Analysis of Exotic Cat Vocalizations and

Middle-Ear Properties

by

Kevin N. O'Connor

Submitted to the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January 10, 2001

© Massachusetts Institute of Technology, 2001. All Rights Reserved.
Analysis of Exotic Cat Vocalizations and Middle-Ear Properties

by

Kevin N. O'Connor

Submitted to the
Department of Electrical Engineering and Computer Science

January 10, 2001

In Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Electrical Engineering and
Computer Science

Abstract

The effects of a peculiar feature of the middle ears in species of the cat family on the reception of intraspecies vocalizations are examined in two species: bobcat and cheetah. Vocalizations of both species are analyzed, and computer-based middle-ear models for each species are generated. The vocalizations often exhibit harmonic structure and the middle ears act as notch filters when transmitting sound from the ear canal to the cochlea. An hypothesis is tested concerning the implications of these features on vocal communication among members of the same species; specifically, does the middle-ear notch have a perceptually salient effect on the vocalizations when it attenuates one of the harmonics? The main finding is that, for the available data, the middle-ear notch does not affect the vocalizations in a manner that could be exploited to enhance vocal communication, and other “adaptive” explanations for the middle-ear structure are more promising.

Thesis Supervisor: William T. Peake
Title: Professor of Electrical and Bio-Engineering
# Table of Contents

Introduction ......................................................................................................................... 10

Part I: Vocalization Descriptions

Background ............................................................................................................................. 18

Chapter 1: Bobcat Vocalizations .......................................................................................... 20

1.1 Methods ......................................................................................................................... 20

1.2 Results .......................................................................................................................... 29

1.3 Summary of Features ................................................................................................... 39

1.4 Discussion and Conclusions ......................................................................................... 42

Chapter 2: Cheetah Vocalizations ....................................................................................... 46

2.1 Methods ......................................................................................................................... 46

2.2 Results .......................................................................................................................... 47

2.3 Summary of Features ................................................................................................... 52

2.4 Discussion and Conclusions ......................................................................................... 54

Part II: Filter Construction and Hypothesis Testing

Background ............................................................................................................................. 58

Chapter 3: Construction of Middle-Ear Filters ................................................................. 62

3.1 Introduction to Model ..................................................................................................... 62

3.2 Theoretical Model Characteristics ................................................................................ 65

3.3 Filter Implementation ..................................................................................................... 69

3.4 Examination of Filters and Comparisons to Theory ..................................................... 82

3.5 Modifications of Models ................................................................................................. 84

3.6 Summary ......................................................................................................................... 92
Chapter 4: Hypothesized Role of the Middle Ear in Vocal Communication ...... 94

4.1 A Model for the Structure, Perception, and Meaning of Vocalizations .... 94

4.2 Hypothesis of the Middle Ear's Role in Felid Vocal Communication ..... 97

4.3 Proposed Methodology for Testing the Hypothesis ....................... 100

Chapter 5: Tests of the Perceptual Hypotheses ........................................ 102

5.1 Filtering and Preliminary Listening .................................................. 102

5.2 Quantitative Perceptual Tests ......................................................... 117

5.3 Correlations Between Test Results and Vocalization Structure ........... 131

5.4 Perception of Filtered or Unfiltered Noise versus Vocalizations ........... 144

Chapter 6: Overall Conclusions ............................................................. 152

6.1 Summary of Results ........................................................................ 152

6.2 Additional Work ............................................................................. 155

6.3 An Alternate Hypothesis ................................................................. 156

Appendix 1: Bobcat Vocalization Summaries and Spectrograms .......... 158

Appendix 2: Cheetah Vocalization Summaries and Spectrograms .......... 190

Appendix 3: Matlab Code for Chapter 3 .................................................. 214

Appendix 4: Matlab Code for Chapter 5 .................................................. 228

Bibliography ........................................................................................... 231
List of Figures

Figure 0.1: Vocal Communication Chain ................................................................. 11
Figure 0.2: Felid Middle-Ear Structure, Cross-Sectional View ............................ 13
Figure 0.3: Measured Middle-Ear Cavity Gain for Domestic Cat and Lion .......... 14
Figure 0.4: Cheetah Vocalization, to Illustrate Harmonic Structure .................... 15
Figure 1.1: Examples of Observation and Measurement Categories .................. 28
Figure 1.2: Example of Pulsed Amplitude Modulation ............................................ 29
Figure 2.1: Example of Pulsed Amplitude Modulation in a “Stutter” .................... 48
Figure 2.2: Middle-Ear Circuit Model Superimposed Upon the Structure it Represents .. 59
Figure 3.1: Felid Middle-Ear Model ........................................................................ 62
Figure 3.2: Low-Frequency Middle-Ear Model ....................................................... 65
Figure 3.3: Linear Plots of $G_{CAV}$ for Left and Right Bobcat Middle Ears ........ 71
Figure 3.4: Bode Plots of $G_{CAV}$ for Left and Right Bobcat Middle Ears ......... 72
Figure 3.5: Linear Plots of $G_{CAV}$ for Left and Right Cheetah Middle Ears ....... 73
Figure 3.6: Bode Plots of $G_{CAV}$ for Left and Right Cheetah Middle Ears ......... 74
Figure 3.7: Illustration of Mapping from Unit Circle to ifft Vector with $N = 8$ ...... 76
Figure 3.8: First 100 Points of the Bobcat and Cheetah Impulse Responses ......... 78
Figure 3.9: Left Bobcat ME: Comparison Betw. Orig. and Truncated Versions ....... 81
Figure 3.10: Left Cheetah ME: Comparison Betw. Orig. and Truncated Versions .... 82
Figure 3.11: Middle-Ear Model with a Plugged Foramen ........................................ 85
Figure 3.12: Middle-Ear Model with Removed Septum .......................................... 86
Figure 3.13: Left Bobcat Ear: Linear Plot Comparison of Original to Plugged Foramen and Removed Septum .......................................................... 87
Figure 3.14: Left Bobcat Ear: Bode Plot Comparison of Original to Plugged Foramen and Removed Septum .................................................................88

Figure 3.15: Left Cheetah Ear: Linear Plot Comparison of Original to Plugged Foramen and Removed Septum .................................................................89

Figure 3.16: Left Cheetah Ear: Linear Plot Comparison of Original to Plugged Foramen and Removed Septum .................................................................90

Figure 4.1: A Model of Felid Perception of Vocalizations .................................................................97

Figure 4.2: Hypothetical Role of the Middle Ear in Vocal Communication .........................................99

Figure 5.1: Overall Scores of Identification and Discrimination Listening Tests .............................122

Figure 5.2: Test Scores for Individual Tokens .....................................................................................123

Figure 5.3: Scores of Individual Test Subjects .....................................................................................124

Figure 5.4: Token 1 H2 and H3 Freqs and Amps Before and After Filtering ........................................134

Figure 5.5: Token 2 H2 and H3 Freqs and Amps Before and After Filtering ........................................135

Figure 5.6: Token 3 H2 and H3 Freqs and Amps Before and After Filtering ........................................136

Figure 5.7: Token 4 H2 and H3 Freqs and Amps Before and After Filtering ........................................137

Figure 5.8: Token 5 H2 and H3 Freqs and Amps Before and After Filtering ........................................138

Figure 5.9: Token 6 H2 and H3 Freqs and Amps Before and After Filtering ........................................139

Figure 5.10: Plots of Test Scores and Structural Measurements on Same Axes ..............................141

Figure 5.11: Listening Test Results for Vocalization and Noise Tokens ............................................149
List of Tables

Table 1.1: Summary of Digitized Vocalization Collections ...........................................22
Table 1.2: Summary of Bobcat Vocalization Features ..................................................41
Table 1.3: “Alpha” Significance Levels based on a 2-sided T-test .................................44
Table 2.1: Summary of Cheetah Descriptions (by Collection) ....................................52
Table 2.2: Summary of Cheetah Descriptions (by Type) .............................................54
Table 3.1: Experimentally Derived Element Values .....................................................64
Table 3.2: Predicted Low-Frequency Behavior of Middle Ears ..................................66
Table 3.3: Calculated Center Frequencies and Bandwidths of the $G_{CAV}$ Dip ........68
Table 3.4: Comparisons of $G_{CAV}$ Implementation to Theory ..................................83
Table 5.1: Summary of Symbols used to Classify Perceived Pre-Post Differences ......105
Table 5.2: Preliminary Perceptual Judgments for the BC-I Collection ..........................107
Table 5.3: Preliminary Perceptual Judgments for the BC-II Collection .........................108
Table 5.4: Preliminary Perceptual Judgments for the BC-III Collection .......................108
Table 5.5: Preliminary Perceptual Judgments for the BC-IV Collection .......................109
Table 5.6: Preliminary Perceptual Judgments for the BC-V Collection .......................109
Table 5.7: Preliminary Perceptual Judgments for the CH-I Collection .........................110
Table 5.8: Preliminary Perceptual Judgments for the CH-II Collection .......................111
Table 5.9: Preliminary Perceptual Judgments for the CH-III Collection .......................112
Table 5.10: Summary of Perceptual Judgments for all 8 Vocalization Collections .......113
Table 5.11: Listening Test Tokens ..............................................................................119
Table 5.12: Identification Test Results ......................................................................120
Table 5.13: Discrimination Test Results ...................................................................121
Table 5.14: Alpha Values for Identification Scores (the Probability that Scores Could be Equalled or Exceeded by Chance Alone) ........................................................ 126

Table 5.15: Alpha Values for Discrimination (ABX) Test Scores (the Probability that Scores Could be Equalled or Exceeded by Chance Alone) ..............................127

Table 5.16: Structural Measurements and Correlations to Listening Test Results ........140

Table 5.17: Noise and Vocalizations Selected for Listening Test .................................147

Table 5.18: Noise vs. Vocalization Listening Test Results ............................................148
Introduction

Vocalizations are a common way for mammals to communicate with one another. Researchers studying vocal communication in humans have broken the process down into a series of stages called the “Speech Communication Chain”. A vocalization is thought to begin when the sender’s brain initiates the elaborate control of the mechanisms in the chest, glottis, and vocal tract necessary to transform it from a thought into a sound. As a sound travelling through the air, a vocalization is subject to environmental factors that may, for instance, affect its amplitude, frequency content, and noisiness. Eventually the vocalization enters the ear of the receiver, passing through the outer ear, middle ear, and the inner ear, where it is transformed into nerve signals that are sent to the receiver’s brain to be perceived.

This thesis involves vocal communication in two felid species, the bobcat and the cheetah, and focuses on the transmission of sound through the middle ear. Figure 0.1 shows a rendition of the vocal communication chain for felids, depicting a sender, a receiver, and an enlarged view of the receiver’s outer, middle, and inner ear. The middle ear in the figure contains a linkage of small bones (called ossicles) that couple motion of the tympanic membrane to motion at the entrance to the cochlea (called the oval window). In addition to these bones, the middle ear features two air cavities connected by a small opening. The cavity containing the ossicles, called the tympanic cavity, is connected to an empty cavity, called the bullar cavity, by a short and narrow airspace called the foramen. A bony septum serves to separate the two air cavities.

The reason that this thesis focuses on the middle ears of cats is that all 37 felid species feature the characteristic structure described above, and that in many cases this structure has an interesting effect on the transmission of an acoustic signal from the ear canal to the cochlea.
Stages of vocal communication in felids: a vocalization leaves the mouth of the sender, travels through the air and passes through the ear of the receiver. An enlarged view of the receiver’s ear is supplied, showing the basic features of the outer, middle, and inner ear. The middle ear is the region between the tympanic membrane and the entrance to the cochlea (oval window). In cats it consists of two air spaces (tympanic cavity and bullar cavity) separated by a bony septum and connected by a small opening (foramen). A linkage of small bones (ossicles) connects the tympanic membrane to the oval window, transmitting sound from the outer ear to the inner ear.

Figure 0.1: Vocal Communication Chain
Figure 0.2 depicts an enlarged cross-sectional view of the felid middle ear, without the ossicles or oval window shown. The transfer function from the ear-canal pressure ($P_{TM}$) to the pressure difference across the tympanic membrane ($P_{TM} - P_{CAV}$) is called the “cavity gain” ($G_{CAV}$), and is equal to ($P_{TM} - P_{CAV}$)/$P_{TM}$. The cavity gain represents the first stage in the transmission of sound through the middle ear, and in many felid species it acts like a notch filter, attenuating the transmission of a narrow band of frequencies. See Figure 0.3 for a Bode plot showing the notch filter behavior of the middle ears of domestic cat and lion.

The cavity gain is viewed in this thesis as a filter that affects the structure of a vocalization on its way from the ear canal to the cochlea in a manner that depends on the mechanical properties of the tympanic membrane, cochlea, and ossicles, as well as the acoustic properties of the air spaces in-between.

