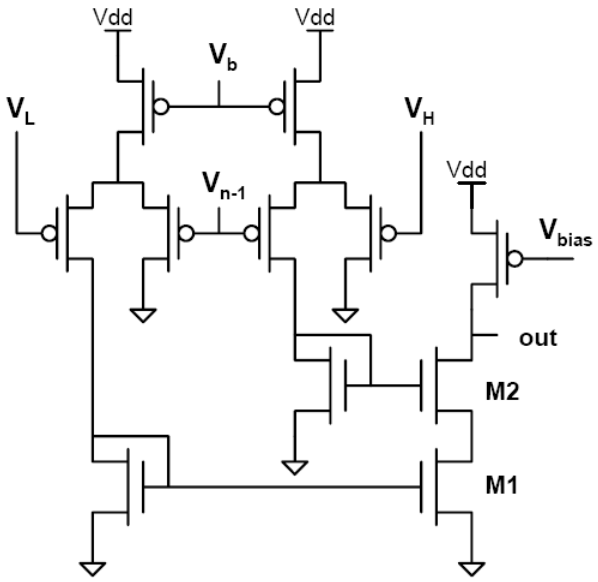

Figure 7.13: Interval Limiter. If the stored interval $V_{n-1}$is valid, both M1 and M2 will be turned on and out will be low indicating a valid interval.

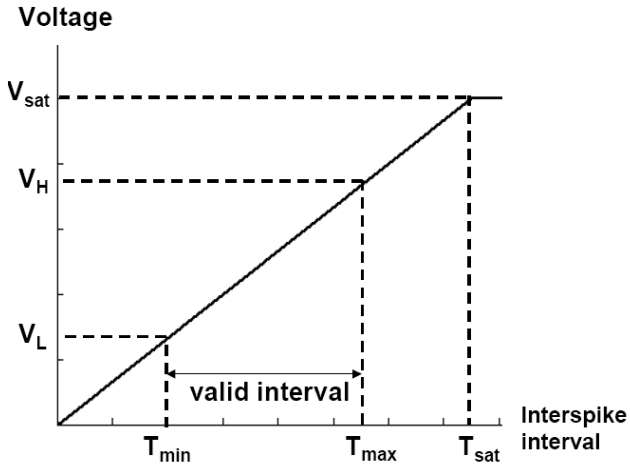

Figure 7.14: The interval-to-voltage transformation and the "valid interval" range as determined by V_L and V_H.

7.3.6 Interval Comparator

The interval comparator is the heart of the decision making process. It must decide whether the two consecutive interspike intervals, T_n and T_{n-1} match; a match being defined as satisfying the inequality $|T_n - T_{n-1}| < \varepsilon_t$, where ε_t is the accuracy (or tolerance) of the comparison. Equivalently in the voltage domain is $|V_n - V_{n-1}| < \varepsilon_v$ where $\varepsilon_v = m\varepsilon_t$ and m is the slope of the transformation. Fig. 7.15 is the circuit that evaluates this inequality. V_b sets a current I_b in the differential pair that is split into I_1 and I_2 which are mirrored to the two outer branches. V_{th} sets a current $I_{th} < I_b/2$. When the input voltages match, both I_1 and I_2 are strong enough to hold V_x and V_y low, thereby turning off M1 and M2 and the output is high, indicating matched intervals. When the inputs do not match, most of I_b is steered in one of the two branches, I_{th}, in turn, pulls either V_x or V_y high and the output voltage is pulled down, indicating non-matching intervals. The decision must be completed before the onset of the reset pulse (i.e. in less than $200\mu s$). This causes ε_v to be a function of V_{th2}. Table 7.1 shows the measured threshold ε_v as a function of $(V_{th} - V_b)$. Since V_n eventually becomes V_{n-1}, the output of this circuit spends most of its time in the matching (high) state. This output is only used, however, after the sampling process has ended and during the time when the reset pulse is high. When the reset pulse is high, if the comparator decides that the two intervals do not match, all five shift registers are reset. The value of V_{th2} also affects the time that it takes for the output to switch back to its high state.

Table 7.1: Accuracy (ε_v) of the comparator (for fixed V_b and V_{th2}) as a function of $\Delta V = V_{th} - V_b$.

$\Delta V(mV)$	49	61	72	83	96	105	114
$\varepsilon_v(mV)$	30	50	73	86	106	125	144

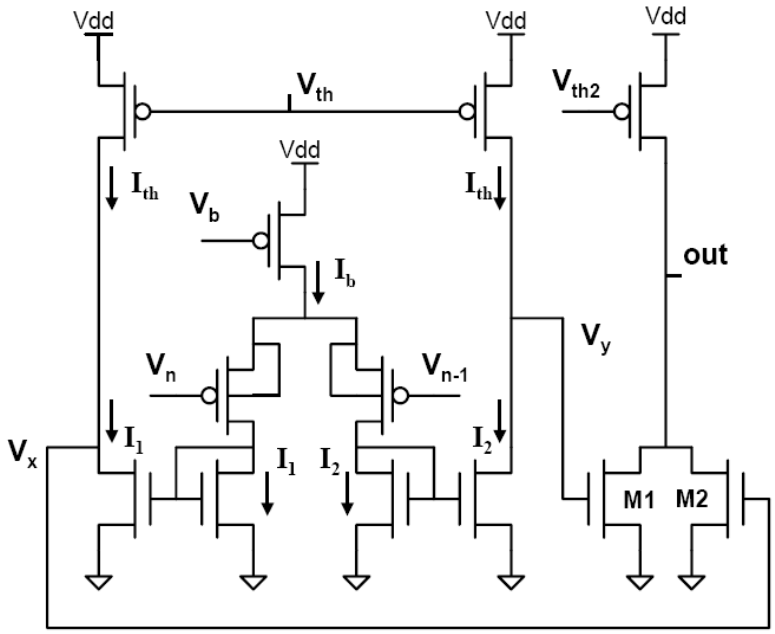

Figure 7.15: Interval Comparator. If the two consecutive intervals do not match, I_b is steered in one of the branches of the differential amplifier, V_x (or V_y) is pulled up high turning on M2 (or M1) pulling down *out*.

7.3.7 Shift Register and Shift Logic

The shift register block shown in Fig. 7.1 is a five-state shift register. When two consecutive intervals match and satisfy the interval-limiter condition we call it a "hit". If all five registers contain zeros (i.e. 'reset'), a "one" gets loaded into the first register only when a "hit" occurs. The "shift logic" block will either generate a clock signal advancing the "one" to a new state if a new "hit" occurs, otherwise, it will generate a reset pulse clearing the entire shift register.

7.3.8 Power Consumption

The power consumption of the chip was measured in two different modes; the "testing" mode and the "stand-alone" mode. The on-chip instrumentation circuitry used to monitor various internal signals was powered in the testing mode but was turned off in the stand-alone mode. In both cases, the power was measured with a band-limited (at $1kHz$) white Gaussian noise input and a $40Hz$ sinusoidal input. In the testing mode, the power consumption was approximately $15\mu W$ and $21\mu W$ for the noise and sinusoidal inputs respectively. In the stand-alone mode, the circuit's normal mode of operation, the power consumption was approximately $1.8\mu W$ and $0.9\mu W$ for the noise and sinusoidal inputs respectively.

7.4 Mathematical Analysis and Modeling

Similar to a passive sonar detection system where the receiver listens for the acoustic signature of engines, propellers and other elements in an underwater

vehicle, this problem can be modeled as detecting a sample function from a random process of target sounds buried in noise [113]. Regardless of whether the received signal is noise only or signal plus noise, the only information we have about the received signal after peak detection lies in the interspike intervals. The decision procedure must be based upon the difference in the statistical properties of these interspike intervals. In the following analysis, we use real data measured from our chip. Signals were generated using MATLAB (Mathworks Inc.) and presented to the chip through a sound card. The MATLAB Instrument Control Toolbox was used to capture the desired data from the oscilloscope. Fig. 7.16 illustrates the measured pdf (probability distribution function) of the interspike intervals for two different cases. In the first case, white Gaussian noise was low-pass filtered at $1000Hz$ and presented to the chip (relative noise power was -28dB). In the second case, a $20Hz$ sine wave was added to the background noise at a SNR (Signal-to-Noise Ratio) of zero dB. In both cases the refractory period was $11.8ms$ and the duration of the signal was 80 seconds. As the SNR increases, the pdf of interspike intervals for the signal plus noise case converges to an impulse located at the period of the sine wave.

The algorithm that we have adopted to distinguish between the two types of inputs is based upon the assumption that the interspike intervals generated from the peaks of the envelope of the signal of interest will be centered at the period of the envelope and that the deviation from this center will be relatively small and will decrease with the increase of the signal-to-noise ratio. The chip counts what we have called "hits". A hit is the simultaneous occurrence (intersection) of two

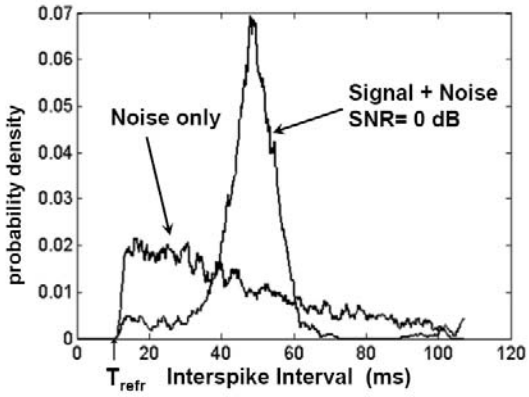

Figure 7.16: Measured pdf of two stimuli: a) Noise only. b) $20Hz$ sinewave signal plus noise. The refractory period was $11.8ms$.

events, the first, denoted as "A", is the event that the stored interval V_{n-1} lies within the acceptance range defined by V_L and V_H. That is

$$A = \{V_L < V_{n-1} < V_H\} \tag{7.6}$$

The second event, denoted as "B", is the event that the two consecutive intervals V_n and V_{n-1} are equal to within a constant ε_v, that is

$$B = \{|V_n - V_{n-1}| < \varepsilon_v\} \tag{7.7}$$

We can now define the probability of a hit as the probability of the intersection of the two events A and B

$$p(hit) = p(A \cap B). \tag{7.8}$$

It is straightforward to show that the probability of a hit is upper bounded by the smaller probability of the two events A and B, that is

$$p(hit) \leq \min(p(A), p(B)). \tag{7.9}$$

As more and more interspike intervals fall in the acceptance range, $p(A)$ approaches unity and $p(hit)$ approaches $p(B)$. The probability of the event "A" can be computed from the pdf of interspike intervals by evaluating the integral

$$p(A) = \int_{T_{\min}}^{T_{\max}} f(x)dx \tag{7.10}$$

where $f(x)$ is the pdf of the interspike intervals.

Before we move on to analyze the effect of different circuit parameters, we introduce a simple dimensionless measure of variability in the interspike interval distribution known as the coefficient of variation C_v [114][115], defined as,

$$C_v = \frac{\sigma_{T_n}}{E(T_n)} \tag{7.11}$$

where $\sigma_{T_n}$is the standard deviation of the interspike interval distribution and $E(T_n)$ is the mean interspike interval. The interspike intervals in our chip obey the inequality:

$$T_{refr} \leq T_n \leq T_{sat} \tag{7.12}$$

Taking the expectation of this inequality we get:

$$T_{refr} \leq E(T_n) \leq T_{sat} \tag{7.13}$$

where T_{refr} and T_{sat} are constant. Since $E(T_n)$ is upper bounded, it follows that the probability of the event "B" is related to C_v, in the following sense:

$$C_v \to 0 \Longrightarrow \sigma_{T_n} \to 0 \Longrightarrow p(B) \to 1. \tag{7.14}$$

We begin by first studying noise.

7.4.1 Noise Only

The distribution of interspike intervals of noise that are generated from the peak-detection process depend on four parameters: refractory period, bandwidth, linear decay rate of the peak detector, and noise power. The effect of these parameters on the measured pdf of the interspike intervals will be shown.

Refractory Period

The effect of the refractory period, T_{refr}, on the pdf of noise interspike intervals is shown in Fig. 7.17.

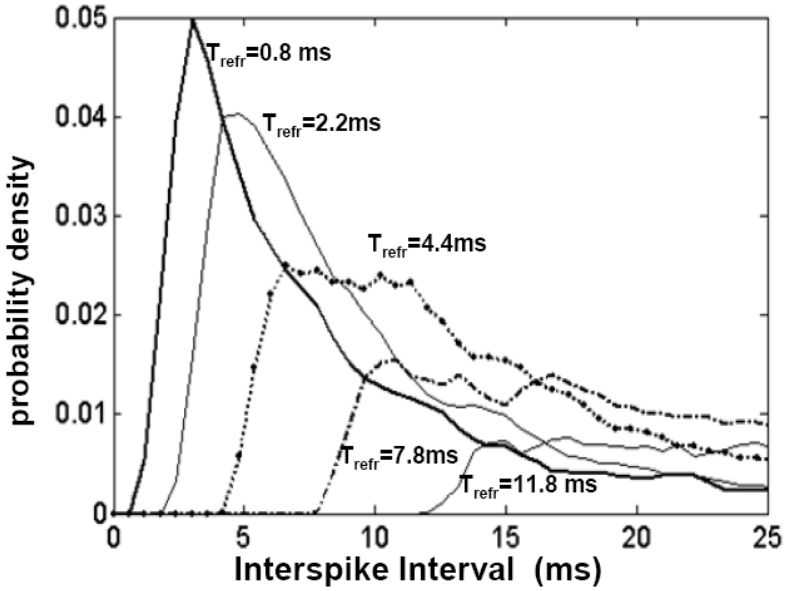

Figure 7.17: The effect of the refractory period on the pdf of interspike intervals.