The particular range of frequencies attenuated by the middle ear depends on the volumes of the cavities and dimensions of the foramen; this fact makes it possible to construct a model of the middle-ear cavity gain based on measurements of these quantities (see Chapter 3). Since the cavity volumes and foramen dimensions vary across species, one would expect the location of the notch to be species-dependent, as is indicated in Figure 0.3 for lion and domestic cat.
Cross-sectional view of the felid middle-ear airspaces. The ear canal would be attached to the right of the tympanic membrane and the oval window to the cochlea would lie somewhere on the left wall of the tympanic cavity. The ossicles have been omitted from the picture. The two pressures, $P_{TM}$ and $P_{CAV}$ represent the pressures in the ear canal adjacent to the tympanic membrane and inside the tympanic cavity respectively. $P_{CAV}$ depends not only on the airspaces, but also on the mechanical properties of the tympanic membrane, ossicles, and cochlea. These two pressures are used for computing the cavity gain, a quantity that models one step in the transmission of sound through the middle ear (see Chapter 3). [Adapted from Huang, et al., 2000]
Figure 0.3: Measured Middle-Ear Cavity Gain for Domestic Cat and Lion

Bode plots of experimentally determined cavity gains for a domestic cat and lion ear. Sound transmission for low frequencies is not substantially affected by the middle ear, but for a narrow band of frequencies in the midrange the middle ear strongly attenuates sound transmission. The center frequency and bandwidth of the notches is dependent on the structure of the middle ear (see Chapter 3), and is species-dependent. [From Huang et al., 1997b]

A question that this thesis addresses is: does the middle-ear notch play an important role in the communication between members of a given species? To see why this appears worth investigating, refer to the spectrogram of a cheetah vocalization shown in Figure 0.4. Notice the dark horizontal bands, harmonics, present in the spectrogram. The harmonics occur with a regular spacing equal to the fundamental frequency, which in this case is the frequency of the lowest harmonic. Thus, if the fundamental frequency changes, the frequencies of the higher harmonics will change by an integer multiple of that change.
Figure 0.4: Cheetah Vocalization, to Illustrate Harmonic Structure

Spectrogram of a vocalization by an adult male cheetah (see Chapters 1 and 2) with four clearly visible harmonics. The horizontal axis is time, the vertical axis is frequency, and the darkness of the image represents magnitude. Thus, the spectrogram displays the time-varying frequency content of the vocalization. The frequency of the lowest harmonic, called the fundamental frequency, is also the spacing between adjacent higher harmonics. Consequently, changing the fundamental frequency affects the frequency and spacing of all higher harmonics.

One might imagine a scenario in which two similar-sounding vocalizations, A and B, with regular harmonic structure, vary slightly in fundamental frequency such that A features a harmonic located at the frequency of the notch, but the analogous harmonic in B lies below the notch. The middle ear would then attenuate a harmonic in A but leave the harmonics in B undiminished. If the version of A with the attenuated harmonic sounds significantly different from the original version, then (considering that A and B sounded similar initially) one would expect the middle-ear filtered versions of A and B to be significantly more discernible than the unfiltered versions. In such a scenario the middle-ear notch could be exploited by the sender as a mechanism for creating recognizably different-sounding vocalizations through small variations of the fundamental frequency. As
such, the sender could use this mechanism to its advantage, turning on or off the effects of the dip to fit different expressive needs (see Chapter 4).

This thesis seeks to test the hypothesis that the notch found in the middle-ear transmission of many felid species enhances the discriminability of the vocalizations produced by the same species (see Chapters 5 and 6). To go about testing this, one must have recordings available of vocalizations made by individuals of the species of interest, as well as enough information about the species’ middle-ear structure that the filtering effects can be simulated and the results studied. The two felid species studied here, the bobcat and the cheetah, were selected because these necessary pieces of information were available.

Part I of this thesis (Chapters 1 and 2) contains descriptions and summaries of the structure of the available bobcat and cheetah vocalizations, along with information that could be useful to someone interested in the vocalizations. Some of this information will be referred to in later chapters.

Part II (Chapters 3-6) covers the construction of the middle-ear models and filter simulation, an elaboration of the communication hypothesis, tests of the hypothesis using the recordings and simulated middle ears, and a discussion of the results.
Part I: Vocalization Descriptions

Background
In a paper from the late 1970’s, six types of domestic cat vocalization were each found to result from different behavioral situations (Brown and Buchwald 1978). Quantitative measurements and sample spectrograms were included to show structural differences between the vocalization types and efforts were made to discover how vocalizations changed with the age of the cats. According to a summary of papers compiled by Peters and Wozencraft in 1989, of 13 papers written before 1989, two offered a comprehensive treatment of domestic cat vocalizations. Considerably fewer papers have been written on the vocalizations of bobcat (*Lynx rufus*) and cheetah (*Acinonyx jubatus*), the two species dealt with in this thesis: I have identified two non-comprehensive papers on bobcat vocalizations, and five non-comprehensive papers on cheetah vocalizations. One of the papers looked carefully at eurasian lynx vocalizations, but included comparisons to bobcat vocalizations (Peters, 1987). This paper included descriptive names of the vocalization types (such as mew, purr, and yowl) as well as a few quantitative measurements, spectrograms, and summaries of the behavioral and functional contexts of the vocalizations to serve as a more objective basis for the categories that were defined. In a more recent paper on the cheetah (Ruiz-Miranda et al, 1998), researchers studied the vocalizations of male cheetahs during separation and reunion experiments, including some quantitative measurements and sample spectrograms (Ruiz-Miranda et al., 1998).

This thesis uses a series of recordings made and supplied by Dr. G. Peters (Bonn, Germany) of captive bobcats, which appear to be the same vocalizations he called “mews” in his 1987 paper, as well as a collection of vocalizations of captive cheetahs called “stutters”, “eeaows”, and “chirps” supplied by Stuart Wells (National Zoo, Washington, D. C.).
The recordings of bobcat include some information about the age and sex of the cats making the vocalizations, though specific information about their behavioral context is not available. Some contextual information was available for the cheetah vocalizations.

The goals of Part I are to present vocalization spectrograms, describe their features, and determine measures of some features in the structure of bobcat and cheetah vocalizations, as well as to familiarize the reader with vocalization structure well enough that aspects of it can be referred to in Part II.
Chapter 1

Bobcat Vocalizations

1.1 Methods
The methods used in studying the bobcat vocalizations are described in Sections 1.1.1-1.1.4 below, and are essentially the same as those used to study the cheetah vocalizations. Any differences in procedure for cheetah will be noted in Chapter 2.

1.1.1 Digitizing
To allow the vocalizations to be analyzed on a computer, the tape recordings needed to be digitized. Each recorded series of vocalizations was maintained intact in a separate digitized file; the boundaries of these series were often easily identified from audible clicks made when the recorder turned on and off. Digitization was accomplished with equipment at the Speech Communication Lab at MIT’s Research Laboratory for Electronics (RLE), which is normally used for digitizing human speech. The vocalizations were passed through an antialiasing filter with a cutoff frequency of just under 8000 Hz and sampled at a rate of 16,129 samples per second. The digitized recordings were saved in a file format called Klatt .wav which is used often in the Speech Communication Lab.

1.1.2 Cataloging Scheme
The bobcat recordings were supplied by Gustav Peters from the Zoologisches Forschungsinstitut und Museum Alexander Koenig in Bonn, Germany. Five vocalization series were digitized from these recordings. These collections were named BC-I, BC-II, BC-III, BC-IV, and BC-V. “BC” denotes bobcat and Roman numeral designations were chosen arbitrarily to denote each collection. Identification of a particular vocalization is accomplished by appending an Arabic numeral to the collection name. For example, the first vocalization in the BC-I collection is called BC-I-1, and the second, fourth and fifth vocalizations in the BC-II collection can be referred to at once as BC-II-(2, 4, 5). The same
style of identification is used for the three cheetah collections covered in Chapter 2, with
the two letter prefix “CH” used instead of “BC” (CH-I, CH-II, and CH-III).

Table 1.1 summarizes the bobcat and cheetah materials examined in Part I. The table
lists the length (in seconds) of each collection, which was determined (approximately) by
measuring the number of seconds from the time the recorder was turned on to the time it
was turned off, based on the audible clicks in the recordings. The number of vocalizations
was determined by counting how many clearly audible vocalizations were present in each
collection. Information about the sex and identity of the cats responsible for vocalizations
in each collection was based on recorded comments on the bobcat tape and on communi-
cations with the source of the cheetah tape (S. Wells). According to the comments on the
bobcat tape, the same female is present in BC-I, II, IV, and V. The same male is present in
BC-IV and V, but whether the male in BC-III is the same is not known. Additionally,
Peters was not able to attribute individual sounds in BC-IV and V to either the female or
the male in those collections. See Chapter 2 for discussion of the cheetah collections.

Table 1.1 also supplies statistical summaries of the vocalization durations (which are
listed in milliseconds) including the mean, standard deviation, smallest value, and largest
value. The “silent interval,” the time elapsed the end of one vocalization to the beginning
of the next, influences the vocalization rate. The same statistical measures are supplied for
the silent interval. For a visual summary of the temporal features of the collections, indi-
cating graphically the approximate duration and temporal placement of each vocalization
on a timeline, refer to Appendix 1 for bobcat and Appendix 2 for cheetah.
<table>
<thead>
<tr>
<th>Measures</th>
<th>BC-I</th>
<th>BC-II</th>
<th>BC-III</th>
<th>BC-IV</th>
<th>BC-V</th>
<th>CH-I</th>
<th>CH-II</th>
<th>CH-III</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Length (s)</td>
<td>26</td>
<td>31</td>
<td>16</td>
<td>35</td>
<td>50</td>
<td>39</td>
<td>22</td>
<td>47</td>
</tr>
<tr>
<td>2. # of Vocs.</td>
<td>13</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>17</td>
<td>12</td>
<td>30</td>
</tr>
<tr>
<td>3. All Vocs Within (s)</td>
<td>15</td>
<td>17</td>
<td>13</td>
<td>33</td>
<td>42</td>
<td>34</td>
<td>18</td>
<td>47</td>
</tr>
<tr>
<td>4. # seconds / Vocalization</td>
<td>1.2</td>
<td>1.9</td>
<td>1.6</td>
<td>4.7</td>
<td>6.0</td>
<td>2.0</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td>Cats Present (#-sex)</td>
<td>1-f</td>
<td>1-f</td>
<td>1-m</td>
<td>1-m, 1-f</td>
<td>1-m, 1-f</td>
<td>1-m</td>
<td>1-m (or more)</td>
<td>1-m</td>
</tr>
<tr>
<td>Mean</td>
<td>453</td>
<td>530</td>
<td>766</td>
<td>2760</td>
<td>3142</td>
<td>569</td>
<td>405</td>
<td>494</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>90</td>
<td>149</td>
<td>407</td>
<td>2055</td>
<td>2242</td>
<td>348</td>
<td>135</td>
<td>81</td>
</tr>
<tr>
<td>Min.</td>
<td>246</td>
<td>333</td>
<td>359</td>
<td>441</td>
<td>471</td>
<td>45</td>
<td>240</td>
<td>311</td>
</tr>
<tr>
<td>Max.</td>
<td>560</td>
<td>791</td>
<td>1752</td>
<td>5212</td>
<td>6206</td>
<td>1300</td>
<td>711</td>
<td>702</td>
</tr>
</tbody>
</table>

| Duration (ms) | | | | | | | | |
| 1-f | 1-f | 1-m | 1-m, 1-f | 1-m, 1-f | 1-m | 1-m (or more) | 1-m |
| Mean | 706 | 1482 | 838 | 2088 | 3271 | 1521 | 1109 | 1050 |
| Std. Dev. | 285 | 1002 | 193 | 2068 | 2342 | 1709 | 487 | 493 |
| Min. | 494 | 631 | 527 | 223 | 909 | 58 | 388 | 362 |
| Max. | 1519 | 3611 | 1143 | 6494 | 6342 | 5784 | 1857 | 2047 |

**Table 1.1: Summary of Digitized Vocalization Collections**

Each of the five bobcat collections (BC-I, II, III, IV, and V) and three cheetah collections (CH-I, II, and III) is summarized in terms of: the overall collection length (in seconds), the number of vocalizations in the collection, the time range within which all vocalizations fall (row 3), the average number of seconds for each vocalization (row 4 = row 1 / row 2), the sex of each cat vocalizing in the collection, descriptive statistics concerning the vocalization durations within the collection (all measured in milliseconds), and the same statistical measures on the lengths of the silent intervals found between adjacent vocalizations on the original recordings.
1.1.3 Analysis Software
Analyses of vocalization structure were performed with a software package developed for human speech at the RLE Speech Communication Lab, called “xkl”. This software displays the waveform, the Fourier transform of a windowed section of the waveform, and spectrograms, and includes audio playback capabilities. It provides a convenient way to navigate through a collection of vocalizations, to listen to any subportion of the collection, to display the frequency behavior over time in the spectrogram, and to focus more precisely on the frequency content of a particular time window in the Fourier transform display.

1.1.4 Methods of Describing Vocalization Structure
The first step taken in studying the structure of vocalizations was to listen to them a few times to get a feel for their durations, pitch, placement relative to one another, timbre, etc. Next I listened to them while looking at spectrograms of each vocalization to see what types of connection could be made between auditory perception and the spectrogram’s structure. The analysis methods I ended up using, described below, are an effort to summarize a variety of structural features in a way that would make it possible to compare the features for different vocalizations, collections, and species. Appendices 1 and 2 contain a compilation of spectrograms for all vocalizations studied. The remainder of this section contains explanations of the six kinds of measurement and observation that were made on the vocalizations (usually by examining the spectrograms). Summaries of the actual observations and measurements can be found in Section 1.2 for bobcat (2.2 for cheetah), Table 1.2 for bobcat (2.1 for cheetah), as well as Appendix 1 for bobcat (Appendix 2 for cheetah). In the catalog of spectrograms found in Appendices 1 and 2, the margins contain miscellaneous observations and measurements made on individual vocalizations. Figures 1.1 and 1.2 below are intended to indicate the meaning of the measurements and observations with respect to the spectrograms.
Measurements and Observations made on Bobcat and Cheetah Vocalizations:

1. **Length**: Number of seconds in a collection.

2. **Duration**: For each digitized collection, the starting time and ending time of each vocalization was measured (in milliseconds). In cases where the beginning or ending appeared faint in the spectrogram, or unclear in the time waveform, it was estimated. Duration, then, is the difference between the ending time and beginning time of each vocalization. See Figure 1.1 for a visual indication of vocalization duration.

3. **Frequency Range**: This was judged by looking at the spectrograms of the collection and roughly noting the lowest and highest frequency between which the spectral content of the vocalizations was visible. See Figure 1.1 for a visual indication of the frequency range of a single vocalization.

4. **Pitch Contour**: This was a harder concept to define precisely. It refers to the way the frequency of the lowest visible harmonic varies with time. The difficulty came when the lowest harmonic failed to form a continuous line but instead became invisible (or indistinguishable from noise) over some portion of its length. Because of such cases it was not always possible to describe the pitch contour by looking only at the lowest harmonic, but often its overall trend could be determined by observing the path taken by the higher harmonics. Instead of making a large number of measurements to graphically summarize the pitch contours, a system was devised for notating their general behavior. A “U” is used if a contour moves upward during a time interval, a “C” is used if it flattens out and becomes constant for some interval, and a “D” is used if it moves downward. With these symbols, pitch contours were reduced to a sequence of the symbols “U”, “C”, and “D”. See Figure 1.1 for a visual indication of a pitch contour. While the “U”, “C”, “D” notation is neither precise nor quantitative, it is a simple summary of pitch contour.
Because many pitch contours reached a clear high point, or “peak” at some time along their length, an effort was made to measure the time and frequency of these highpoints using the following notation:

\[ H1 \text{ [or F0]} \at \text{NNN ms} = \text{MMM Hz [peak]} \]

Depending on the harmonic structure of a given vocalization, the lowest visible harmonic may or may not appear to be the fundamental frequency. “H1” or “F0” is used accordingly. “@ NNN ms” indicates that the measurement was made at the time point of NNN ms in the digitized collection, and “= MMM Hz” means that the value of the measurement is MMM Hz. When the word “peak” appears at the end, the measurement constitutes the highpoint of the pitch contour. In some cases additional measurements were made at relatively flat segments of the pitch contour other than the peak. In such cases the same notation as above is used, omitting the word “peak”. Measurements of this kind can be found in the margins of the spectrogram catalog of Appendices 1 and 2. See Figure 1.1 for a visual indication of a peak in a pitch contour.

Often an attempt is made to determine where, relative to the center of a vocalization, the peak of the pitch contour appears to occur. Specifically, I attempt to point out cases when the peak appears to occur earlier, later, or near the center of a vocalization.

5. **Frequency Content**: This category describes the amplitudes of the frequency components within a vocalization, as judged by the darkness of the spectrogram images. The amplitude-to-darkness mapping for the spectrograms within a given collection is the same, so regions of equal darkness should correspond to equal sound pressure level. When comparing vocalizations from different collections, however, equal darkness probably does not correspond to equal sound pressure level, because recording conditions and level settings are likely to be different for each collection. Additionally, the spectrograms only make visible spectral content above a particular amplitude threshold. Consequently, some weak
vocalization features may not be visible in the spectrograms, but they may still be visible in conventional Fourier transforms of windowed portions of the vocalization. Most observations concerning the frequency content of vocalizations were made by looking at their spectrograms, but on some occasions I referred to a Fourier transform to clarify features that were difficult to see in the spectrogram.

In the spectrograms, it is often possible to see dark horizontal bands, or harmonics, within the vocalizations. Sometimes these harmonics are continuous and smoothly-varying in darkness, suggesting that their amplitudes do not change quickly over their length. In other cases they feature sudden darkness changes, suggesting that their amplitudes vary abruptly.

Many vocalizations feature regions that lack clear structure and consist of various shades of gray. When the grayness appears only within the vocalization (as opposed to in the background surrounding the vocalization), I refer to it as “vocalized noise”. See Figure 1.1 for an example. Sometimes vocalized noise has a high-enough amplitude to obscure the harmonic structure of a vocalization (see the BC-I collection especially), but other times it does not seriously hinder the visibility of harmonics (see BC-II for example). For vocalizations that appear to contain a large amount of vocalized noise I often comment to that effect in the margin notes of Appendices 1 and 2. Another type of noise present in the spectrograms is background noise. It can be distinguished from vocalized noise in that it is present in portions of the spectrogram where there are no vocalizations. Sometimes the background noise looks like a band of random gray spectral content spread over a prolonged length of time and sounds as though it is caused by wind, traffic, or the roar of airplanes for example. This type of background noise is present in all of the collections to varying degrees, though especially obvious examples of it can be found in the BC-III collection. When the spectrogram image of background noise appears to have a non-random
structure, that may indicate that it was produced by events such as birds singing, objects hitting the microphone, or pieces of metal clanging. An example of this type of background noise is indicated on the right-hand side of Figure 1.1.

6. **Additional Features:** I use this observation category to point out structural features of vocalizations that seem worthy of special mention. For instance, I might point out cases in which a vocalization features regular pulsed amplitude modulation and mention the approximate rate, or where some vocalizations sound qualitatively different from the others in the collection. See Figure 1.2 for a visual indication of pulsed amplitude modulation in a bobcat vocalization (for more elaborate examples in cheetah vocalizations, see the CH-I and II collections).
Figure 1.1: Examples of Observation and Measurement Categories

Spectrogram of a bobcat vocalization (BC-I-8) with visual pointers to examples of various observation and measurement categories. The vocalization duration is indicated by a double-sided arrow extending from the beginning to the end of the vocalization. The frequency range is indicated by a double-sided arrow extending from the lowest to the highest frequency containing the vocalization's image. The pitch contour is indicated by dividing the time scale into regions for which the lowest harmonic is increasing in frequency with time ("U" for upward), remaining roughly constant ("C"), and decreasing with time ("D"). The highpoint of the pitch contour is indicated by the "H1 peak" label. Two examples of background noise are labeled and encircled. The rightmost example sounds like something is hitting the microphone. Finally, an example of vocalized noise is encircled in white near the top of the vocalization. The vocalized noise may be present over a wide frequency range, but it is easier to see by itself in the higher frequency range where the harmonics (dark horizontal bands) are weaker.
Figure 1.2: Example of Pulsed Amplitude Modulation

Spectrogram of a bobcat vocalization (BC-I-12) with a line surrounding the region containing pulsed amplitude modulation. Approximately 9 pulses occur within a 100 ms interval, for a rate of around 90 pulses per second.

1.2 Results

The comments contained in the margins of Appendices 1 and 2 are not complete descriptions of the vocalizations; they are intended to guide one’s examination of the spectrograms.

At the beginning of each collection in the Appendices is a summary of the observations and measurements made on the vocalizations in that collection (see Section 1.1.4), as well as a graphical rendition of the collection as a timeline with symbols representing vocalizations placed and scaled (within 100 milliseconds) in their appropriate time intervals within the recording.
1.2.1 BC-I Collection
(see pp 159-164 of Appendix 1)

The 13 vocalizations found in this 26 second collection were produced by an adult female bobcat (the same female represented in BC-II, IV, and V). All 13 vocalizations occur within a 15 second interval (see timeline on page 160 of Appendix 1), with the spacing between vocalizations ranging from 494 to 1449 ms (mean: 706, standard deviation: 285). Vocalization duration ranges from 246 to 560 ms (mean: 453, standard deviation: 90). Spectrograms of four of the vocalizations in this collection (BC-I-(9-12)) can be found in Figure 2 (p. 321) of Peters, 1987, where the caption referring to vocalizations 9 and 10 reads “medium intensity mews by a female”, and for 11 and 12 reads “mews of low and medium intensity by a female. The calls show rhythmical amplitude modulation”. In voice comments on the tape, Peters describes the vocalizations in this collection as being recorded at a high level with heavy background noise.

Frequency Content

The spectrograms of this collection (see pp 161-164 of Appendix 1) display frequency content ranging from 0.6 to 7 kHz. Harmonic structure is evident in the presence of dark horizontal bands in the spectrograms, but it is usually obscured by the large amount of vocalized noise present in the vocalizations of this collection. The lowest visible harmonic typically occurs around 1 kHz, and it is often possible to find higher harmonics whose frequencies appear to occur at (nearly) integer multiples of this lowest harmonic frequency. However, apparent harmonic content is usually visible between these widely spaced harmonics for at least part of the vocalization, suggesting that the lowest visible harmonic is actually not the fundamental frequency (F0), but probably two times F0. For this reason I refer to the lowest harmonic as H1 rather than F0 when I make measurements on it. Many of the harmonics, when they are visible, vary irregularly and abruptly in amplitude.
A large amount of “vocalized” noise is present in vocalizations of this collection, evidenced by the dark gray content filling the frequency range and the spaces between harmonics. I attribute much of my difficulty in clearly seeing the harmonic structure to the presence of this noise.

Vocalizations BC-I-(1, 2, and 3) are fairly weak, based on the faintness of their spectrogram images relative to the others in this collection (they sound weaker as well). Vocalizations BC-I-(4-9) are stronger and appear similar to one another: they have a large frequency range, contain “vocalized” noise, and their harmonic amplitudes are irregular, judging by the splotchiness of their spectrogram images. Vocalizations BC-I-(10-12) have a narrower frequency range, with widely spaced harmonics at the beginning, and in each spectrogram it appears that two narrow-band components just above H1 begin 50 and 150 ms after the beginning. The darkest content in this collection generally appears between 1 and 3 kHz, with the lowest harmonic sometimes, but not always the most prominent.

**Pitch Contour**

The lowest harmonic in all of these vocalizations rises in frequency at the beginning and falls in frequency at the end. In between it typically rises to peak relatively close to the beginning, flattens momentarily before falling to a lower flat point, or flattens for a good portion of the vocalization before falling at the end. In terms of the pitch contour notation, these vocalizations are called either “UCD” or “UCDCD”. The fact that the peak often occurs near the beginning of the vocalizations makes the pitch contours appear skewed toward the beginning of the vocalizations. The audible result of this is that the pitch of the vocalizations rises abruptly near the beginning then decreases a little more gradually afterward. The H1 peak frequency measurements ranged from 1040 to 1449 Hz, (mean: 1294, standard deviation: 105).
**Additional Features**

Vocalizations BC-I-(10, 11, and 12) feature pulsed amplitude modulation near their beginnings, indicated by the periodic amplitude variation visible in the spectrograms. Only around 2 pulses are clearly visible in BC-I-10, around 6 in BC-I-11, and around 9 in BC-I-12. The pulse rate is roughly 90 per second in all three cases.

**1.2.2 BC-II Collection**
(see pp 165-169 of Appendix 1)

The 9 vocalizations found in this 31 second collection were produced by an adult female bobcat (the same female represented in BC-I, IV, and V). All 9 vocalizations occur within one 17 second interval (see timeline on p. 166 of Appendix 1), with the spacing between vocalizations ranging from 631 to 3611 ms (mean: 1482, standard deviation: 1002). Vocalization duration ranges from 333 to 791 ms (mean: 130, standard deviation: 149). On the tape Peters mentions that the vocalizations in this collection were recorded at a high level with heavy background noise. No spectrograms from this collection are presented in Peters, 1987.

**Frequency Content**

The spectrograms in this collection (see pp 167-169 of Appendix 1) display frequency content ranging from 0.6 to 7 kHz (as in BC-I). Harmonic structure is evident, perhaps more clearly than in BC-I. As in BC-I the lowest visible harmonic appears to be twice the fundamental frequency. Some harmonics form smooth continuous lines and others appear to cut in and out abruptly. There appears to be vocalized noise present, though perhaps less overall than BC-I. The darkest spectrogram content typically occurs between 1.2 and 3 kHz, and the lowest harmonic is never the strongest.
Pitch Contour

As in BC-I, the lowest harmonic rises at the beginning and falls at the end, leading to several vocalizations with “UCD” contours. Others follow a more complicated path in between, such as “UCDCUCD” (BC-II-1), or “UCUCDCD” (BC-II-7). The contours also differ from BC-I in that many of the peaks occur near the center (BC-II-(4, 5, 8, and 9)) or to the right of center (BC-II-(2 and 3)) of the vocalizations. The H1 peak measurements ranged from 945 to 1166 Hz (mean: 1078 Hz (close to a minor third below BC-I’s mean), standard deviation: 78 Hz).

Additional Features

Pulsed amplitude modulation occurs at the beginning of vocalization BC-II-7. Around 6 pulses are visible, with a rate of approximately 85 pulses per second. Vocalization BC-II-9 appears to have around 17 closely spaced pulses in its middle section, for a rate of around 140 pulses per second.

Comparisons to BC-I

The same adult female bobcat was responsible for the vocalizations in BC-I and BC-II. Compared to BC-I, BC-II has the same frequency range, a slightly longer average duration (77 ms longer on average), and a considerably longer average silent interval (776 ms longer).

The harmonic amplitudes in both collections tend to vary irregularly, though in BC-II the harmonics are generally easier to see. Both collections feature a fair amount of vocalized noise.

The pitch contours in BC-II, like those in BC-I, follow a “UCD” pattern or a variant with more complicated pitch changes in the middle. Unlike BC-I, the H1 peaks in BC-II tend to occur closer to the center or end of the vocalization (as opposed to occurring closer
to the beginning). The average peak frequency of H1 for BC-II is 216 Hz lower than that of BC-I (comparable to the musical interval of a minor third). Both of these differences are audible when listening to these two collections.