The first effect is that the interval distribution is zero for intervals less than T_{refr}. The second effect is that the pdfs spread out as T_{refr} is increased and are

not just shifted versions of one another. This can be attributed to the behavior of the refractory circuit; after a spike is generated, the circuit will remain in its refractory state (thus longer intervals) if subsequent inter-peak intervals are less than the refractory period. As a result, stimuli that contain many short inter-peak intervals will have a broader, flatter pdf as these short intervals get connected onto other intervals. Fig. 7.18 illustrates how the increase of T_{refr} affects several parameters. On one hand, more of the noise intervals fall in the valid interval region increasing $p(A)$. On the other hand, however, the increase in the spread of the distribution means fewer consecutive intervals will match resulting in the decrease of $p(B)$. The net effect on $p(hit)$ is also shown. The decrease in C_v is in agreement with the intuitive idea that the spike train becomes more regular as the refractory period is increased.

Bandwidth

Decreasing the bandwidth of the noisy input means that the envelope of the waveform will become "smoother" and the probability of generating multiple spikes off the same peak decreases. Fig. 7.19 shows that for bandwidths greater than $400Hz$, there is no significant change in the pdf of the interspike intervals (for this set of parameters).

Decay Rate of Peak Detector

Qualitatively, increasing the decay rate of the peak detector causes the peak detector's output to encounter a peak earlier than it would have with a smaller decay.

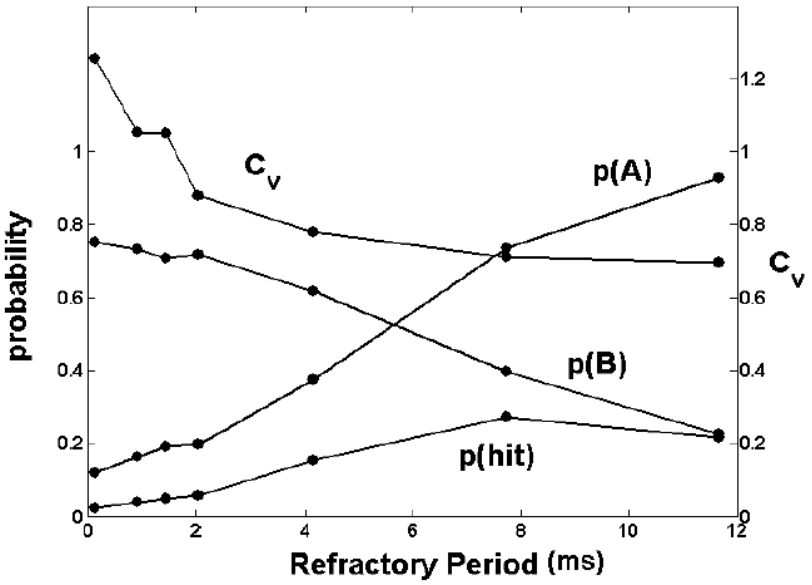

Figure 7.18: The effect of the refractory period on $p(hit)$ and C_v. The greater the refractory period the more regular the spike train the smaller C_v is. As more of the interspike intervals fall in the valid interval region, $p(A)$ approaches unity and $p(hit)$ approaches $p(B)$.

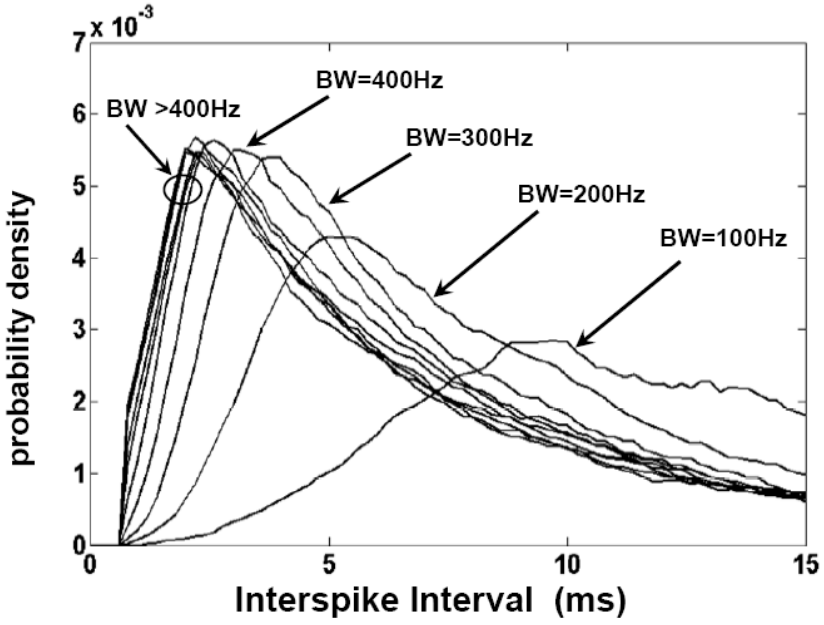

Figure 7.19: The effect of bandwidth on the pdf of noise interspike intervals. No significant change in the pdf for bandwidths greater than $400Hz$.

The effect this has on the pdf of the interspike intervals is illustrated in Fig. 7.20. Decreasing the decay rate causes more intervals to fall in the valid interval range increasing $p(A)$. The spread in the pdf, however, causes a decrease in $p(B)$. The maximum $p(hit)$ that was measured for this set of data was 0.11.

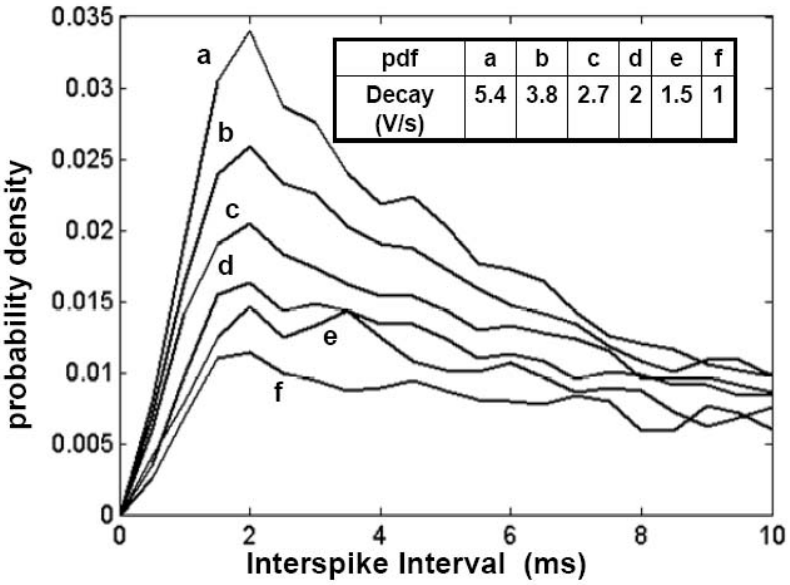

Figure 7.20: The effect of the decay rate (in V/s) on the pdf of noise interspike intervals.

Noise Power

Increasing the noise power increases the probability of occurrence of large impulses of noise. Because the fixed linear decay rate of the peak detector does not scale with amplitude, the time it takes for the peak voltage to encounter another peak of the envelope and thus generate a spike increases. Fig. 7.21 shows that increasing the noise power increases the relative probability of getting a long interval.

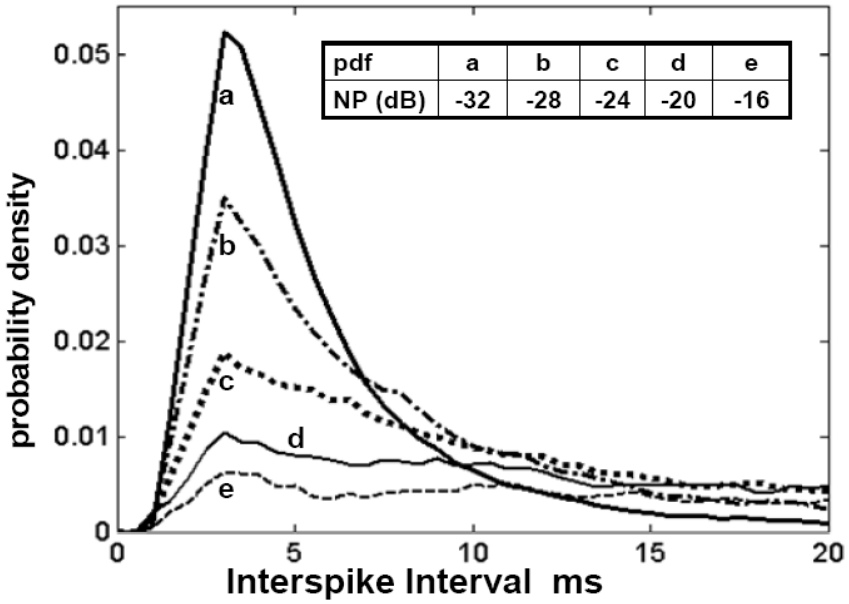

Figure 7.21: The effect of noise power (NP) on the pdf of noise interspike intervals.

7.4.2 Signal Plus Noise

For simplicity we will use a sinusoidal waveform as our signal of interest. The input filtering bandwidth must be wide enough to pass our signal of interest. The period T of our signal must satisfy the following inequality $T_{refr} < T_{min} < T < T_{max}$. To successfully detect periodicity, only one spike should be generated at the peak of the envelope. Without a refractory period, it is possible to get more than one spike at the peaks of the signal. This is likely to occur in a noisy environment but can occur even when the SNR is relatively high. Fig. 7.22 illustrates the effect of the refractory period on the pdf of interspike intervals of a $20Hz$ sinusoidal signal with a SNR of 15dB. In the absence of a refractory period (other than the $0.3ms$ refractory period caused by the width of the spike itself), multiple

spikes are produced in the region where the peak-voltage "tracks" the peak of the sinusoid; this explains the non-zero probability of getting small intervals. In addition, any long intervals obtained will tend to be smaller than the actual period of the sinusoid which explains why the small "bump" in the pdf is not centered at $50ms$.

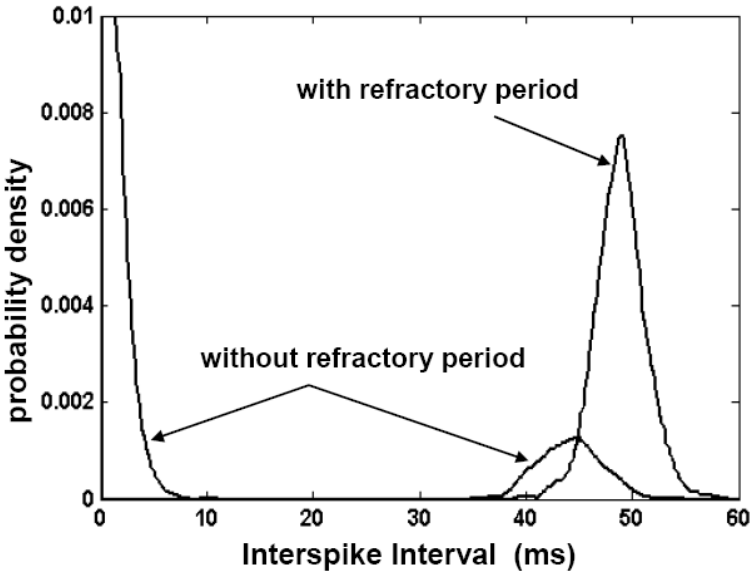

Figure 7.22: Effect of the refractory period (T_{refr}=11.8ms) on the pdf of interspike intervals in the signal plus noise case. The signal is a $20Hz$ sinusiod (SNR= 15dB).

Having a refractory period (11.8ms in this case) suppresses spiking for a short time after a spike has been generated. The effect this has on the pdf of interspike intervals is also shown in Fig. 7.22. Heuristically, if we assume that the peak detector will track the sinusoidal input for one eighth of the period, then, in order for the refractory period to be effective in suppressing spiking T must satisfy

the inequality $8T_{refr} > T$. This may set a larger lower-bound on the frequency range than that imposed by T_{max}.

7.4.3 Modeling the Counter

The shift-register-based counter used in our circuit (5 registers) can be modeled as a finite Markov Chain. There are six states $X0, X1...X5$. The $X0$ state corresponds to all five registers being reset. The states $X1, X2 \ldots X5$ correspond to registers 1 to 5 respectively. When a hit occurs, the first register (state $X1$) is loaded with a logic 1, if another hit occurs, the logic 1 is shifted to the second register (state $X2$) and so on, if a spike occurs and there is no match, the whole counter is reset back to zero (state $X0$). The counter can count up to five consecutive hits. The state diagram for this chain is shown in Fig. 7.23.