Vocalization BC-II-7 features pulsed amplitude modulation, at a rate comparable to that of vocalizations BC-I-(10, 11, and 12).

See Table 1.2 for comparisons among all 5 bobcat collections.

1.2.3 BC-III Collection
(see pp 170-174 of Appendix 1)

The 8 vocalizations found in this 16 second collection were produced by an adult male bobcat. All 8 vocalizations occur within one 13 second interval (see timeline on p. 171 of Appendix 1), with the spacing between vocalizations ranging from 527 to 1143 ms (mean: 838, standard deviation: 193). Vocalization duration ranges from 359 to 1752 ms (mean: 766, standard deviation: 407). BC-III-3 is too weak to analyze, so only 7 vocalizations were studied. In Table 1 of Peters, 1987 (p. 321), Peters refers to a collection of 7 male bobcat vocalizations, which probably is the same as these 7. The spectrograms of BC-III-(1 and 2) appear in Figure 2 on p. 321 of Peters, 1987. The caption reads: “Mews of low intensity by a male”. Peters states on the tape that it was recorded at a high level with very heavy background noise.

Frequency Content

The spectrograms of this collection (see pp 172-174 of Appendix 1) display frequency content ranging from 0.5 to 4 kHz (lower on both ends than BC-I and II), with occasional content reaching as high as 7 kHz. Harmonic structure is clearly evident, although the harmonics are difficult to see at times due to their weak overall amplitudes. Harmonics, when strong enough to be visible, appear smooth and continuous. “Vocalized” noise appears to be minimal, though this is difficult to tell with the very heavy background noise.
added to the vocalizations. Harmonic spacing is at integer multiples of the lowest frequency, so the lowest harmonic occurs at the fundamental frequency (F0). Often the spectrogram image is darkest near the end of the vocalization, e.g. BC-III-(2, 3, 5, 6, 7, and 8), at which point as many as seven harmonics may become visible. The darkest band of the vocalizations usually occurs between 1 and 3 kHz.

**Pitch Contour**

As in BC-I and II the contour rises at the beginning and falls at the end. Most vocalizations, then, are “UCD”, though BC-III-(1 and 4) are more complicated (“UCDCD” and “UCDCDCD” respectively). As the harmonics were generally weak and hard to see in the vocalizations, I could not always find definitively where the peaks occur. Since most vocalizations were relatively flat in frequency versus time over their highest range, the measurements I included with the spectrograms in Appendix 1 still give a good feel for the upper limit of the pitch contour. From what I can tell the peaks occur closer to the beginning in BC-III-(1, 4, and 5), and near the center in BC-III-(2, 3, 6, 7, and 8). F0 measurements ranged from 819 to 914 Hz (mean: 869 Hz (a little less than a perfect fifth below the BC-I mean), standard deviation: 29 Hz).

**Additional Features**

Vocalization 4 is significantly longer and more intense than the others in this collection.

**Comparisons to BC-I and BC-II**

A single adult male bobcat was responsible for all of the vocalizations in BC-III, whereas the vocalizations in BC-I and II were made by an adult female bobcat. The frequency range of BC-III is slightly lower at the lower limit and usually around 3 kHz lower at the upper limit than BC-I and II. Average vocalization duration is considerably longer.
than BC-I and II (300 and 236 ms longer respectively), and the mean silent interval falls between that of BC-I and II (154 ms longer than BC-I and 644 ms shorter than BC-II).

The harmonic amplitudes in BC-III are smoother than those in BC-I and II, but the harmonics are often too weak to see in the spectrograms relative to the background noise. The background noise in this collection is substantially heavier than BC-I and II. There is probably less vocalized noise in BC-III than BC-I and II based on the smoother sound of its vocalizations and the smoother-looking harmonics in the spectrograms.

Unlike BC-I and II, the lowest harmonic in the BC-III vocalizations occurs at the fundamental frequency. Pitch contours follow the same sort of “UCD” pattern in all three collections, though the average peak F0 frequency in BC-III was lower than both BC-I and II (by 425 Hz (just under a perfect fifth) and 209 Hz (just under a major third) respectively).

None of the vocalizations in BC-III featured pulsed amplitude modulation. See Table 1.2 for further comparisons.

1.2.4 BC-IV Collection
(see pp 175-181 of Appendix 1)

The seven vocalizations found in this 35 second collection were produced by both an adult male bobcat (who is also in BC-V but may or may not be the same male found in BC-III) and an adult female bobcat (the same as found in BC-I, II, and V). All vocalizations occur within a 33 second interval (see timeline on page 176 of Appendix 1), with spacing between vocalizations ranging from 223 to 6494 ms (mean: 2088, standard deviation: 2068). Vocalization durations vary from 441 to 5212 ms (mean: 2760, standard deviation: 2055). As the recording was made without watching the cats, it was not possible for Peters to attribute individual sounds to individual bobcats. Several different types of vocalization are present in this collection, including growls, hisses, and prolonged calls with varied, complex structures. Peters’ comments on the tape recording refer to this and
the following collection as “diverse sounds”, and the prolonged, mew-like sounds are referred to as having a “growly superposition of varying degrees; for some it is relatively strong.” The background noise is mild.

**Frequency Content**

The spectrograms of this collection (see pp 177-181 of Appendix 1) display frequency content ranging from 0.1 to over 7 kHz. Harmonic structure is evident in BC-IV-(1, 2, 3, 4, and 7). The harmonic amplitudes are often pulse-modulated over a significant portion of the vocalization duration (e.g. BC-IV-4c and BC-IV-7c). Vocalizations on the whole appear to contain a good amount of “vocalized” noise.

**Pitch Contour**

No attempt was made to completely summarize the pitch contours of these prolonged vocalizations with the “U”, “C”, and “D” symbols. The contours appear on the whole to be flat during the prolonged episodes (e.g. BC-IV-(3, 4, 7b-7d)), and sometimes near the beginning feature a “UCD” type behavior (e.g. BC-IV-7a). Some F0 measurements were made over the course of the vocalizations (see spectrograms in Appendix 1) but they were not necessarily of peaks in the pitch contour.

**Additional Features**

Vocalizations 1, 2, 3, 4, and 7 are heterogeneous mew-like sounds with growly parts. Examples of growly parts are BC-IV-(4b-4d) and BC-IV-(7c-7d). Vocalization 5 sounds like breathing, and 6 sounds like a short growl followed by a “huff”.

**Comparisons with BC-I, II, and III**

The vocalizations in BC-IV were produced by two bobcats rather than a single individual as in BC-I, II, and III. The range of vocalization duration in BC-IV (441 to 5212 ms) is far greater than that of the other collections, and both the mean and standard deviation are
high (mean: 2760 ms, std. dev: 2055 ms) due to the fact that BC-IV contains three vocalizations that exceed four seconds in length and four vocalizations that are under two seconds in length. The prolonged vocalizations in BC-IV feature harmonic structure, a growly sound during parts of their duration, the appearance of vocalized noise, and generally a complex, time-varying structure. These vocalizations differ substantially in structure from those of BC-I, II, and III, as is apparent by comparing the spectrograms in Appendix 1. BC-IV-(5 and 6) are short noisy sounds that differ from any of the other vocalizations shown so far.

See Table 1.2 for further comparisons among bobcat collections.

1.2.5 BC-V Collection
(see pp 182-188 of Appendix 1)

The seven vocalizations found in this 52 second collection were made by the same adult male as in BC-IV, and the same adult female as in BC-I, II, and IV. All vocalizations occurred within a 42 second interval (see timeline on page 183 of Appendix 1), with spacing between vocalizations ranging from 909 to 6042 ms (mean: 3271, standard deviation: 2342). Vocalization durations varied from 471 to 6206 ms (mean: 3142, standard deviation: 2242). As in BC-IV, Peters was not able to attribute individual sounds to individual cats. Diverse vocalizations are present, including snorts, growls, and mew-like sounds. Background noise is mild.

Frequency Content

The spectrograms of this collection (pp 184-188 of Appendix 1) display frequency content ranging from 0.1 to around 7 kHz, though most content lies below 4 kHz. Harmonics are evident in BC-V-(1, 2, 3, and 6). As in BC-IV, harmonic amplitudes often feature pulsed amplitude modulation over some of their length (e.g. BC-V-2d and BC-V-3c).
Frequency content on the whole appears to have a good amount of “vocalized” noise added in.

Pitch Contour

The harmonic frequencies for the mew-like sounds of this collection, BC-V-(1, 2, 3, 6), appear flat for most of their length, with perhaps a small amount of curvature at times. A small number of F0 measurements were made to give an idea of the frequency of the flat portions (See the right-hand margins of the spectrograms displayed in Appendix 1).

Additional Features

BC-V-(1, 2, 3, and 6) have harmonic structure and contain growly parts, 4 is a “snort-like” sound, 5 is a “hissy-growl-like” sound, and 7 is a growl.

Comparisons with BC-I, II, III, and IV

As in BC-IV there are two bobcats responsible for the vocalizations and the collection contains a number of vocalizations that last for several seconds. It also contains three vocalizations that are substantially shorter. As a result, the range of durations (471 to 6206 ms) is large and both the mean and standard deviation are also high (mean: 3142, std. deviation: 2242 ms). The prolonged vocalizations in BC-V resemble the prolonged vocalizations in BC-IV, and the short sounds in BC-V are noisy and have unusual structures like the short sounds in BC-IV.

See Table 1.2 for additional comparisons.

1.3 Summary of Features

Table 1.2 below summarizes 15 features described in the Results section above, and facilitates making comparisons among the five collections. The first column lists abbreviated names of the various features that were measured and described for the vocalizations, and the remaining 5 columns contain the appropriate entries for each of the 5 bobcat collec-
The first 6 features are reproduced from Table 1.1. **Frequency Range** (feature 7) indicates approximately the lowest and highest frequency reached by the vocalizations in a collection. The number in parenthesis for BC-III indicates the highpoint that best fits most of the vocalizations within the collection. **Darkest Range** (feature 8) indicates approximately the range of frequencies within which the highest harmonic amplitudes occur. **Harmonic Content** (feature 9) indicates whether the vocalizations within a collection contain a significant amount of vocalized noise (“noisy”) or not (“smooth”), or whether there is substantial variation in the amount of vocalized noise present (“varied”). **Fundamental Present?** (feature 10) indicates whether the lowest harmonic is the fundamental frequency (“yes”) or whether the lowest harmonic does not seem to occur at the fundamental frequency (“no”). **Peak F0/H1** (feature 11) indicates statistics for the measurements of the peak frequency of the lowest harmonic, based on data available in Appendix 1. No peak measurements were made on the vocalizations in BC-IV and V. **Pitch Contours** (feature 12) summarizes the common pitch contours of the vocalizations in a collection. Most entries are written in “U”, “C”, “D” notation (see Section 1.1.4) and are followed by a number in parenthesis indicating the number of vocalizations with the preceding contour. **Mew-Like / Growly / Other** (features 13-15) indicate how many vocalizations are similar to “mews” as described by Peters, how many contain a substantial amount of growliness (as heard upon listening), and how many do not fit either of these descriptions.
<table>
<thead>
<tr>
<th>Feature</th>
<th>BC-I</th>
<th>BC-II</th>
<th>BC-III</th>
<th>BC-IV</th>
<th>BC-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. # Vocalizations</td>
<td>13</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>2. Length (s)</td>
<td>26</td>
<td>31</td>
<td>16</td>
<td>35</td>
<td>50</td>
</tr>
<tr>
<td>3a. All Vocs w/in (s)</td>
<td>15</td>
<td>17</td>
<td>13</td>
<td>33</td>
<td>42</td>
</tr>
<tr>
<td>3b. # seconds per Voc</td>
<td>1.2</td>
<td>1.9</td>
<td>1.6</td>
<td>4.7</td>
<td>6.0</td>
</tr>
<tr>
<td>4. Cats Present</td>
<td>1-f</td>
<td>1-f</td>
<td>1-m</td>
<td>1-f, 1-m</td>
<td>1-f, 1-m</td>
</tr>
<tr>
<td>5. Duration: Mean</td>
<td>453</td>
<td>530</td>
<td>766</td>
<td>2760</td>
<td>3142</td>
</tr>
<tr>
<td>Standard Dev</td>
<td>90</td>
<td>149</td>
<td>407</td>
<td>2055</td>
<td>2242</td>
</tr>
<tr>
<td>Minimum</td>
<td>246</td>
<td>333</td>
<td>359</td>
<td>441</td>
<td>471</td>
</tr>
<tr>
<td>Maximum</td>
<td>560</td>
<td>791</td>
<td>1752</td>
<td>5212</td>
<td>6206</td>
</tr>
<tr>
<td>6. Silent Int.: Mean</td>
<td>706</td>
<td>1482</td>
<td>838</td>
<td>2088</td>
<td>3271</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>285</td>
<td>1002</td>
<td>193</td>
<td>2068</td>
<td>2342</td>
</tr>
<tr>
<td>Minimum</td>
<td>494</td>
<td>631</td>
<td>527</td>
<td>223</td>
<td>909</td>
</tr>
<tr>
<td>Maximum</td>
<td>1519</td>
<td>3611</td>
<td>1143</td>
<td>6494</td>
<td>6342</td>
</tr>
<tr>
<td>7. Freq. Range (kHz)</td>
<td>0.6-7</td>
<td>0.6-7</td>
<td>0.5-7 (4)</td>
<td>0.1-7</td>
<td>0.1-7</td>
</tr>
<tr>
<td>8. Darkest Rng. (kHz)</td>
<td>1-3</td>
<td>1.2-3</td>
<td>1-3</td>
<td>&lt;4</td>
<td>&lt;4</td>
</tr>
<tr>
<td>9. Harmonic Content</td>
<td>noisy</td>
<td>noisy</td>
<td>smooth</td>
<td>varied</td>
<td>varied</td>
</tr>
<tr>
<td>10. Fund. Present?</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>11. Peak F0/H1: Mean</td>
<td>1294</td>
<td>1078</td>
<td>869</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>105</td>
<td>78</td>
<td>29</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Minimum</td>
<td>1040</td>
<td>945</td>
<td>819</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Maximum</td>
<td>1449</td>
<td>1166</td>
<td>914</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>12. Pitch Contours</td>
<td>UCD(9)</td>
<td>UCD(6)</td>
<td>UCD(5)</td>
<td>~flat</td>
<td>~flat</td>
</tr>
<tr>
<td></td>
<td>UCDCD(4)</td>
<td>other(3)</td>
<td>other(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Mew-Like</td>
<td>13</td>
<td>9</td>
<td>8</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>14. Growly</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>15. Other</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1.2: Summary of Bobcat Vocalization Features
1.4 Discussion and Conclusions
The 5 bobcat collections were produced by at least 2 different bobcats (one male and one female), and they contain a reasonable variety of sounds. BC-I and II were produced by the same adult female, and appear and sound like they have much in common. The vocalized noise, irregular harmonic amplitudes, missing fundamental harmonic, identical frequency ranges, and UCD (or some variant) pitch contours all make them similar to each other. There are a few noticeable differences, however. The pitch contours in BC-I tend to peak near the beginning of the vocalization (often with a large positive slope at the beginning and a smaller negative slope at the end), whereas those of BC-II tend to peak closer to the center or end of the vocalization (often with a positive slope at the beginning that is equal or less than the magnitude of the negative slope at the end), a difference which is audible. The H1 peak frequency is higher for BC-I than BC-II, the rate of vocalization is higher in BC-I, and the vocalized noise in BC-II does not appear to be as dense as in BC-I, all three of which are audible differences. Qualitatively, BC-II does not sound as harsh as BC-I.

BC-III was produced by a male bobcat. The vocalizations are lower in frequency than BC-I and II, are weaker (within very heavy background noise), and appear to lack any significant amount of vocalized noise. The pitch contours do not tend to exhibit large slopes, and generally vary gradually in frequency. Qualitatively, BC-III sounds substantially calmer than either BC-I or II.

BC-IV and V were both produced by a male and a female bobcat. Both of these collections differ substantially from BC-I, II, and III in that the (tonal) vocalizations are much longer in duration, and far more varied in structure. The short noisy vocalizations present in these 2 collections are also different from anything found in the first 3 collections. Growliness is prevalent in the prolonged vocalizations.
**Statistical Tests**

A two-sided t-test was used to compare the means of the measurements of Duration, Silent Interval, and H1/F0 Peak measurements in the 5 collections. The test produces a significance level, called “alpha”, that can have a value between 0 and 1 and indicates a higher significance level when it is smaller. For many applications the result of alpha being less than 0.05 is deemed “statistically significant”, which is the rule of thumb that will be used here. The results of the t-tests are found in Table 1.3.

In terms of duration, vocalizations in the first 3 collections were significantly shorter than those in the last 2 collections. The mean increased for each collection beginning with BC-I and ending with BC-V. Using the t-test to compare the populations of duration measurements for the vocalizations from BC-I, II, and III, the alpha levels were all greater than 0.082, but comparing BC-I, II, and III with BC-IV and V the alpha levels were all less than or equal to 0.055 (see “Duration” in Table 1.3). The t-test, then, basically demonstrates that the durations of the first 3 collections differ significantly from the durations of the last 2 collections.

Comparison of the silent intervals between sections of the various collections shows that none of the alpha levels is below 0.05 (see “Interval” in Table 1.3). However, comparing BC-I to BC-III gave an alpha of 0.082 and comparing BC-I to BC-V gave an alpha of 0.058, two scores that are not far from being considered significant.

The peak H1/F0 measurements that were made on the BC-I, II, and III collections (see notes in the margins of Appendix 1) were also compared using a two-sided t-test. In this case, however, the alpha levels were all below $4.0 \times 10^{-5}$ (see “H1 Max” in Table 1.3), suggesting that these 3 collections feature significantly different high-points in their pitch contours. This difference is easily audible when listening to one collection after another:
perceptually, BC-I has the highest overall pitch, BC-II is somewhat lower, and BC-III lower still.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Duration</th>
<th>Interval</th>
<th>Max H1</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC-I and II</td>
<td>0.21</td>
<td>0.082</td>
<td>4.0x10^{-5}*</td>
</tr>
<tr>
<td>BC-II and III</td>
<td>0.18</td>
<td>0.14</td>
<td>2.8x10^{-5}*</td>
</tr>
<tr>
<td>BC-I and III</td>
<td>0.082</td>
<td>0.27</td>
<td>1.2x10^{-9}*</td>
</tr>
<tr>
<td>BC-I and IV</td>
<td>0.03*</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>BC-I and V</td>
<td>0.026*</td>
<td>0.058</td>
<td></td>
</tr>
<tr>
<td>BC-II and IV</td>
<td>0.038*</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>BC-II and V</td>
<td>0.029*</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>BC-III and IV</td>
<td>0.055</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>BC-III and V</td>
<td>0.041*</td>
<td>0.068</td>
<td></td>
</tr>
<tr>
<td>BC-IV and V</td>
<td>0.76</td>
<td>0.42</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.3: "Alpha" Significance Levels based on a 2-sided T-test

Significance levels for two-sided T-tests between measurement populations from the 5 bobcat collections. Duration and silent interval measurements were compared for all 5 collections, but the maximum H1 frequency was compared for the first 3 collections only (maximum H1 measurements were not made for the vocalizations in the BC-IV and V collections). The numerical result of each test is referred to as “alpha”, which can have a value between 0 and 1 and for which smaller values are considered more significant. Here an alpha level that is less than ~0.05 is considered “statistically significant”, and is indicated with an asterisk.

Given the small number of individual bobcats represented in these 5 recordings (probably two), and the small number of different recordings, it is not possible to know in what way the differences between the vocalizations in the 5 collections are related to sex, the particular individual vocalizing, or the behavioral situation the cat (or cats) were experiencing.
Chapter 2

Cheetah Vocalizations

2.1 Methods
The methods used in studying the cheetah vocalizations are essentially those given for bobcat in Chapter 1. Digitizing (Section 1.1.1), Analysis and Software (1.1.3), and the Methods for Describing Vocalizations (1.1.4) sections are all equally applicable to the cheetah vocalizations. With regard to the cataloging scheme, as was mentioned in Section 1.1.2, three collections were digitized for cheetah, referred to as CH-I, CH-II, and CH-III. “CH” denotes cheetah and the Roman numeral designates a particular collection. Individual vocalizations, again, are indicated by appending an Arabic numeral to the end of the collection name (see Section 1.1.2). Refer to Table 1.1 for information about the three cheetah collections (length, number of vocalizations, cats present, vocalization duration statistics, and silent interval statistics as compared to the bobcat collections). All 3 collections contain vocalizations by the same male cheetah (known as “Norok”). The vocalizations were supplied by Stuart Wells, Assistant Curator of the National Zoo in Washington, D.C., a co-author of the Ruiz-Miranda et al., 1998, paper on cheetahs. Wells called the vocalizations in CH-I, II, and III “stutters”, “eeaows”, and “chirps” respectively, and I will use the same names in the descriptions that follow. In CH-II I thought that two or more individuals were vocalizing because two distinct call types were present and at times one seemed to follow another of a different type too quickly to be made by one individual (such as the eeaow CH-II-4 following the stutter CH-II-3). Additionally, the stutters sounded as though they were produced closer to the microphone than the eeaows. Wells confirmed that more than one cheetah was sometimes on the premises during the recordings, but he did not indicate specifically that some of the CH-II vocalizations were made
by an individual other than Norok.

The spectrograms for cheetah vocalizations are formatted and commented in the same manner as for bobcat. Refer to Appendix 2 (pp 191-213) for complete summaries, timelines, spectrograms, and comments of the three cheetah collections.

2.2 Results

2.2.1 CH-I Collection

(see pp 191-197 of Appendix 2)

The 17 vocalizations found in this 39 second collection were produced by one adult male cheetah. All 17 vocalizations occur within a 34 second interval (see timeline on page 193 of Appendix 2), with the spacing between vocalizations ranging from 58 to 5784 ms (mean: 1521, standard deviation: 1709). Vocalization duration ranged from 45 to 1300 ms (mean: 569, standard deviation: 348). The vocalizations are described (with some exceptions) as “stutters” by S. Wells. G. Peters reported a similar type of vocalization which he called “gurgles” in his 1983 paper. Background noise is heavy and the sound quality is relatively poor compared to the bobcat recordings (i.e. the cheetah recordings have less clarity than the bobcat recordings).

Frequency Content

The spectrograms of this collection (see pp 194-197) display frequency content between 0.4 and 7 kHz. The vocalizations CH-I-(1, 2, 3, 4, 6, 7, 8, 9, 14, 16, and 17) are all “stutters”, characterized by the strong presence of pulsed amplitude modulation in their spectrograms (see Figure 2.1 for a visual indication of this). The pulse rate is approximately 20 per second. The structure of individual pulses, which have a duration on the order of 25 ms, appears to be harmonic (judging by the presence of regularly-spaced horizontal stripes visible in the spectrograms of many pulses), probably with vocalized noise added. The lowest harmonic frequency is the fundamental. Often the darkest harmonics
are the first and third. The remaining vocalizations, CH-I-(5, 10, 11, 12, 13, and 15) are not stutters (they lack the pulsed amplitude modulation), and have varied structures that are best appreciated by consulting the spectrograms.

**Figure 2.1: Example of Pulsed Amplitude Modulation in a “Stutter”**

Example of a cheetah “stutter” (CH-I-4), characterized by the pulsed amplitude modulation that is clearly visible in the spectrogram. Approximately 21 pulses are visible within the indicated range, with the first and last very faint. Three harmonics are indicated for one of the pulses.

**Pitch Contour**

The pitch contour for stutters is essentially constant over the whole vocalization. The contours of individual pulses appear to be “UCD”. F0 measurements for stutters range from 473 to 630 Hz (mean: 504, standard deviation: 42) (based on data available in Appendix 2).

The non-stutters have either an essentially flat pitch contour (vocalizations 10 and 15), an unclear pitch contour (10, 12, and 13), or an unusual pitch contour (5 has a CUCDCD contour) (see Appendix 2 for any relevant F0 measurements on non-stutters).
Additional Features

The stutter repetition rates for this collection range from 18 to 29 pulses per second (mean: 22, standard deviation: 4). See Table 2.1 and Section 2.3 for cumulative statistics on all stutters found in CH-I and II. As mentioned earlier, vocalizations CH-I-(5, 10, 11, 13, and 15) are sounds other than stutters. Their structures are varied and do not follow a regular pattern. The margin comments in Appendix 2 offer brief descriptions of each vocalization.

2.2.2 CH-II Collection

(see pp 198-203 of Appendix 2)

The 12 vocalizations found in this 22 second collection were produced by the same male cheetah as CH-I, as well as possibly another cheetah. All 12 vocalizations were made within an 18 second interval (see timeline on page 200 of Appendix 2), with the spacing between vocalizations ranging from 388 to 1857 ms (mean: 1109, standard deviation: 487). Vocalization durations ranged from 240 to 711 ms (mean: 405, standard deviation: 135). Of the 12 vocalizations, 3 are stutters (3, 10, and 11) and the rest are referred to as “eeaows” by S. Wells.

When listening to the tape it sounds as though two cats are present, one responsible for the stutters and the other for the non-stutters: the stutters sound as though they are produced closer to the microphone than the eeaows, and the eeaow CH-II-4 follows stutter CH-II-3 quickly enough that on the recording it sounds as though they are produced by different cats. Like CH-I, recording quality is relatively poor and background noise is heavy.

Frequency Content

The spectrograms of this collection (see pp 201-203 of Appendix 2) display frequency content between 0.5 and 5 kHz, with most of the content occurring below 2 kHz. Stutter
structure is essentially the same as in CH-I, though there may be a little more additive noise. Of the eeaows, harmonic structure is evident, with the lowest harmonic at the fundamental frequency. These vocalizations are relatively weak, with usually only 2 harmonics visible. The background noise is too heavy to tell if “vocalized” noise is present. Occasionally, especially near the end of a vocalization, 5 or more harmonics can become visible, e.g. CH-II-(1, 6, 7, and 8).

**Pitch Contour**

The stutters, again, have flat pitch contours, which in this collection have F0 ranging from 410 to 473 Hz (mean: 441, standard deviation: 26). See Table 2.1 and Section 2.3 for a cumulative summary of the stutters in CH-I and II. The eeaows typically have a “UCDC” contour that sometimes also dips near the end (“UCDCD”). As most vocalizations were rather weak, it was not always possible to see the full contour, so some notations are estimated. Additionally, it was not always possible to measure the peak F0, and the measurements were sometimes made on the trailing constant region. So, these measurements do not necessarily reflect an upper bound on F0. Nonetheless, the F0 measurements for eeaows range from 693 to 788 Hz (mean: 735, standard deviation: 33).