The chain can be characterized by its state transition matrix, P. The entry in the matrix denotes the probability of transitioning from state i to state k. Two assumptions have been made regarding this model. The first assumption is that the chain is homogenous, that is, the transition probabilities are independent of the time step n. The second assumption, which is generally not true but has been made to simplify the analysis, is that all the forward transition probabilities have been assumed to be equal:

$$p_{ik} = p \qquad i = 0, 1...4; \quad k = i + 1 \tag{7.15}$$

where $p = p(hit)$. The chain is irreducible and aperiodic, and a theorem [116]

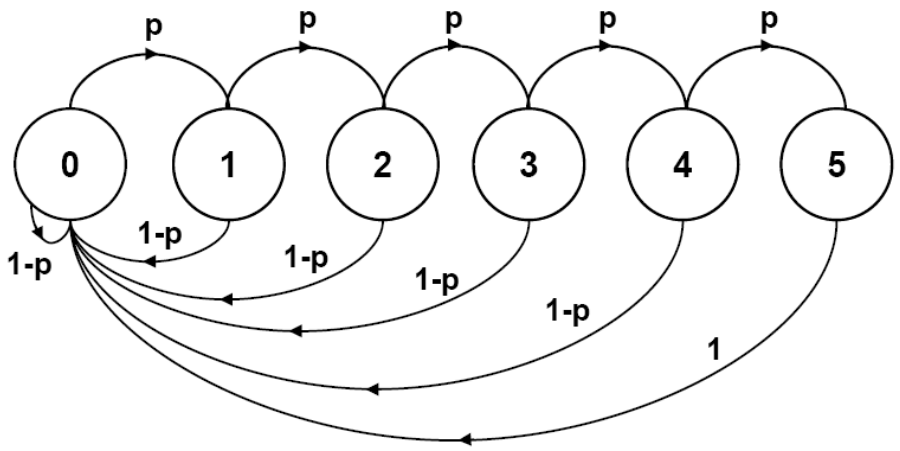

Figure 7.23: State diagram of the 5 shift-register-based counter used to count consecutive hits.

holds that:

$$p_{ik}(n) \rightarrow \frac{1}{\mu_k} \quad \text{as } n \rightarrow \infty, \text{ for all } i \text{ and } k \tag{7.16}$$

where μ_k is the mean recurrence time for state k. Therefore, $p_{ik}(\infty)$ can thus be considered as the normalized rate of hits reaching the k^{th} state for $k > 0$. Using this model, p was scanned from 0 to 1 and $p_{ik}(\infty)$ was computed numerically in each case by evaluating P^n for a large value of n (all entries in the k^{th} column become equal). The simulation results are presented in Fig. 7.25. Note that when $p = 1$, the counter behaves like a six-state ring counter, and the normalized rate of hits for all states equals 1/6.

7.5 Results

7.5.1 Testing with Noise and Sinewaves

Fig. 7.24 illustrates the measured relationship between the coefficient of variation, C_v, and the probability of getting two consecutive intervals to match $p(hit)$. Both C_v and $p(hit)$ were computed from chip data. In some cases the input was noise only with different settings for bandwidth, decay rate, noise power and refractory period (as described in section IV). This explains the wide variation in C_v. In other cases the input had a $20Hz$ sinusoid added to the background noise with different signal-to-noise ratios. The increase in $p(hit)$ as C_v decreases indicates that the algorithm is successful in predicting regularity in the spike train.

Fig. 7.25 illustrates the mean rate of hits measured at the outputs of the

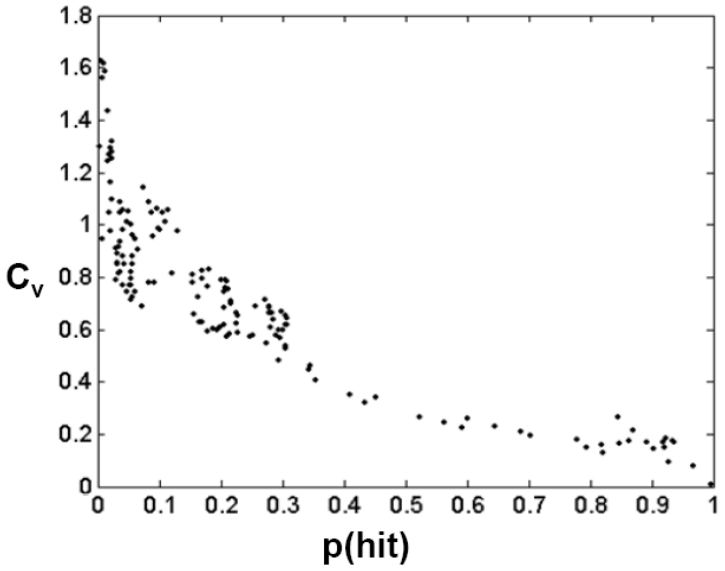

Figure 7.24: Coefficient of Variation (C_v) versus $p(hit)$. The more regular the spike train the smaller the coefficient of variation and the higher the probability of getting two consecutive intervals to match.

third and fifth registers (states) plotted as a function of p(hit). The individual dots are measurements from the chip while the solid lines are from the Markov model simulation where the normalized rate was multiplied by 20 to account for the $20Hz$ frequency of the sinusoidal input. We notice that for medium values of $p(hit)$ our measurements are greater that the simulated values. This could be attributed to the fact that the transitional probabilities depend on the instantaneous value of the inter-spike interval; therefore, the chain is not homogeneous. This also means that the transitional probabilities are not equal. Another possibility is that due to the inability to directly measure the accuracy of the interval comparator circuit and device mismatch, we may have underestimated the value of $p(hit)$ computed from our measurements. Note that as $p(hit)$ approaches 1, the rate of hits measured at the third and fifth states converge to 3.33 ($20Hz/6$) confirming the six-state ring counter behavior.

Figs. 7.24 and 7.25 serve to illustrate that the chip has successfully implemented the algorithm; higher regularity in the spike trains results in a higher probability of getting a hit which in turn results in a higher rate of measuring hits at the third and fifth outputs of the shift register. Fig. 7.26 shows the measured rate of hits at the output of the fifth register as a function of SNR. The input is a $20Hz$ sinusoidal input added to a background white Gaussian noise (relative noise power is -28dB), low-pass filtered at $1000Hz$. A refractory period of $11.8ms$ was used. It is predicted from the plot that the sinusoid with a SNR as low as -5dB can be detected.

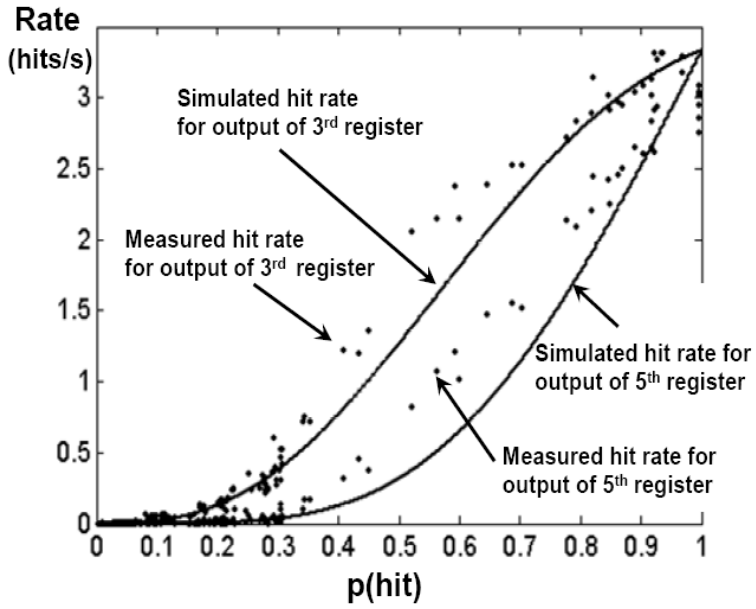

Figure 7.25: Measured and simulated rate of hits at the outputs of the third and fifth taps of the shift register.

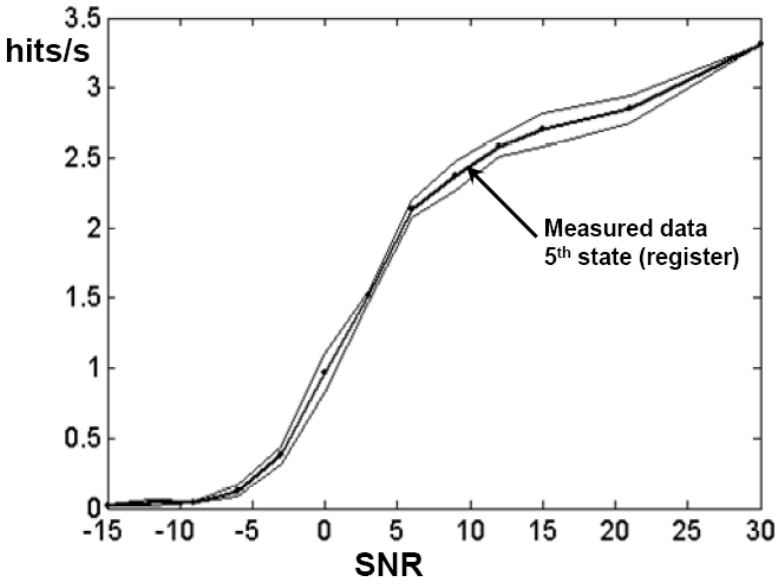

Figure 7.26: Measured rate (dark line) of hits at the output of the fifth register as a function of SNR for a $20Hz$ sinusoidal input with T_{refr}=11.8ms. The two thinner lines represent the one standard deviation boundary.

7.5.2 Real World Sounds

We have tested the periodicity detection chip tuned for voice frequencies and for vehicle engine frequencies. For the voice testing, the refractory period was $1.8ms$, the measured minimum and maximum allowable frequencies (set by V_H and V_L) were $38Hz$ and $248Hz$ respectively. The core test set of the TIMIT speech database [117] was used to evaluate the performance of the chip as a voice detector. One male speaker and one female speaker from each of the eight dialects were selected providing a test set of sixteen speakers. The total number of sentences used in the test was one hundred and sixty (ten per speaker). The mean duration of the sentences was approximately three seconds. The sentences were down-sampled to $2kHz$ and repeated five times. The number of three-consecutive hits and five-consecutive hits was recorded for each trial and the mean rate of hits was computed for each speaker. Fig. 7.27 represents the mean rate of five-consecutive hits for each speaker. Speakers 1 through 8 were male speakers while speakers 9 through 16 were female speakers. Because of their higher pitch, female speakers tend to produce a greater number of hits than males. The quality of the recording was such that no false alarms were generated for a noise signal with the same power as that of the background noise in the recording. Therefore, by setting the threshold rate sufficiently low, 100% of the speakers can be detected with no false alarms. Fig. 7.28 shows the effect of adding noise to four different sentences spoken by four different speakers. The speakers that were chosen were the two males and two females that had the greatest and smallest rate of five-

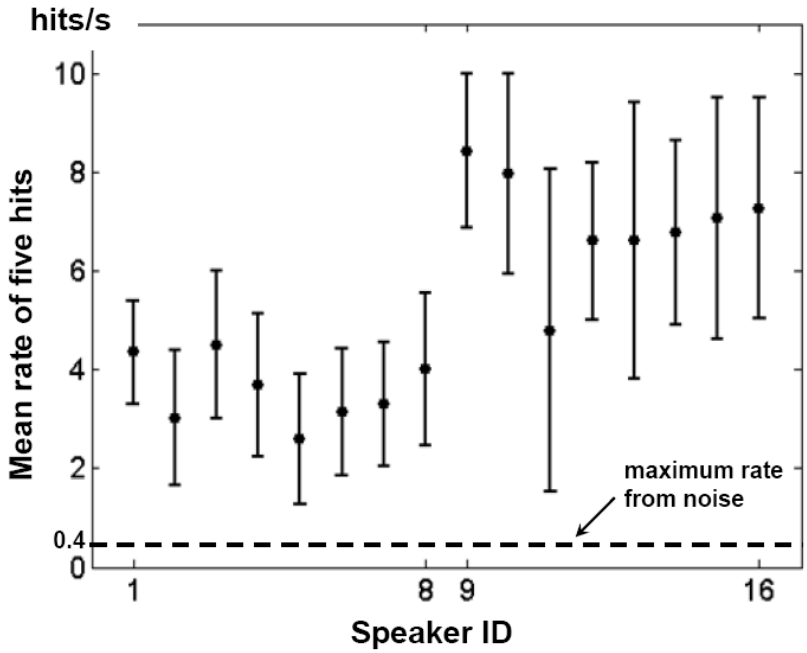

Figure 7.27: Mean rate of five hits for test sentences taken from the TIMIT speech database. Speakers 1 through 8 are male speakers while speakers 9 through 16 are female speakers.

consecutive hits. With a threshold of as low as 0.5 hits/s, the sentences can be detected down to a SNR of 15dB. To compare the response to noise input, the chip was tested with white Gaussian noise having a bandwidth of $1kHz$ and a duration of 20 seconds. The noise power was scanned from -60dB to -5dB in steps of 5dB, it was also tested at -3dB and -1dB. No spikes were observed for noise powers less than -40dB. The greatest recorded false-alarm rate for the five-consecutive hit output was 0.4 hits/s. Fig. 7.29 shows the response of the

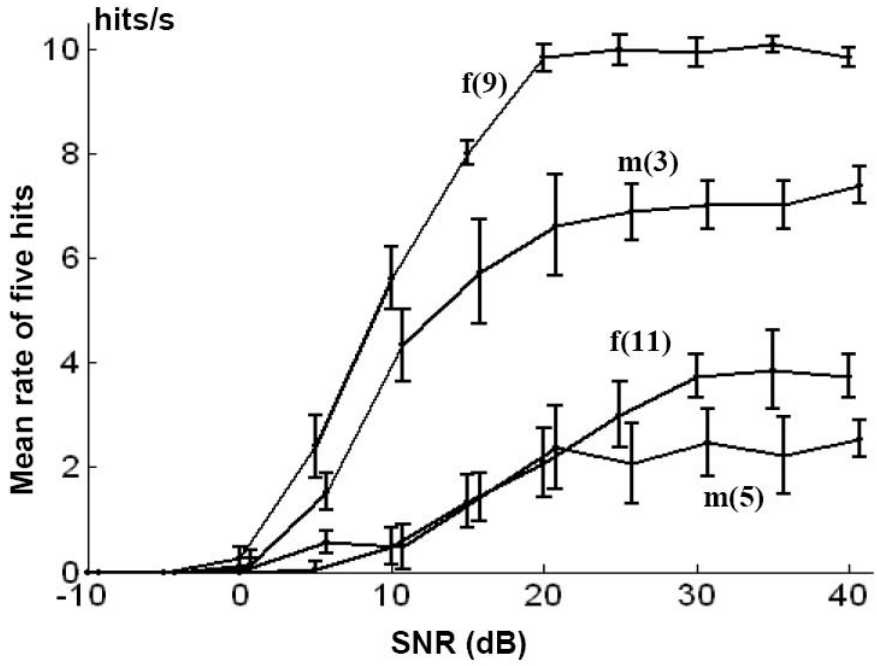

Figure 7.28: The effect of adding noise to four different sentences spoken by four different speakers. "m" and "f" denote gender while the number in the parenthesis is the speaker ID. With a threshold of 0.5 hits/s, detection is possible down to a SNR of 15dB.

chip to the speech utterance "zero", the three-consecutive hit output is shown to

trigger in the voiced segments of the utterance and not in the noisy onset. Because the Sampled Interval voltage V_n is directly related to the interspike (inter-peak) interval, the pitch in the voiced segments can be computed from V_n.