**Additional Features**

The repetition rates of the three stutters (CH-II-(3, 10, and 11)) range from 20.7 to 21.8 pulses per second (mean 21.2, standard deviation 0.4). The cumulative stutter rates (based on all stutters from CH-I and II) range from 18.2 to 29.1 pulses per second (mean: 21.6, standard deviation: 3.2). The cumulative stutter durations range from 103 to 1029 ms (mean: 526, standard deviation: 253), and the fundamental frequency measurements range from 410 to 630 (mean: 491, standard deviation: 47).
2.2.3 CH-III Collection
(see pp 204-213 of Appendix 2)

The 30 vocalizations found in this 47 second collection were made by the same adult male cheetah as CH-I and II. The vocalizations are spread out within the entire 47 seconds of the collection (see timeline on page 205 of Appendix 2), with the spacing between vocalizations ranging from 362 to 2047 ms (mean: 1050, standard deviation: 493). Vocalization durations range from 311 to 702 ms (mean: 494, standard deviation: 81). These vocalizations were described as chirps by S. Wells. The sound quality on the recording is relatively poor and background noise is fairly heavy.

Frequency Content

The spectrograms of this collection (see pp 206-213 of Appendix 2) display frequency content between 0.1 and 7.5 kHz. All vocalizations display harmonic structure, with the lowest harmonic the fundamental. At least 3 harmonics are visible in each vocalization with portions of some vocalizations showing up to 11 harmonics (e.g. CH-III-(1, 7, 15, and 29)), often near the beginning or end. Vocalizations appear to contain a substantial amount of “vocalized” noise, usually concentrated closer to the beginning of the vocalization. At the end of many vocalizations only the first 2 harmonics remain visible.

Pitch Contour

Most of the vocalizations in this collection feature a “UCDC” pitch contour. A few are slight variants of this, but all feature a rising F0 at the beginning and a constant F0 at the end. The peak frequency usually occurs to the left of the vocalization’s midpoint. Peak measurements range from 662 to 1040 Hz (mean: 882, standard deviation: 94).

Additional Features

None.
2.3 Summary of Features

Table 2.1, like Table 1.2 for bobcat, summarizes an assortment of features of the vocalizations in the three cheetah collections. The first column lists abbreviated names of the various features that were measured and described for the vocalizations, and the remaining 3 columns contain the appropriate entries for each of the 3 cheetah collections.

Entries 1-12 of the table have the same meaning as in Table 1.2 (refer to Section 1.3). Entries 13-16 indicate how many vocalizations within each collection are “stutters”, “eeaows”, “chirps”, or none of the above (other). Entry 17 contains statistics for stutter rate (# pulses / duration of stutter) for the stutters within each collection.

In Table 2.2 vocalizations of the same type are combined and a few features of these groups are displayed. # Vocalizations indicates how many vocalizations of the given type were available. Collections indicates which cheetah collections (I, II, or III) vocalizations of the given type were taken from. Duration and F0 Measurement indicate duration and F0 statistics for the vocalizations of each type. In addition to the Mean, Standard Deviation, Maximum, and Minimum, the Standard Errors of the Mean (Std. Dev / sqrt(N)) are also computed. In the case of stutters and eeaows the F0 measurements were not of a peak in the pitch contour, but in the case of chirps the measurements were of a peak. Stutter Rate indicates statistics for the pulse rate of the complete ensemble of stutters.

<table>
<thead>
<tr>
<th></th>
<th>CH-I</th>
<th>CH-II</th>
<th>CH-III</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. # Vocalizations</td>
<td>17</td>
<td>12</td>
<td>30</td>
</tr>
<tr>
<td>2. Length (s)</td>
<td>39</td>
<td>22</td>
<td>47</td>
</tr>
<tr>
<td>3a. All Vocs w/in (s)</td>
<td>34</td>
<td>18</td>
<td>47</td>
</tr>
<tr>
<td>3b. # seconds per Voc.</td>
<td>2</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td>4. Cats Present</td>
<td>1-m</td>
<td>1-m (or more)</td>
<td>1-m</td>
</tr>
</tbody>
</table>

Table 2.1: Summary of Cheetah Descriptions (by Collection)
<table>
<thead>
<tr>
<th></th>
<th>CH-I</th>
<th>CH-II</th>
<th>CH-III</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Duration: Mean</td>
<td>569</td>
<td>405</td>
<td>494</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>348</td>
<td>135</td>
<td>81</td>
</tr>
<tr>
<td>Minimum</td>
<td>45</td>
<td>240</td>
<td>311</td>
</tr>
<tr>
<td>Maximum</td>
<td>1300</td>
<td>711</td>
<td>702</td>
</tr>
<tr>
<td>6. Silent Int.: Mean</td>
<td>1521</td>
<td>1109</td>
<td>1050</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>1709</td>
<td>487</td>
<td>493</td>
</tr>
<tr>
<td>Minimum</td>
<td>58</td>
<td>388</td>
<td>362</td>
</tr>
<tr>
<td>Maximum</td>
<td>5784</td>
<td>18</td>
<td>2047</td>
</tr>
<tr>
<td>7. Freq. Range (kHz)</td>
<td>0.45-7</td>
<td>0.5-5.5</td>
<td>0.1-7.5</td>
</tr>
<tr>
<td>8. Darkest Rng. (kHz)</td>
<td>&lt; 2</td>
<td>&lt; 2</td>
<td>&lt; 3</td>
</tr>
<tr>
<td>9. Harmonics Desc.</td>
<td>some noise</td>
<td>weak, smooth</td>
<td>noisy, smooth</td>
</tr>
<tr>
<td>10. Fund Present?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>11. F0/H1: Mean</td>
<td>515</td>
<td>662</td>
<td>882</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>50</td>
<td>131</td>
<td>94</td>
</tr>
<tr>
<td>Minimum</td>
<td>473</td>
<td>410</td>
<td>662</td>
</tr>
<tr>
<td>Maximum</td>
<td>630</td>
<td>788</td>
<td>1040</td>
</tr>
<tr>
<td>12. Pitch Contours</td>
<td>flat (stutters)</td>
<td>mostly UCDC (eeaows)</td>
<td>mostly UCDC</td>
</tr>
<tr>
<td>13. # Eeaows</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>14. # Chirps</td>
<td>0</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>15. Other</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16. Stutters</td>
<td>11</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>17. Stutter Rate: Mean</td>
<td>21.7</td>
<td>21.2</td>
<td>--</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>3.6</td>
<td>0.4</td>
<td>--</td>
</tr>
<tr>
<td>Minimum</td>
<td>18.2</td>
<td>20.7</td>
<td>--</td>
</tr>
<tr>
<td>Maximum</td>
<td>29.1</td>
<td>21.8</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 2.1: Summary of Cheetah Descriptions (by Collection)
Table 2.2: Summary of Cheetah Descriptions (by Type)

<table>
<thead>
<tr>
<th></th>
<th>Stutters</th>
<th>Eeaows</th>
<th>Chirps</th>
</tr>
</thead>
<tbody>
<tr>
<td># Vocalizations</td>
<td>14</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td>Collections</td>
<td>I, II</td>
<td>II</td>
<td>III</td>
</tr>
<tr>
<td>Duration: Mean</td>
<td>526</td>
<td>367</td>
<td>494</td>
</tr>
<tr>
<td></td>
<td>253</td>
<td>109</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>103</td>
<td>240</td>
<td>311</td>
</tr>
<tr>
<td></td>
<td>1029</td>
<td>594</td>
<td>702</td>
</tr>
<tr>
<td></td>
<td>68</td>
<td>36</td>
<td>15</td>
</tr>
<tr>
<td>Minimum</td>
<td>103</td>
<td>240</td>
<td>311</td>
</tr>
<tr>
<td>Maximum</td>
<td>1029</td>
<td>594</td>
<td>702</td>
</tr>
<tr>
<td>Std. Error</td>
<td>68</td>
<td>36</td>
<td>15</td>
</tr>
<tr>
<td>F0 Meas.: Mean</td>
<td>491</td>
<td>735</td>
<td>882</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>33</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>410</td>
<td>693</td>
<td>662</td>
</tr>
<tr>
<td></td>
<td>630</td>
<td>788</td>
<td>1040</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>Stutter Rate: Mean</td>
<td>21.6</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>3.2</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>18.2</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>29.1</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Std. Error</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.4 Discussion and Conclusions
The vocalizations within the three cheetah collections were more varied than the bobcat vocalizations. Three distinct cheetah call types were represented (called “stutters”, “eeaows”, and “chirps” after notation presented in Ruiz-Miranda et al. 1998) in the recordings, whose features are easily distinguishable. Additionally, a handful of unclassified vocalizations were present in the CH-I collection, showing further variety in call
The stutters tend to last for around half a second, and feature pulsed amplitude modulation with a rate that sticks pretty close to 22 pulses per second. The standard error for duration in stutters is 68 ms, larger than that of eeaows and chirps. The standard error for pulse rate was less than 1, indicating that the stutter rate does not vary much. The individual pulses feature harmonic structure, with the lowest harmonic of each pulse sticking pretty close to 490 Hz, with standard error of 13 Hz. There appears to be vocalized noise present in the pulses. Of the 3 cheetah collections the silent interval in CH-I is the largest, yielding an average time per vocalization of 2 seconds, compared to 1.5 for CH-II and 1.6 for CH-III.

The eeaows are the shortest and the weakest of the cheetah vocalizations. They are difficult to observe and measure due to their faintness in the spectrograms. Nonetheless, it is possible to see that their pitch contour is roughly “UCDC”. Only two harmonics are usually visible, except near the end of some vocalizations, at which point several harmonics sometimes become visible for a short while. Vocalized noise is minimal. Qualitatively the eeaows sound gentle and smooth.

The chirps are the most numerous of any of the types, featuring a fairly large amount of vocalized noise, prominent harmonic structure (3 or more harmonics are visible), and typically a “UCDC” pitch contour. At the beginning of many of the chirps the pitch rises rapidly to the peak, relative to the other chirps. In many chirps there are short times when many harmonics become visible. Qualitatively many of the chirps, especially those with the sudden pitch rise at the beginning and a large amount of vocalized noise, sound almost like barks. They sound harsh and agitated. Chirps with more gradual pitch rises and less vocalized noise have a somewhat gentler, less urgent sound.
Statistical comparisons for the cheetah vocalization measurements were not performed, since the utility of such comparisons is not great given that the vocalizations are so different to begin with, as well as the fact that F0 peak measurements were only made for chirps.
Part II: Filter Construction and Hypothesis Testing

Background

In Part I we investigated the acoustic structure of a series of bobcat and cheetah vocalizations. In Part II we look at the middle ears of these two species (see Chapter 3) and test an hypothesis that there is a relationship between vocalization structure and middle-ear structure that enhances vocal communication for these two cat species (see Chapters 4, 5, and 6).

Researchers have established that the acoustic properties of domestic cat middle ears can be modeled with a 6 element lumped parameter circuit model (see Figure 2.1) for which current is analogous to volume velocity and voltage is analogous to pressure. Acoustic measurements on ears of anaesthetized and deceased cats have established the applicability of this model to a few cat species. Such a model is used in this thesis research to represent the acoustic and mechanical properties of bobcat and cheetah middle ears, with element values derived from experiments on ears of these two species (from Taber-ner, 1999, and Huang, 1997a. See Table 3.1).

With the middle-ear model and experimentally derived element values it is possible to simulate on a computer the Cavity Gain ($G_{CAV}$), i.e. the function that describes the transmission of sound from the ear canal to the ossicular chain (see Introduction and Chapter 3). By using the Cavity Gain simulations and digitized versions of the vocalizations shown in Part I, it is possible to filter the vocalizations with the Cavity Gain and thus produce a simulated version of the sound that would drive the ossicular chain of a bobcat or a cheetah (see Chapter 3).

As was shown in the Introduction (Figure 0.3) for lion and domestic cat, the magnitude of the Cavity Gain for many felid species acts as a notch filter, sharply attenuating the
transmission of a narrow range of frequencies. This fact combined with the structural observations of the vocalizations in Part I leads to an hypothesis concerning a possible function of the $G_{CAV}$ structure in vocal communication (see Chapter 4).

**Figure 2.2: Middle-Ear Circuit Model Superimposed Upon the Structure it Represents**

Electrical circuit model of the felid middle ear, with circuit elements superimposed on the structural features they model. $Y_{TM}$ represents the admittance seen looking into the middle ear from the ear canal. $P_{TM}$ represents the sound pressure in the ear canal next to the tympanic membrane, and $P_{CAV}$ represents the pressure inside the tympanic cavity. The transfer function $(P_{TM} - P_{CAV})/P_{TM}$ represents $G_{CAV}$, a factor in the transmission of sound from the ear canal to the cochlea. $R_{TOC}$ and $C_{TOC}$ represent, respectively, the resistance and mechanical compliance of the tympanic membrane, ossicles, and cochlea as seen at the tympanic membrane. $C_{TC}$ and $C_{BC}$ represent the tympanic cavity and bullar cavity acoustic compliances respectively. $M_F$ and $R_F$ represent the acoustic mass and resistance of the foramen respectively. [From Huang et al., 2000]
Once filtered versions of the vocalizations are available, it is possible to devise listening experiments to test whether differences perceived (by humans) between the filtered and unfiltered vocalizations lend any support to the hypothesis that the middle ears and vocalizations of the bobcat and cheetah are structured in a way that enhances vocal communication (see Chapters 5 and 6).
Chapter 3

Construction of Middle-Ear Filters

3.1 Introduction to Model

This chapter uses the lumped element middle-ear model shown in Figure 3.1 to simulate the transfer of sound through the middle ears of bobcats and cheetahs.

![Figure 3.1: Felid Middle-Ear Model](image)

Lumped-element circuit model for the felid middle ear (see Figure 2.1 for an indication of how these elements relate to middle-ear structure). The impedances modelling the tympanic-ossicular chain are grouped as $Z_{TOC}$, the impedance of the tympanic cavity is grouped as $Z_{TC}$, and the impedance representing the foramen and bullar cavity is grouped as $Z_{BC}$. The pressure at the tympanic membrane is marked as $P_{TM}$, the pressure in the tympanic cavity is marked as $P_{CAV}$, and the pressure across the tympanic-ossicular chain is marked $P_{TOC}$. The Cavity Gain, $G_{CAV}$, is $P_{TOC}/P_{TM}$, which is shown in Equation 3.2 in terms of the impedances of this figure.

In the Figure 3.1, $R_{TOC}$ and $C_{TOC}$ represent the mechanical resistance and compliance of the tympanic membrane and ossicular chain. Their combined impedance, $Z_{TOC}$, is equal to $R_{TOC} + 1/(j\omega C_{TOC})$. $C_{TC}$ represents the acoustic compliance of the tympanic cavity, whose impedance ($Z_{TC}$) is equal to $1/(j\omega C_{TC})$. $R_F$ and $M_F$ represent the acoustic
resistance and mass respectively of the foramen, and $C_{BC}$ represents the acoustic compliance of the bullar cavity. The combined impedance of the foramen and bullar cavity, $Z_{BC}$, is equal to $R_F + j\omega M_F + 1/(j\omega C_{BC})$. The function representing a factor in the transfer of sound to the tympanic membrane, the Cavity Gain ($G_{CAV}$), is equal to the transfer function from the pressure in the ear canal ($P_{TM}$) to the pressure across the tympanic-ossicular chain ($P_{TOC} = P_{TM} - P_{CAV}$):

$$G_{CAV} = (P_{TM} - P_{CAV})/P_{TM} = P_{TOC}/P_{TM}.$$  \hspace{1cm} (3.1)

In terms of the impedances shown in Figure 3.1, $G_{CAV}$ can be found using the voltage divider relationship between $P_{TOC}$ and $P_{TM}$:

$$G_{CAV} = Z_{TOC} / (Z_{TOC} + (Z_{TC} \parallel Z_{BC})).$$  \hspace{1cm} (3.2)

$G_{CAV}$ is simulated using the experimentally derived values for the circuit parameters shown in Table 3.1 for bobcat and cheetah. All six parameters for the cheetah middle ear were experimentally determined by Taberner (1999), on the left and right ears of one deceased cheetah. For the bobcat middle ear $C_{TC}$, $C_{BC}$, $M_F$ and $R_F$ were also experimentally determined by Taberner on the left and right ears of one deceased bobcat, but measurements of $C_{TOC}$ and $R_{TOC}$ were not available for those ears. Instead, values of $C_{TOC}$ and $R_{TOC}$ for bobcat were estimated based on Figures 3A and 4B of Huang, 1997a.

The impedance of the foramen resistance $R_F$, i.e. the acoustic resistance of a narrow tube, is modeled as being proportional to the square root of frequency (see Beranek, 1996, p. 137). The proportionality constant was determined by Taberner, 1999, for the models used in this thesis. All element values in Table 3.1 are given in standard MKS units (meters, kilograms, seconds) for acoustical elements.
<table>
<thead>
<tr>
<th>Element</th>
<th>BCL</th>
<th>BCR</th>
<th>CHL</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{TOC}$</td>
<td>$5.305 \times 10^{-12}$</td>
<td>$5.305 \times 10^{-12}$</td>
<td>$3.3 \times 10^{-12}$</td>
<td>$3.61 \times 10^{-12}$</td>
</tr>
<tr>
<td>$C_{TC}$</td>
<td>$2.57 \times 10^{-12}$</td>
<td>$2.28 \times 10^{-12}$</td>
<td>$5.32 \times 10^{-12}$</td>
<td>$4.98 \times 10^{-12}$</td>
</tr>
<tr>
<td>$C_{BC}$</td>
<td>$0.94 \times 10^{-11}$</td>
<td>$1.0 \times 10^{-11}$</td>
<td>$2.45 \times 10^{-11}$</td>
<td>$2.49 \times 10^{-11}$</td>
</tr>
<tr>
<td>$R_{TOC}$</td>
<td>$5.5 \times 10^{7}$</td>
<td>$5.5 \times 10^{7}$</td>
<td>$2 \times 10^{7}$</td>
<td>$2 \times 10^{7}$</td>
</tr>
<tr>
<td>$R_{F}$</td>
<td>$5 \times 10^{4} \sqrt{f}$</td>
<td>$5 \times 10^{4} \sqrt{f}$</td>
<td>$3 \times 10^{4} \sqrt{f}$</td>
<td>$3 \times 10^{4} \sqrt{f}$</td>
</tr>
<tr>
<td>$M_{F}$</td>
<td>1622</td>
<td>1622</td>
<td>1380</td>
<td>1380</td>
</tr>
</tbody>
</table>

**Table 3.1: Experimentally Derived Element Values**

Values of the 6 circuit elements for left and right bobcat ears (BCL and BCR respectively) as well as left and right cheetah ears (CHL and CHR respectively). Acoustic compliances ($C_{TOC}$, $C_{TC}$, and $C_{BC}$) have the units [meters$^3$/Pascal]. Acoustic masses ($M_{F}$) have the units [Pascal*seconds/meters$^3$]. Acoustic resistances ($R_{TOC}$ and $R_{F}$) have the units [kilograms/(seconds*meters$^4$)]. $R_{F}$ is proportional to the square root of frequency to better represent the acoustic loss in a narrow tube (see Beranek, 1996, page 137). [Values were taken from Taberner, 1999, and adapted from Huang et al., 1997a).

In this chapter the acoustic model will be examined first from a theoretical standpoint, using the circuit model to predict the behavior of the transfer function. Next the procedure for implementing the model as a filter in Matlab will be presented in detail, followed by comparisons between the theoretical models and the implementations to verify their accuracy. As will be shown in Section 3.3.2, the left and right Cavity Gains are very similar for both species, so only the left ear Cavity Gains will be used for most of the analysis and actual filtering in this thesis. Finally, the relationship between middle-ear structural features and some features of $G_{CAV}$ (dip center frequency, bandwidth, Quality Factor (Q), and low-frequency gain) will be clarified by showing how structural modifications can affect $G_{CAV}$.
3.2 Theoretical Model Characteristics

This section uses the circuit model in Figure 3.1, the element values in Table 3.1, and the $G_{CAV}$ expression in Equation 3.2 to estimate the behavior of $G_{CAV}$, both for the purpose of building familiarity with the circuit's behavior, and for establishing a way of testing the behavior of the Matlab implementation that will be presented in Section 3.3. The low-frequency gain, center frequency and bandwidth of the dip, and high-frequency gain of $G_{CAV}$ will be approximated from analysis of the circuit model.

3.2.1 Low-Frequency Behavior

At low frequencies the impedance of an inductor decreases, and that of a capacitor increases. At sufficiently low frequencies, then, $M_F$ looks like a short circuit, and $R_F$ and $R_{TOC}$ are small in magnitude compared to the impedance of their respective series capacitors, such that they can be replaced with short circuits. The circuit, then, becomes as shown in Figure 3.2.

Figure 3.2: Low-Frequency Middle-Ear Model

Felid middle-ear model approximated for low frequencies. As the frequency becomes small, capacitive elements dominate over series resistive and inductive elements.
As the frequency approaches zero, the cavity gain becomes (based on the voltage divider relationship between $P_{TM}$ and $P_{TOC}$ of Figure 3.2):

$$G_{CAV} = \frac{C_{TC} + C_{BC}}{C_{TC} + C_{BC} + C_{TOC}}$$

(3.3)

This result, in dB, is shown for the four ears in Table 3.2:

<table>
<thead>
<tr>
<th>Low-Freq. Gain (dB)</th>
<th>BCL</th>
<th>BCR</th>
<th>CHL</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3.194</td>
<td>-3.119</td>
<td>-0.912</td>
<td>-0.991</td>
</tr>
</tbody>
</table>

Table 3.2: Predicted Low-Frequency Behavior of Middle Ears

Estimates of the DC (frequency = 0) values of $G_{CAV}$ based on the low-frequency approximation given by Equation 3.3. The numbers are in decibels.

The cavity gain at DC only depends on the capacitor values. As $C_{TC} + C_{BC}$ becomes large relative to $C_{TOC}$, the DC gain approaches 1. As $C_{TOC}$ becomes large relative to $(C_{TC} + C_{BC})$, the DC gain decreases. Since acoustic compliances are proportional to cavity volume (i.e. $C = V_0/(\rho_0 * c^2)$, $V_0 =$ cavity volume, $\rho_0 = 1.18$ kg/m$^3$, $c = 345$ m/s), increasing the volumes of the tympanic cavity or bullar cavity will increase the low-frequency gain (see Section 3.5).

3.2.2 Frequency and Bandwidth of $G_{CAV}$ Dip

As was mentioned in the introduction, the transfer function for many felid middle ears features a narrow notch at a frequency that depends on the middle-ear structure. Referring to the definition of $G_{CAV}$ as $(P_{TM} - P_{CAV})/P_{TM}$ (Equation 3.1), the gain could approach 0 if $P_{CAV}$ were approximately equal to $P_{TM}$. Such a case occurs when the impedance of the cavities and foramen, $Z_{TC} || Z_{BC}$, becomes very large relative to $Z_{TOC}$, thus driving the expression for $G_{CAV}$ in Equation 3.2 toward zero. Equation 3.4 shows the value of $Z_{TC} || Z_{BC}$ in terms of the grouped impedances shown in Figure 3.1.

$$Z_{TC} || Z_{BC} = Z_{TC}Z_{BC} / (Z_{TC} + Z_{BC})$$

(3.4)
Clearly this impedance reaches its maximum value when \( Z_{TC} + Z_{BC} \) reaches its minimum value, such that the denominator of Equation 3.4 is at its smallest. In Equation 3.5 the denominator is expressed in terms of the individual elements of Figure 3.1:

\[
Z_{TC} + Z_{BC} = 1/(j\omega C_{TC}) + 1/(j\omega C_{BC}) + j\omega M_F + R_F. \tag{3.5}
\]

Combining the capacitors together, Equation 3.5 becomes:

\[
Z_{TC} + Z_{BC} = (C_{TC} + C_{BC})/(j\omega (C_{TC}*C_{BC})) + j\omega M_F + R_F. \tag{3.6}
\]

Ignoring the frequency-dependence of \( R_F \) for the moment, the smallest value of this impedance occurs at the frequencies when the capacitor and inductor impedances are equal in magnitude but opposite in sign such that they cancel each other out. This frequency, \( \omega_0 \), is:

\[
\text{Center frequency (radians/second)} = \omega_0 = 1 / \sqrt{M_F*(C_{TC}*C_{BC})/(C_{TC} + C_{BC})}. \tag{3.7}
\]

At this frequency (again, assuming that \( R_F \) does not vary with frequency) the magnitude of the impedance of the cavities reaches its maximum value and \( G_{CAV} \) reaches its minimum value. Consequently, the center frequency of the dip in \( G_{CAV} \) (in radians per second) should be \( \omega_0 \). In the Matlab implementation of \( G_{CAV} \) (see Section 3.3) \( R_F \) will be frequency-dependent, so the dip center frequency will not be exactly equal to those computed theoretically according to Equation 3.7.

At \( \omega_0 \), since the impedance of the cavities \( (Z_{TC}||Z_{BC}) \) reaches its maximum value, the volume velocity entering the cavities through \( Z_{TOC} \) will be small. If we assume this volume velocity is zero, the cavities and foramen form a series “RLC” circuit. The bandwidth of a series RLC circuit is:
Bandwidth (radians/second) = R/L. \hspace{1cm} (3.8)

Here, R would be $R_F$ and L would be $M_F$. For the purpose of making calculations I converted radian frequencies to Hz frequencies (dividing by $2\pi$) and evaluated the frequency-dependent $R_F$ at the center frequency $f = 2\pi\omega_0$. Table 3.3 shows the dip frequencies (in Hz) and bandwidths (in Hz), as well as the Quality Factor, or $Q$ (equal to the center frequency divided by the bandwidth), calculated for $G_{CAV}$ of the four ear models.

<table>
<thead>
<tr>
<th></th>
<th>BCL</th>
<th>BCR</th>
<th>CHL</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dip Center Frequency (Hz)</td>
<td>2783</td>
<td>2900</td>
<td>2049</td>
<td>2103</td>
</tr>
<tr>
<td>Bandwidth (Hz)</td>
<td>259</td>
<td>264</td>
<td>157</td>
<td>159</td>
</tr>
<tr>
<td>$Q$</td>
<td>10.7</td>
<td>11.0</td>
<td>13.1</td>
<td>13.2</td>
</tr>
</tbody>
</table>

Table 3.3: Calculated Center Frequencies and Bandwidths of the $G_{CAV}$ Dip

Calculated center frequencies and bandwidths (in Hz), for $G_{CAV}$ of the four ear models, computed using Equations 3.7 and 3.8 respectively (and converting to Hz). The center frequencies were calculated while assuming that $R_F$ was not frequency-dependent. Consequently the values in the implementation (see Section 3.3) should differ somewhat from these calculations. For the bandwidth calculations a value of the frequency-dependent $R_F$ evaluated at the appropriate center frequency was used. The “quality factor”, $Q$, was computed by dividing the center frequency by the bandwidth.