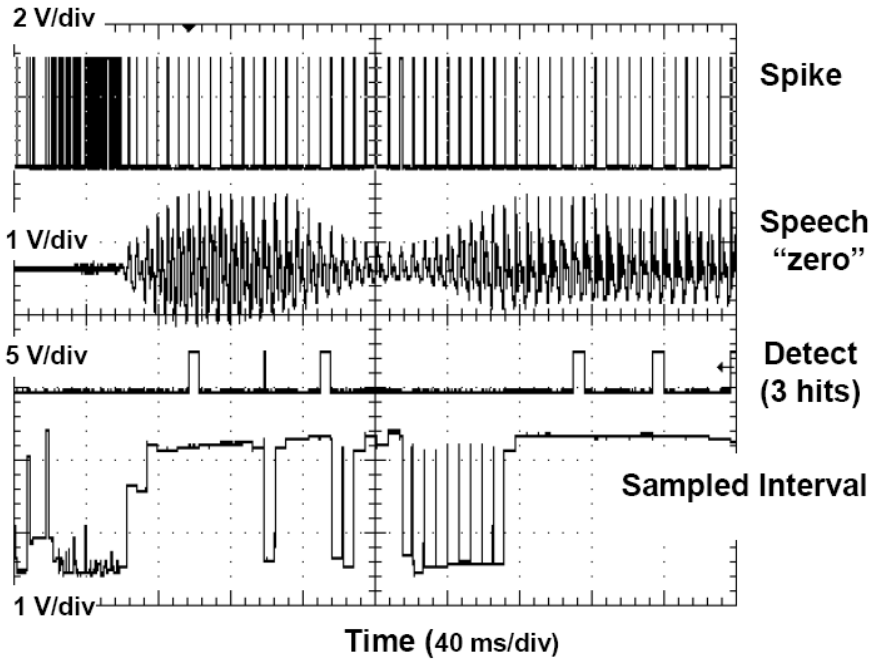

Figure 7.29: Response to the speech utterance "zero". No detection during the "noisy" onset. Detects the voiced segments. Sampled Interval can be used to estimate pitch.

In the vehicle detection case, although extensive testing was not done, vehicle detection is possible based on observations of the system response to the recordings of vehicles. We show that the chip functions properly in this low frequency range. The chip parameters were set so that the minimum and maximum allowable frequencies would be around $7Hz$ and $45Hz$ respectively. Fig. 7.30 shows the response of the chip to a recording of the sound of a vehicle. The recording had a sampling frequency of $1kHz$. White Gaussian noise was added

to the recording at a SNR of 20dB. The pitch of the envelope in this example as estimated from the inter-spike interval is approximately $35Hz$. The spectrum (not shown) of the signal shows a peak at $70Hz$ and $105Hz$, thus the fundamental frequency ($35Hz$) is missing.

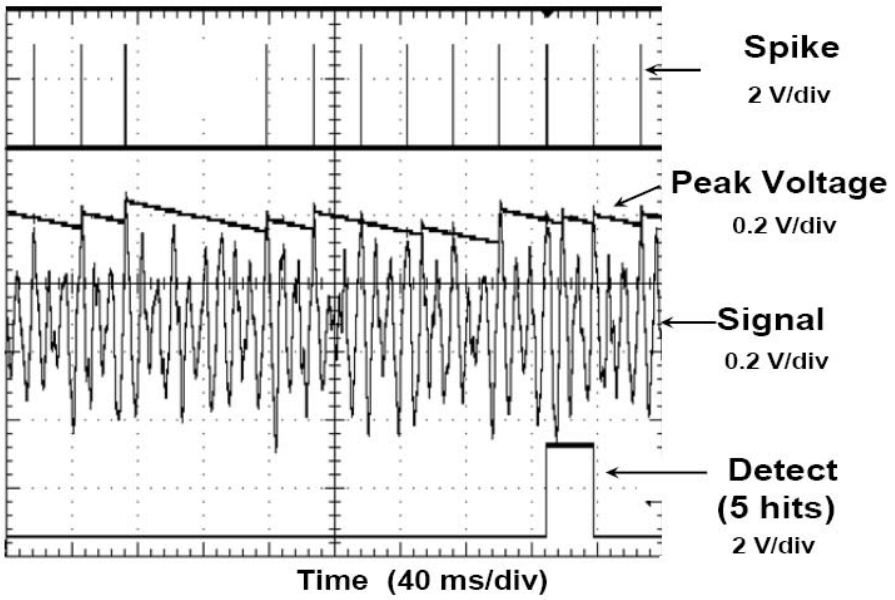

Figure 7.30: Response to a recording of the sound of a vehicle.

7.6 Conclusion

An analog VLSI low power circuit has been designed that successfully implements an algorithm for detecting periodicities in the time-domain envelope of a signal. We have tested the chip with speech, pure tones, and informally tested it with vehicle signals. The chip's input was band-limited at $1kHz$ and was tested with target signals having pitches as low as $7Hz$ and as high as $250Hz$. The overall

system has been mathematically modeled and the chip measurements were compared against the model. The chip was fabricated in a commercially-available 2-poly $1.5\mu m$ CMOS process and occupies an area of about 0.242 mm^2. The chip was operated at a supply voltage (Vdd) of three volts and consumes less than $1.8\mu W$. It consists of 215 transistors and nine capacitors. The analog processing blocks of the circuit account for 92 of the transistors; the logic and shift register blocks account for the rest.

Chapter 8

Conclusions and Future Work

8.1 Dissertation Contributions

In this dissertation we have presented two spike-based auditory signal processing systems. The first system (presented in Chapter 7) is a low-power analog VLSI envelope periodicity detector that detects periodicity in the time-domain envelope of a signal. The system is a single chip that was fabricated in a 2-poly, $1.5\mu m$ CMOS process.

The second system, (presented in Chapters 2 through 6) which is the main focus of this dissertation, is a wideband neuromorphic VLSI bat echolocation system that simultaneously estimates the elevation and azimuth of a sound source. The main contributions of this second system are summarized next.

An artificial bat head was designed using 3D CAD software and was fabricated using a 3D printer. We measured the response of the artificial bat head to hyperbolic FM chirps from 703 directions in the frontal hemisphere. To our

knowledge, our bat head is the smallest artificial bat head that has been fabricated.

We designed and fabricated an ultrasonic VLSI cochlea chip with 16 cochlear filters and 128 spiking neurons. The cochlear filters were designed with a moderate quality factor. Each cochlear filter connects to eight neurons; each neuron can be set to have a different threshold. The different thresholds model the different auditory nerve fibers that connect to a single inner hair cell. To our knowledge, our electronic cochlea chip is the first to model the different thresholds of the auditory nerve fibers. Our localization system used two cochlea chips to simulate the right and left ears.

Because of the short duration of the echolocation calls (generally<20ms), the rate coding of intensity is not a viable code. Instead, in our system, the intensity of the sounds recorded from the artificial bat head is encoded in the spike latency of the individual cochlear neurons. We showed that the ILD cue is encoded in the difference between the spike latencies of the right and left cochlear neurons.

We studied the mismatch in the quality factor (Q) of the cochlear filters, and used the lognormal distribution to explain qualitatively (and quantitatively to some extent) why the mismatch in Q increases as Q itself increases.

We designed and fabricated a binaural VLSI interaural level difference (ILD) chip with 32 spiking neurons that can extract the binaural ILD cues from the spike latencies of the right and left cochlea chips. We also designed and fabricated a monaural VLSI spectral difference chip with 240 spiking neurons. The spectral difference chip extracts the monaural spectral cues from the spiking output of a single cochlea chip. The cochlea chip, ILD chip, and spectral difference chip were

fabricated in a 2-poly, 3-metal, $0.5\mu m$ CMOS process.

We demonstrated that the feature extraction chips (the binaural ILD chip and the monaural spectral difference chip) can extract the localization cues from the spiking cochlear output and that the output of the feature extraction chips can be decoded to simultaneously estimate the elevation and azimuth of a sound source. We have also demonstrated that our hardware system can localize with as few as 32 spikes.

8.2 Future Work

In this work we designed and fabricated an artificial bat head as well as several integrated circuits that model some of the processing that takes place in the bat's auditory system and demonstrated that the hardware system can be used to localize sounds. To extend this research we propose several experiments that can be done with the existing hardware system as well as suggest some modifications and improvements to the hardware system.

8.2.1 Further Experiments with the Existing System

Even though studies have shown that the big brown bat compensates for some of the range-dependent changes in intensity; the bat will, nevertheless, experience changes in the absolute value of echo intensity because of the different target sizes as well as dynamic changes in the target's cross-section (for fluttering insects). It is therefore, necessary to investigate how robust the localization system is to

changes in intensity.

Another set of experiments is to investigate how changing some of the system parameters affects the localization acuity and the robustness of the system to intensity changes. Some of the parameters than can be changed are: the quality factor Q of the cochlear filters, the bias currents of the half-wave rectifiers, etc. The individual cochlear filters can also be tuned to have an asymmetric tuning in frequency. The voltage bias V_{ref_hwr} of the half-wave rectifier can also be changed to allow spontaneous firing for some of the cochlear neurons.

In our controlled setup, the spectrum of the sound was identical for all directions. It can be expected that, in a more realistic setting, changes in the spectrum of the echo will occur as a result of the interaction of the incident sound with the complex shape of realistic targets, as well as changes in the intensity of the echo. It seems that the lower frequencies (longer wavelengths) in the FM sweep may be less affected by some of these changes. Since the frequency span of the cochlear filterbank can be modified, the question of whether the system can maintain the same localization acuity while focusing on the lower frequencies of the FM sweep can be addressed by the current system.

8.2.2 Modifications to the Existing System

There are several improvements and additions that can be made to the existing hardware system.

Vocalizations, Middle ear, and Intensity Compensation

As mentioned in Chapter 1, as a bat closes in on its target, the absolute intensity of the received echo return increases because of the decrease in spreading and atmospheric losses. Studies have shown that bats employ several strategies to compensate for intensity increases such as lowering the intensity of the emitted sound, and by contracting their middle ear muscles with each vocalization (which results in range-dependent change in their hearing threshold).

In this dissertation we did not model the outgoing sound nor have we modeled the effect of the middle ear muscles. Modeling of the outgoing vocalization requires the design of a speaker that is small enough to fit in the artificial bat's mouth, and powerful enough to get a strong echo off of a target. Modeling the attenuation of the middle-ear muscle can be achieved either by modulating the gain of the microphone amplification stage or by adding a separate circuit.

Pinna Design

In this dissertation, the pinna that was fabricated has a simple geometric design that is symmetrical about its center plane. One possible line of research is to explore the effect of more complex pinna designs by designing the pinna to be asymmetric as well as to include other parts of the pinna that have not be modeled in this work such as the tragus. One unanswered question is what pinna design feature gives rise to the spectral notch characteristic that has been observed in both cats and bats (see Chapter 2).

Cochlea Chip

One of the improvements that can be made to the cochlea chip is to increase the dynamic range of its input. Another design modification that may improve the quality factor matching of the cochlear filters is to use the same technique used by van Schaik [29] to improve Lyon's original cochlea design, namely, the use of Compatible Lateral Bipolar Transistors [41] to bias the cochlear filters. An alternative way to improve matching is to use floating gate structures to program the different current sources (biases) to their desired values.

Feature Extraction Chips

The ILD chip presented in this work consisted of 16 ILD blocks, one for each cochlear filter. Because of the limited number of pins, the bias settings for all 16 ILD blocks are identical and, therefore, all blocks have the same ILD threshold. Although it is possible with the current hardware system to re-route the address-event spikes from the cochlea chips so that a single ILD chip is dedicated to the ILD processing of a single frequency channel, the ILD neurons would have to be modified first so as to be able to set different ILD thresholds for the 16 ILD blocks. This may be achieved by using a floating gate structure to program the different biases. The same solution also applies to the spectral difference chip, where it is desirable to have different neuron/synapse settings according to the frequency channels being compared.

Static Power Consumption

A significant part of the static power consumption of the localization system is consumed by the artificial neurons. There are 128, 240, and 32 neurons in each cochlea chip, spectral difference chip, and ILD chip respectively. At the heart of the neural circuit is a differential pair that has been operated as a voltage comparator. The disadvantage of using the differential pair is that it is powered by a constant current source. An alternative approach that would reduce the static power consumption would be to use an inverter. The static power consumption may also be lowered by reducing the bias current of the half-wave rectifiers, however, its effect on localization must first be studied.

It should be noted that the microcontroller used in the merge board requires a DC power supply current of approximately $100mA$ which is three orders of magnitude greater than the rest of our system. Nevertheless, the need for the microcontroller can be eliminated by either integrating the entire system on a single chip or by modifying the ILD chip by providing separate inputs for the right and left address-event cochlea spikes, as well as implementing the full address-event protocol between the cochlea chips and the feature extraction chips.

Localization Algorithm

Recall that in our system the spectrum of the received sound (echo) is transformed into a binary code. In our localization procedure we assumed that all bits in the code word have an equal weight and when studying the effect of errors in the code

on localization we assumed that the bit errors could occur randomly in the code.

As mentioned above, it is expected that the lower frequencies in the FM sweep may be less affected by the interaction of the incident sound with the complex shape of realistic targets. Therefore, the localization algorithm should be modified by putting more emphasis (weight) on the lower frequencies.

8.3 Suggested Biological Experiments

It is interesting (and puzzling) to note that when the passive localization ability of the big brown bat *Eptesicus fuscus* was tested using its own vocalization, the bat's localization acuity was only 15 degrees compared to its 1.5° horizontal angle acuity while echolocating. A possible explanation is that as the big brown bat closes in on its target it increases the repetition rate and decreases the duration of its FM sweep. It is therefore reasonable to assume that localization requires knowing the exact shape of the vocalization which is linked to the range of a target. This idea can be tested by passively sending two vocalizations whose time separation and shape matches that which has been observed during the bat's normal behavior and observe whether there is any improvement in the passive localization accuracy.

Another experiment is to test the localization accuracy of the big brown bat *Eptesicus fuscus* at its maximum range (i.e. when it first detects a target) and to test whether the bat's localization acuity is a function of range.

As discussed in Chapter 2, the type IV neuron of the dorsal cochlear nucleus (DCN) has been the focus of research in spectral processing for localization. The

DCN of the big brown bat *Eptesicus fuscus* is similar to that in other mammals in that it has a laminated structure; however, it has a poorly laminated fusiform cell layer [92]. The implication of this observation is not known.

Recall that, echolocation is an active form of localization where the sound being localized is an echo of the sound emitted by animal itself as opposed to the passive localization in cats (and other mammals). It might be speculated that the processing of spectral cues in bats is different from that in other mammals. One of the many hypotheses that have been put forward to explain the role of the MSO in bats is that monaural MSO neurons may extract the monaural spectral cues [12]. This hypothesis can be tested by playing sounds to a bat from different directions while recording from MSO neurons and determine whether there is any spatial pattern in the firing of the neuron.

8.4 Final Thoughts

Designing and fabricating VLSI chips is a costly and time consuming process. It seems imperative to make use of emerging technologies like floating gate analog arrays (FPAAs) to test new ideas. One problem that we encountered time after time is the limited number of chip pins, with the address-event system occupying 1/4 to 1/2 of the available pins; this resulted in neural arrays that have identical bias settings (except for mismatch). Therefore, it seems imperative to incorporate floating gate technology into our chips to have more flexibility in setting the parameters of our systems.

Appendix A

Artificial Bat Head Design

The artificial bat head was designed using the SolidWorks^{TM} CAD software. The bat head was broken down into two parts that were designed separately and assembled (in software). The design was then converted to the STL (Stereolithography) format and fabricated using a 3D printer.

A.1 Head/Body Design

Fig. A.1 illustrates some of the steps in constructing the artificial bat head. An ellipse (width=25mm, height=10mm) was extruded along ellipsoidal guide lines to form the shape shown in Fig. A.1a. The length of the head from tip to base is 10mm. The elliptic base of the head was extruded backwards for an additional 5mm with no change in cross section. A square (side=8mm) was extruded (length=25mm) to form a holding post as shown in Fig. A.3b. Fig. A.1c shows the extruded cuts that were used to form the ear canal holes extending to the back

of the head where the microphones were inserted. A shallow rectangular notch was formed on the surface of the head to assist in assembling the pinna to the head. Fig. A.1d shows the holes that were made in the holding post in which a metallic rod was inserted.

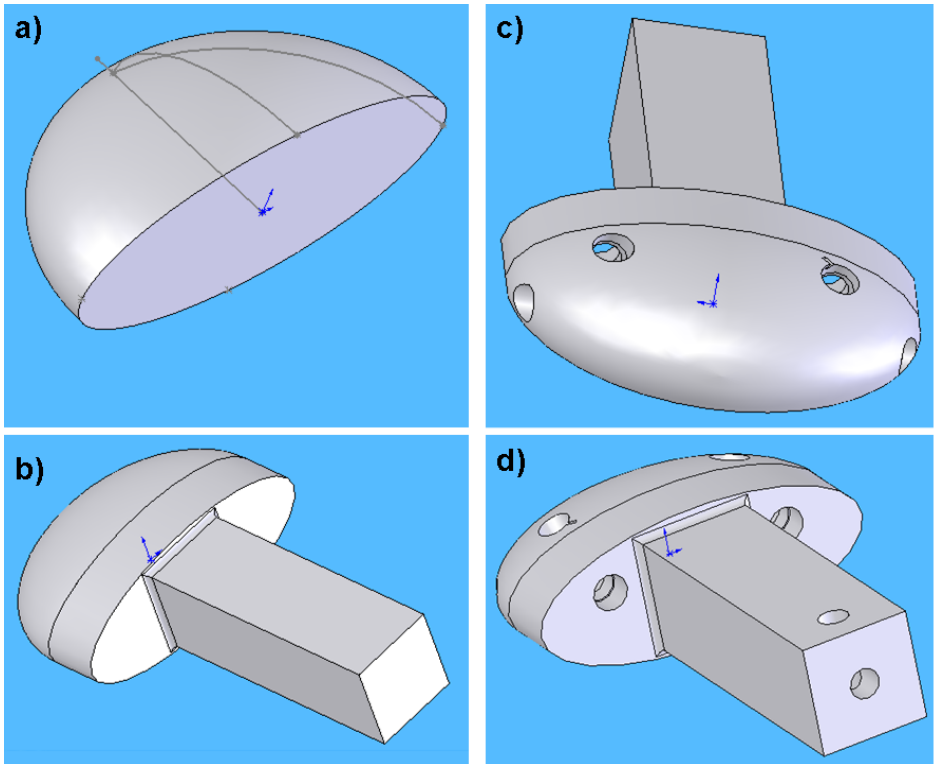

Figure A.1: Constructing the bat head.

A.2 Pinna Design and Assembled Design

Fig. A.2 illustrates some of the steps in constructing the bat pinna. The rectangular block (height=14mm, width=10mm, thickness=3.5mm) shown in Fig. A.2a is the starting point of the pinna design. An extruded cut using a half ellipsoid

(width=9mm, height=5mm) and ellipsoidal guide curves (not shown) formed the inner shape of the pinna shown in Fig. A.2b, the pinna has a maximum height of 12.6mm, maximum depth of 2.5mm (at its base) and a maximum with of 9mm (at its base). Figs. A.2c,d show the removal of the unwanted material using extruded cuts. The Fig. A.3 shows the assembled bat head. The two pinna are pointed outwards by 45 degrees and tilted forward by 20 degrees.

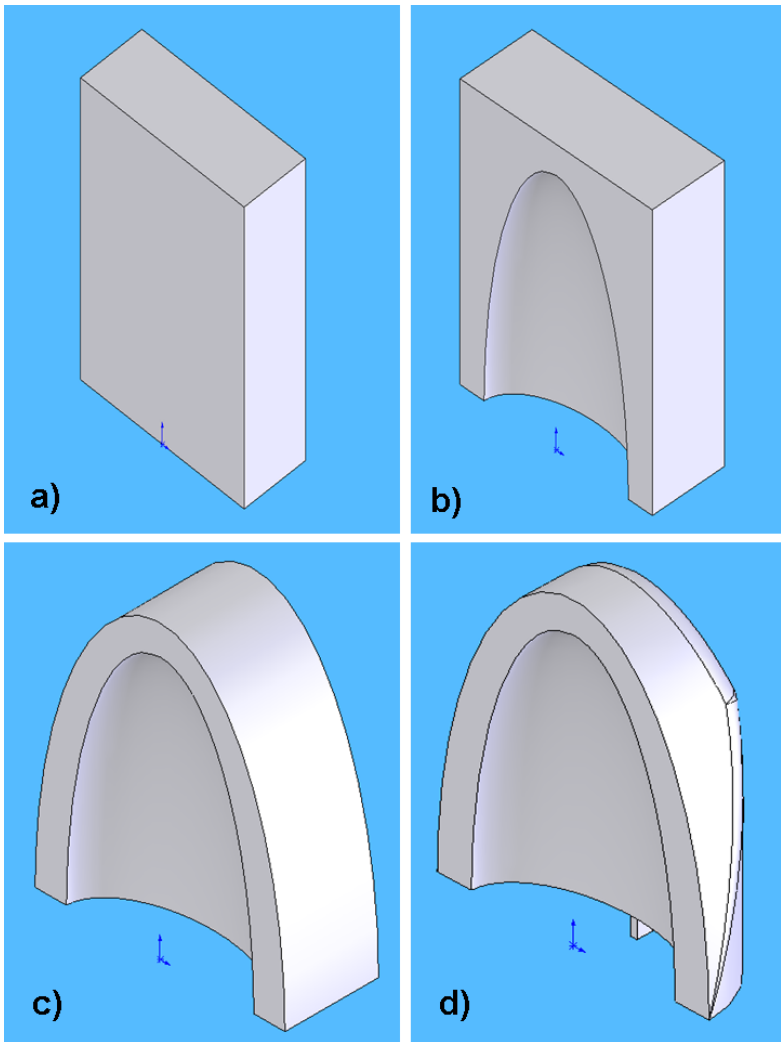

Figure A.2: Constructing the bat pinna. The pinna has a maximum height of 12.6mm, maximum depth of 2.5mm, and a maximum width of 9mm.

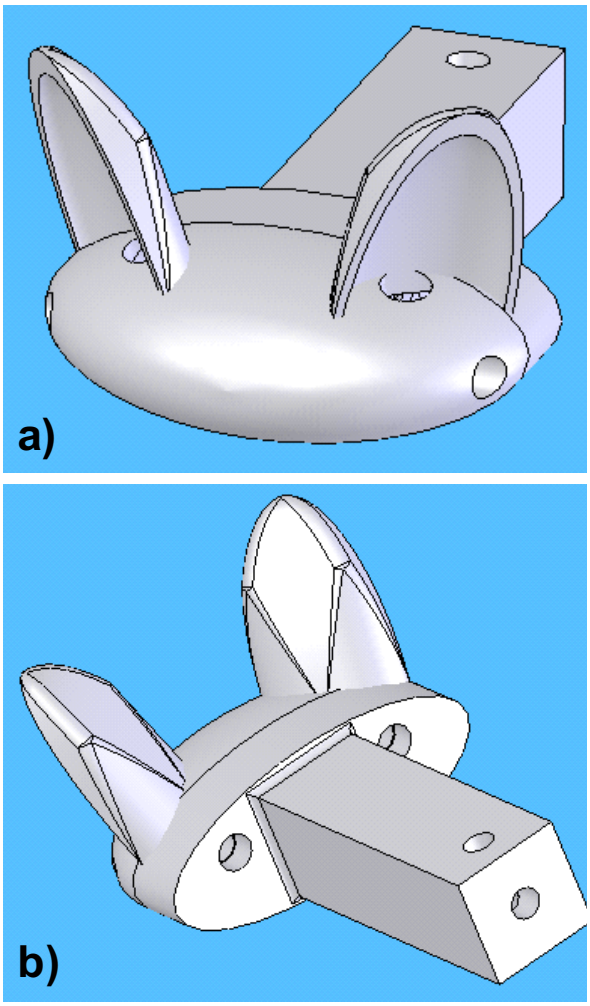

Figure A.3: Assembled bat head. a) Front view. b) Rear view. The two pinna are pointed outwards by 45 degrees and tilted forward by 20 degrees.

Appendix B

Sallen Key Filter

The transfer function of the second order Sallen Key highpass filter shown in Fig. B.1, taking into account the open-loop gain of the operational amplifier (*OpAmp*) can be derived as follows: By definition

$$V_{out} = A(V_1 - V_2) \tag{B.1}$$

where A is the open loop gain of the operational amplifier. The open loop gain of the operational amplifier can be modeled as a single pole lowpass filter with a transfer function

$$A(s) = \frac{A_o}{1 + s\tau} \tag{B.2}$$

where A_o is the DC gain of the operational amplifier and τ is the time-constant of the pole. To simplify the notation, however, the s will be dropped from $A(s)$. Applying the voltage divider rule at the *OpAmp's* $-ve$ input terminal

$$V_2 = \frac{R_3}{R_3 + R_4} V_{out} = \frac{V_{out}}{k} \tag{B.3}$$

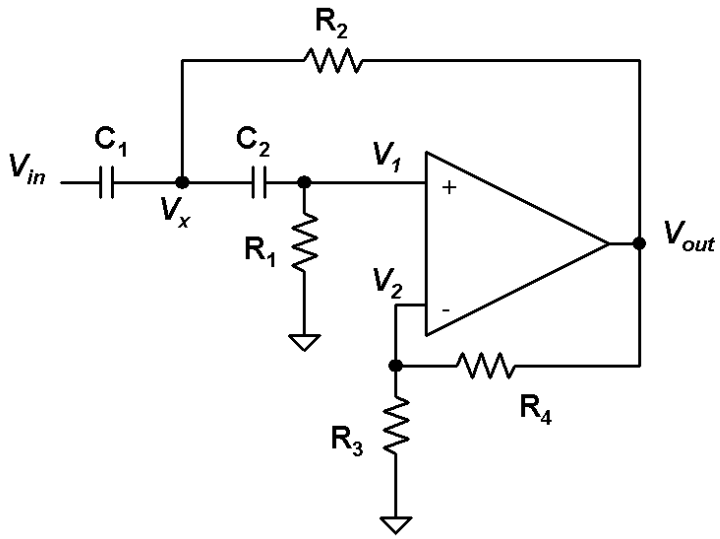

Figure B.1: Schematic of the second order Sallen Key highpass filter.

where k is defined as

$$k = \frac{R_3 + R_4}{R_3} \tag{B.4}$$

Applying the voltage divider rule at the *OpAmp's +ve* input terminal

$$V_1 = \frac{R_1}{R_1 + 1/sC_2} V_x = \frac{V_x}{1 + sR_1C_2} sR_1C_2 \tag{B.5}$$

Applying KCL at the node V_x we get

$$(V_x - V_{in})sC_1 + \frac{V_x}{R_1 + 1/sC_2} + \frac{(V_x - V_{out})}{R_2} = 0$$

$$V_x(sC_1 + \frac{sC_2}{1 + sR_1C_2} + \frac{1}{R_2}) = \frac{V_{out}}{R_2} + sC_1V_{in}$$

$$\frac{V_x}{1 + sR_1C_2}(sR_2C_2 + (1 + sR_2C_1)(1 + sR_1C_2)) = V_{out} + sR_2C_1V_{in} \tag{B.6}$$

Let

$$D = (sR_2C_2 + (1 + sR_2C_1)(1 + sR_1C_2)) \tag{B.7}$$

Substituting B.7 in B.6 and B.6 in B.5 we get

$$V_1 = \frac{sR_1C_2V_{out} + sR_1R_2C_1C_2V_{in}}{D} \tag{B.8}$$

Substituting B.3 and B.8 in B.1 we get

$$\frac{k}{k}V_{out} = A(\frac{sR_1C_2V_{out} + sR_1R_2C_1C_2V_{in}}{D} - \frac{Vout}{k})$$

$$V_{out}(\frac{k+A}{k} - A\frac{sR_1C_2}{D}) = A\frac{sR_1R_2C_1C_2}{D}V_{in} \tag{B.9}$$

Multiplying by $kD/(k+A)$

$$V_{out}(D - \frac{kA}{k+A}sR_1C_2) = \frac{kA}{k+A}sR_1R_2C_1C_2V_{in} \tag{B.10}$$

Let

$$k//A = \frac{kA}{k+A} \tag{B.11}$$

which resembles the expression for the equivalent resistance of two parallel resistors k and A. The transfer function of the filter can then be written as

$$H(s) = \frac{V_{out}}{V_{in}} = \frac{(k//A)sR_1R_2C_1C_2}{s^2R_1R_2C_1C_2 + s(R_2C_2 + R_2C_1 + R_1C_2(1 - k//A)) + 1} \tag{B.12}$$

The transfer function in equation B.12 can be written as

$$H(s) = \frac{ks}{s^2 + s(\frac{\omega_o}{Q}) + \omega_o^2} \tag{B.13}$$

where ω_o is the cutoff frequency and Q is the quality factor. Equating B.12 and B.13:

$$\begin{aligned} \omega_o &= \frac{1}{\sqrt{R_1R_2C_1C_2}} \\ Q &= \frac{\sqrt{R_1R_2C_1C_2}}{R_2C_2 + R_2C_1 + R_1C_2(1 - k//A)} \end{aligned} \tag{B.14}$$

The transfer function shown in Fig. B.2 illustrates why it was necessary to take into account the open-loop gain of the OpAmp in the filter design process. The filter was specified to have a gain k of 60dB (1000), a cutoff frequency f_o of 1MHz, and a Q of 0.5. The design values are:

C_1	C_2	R_1	R_2	R_3	R_4
$470pF$	$470pF$	14.8Ω	$7.73k\Omega$	9.9Ω	$9.88k\Omega$

The implemented design in has $k = 999$, $f_o = 1$MHz, and $Q = 0.49$. In the case of an ideal (*simulated*) operational amplifier the transfer function is in agreement with the design. However, when implemented using an OP27 operational amplifier, which has an 8MHz gain-bandwidth product, the transfer function has a much lower cutoff frequency and the maximum gain is off by more than 50dB than the specified gain. The $k//A$ factor shows up in both the numerator and denominator of equation B.12. In the numerator, $k//A$ is a gain factor for the entire transfer function setting an upper limit on the gain. In the absence of the $k//A$ term in the denominator, the transfer function would have continued to increase (see Fig. B.2) until it approached the open-loop gain after which it would have followed it.

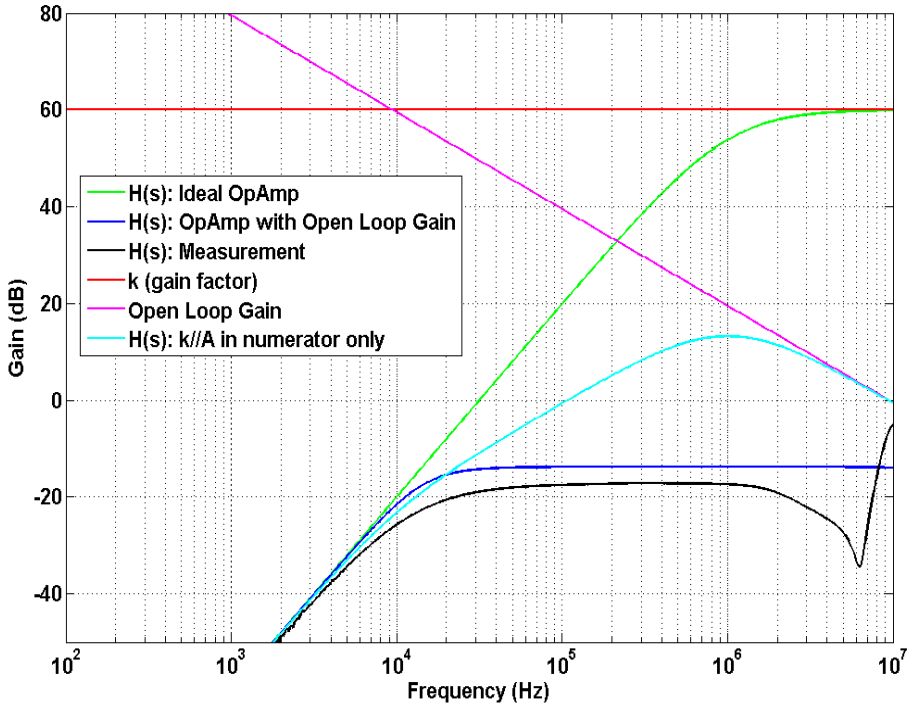

Figure B.2: The transfer function of the second order Sallen-Key highpass filter. The measured transfer function (black) is shown along with both the ideal OpAmp response (green) and when the OpAmp's open-loop gain is included (blue). The transfer function begins to deviate from the ideal behavior when the OpAmp's open-loop gain (magenta) approaches the desired gain k (red).

Appendix C

Transconductance Amplifier

Shown in Fig. C.1 is the schematic of the source-degenerated transconductance amplifier used in the cochlear filter. The transconductance of this circuit is derived as follows: By applying KCL at the node Vc

$$I_b = I_1 + I_2 \tag{C.1}$$

The current I_1 is mirrored by M5 and M6. By applying KCL at the drain of M6 we get

$$I_{out} = I_1 - I_2 \tag{C.2}$$

The currents I_1 and I_2 are given by

$$I_1 = I_o \exp(\frac{\kappa V2 - Vc}{V_T}), \quad I_2 = I_o \exp(\frac{\kappa V4 - Vc}{V_T})$$

Dividing I_1 and I_2 we get

$$\frac{I_1}{I_2} = \exp(\frac{k}{V_T}(V2 - V4)) \tag{C.3}$$

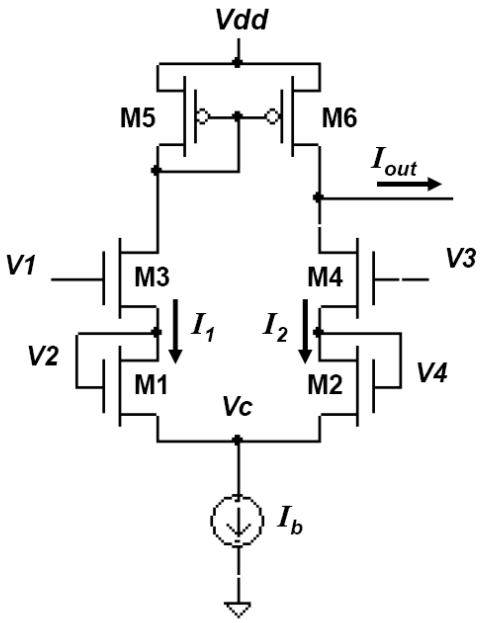

Figure C.1: Schematic of a source-degenerated transconductance amplifier.

By equating the current flowing in M1 and M3 it can be shown that

$$V2 = \frac{\kappa}{1+\kappa}V1 + \frac{Vc}{1+\kappa} \tag{C.4}$$

Similarly, by equating the current flowing in M2 and M4 it can be shown that

$$V4 = \frac{\kappa}{1+\kappa}V3 + \frac{Vc}{1+\kappa} \tag{C.5}$$

Substituting C.4 and C.5 in C.3 we get

$$\begin{aligned} \frac{I_1}{I_2} &= \exp(\frac{k^2}{(1+k)V_T}(V1 - V3)) \\ &= \exp(\frac{k^2}{(1+k)V_T}V_{in}) \\ &= \exp(x) \end{aligned} \tag{C.6}$$

where V_{in} is defined as

$$V_{in} = V1 - V3 \tag{C.7}$$

and x is defined as

$$x = \frac{k^2}{(1+k)V_T} V_{in} \tag{C.8}$$

Manipulating equation C.6 results in

$$\frac{I_1 - I_2}{I_1 + I_2} = \frac{e^x - 1}{e^x + 1} = \frac{e^{x/2} - e^{-x/2}}{e^{x/2} + e^{-x/2}} = \tanh(x/2) \tag{C.9}$$

Substituting equations C.1, C.2, and C.8 in C.9 we get

$$I_{out} = I_b \tanh(x/2) = I_b \tanh(\frac{k^2}{(1+k)2V_T} V_{in}) \tag{C.10}$$

The transconductance G_m of this circuit is defined as

$$\begin{aligned} G_m &= \frac{dI_{out}}{dV_{in}} \qquad at \;\; V_{in} = 0 \\ &= \frac{k^2}{(1+k)2V_T} I_b \\ &= G_{\max} \end{aligned}$$

Fig. C.2 shows I_{out} (normalized with respect to I_b) as a function of V_{in}. If there is a DC offset between the input terminals then the transconductance will be smaller than G_{max}. Fig C.3 shows the transconductance (G_m) of the source degenerated transconductance amplifier (normalized with respect to G_{max}) as a function of the DC offset voltage V_{offset}.

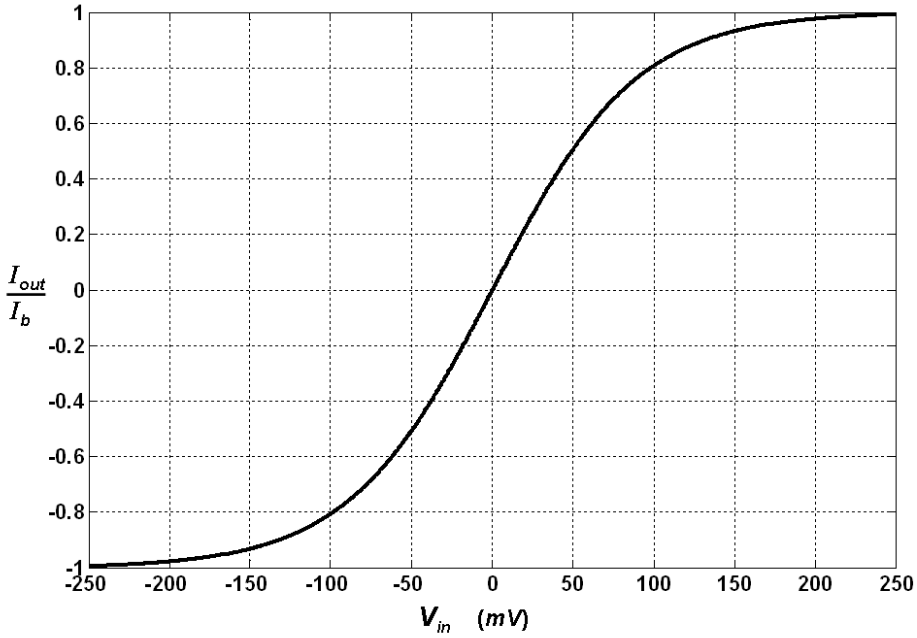

Figure C.2: I_{out} (normalized with respect to I_b) as a function of V_{in}. For this plot a value of 0.7 was used for κ.

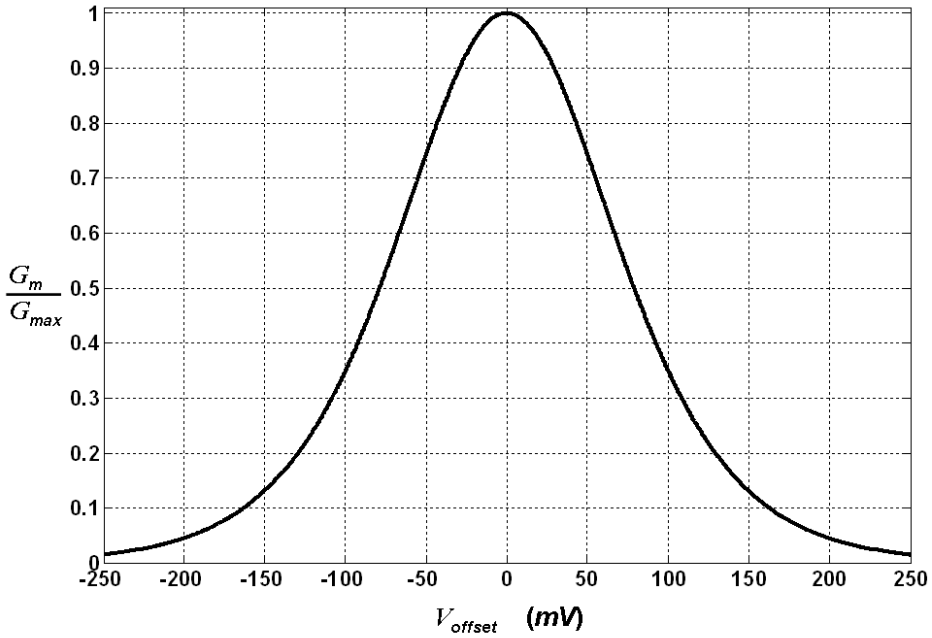

Figure C.3: Transconductance (G_m) of the source degenerated transconductance amplifier as a function of the DC offset voltage V_{offset}.

Appendix D

Filterbank Power Consumption

The expression for the power consumption of the cochlear filterbank can be derived as follows: the i^{th} cochlear filter ($i = 1, 2, ...16$) has four transconductance amplifiers, three of which have the same bias current $I_L(i)$ controlled by a voltage $V_L(i)$ and one has a bias current $I_f(i)$ controlled by a voltage $V_f(i)$. The voltages $V_L(i)$ are generated using a uniform polysilicon resistor ladder, the end voltages of this ladder are V_{L1} and V_{L2}. The voltages $V_f(i)$ are generated using a uniform polysilicon resistor ladder, the end voltages of this ladder are V_{f1} and V_{f2}.The power consumption of a single cochlear filter is given by

$$P(i) = V_{dd}(3I_L(i) + I_f(i)) \tag{D.1}$$

The currents I_L and I_f are generated by the source degenerated circuit

$$\begin{aligned} I_L(i) &= I_o e^{\beta V_L(i)} \\ I_f(i) &= I_o e^{\beta V_f(i)} \\ \beta &= \frac{\kappa^2}{(1+\kappa)2V_T} \end{aligned} \tag{D.2}$$

The voltages $V_L(i)$ and $V_f(i)$ can be written as

$$\begin{aligned} V_L(i) &= V_{L1} + \frac{(V_{L2} - V_{L1})}{15}(i-1) \\ V_f(i) &= V_{f1} + \frac{(V_{f2} - V_{f1})}{15}(i-1) \end{aligned}$$

Notice that

$$\begin{aligned} V_L(1) &= V_{L1}, \quad V_L(16) = V_{L2} \\ V_f(1) &= V_{f1}, \quad V_f(16) = V_{f2} \end{aligned}$$

The filterbank is biased such that$V_f(i) - V_L(i)$ is constant. The power consumption of the entire filterbank is given by

$$P_T = \sum_{i=1}^{16} P(i) \tag{D.3}$$

Substituting equations D.1 and D.2 in D.3

$$\begin{aligned} P_T &= \sum_{i=1}^{16} (V_{dd}(3I_L(i) + I_f(i))) \\ &= V_{dd} I_o \sum_{i=1}^{16} (3e^{\beta V_L(i)} + e^{\beta V_f(i)}) \\ &= V_{dd} I_o (3e^{\beta V_{L1}} \sum_{i=0}^{15} e^{\beta \frac{\triangle}{15} i} + e^{\beta V_{f1}} \sum_{i=0}^{15} e^{\beta \frac{\triangle}{15} i}) \\ &= V_{dd} I_o (3e^{\beta V_{L1}} + e^{\beta V_{f1}}) \sum_{i=0}^{15} e^{\beta \frac{\triangle}{15} i} \\ & \quad V_{dd} I_o (3e^{\beta V_{L1}} + e^{\beta V_{f1}}) \frac{e^{\beta \frac{16}{15} \triangle} - 1}{e^{\beta \frac{1}{15} \triangle} - 1} \end{aligned}$$

Bibliography

[1] T. C. Yin, *Neural Mechanisms of Encoding Binaural Localization Cues in the Auditory Brainstem. In: Integrative Functions in the Mammalian Auditory Pathway*, 1st ed., D. Oertel, A. Popper, and R. Fay, Eds. New York: Springer-Verlag, 2002.

[2] J. C. Middlebrooks, J. C. Makous, and D. M. Green, "Directional sensitivity of sound-pressure levels in the human ear canal," *J. Acoust. Soc. Am.*, vol. 86, pp. 89–108, Jul 1989.

[3] J. C. Middlebrooks and D. M. Green, "Sound localization by human listeners," *Annu. Rev. Psychology*, vol. 42, pp. 135–159, 1991.

[4] K. Maki and S. Furukawa, "Acoustical cues for sound localization by the Mongolian gerbil, Meriones unguiculatus," *J. Acoust. Soc. Am.*, vol. 118, pp. 872–886, Aug 2005.

[5] M. K. Obrist, M. B. Fenton, J. L. Eger, and P. A. Schlegel, "What ears do for bats: a comparative study of pinna sound pressure transformation in chiroptera," *J. Exp. Biology*, vol. 180, pp. 119–152, Jul 1993.

[6] J. Rayleight, *The Theory of Sound*, 2nd ed. New York, NY, USA: Dover Publications Inc., 1896.

[7] G. Koay, D. Kearns, H. E. Heffner, and R. S. Heffner, "Passive sound-localization ability of the big brown bat (Eptesicus fuscus)," *Hearing Research*, vol. 119, pp. 37–48, May 1998.

[8] D. W. Batteau, "The role of the pinna in human localization," *Proc. R. Soc. London, B, Biol. Sci.*, vol. 168, pp. 158–180, Aug 1967.

[9] E. A. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," *J. Acoust. Soc. Am.*, vol. 44, pp. 240–249, Jul 1968.

[10] A. D. Grinnell, *Hearing in Bats: An Overview. In: Hearing by Bats*, 1st ed., A. Popper and R. Fay, Eds. New York: Springer-Verlag, 1995.

[11] M. Fenton, *Natural History and Biosonar Signals. In: Hearing by Bats*, 1st ed., A. Popper and R. Fay, Eds. New York: Springer-Verlag, 1995.

[12] E. Covey and J. Casseday, *The Lower Brainstem Auditory Pathways. In: Hearing by Bats*, 1st ed., A. Popper and R. Fay, Eds. New York: Springer-Verlag, 1995.

[13] J. Simmons, S. Kick, C. Lawrence, C. Bard, and B. Escudie, "Acuity of horizontal angle discrimination by the echolocating bat, Eptesicus fuscus," *J. of Comparative Physiology A*, vol. 153, pp. 321–330, 1983.

[14] B. D. Lawrence and J. A. Simmons, "Echolocation in bats: the external ear and perception of the vertical positions of targets," *Science*, vol. 218, pp. 481–483, Oct 1982.

[15] J. Wadsworth and C. F. Moss, "Vocal control of acoustic information for sonar discriminations by the echolocating bat, Eptesicus fuscus," *J. Acoust. Soc. Am.*, vol. 107, pp. 2265–2271, Apr 2000.

[16] A. Grossetête and C. F. Moss, "Target flutter rate discrimination by bats using frequency-modulated sonar sounds: behavior and signal processing models," *J. Acoust. Soc. Am.*, vol. 103, pp. 2167–2176, Apr 1998.

[17] B. D. Lawrence and J. A. Simmons, "Measurements of atmospheric attenuation at ultrasonic frequencies and the significance for echolocation by bats," *J. Acoust. Soc. Am.*, vol. 71, pp. 585–590, Mar 1982.

[18] S. A. Kick, "Target-detection by the echolocating bat, Eptesicus fuscus," *J. Comparative Physiology. A*, vol. 145, pp. 431–435, Dec 1982.

[19] S. A. Kick and J. A. Simmons, "Automatic gain control in the bat's sonar receiver and the neuroethology of echolocation," *J. Neuroscience*, vol. 4, pp. 2725–2737, Nov 1984.

[20] D. J. Hartley, "Stabilization of perceived echo amplitudes in echolocating bats. I. Echo detection and automatic gain control in the big brown bat, Eptesicus fuscus, and the fishing bat, Noctilio leporinus," *J. Acoust. Soc. Am.*, vol. 91, pp. 1120–1132, Feb 1992.

[21] ——, "Stabilization of perceived echo amplitudes in echolocating bats. II. The acoustic behavior of the big brown bat, Eptesicus fuscus, when tracking moving prey," *J. Acoust. Soc. Am.*, vol. 91, pp. 1133–1149, Feb 1992.

[22] J. Simmons, A. Moffat, and W. Masters, "Sonar gain control and echo detection thresholds in the echolocating bat, Eptesicus fuscus," *J. Acoust. Soc. Am.*, vol. 91, no. 2, pp. 115–1162, Feb 1992.

[23] C. Mead, *Analog VLSI and Neural Systems*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989.

[24] R. Sarpeshkar, R. F. Lyon, and C. Mead, "A low-power wide-dynamic-range analog VLSI cochlea," *Analog Integrated Circuits Signal Processing*, vol. 16, no. 3, pp. 245–274, 1998.

[25] R. Sarpeshkar, "Analog versus digital: extrapolating from electronics to neurobiology," *Neural Computation*, vol. 10, no. 7, pp. 1601–1638, 1998.

[26] R. Douglas, M. Mahowald, and C. Mead, "Neuromorphic analogue VLSI," *Annu. Rev. Neuroscience*, vol. 18, pp. 255–281, 1995.

[27] R. Lyon and C. Mead, "An analog electronic cochlea," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 7, pp. 1119–1134, Jul 1988.

[28] L. Watts, D. Kerns, R. Lyon, and C. Mead, "Improved implementation of the silicon cochlea," *Solid-State Circuits, IEEE Journal of*, vol. 27, no. 5, pp. 692–700, May 1992.

[29] A. van Schaik, E. Fragnire, and E. Vittoz, "Improved silicon cochlea using compatible lateral bipolar transistors," *in Advances in Neural Information Processing Systems 8*, pp. 671–677, 1996.

[30] J. Lazzaro and C. Mead, *A silicon model of auditory localization. In:An introduction to neural and electronic networks*, S. Zornetzer, J. Davis, and C. Lau, Eds. San Diego, CA, USA: Academic Press Professional, Inc., 1990.

[31] ——, *Circuit Models of Sensory Transduction in the Cochlea. In: Analog VLSI Implementation of Neural Networks*, M. Ismail and C. Mead, Eds. Norwell: Kluwer Academic Publishers, 1989.

[32] A. van Schaik, "Analogue VLSI Building Blocks for an Electronic Auditory Pathway," Dissertation, Swiss Federal Institute of Technology, 1997.

[33] ——, "An analog VLSI model of periodicity extraction in the human auditory system," *Analog Integrated Circuits Signal Processing*, vol. 26, no. 2, pp. 157–177, 2001.

[34] A. Mishra and A. Hubbard, "A cochlear filter implemented with a field-programmable gate array," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 49, no. 1, pp. 54–60, Jan 2002.

[35] T. Horiuchi and K. M. Hynna, "A VLSI-based model of azimuthal echolocation in the big brown bat," *Auton. Robots*, vol. 11, no. 3, pp. 241–247, 2001.

[36] T. Horiuchi and K. Hynna, "Spike-based VLSI modeling of the ILD system in the echolocating bat," *Neural Networks*, vol. 14, pp. 755–762, 2001.

[37] R. Z. Shi and T. K. Horiuchi, "A neuromorphic VLSI model of bat interaural level difference processing for azimuthal echolocation," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 54, no. 1, pp. 74–88, Jan. 2007.

[38] R. Shi and T. Horiuchi, "A VLSI model of the bat lateral superior olive for azimuthal echolocation," *Circuits and Systems, 2004. ISCAS '04. Proceedings of the 2004 International Symposium on*, vol. 4, pp. IV–900–3 Vol.4, May 2004.

[39] ——, "A VLSI model of the bat dorsal nucleus of the lateral lemniscus for azimuthal echolocation," *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, pp. 4217–4220 Vol. 5, May 2005.

[40] M. Cheely and T. Horiuchi, "Analog VLSI models of range-tuned neurons in the bat echolocation system," *EURASIP J. Appl. Signal Process.*, vol. 2003, pp. 649–658, 2003.

[41] E. Vittoz, "MOS transistors operated in the lateral bipolar mode and their application in CMOS technology," *Solid-State Circuits, IEEE Journal of*, vol. 18, no. 3, pp. 273–279, Jun 1983.

[42] L. Watts, "Cochlear Mechanics: Analysis and Analog VLSI," Dissertation, California Institute of Technology, 1993.

[43] B. Wen and K. Boahen, "A linear cochlear model with active bi-directional coupling," *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, vol. 3, pp. 2013–2016 Vol.3, Sept. 2003.

[44] ——, *Active Bidirectional Coupling in a Cochlear Chip*, Y. Weiss, B. Schölkopf, and J. Platt, Eds. Cambridge, MA: MIT Press, 2006.

[45] B. Wen, "Modeling the Nonlinear Active Cochlea: Mathematics and Analog VLSI," Dissertation, University of Pennsylvania, 2006.

[46] T. Hamilton, C. Jin, A. van Schaik, and J. Tapson, "A 2-D silicon cochlea with an improved automatic quality factor control-loop," *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*, pp. 1772–1775, May 2008.

[47] ——, "An active 2-D silicon cochlea," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 2, no. 1, pp. 30–43, March 2008.

[48] T. Hamilton, J. Tapson, C. Jin, and A. van Schaik, "Analogue VLSI implementations of two dimensional, nonlinear, active cochlea models," *Biomedical Circuits and Systems Conference, 2008. BioCAS 2008. IEEE*, pp. 153–156, Nov. 2008.

[49] T. Hamilton, C. Jin, and A. van Schaik, "A basilar membrane resonator for an active 2-D cochlea," *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*, pp. 2387–2390, May 2007.

[50] C. Summerfield and R. Lyon, "ASIC implementation of the Lyon cochlea model," *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, vol. 5, pp. 673–676 vol.5, Mar 1992.

[51] M. P. Leong, C. T. Jin, and P. H. W. Leong, "An FPGA-based electronic cochlea," *EURASIP J. Appl. Signal Processing*, vol. 2003, pp. 629–638, 2003.

[52] H. Abdalla and T. Horiuchi, "An ultrasonic filterbank with spiking neurons," *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, pp. 4201–4204 Vol. 5, May 2005.

[53] ——, "Binaural spectral cues for ultrasonic localization," *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*, pp. 2110–2113, May 2008.

[54] K. Boahen, "Point-to-point connectivity between neuromorphic chips using address events," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 47, no. 5, pp. 416–434, May 2000.

[55] B. Fontaine and H. Peremans, "Bat echolocation processing using first-spike latency coding," *Neural Networks*, In Press 2009.

[56] M. Aytekin, E. Grassi, M. Sahota, and 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, Dec 2004.

[57] H. Abdalla and T. Horiuchi, "An analog vlsi low-power envelope periodicity detector," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 52, no. 9, pp. 1709–1720, Sept. 2005.

[58] L. Robles and M. A. Ruggero, "Mechanics of the mammalian cochlea," *Physiological Reviews*, vol. 81, no. 3, pp. 1305–52, 2001.

[59] C. D. Geisler, *From Sound to Synapse: Physiology of the Mammalian Ear*. Oxford, New York: Oxford University Press, 1998.

[60] L. Chittka and A. Brockmann, "Perception space–the final frontier," *PLoS Biol.*, vol. 3, p. e137, Apr 2005.

[61] B. F. Katz, "Boundary element method calculation of individual head-related transfer function. I. rigid model calculation," *J. Acoust. Soc. Am.*, vol. 110, no. 5 Pt 1, pp. 2440–8, 2001.

[62] ——, "Boundary element method calculation of individual head-related transfer function. II. impedance effects and comparisons to real measurements," *J. Acoust. Soc. Am.*, vol. 110, no. 5 Pt 1, pp. 2449–55, 2001.

[63] V. R. Algazi, R. O. Duda, R. Duralswami, N. A. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," *J. Acoust. Soc. Am.*, vol. 112, pp. 2053–2064, Nov 2002.

[64] T. Xiao and Q. H. Liu, "Finite difference computation of head-related transfer function for human hearing," *J. Acoust. Soc. Am.*, vol. 113, pp. 2434–2441, May 2003.

[65] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis," *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 858–67, 1989.

[66] J. J. Rice, B. J. May, G. A. Spirou, and E. D. Young, "Pinna-based spectral cues for sound localization in cat," *Hearing Research*, vol. 58, pp. 132–152, Mar 1992.

[67] E. Young, J. Rice, I. Nelken, and R. Conley, *Head-Related Transfer Functions in cat: neural representation and the effects of pinna movement. In: Binaural and Spatial Hearing in Real and Virtual Environments*, 1st ed., R. Gilkey and T. Anderson, Eds. New Jersey: Erlbaum Associates, 1997.

[68] P. Dallos, "The active cochlea," *J. Neuroscience*, vol. 12, pp. 4575–4585, Dec 1992.

[69] M. Vater and M. Kossl, *The Ears of Whales and Bats. : Echolocation in Bats and Dolphins*, 1st ed., J. Thomas, C. Moss, and M. Vater, Eds. Chicago: University of Chicago Press, 2004.

[70] N. Suga and P. Jen, "Peripheral control of acoustic signals in the auditory system of echolocating bats," *J. Exp. Biology*, vol. 62, no. 2, pp. 277–311, 1975.

[71] Y. Raphael and R. A. Altschuler, "Structure and innervation of the cochlea," *Brain Res. Bull.*, vol. 60, no. 5-6, pp. 397–422, 2003.

[72] N. B. Slepecky, *Structure of the Mammalian Cochlea. In: The Cochlea*, 1st ed., P. Dallos, A. Popper, and R. Fay, Eds. New York: Springer-Verlag, 1996.

[73] D. K. Ryugo, *The auditory nerve: Peripheral innervation, cell body morphology, and central projections. In: The Mammmalian Auditory Pathway: Neuroanatomy*, 1st ed., D. Webster, A. Popper, and R. Fay, Eds. New York: Springer-Verlag, 1992.

[74] A. Hudspeth, "Mechanical amplification of stimuli by hair cells," *Curr. Opin. Neurobiology*, vol. 7, pp. 480–486, Aug 1997.

[75] D. K. Chan and A. J. Hudspeth, "Ca2+ current-driven nonlinear amplification by the mammalian cochlea in vitro," *Nature Neuroscience*, vol. 8, pp. 149–155, Feb 2005.

[76] D. B. Webster, *An Overview of Mammalian Auditory Pathways with an Emphasis on Humans. In: The Mammmalian Auditory Pathway: Neuroanatomy*, 1st ed., D. Webster, A. Popper, and R. Fay, Eds. New York: Springer-Verlag, 1992.

[77] M. A. Reid, J. Flores-Otero, and R. L. Davis, "Firing Patterns of Type II Spiral Ganglion Neurons In Vitro," *J. Neuroscience*, vol. 24, no. 3, pp. 733–742, 2004.

[78] N. Kiang, *Discharge patterns of single fibers in the cat's auditory nerve*, 1st ed. Cambridge, Mass.: MIT Press, 1965.

[79] P. A. Saillant, J. A. Simmons, S. P. Dear, and T. A. McMullen, "A computational model of echo processing and acoustic imaging in frequency-modulated echolocating bats: the spectrogram correlation and transformation receiver," *J. Acoust. Soc. Am.*, vol. 94, pp. 2691–2712, Nov 1993.

[80] N. Suga and P. H. Jen, "Further studies on the peripheral auditory system of 'CF-FM' bats specialized for fine frequency analysis of Doppler-shifted echoes," *J. Exp. Biology*, vol. 69, pp. 207–232, Aug 1977.

[81] M. C. Liberman, "Auditory-nerve response from cats raised in a low-noise chamber," *J. Acoust. Soc. Am.*, vol. 63, no. 2, pp. 442–55, 1978.

[82] ——, "Single-Neuron Labeling in the Cat Auditory Nerve," *Science*, vol. 216, pp. 1239–1241, Jun 1982.

[83] T. Kawase and M. C. Liberman, "Spatial organization of the auditory nerve according to spontaneous discharge rate," *J. Comparative Neurology*, vol. 319, pp. 312–318, May 1992.

[84] M. A. Ruggero, *Physiology and Coding of Sound in the Auditory Nerve. In: The Mammmalian Auditory Pathway: Neurophysiology*, 1st ed., A. Popper and R. Fay, Eds. New York: Springer-Verlag, 1992.

[85] W. S. Rhode and S. Greenberg, *Physiology of the Cochlear Nuclei. In: The Mammmalian Auditory Pathway: Neurophysiology*, 1st ed., A. Popper and R. Fay, Eds. New York: Springer-Verlag, 1992.

[86] B. J. May, "Role of the dorsal cochlear nucleus in the sound localization behavior of cats," *Hearing Research*, vol. 148, pp. 74–87, Oct 2000.

[87] T. J. Imig, N. G. Bibikov, P. Poirier, and F. K. Samson, "Directionality derived from pinna-cue spectral notches in cat dorsal cochlear nucleus," *J. Neurophysiology*, vol. 83, pp. 907–925, Feb 2000.

[88] K. A. Davis, R. Ramachandran, and B. J. May, "Auditory processing of spectral cues for sound localization in the inferior colliculus," *J. Assoc. Res. Otolaryngology*, vol. 4, pp. 148–163, Jun 2003.

[89] E. D. Young and W. E. Brownell, "Responses to tones and noise of single cells in dorsal cochlear nucleus of unanesthetized cats," *J. Neurophysiology*, vol. 39, pp. 282–300, Mar 1976.

[90] E. Young and K. A. Davis, *Circuitry and Function of the Dorsal Cochlear Nucleus. In: Integrative Functions in the Mammalian Auditory Pathway*, 1st ed., D. Oertel, A. Popper, and R. Fay, Eds. New York: Springer-Verlag, 2002.

[91] I. Nelken and E. D. Young, "Two separate inhibitory mechanisms shape the responses of dorsal cochlear nucleus type IV units to narrowband and wideband stimuli," *J. Neurophysiology*, vol. 71, pp. 2446–2462, Jun 1994.

[92] S. Haplea, E. Covey, and J. Casseday, "Frequency tuning and response latencies at three levels in the brainstem of the echolocating bat, Eptesicus fuscus," *J. of Comparative Physiology A*, vol. 174, pp. 671–683, 1994.

[93] L. A. Reiss and E. D. Young, "Spectral edge sensitivity in neural circuits of the dorsal cochlear nucleus," *J. Neuroscience*, vol. 25, pp. 3680–3691, Apr 2005.

[94] B. Grothe, E. Covey, and J. H. Casseday, "Medial superior olive of the big brown bat: neuronal responses to pure tones, amplitude modulations, and pulse trains," *J. Neurophysiology*, vol. 86, pp. 2219–2230, Nov 2001.

[95] B. Grothe and T. J. Park, "Structure and function of the bat superior olivary complex," *Microsc. Res. Tech.*, vol. 51, pp. 382–402, Nov 2000.

[96] T. J. Park, P. Monsivais, and G. D. Pollak, "Processing of Interaural Intensity Differences in the LSO: Role of Interaural Threshold Differences," *J. Neurophysiology*, vol. 77, no. 6, pp. 2863–2878, 1997.

[97] M. C. Reed and J. J. Blum, "A model for the computation and encoding of azimuthal information by the lateral superior olive," *J. Acoust. Soc. Am.*, vol. 88, no. 3, pp. 1442–1453, 1990.

[98] M. Baker and R. Sarpeshkar, "A low-power high-psrr current-mode microphone preamplifier," *Solid-State Circuits, IEEE Journal of*, vol. 38, no. 10, pp. 1671–1678, Oct. 2003.

[99] "http://www.tdt.com/products/EC1.htm," Tucker Davis Technologies, Alachua, FL 32615 U.S.A.

[100] "FG series datasheet," Knowles Acoustics, Itasca, IL 60143 U.S.A.

[101] P. H. Jen and D. M. Chen, "Directionality of sound pressure transformation at the pinna of echolocating bats," *Hearing Research*, vol. 34, pp. 101–117, Jul 1988.

[102] P. Furth, "A continuous-time bandpass filter implemented in subthreshold cmos with large-signal stability," *Circuits and Systems, 1997. Proceedings of the 40th Midwest Symposium on*, vol. 1, pp. 264–267 vol.1, Aug 1997.

[103] M. D. Godfrey, "CMOS device modeling for subthreshold circuits," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 39, no. 8, pp. 532–539, 1992.

[104] D. Culler and H. Mulder, "Smart sensors to network the world," *Scientific American*, pp. 85–91, 2004.

[105] A. V. 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, 2001.

[106] D. Goldberg, P. Pouliquen, A. Andreou, L. Riddle, P. Julian, and R. Rosasco, "A wake-up detector for an acoustic surveillance sensor network: algorithm and VLSI implementation," in *Proc. Information Processing in Sensor Networks,International Symposium on*, April 2004, pp. 134–141.

[107] O. D. Ladislav, "An instantaneous pitch-period indicator," *J. Acoust. Soc. Am.*, vol. 27, no. 1, pp. 67–72, 1955.

[108] F. Miroslav, "Envelope periodicity detection," *J. Acoust. Soc. Am.*, vol. 45, no. 3, pp. 719–732, 1969.

[109] B. Gold and L. Rabiner, "Parallel processing techniques for estimating pitch periods of speech in the time domain," *J. Acoust. Soc. Am.*, vol. 46, pp. 442–448, 1969.

[110] C. Mead, *Analog VLSI and Neural Systems*. Reading, Mass.: Addison-Wesley, 1989.

[111] J. E. Dowling, *Neurons and Networks : An Introduction to Neuroscience*. Cambridge, Mass.: Harvard University Press, 1992.

[112] N. Kumar, G. Cauwenberghs, and A. G. Andreou, "Auditory feature extraction using self-timed, continuous-time discrete-signal processing circuits," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 44, no. 9, pp. 723–728, 1997.

[113] H. L. Van Trees and J. W. . Sons., "Detection, estimation, and modulation theory. part I, detection, estimation, and linear modulation theory," 2001.

[114] C. Koch and I. Segev, *Methods in Neuronal Modeling: From Ions to Networks*, 2nd ed. Cambridge, Mass.: MIT Press, 1998.

[115] D. H. Johnson, "Point process models of single-neuron discharges," *J. of Computational Neuroscience*, vol. 3, no. 4, pp. 275–299, 1996.

[116] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, 2nd ed. Oxford, New York: Oxford University Press, 1992.

[117] "The DARPA TIMIT acoustic-phonetic continuous speech corpus," pp. NIST Speech Disc 1–1.1, NTIS order number PB91–505 065, 1990.