3.2.3 High-Frequency Cavity Gain

At high frequencies the impedance of an inductor becomes large, and that of a capacitor becomes small. Referring to Figure 3.1, in the limit as $\omega \rightarrow \infty$, the inductor $M_F$ looks like an open circuit and the capacitor $C_{TC}$ looks like a short circuit, so $P_{CAV}$ approaches zero. The capacitor $C_{TOC}$ also looks like a short circuit, such that the negative reference node of $P_{TOC}$ becomes grounded. Consequently, at very high frequencies $P_{TOC}$ and $P_{TM}$ approach the same value and $G_{CAV}$ approaches 1.
3.3 Filter Implementation

Filters with frequency-domain behavior described by \( G_{\text{CAV}} \) were implemented in Matlab for each ear as discrete-time impulse responses appropriate for convolving with the samples of the digitized vocalizations described in Part I. The discrete-time implementations are based on the continuous-time model in Equation 3.2.

3.3.1 Generation of Frequency Response Samples

Four functions, one for each ear, were written in Matlab to take as input a vector of frequency values and give as output a vector of \( G_{\text{CAV}} \) evaluated at those frequencies. The four functions, BCL.m, BCR.m, CHL.m, and CHR.m (see pp 214-217 of Appendix 3), contain the values of the lumped elements for their respective ears (see Table 3.1), and perform the transfer function calculation on the vector of input frequencies. Additionally, switches were implemented for turning on or off the frequency dependence of the foramen resistor, as well as simulating other structural modifications to the ears (see Section 3.5). These four functions perform all of the computation necessary to generate the discrete frequency representations of \( G_{\text{CAV}} \). In the following sections I explain how appropriate frequencies for evaluating \( G_{\text{CAV}} \) were chosen, and how the \( G_{\text{CAV}} \) samples were arranged so that the inverse fast-Fourier-transform function (\texttt{ifft}) could be used on them to compute the impulse responses.

3.3.2 Selection of Frequencies for Evaluating \( G_{\text{CAV}} \)

The sampling rate used for the vocalizations is 16,129 samples/second (see Chapter 1). Consequently, the amount of time passing between samples is \( T = 1/16129 = 0.062 \) milliseconds. The discrete-time impulse response to be used with the digitized vocalizations, then, must have the same spacing between its samples. In other words, it must use the same sampling frequency as that used to digitize the vocalizations. To accommodate this, the frequencies used for evaluating \( G_{\text{CAV}} \) must be band-limited within the range (\( +/- \)) \( 1/(2T) = ( +/- ) 16,129/2 = ( +/- ) 8064.5 \) Hz. The number of samples taken, \( N \), was chosen to
be 1024, a number that is large enough to capture most of the details of the frequency-domain behavior (samples are ~16 Hz apart (8064.5/512)). The frequencies for evaluating $G_{CAV}$ were initially chosen from -1/2T to +1/2T, with N+1 samples in-between so that the samples would be symmetric with respect to 0. A vector of these frequencies was used as the input to the functions described in Section 3.3.1, resulting in a vector of $G_{CAV}$ evaluated at those frequencies for the four ears.

Figures 3.3-3.6 show linear plots and Bode (log-log) plots for the left and right bobcat and cheetah ears. As can be seen from the plots, the versions of $G_{CAV}$ for the left and right ears are very similar for both cheetah and bobcat. Consequently, I choose to use only the left ears for most of the remaining analyses, plots, and filtering. The rest of this section covers the implementation details behind transforming this frequency domain behavior into discrete time impulse responses appropriate for convolution with the vocalizations presented in Part I.
Figure 3.3: Linear Plots of $G_{CAV}$ for Left and Right Bobcat Middle Ears

Linear plots of the magnitude and phase of $G_{CAV}$ for the left and right ears of a bobcat. The transfer function of Equation 3.2 was implemented on Matlab using the appropriate bobcat element values in Table 3.1. The $G_{CAV}$ function was evaluated at 1024 equally spaced frequencies between -8064.5 and +8064.5 Hz. Only positive frequencies are shown. The phase is measured in periods.
Figure 3.4: Bode Plots of $G_{CAV}$ for Left and Right Bobcat Middle Ears

Bode plots of the magnitude and phase of $G_{CAV}$ for the left and right ears of a bobcat. The same transfer functions mentioned in Figure 3.3 were evaluated at logarithmically spaced frequencies spanning 2 decades from 100 Hz to 10,000 Hz. The dB magnitude is computed as $20\log_{10}(|G_{CAV}(j\omega)|)$. The phase is in periods.
Linear plots of the magnitude and phase of $G_{CAV}$ for the left and right ears of a cheetah. The transfer function of Equation 3.2 was implemented on Matlab using the appropriate cheetah element values in Table 3.1. The $G_{CAV}$ function was evaluated at 1024 equally spaced frequencies between -8064.5 and +8064.5 Hz. Only positive frequencies are shown. The phase is measured in periods.
Figure 3.6: Bode Plots of $G_{CAV}$ for Left and Right Cheetah Middle Ears

Bode plots of the magnitude and phase of $G_{CAV}$ for the left and right ears of a cheetah. The same transfer functions mentioned in Figure 3.5 were evaluated at logarithmically spaced frequencies spanning 2 decades from 100 Hz to 10,000 Hz. The dB magnitude is computed as $20 \times \log_{10}(|G_{CAV}(j\omega)|)$. The phase is in periods.

3.3.3 Effects of Frequency-Dependent $R_F$ on Inverse Transform Operation

Considering that the frequency-dependency of $R_F$ was implemented using $\sqrt{f}$, when $f$ is negative $R_F$ is imaginary. The different behavior of $R_F$ for positive and negative frequencies destroys the even symmetry in the real part and odd symmetry in the imaginary part of $G_{CAV}$ (Figures 3.3-3.6 only show positive frequencies so the problem is not visible there). Such symmetry is necessary for the inverse transform to be real, and would
in fact be present if not for the \sqrt{f} term. My solution to this problem is to generate the required even symmetry in the real part and odd symmetry in the imaginary part by appropriately copying the $G_{CAV}$ values for positive frequencies to the negative frequencies. In other words, the negative frequencies evaluated from the transfer function implementation are discarded and replaced by the complex conjugate of the positive samples. $N+1$ samples of $G_{CAV}$ result, evaluated from $-1/(2T)$ to $+1/(2T)$ and centered around 0. These $N+1$ frequency samples constitute the information necessary to take the inverse transform, but first they must be arranged in a manner that the inverse-Fourier-transform function in Matlab, called \texttt{ifft}, can use.

\textbf{3.3.4 Use of the \texttt{ifft} Function}

The \texttt{ifft} function expects a vector of $N$ points (a power of two), with the first point corresponding to the frequency 0, “DC”, and the next $(N/2)+1$ points corresponding to positive frequencies up through the highest positive frequency. The remaining $N/2-1$ points are negative frequencies, in reverse order, beginning with the second-most-negative frequency (the most negative frequency shares a spot with the highest positive frequency) and ending with the negative frequency immediately below DC. It is easiest to imagine the points wrapped around a unit circle at the center of a complex plane (the Z plane), whose upper half contains positive frequencies and lower half contains negative frequencies. With DC located at the $(1, 0)$ coordinate, and the highest frequency located at $(-1, 0)$, traveling counterclockwise from $(1, 0)$ over the upper semicircle will visit increasing positive frequencies until the maximum positive frequency is reached at the point $(-1, 0)$. Beginning at $(1, 0)$ and traveling clockwise over the lower semicircle visits negative frequencies increasing in magnitude until the maximum negative frequency is reached at the point $(-1, 0)$. Both the maximum negative frequency and maximum positive frequency correspond to the same point in this arrangement.
Beginning at point (1, 0), N evenly-spaced points are arranged around the unit circle, each of which is associated with the value of $G_{CAV}$ evaluated at that frequency (with negative frequencies modified to preserve symmetry, as described in Section 3.3.3). A vector appropriate as an input for the $\text{ifft}$ function can be created by beginning with the value of $G_{CAV}$ associated with the point (1, 0) (DC) and appending to it each subsequent $G_{CAV}$ value that is encountered while travelling counterclockwise around the circle. This process is illustrated in Figure 3.5 for $N = 8$. In practice the vector used as input to $\text{ifft}$ is generated by using Matlab to appropriately reorder the samples of $G_{CAV}$.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.7.png}
\caption{Illustration of Mapping from Unit Circle to $\text{ifft}$ Vector with $N = 8$}
\end{figure}

Figure 3.7: Illustration of Mapping from Unit Circle to $\text{ifft}$ Vector with $N = 8$

Depiction of how the frequencies in the complex Z plane are arranged as an input into the Matlab $\text{ifft}$ function. The points on the unit circle in the Z plane shown to the left represent the frequencies at which an 8-point discrete Fourier transform is evaluated. The points in the upper half-plane represent positive frequencies (increasing in the counterclockwise direction) and the points in the lower half-plane represent the negative frequencies (becoming more negative in the clockwise direction). The Fourier transform values determined at these points are then stored in a vector ordered according to the right-hand side of the diagram. Basically, points are collected by going in a counterclockwise direction over the unit circle beginning at point 1.
One more issue needs to be addressed before the inverse transform can be performed. I already mentioned that the value of the maximum positive frequency and most negative frequency are treated as the same point in this arrangement (see point 5 in Figure 3.7 for example). As it turns out, the real parts of the function (ignoring the frequency-dependent \( R_F \) problem) are naturally the same at this point, but the imaginary parts are opposite in sign and nonzero. This, of course, is consistent with the fact that the imaginary parts exhibit odd symmetry. However, if I were to allow either the imaginary part of the maximum positive frequency or the imaginary part of the most negative frequency to occupy the maximum-frequency position, the symmetry of the whole collection of samples would be disrupted and the impulse response produced by \( \texttt{IFFT} \) would not be real. Since the impulse response must be real, I decided to preserve the symmetry by changing the imaginary part of the value at the maximum frequency point to zero. The implications of having a discontinuity in the imaginary part of this point are discussed in Section 3.3.6.

3.3.5 Discussion of Impulse Responses

In the inverse Fourier transforms, the impulse responses had non-zero imaginary parts (with amplitudes that were several orders of magnitude smaller than the real parts) due most likely to round-off errors in the computation (without the symmetry adjustments of Section 3.3.5 the imaginary parts had amplitudes similar to those of the real parts). These small terms were eliminated by taking the real part of the impulse responses. Each impulse response is \( N \) samples long, with time between samples of \( T = 0.062 \) milliseconds. Within the first 100 samples of the impulse responses the amplitudes of the samples display most of their variation. See Figure 3.8 for plots of the first 100 points of the impulse responses of bobcat and cheetah \( G_{\text{CAV}} \) implementations (for the left ears).
Impulse responses of $G_{CAV}$ for the left bobcat (top) and left cheetah (bottom) ears. Only the first 100 samples are shown, which for both plots converge toward zero by the end. The decay in the cheetah plot appears to have a larger time constant than that of the bobcat plot, since the amplitude of the ringing does not die down as quickly. The time axis is measured in milliseconds.

The m-files BCear.m and CHear.m (see pp 218-227 of Appendix 3) automatically generate these impulse responses, as well as give options for generating a variety of other plots. As stated earlier, the number of samples, $N$, was set to 1024 for these impulse responses. At the high time-indices of the impulse responses (not shown in Figure 3.8) the impulse responses begin to contain artifacts that cause the amplitudes to randomly rise.

**Figure 3.8:** First 100 Points of the Bobcat and Cheetah Impulse Responses
above zero by a substantial amount. This issue, a consequence of the implementation, is considered in the next section.

3.3.6 Impulse Response Artifacts

The impulse responses generated by the techniques described above should theoretically decay closer and closer to zero as time increases. Near the end of the impulse responses, however, the amplitude begins to stray randomly above zero. These artifacts are a consequence of the manner in which the impulse responses were created from a continuous-time model and do not constitute behavior that we seek to model.

The discrete-time model of $G_{CAV}$ was found by sampling the continuous-time model of $G_{CAV}$ over a limited frequency range. The odd-symmetry of the imaginary part in the original continuous-time model required that, as $\omega \rightarrow (+/-)\infty$ the imaginary parts approach values with the same magnitude but opposite sign. By bandlimiting this continuous-time model and sampling it, however, it is also required that the frequency response be periodic. A convenient way to visualize this is to wrap the samples of the frequency response (evaluated over the frequency range $\omega = -\pi/T$ to $\pi/T$) around the unit circle as shown in Section 3.3.4. In Figure 3.7 the single point $(-1, 0)$ constitutes both the largest positive and the largest negative frequency at which $G_{CAV}$ is to be evaluated. Approaching this point from the upper half of the circle, the imaginary part of $G_{CAV}$ approaches a positive value, but approaching $(-1, 0)$ from the lower half of the circle the imaginary part of $G_{CAV}$ approaches a negative value. These two values are equal in magnitude but opposite in sign, so in order to preserve the odd symmetry of the imaginary part it is necessary to change the imaginary part of this point to zero. This modification allows the impulse response to be real, but it still leaves a sudden discontinuity in the imaginary part at $(-1, 0)$.

To obtain an impulse response, we must find the inverse Fourier transform of our discrete-time version of $G_{CAV}$. Noting that the inverse Fourier transform algorithm is essen-
tially identical to the Fourier transform algorithm aside from a scale factor and sign change (Oppenheim et al., 1999, p. 561), we can imagine the inverse Fourier transform algorithm trying to represent the discontinuity in the imaginary part at (-1, 0) (in the frequency domain) using a series of smoothly varying sinusoids. To represent a sudden change (in the frequency domain) one would expect to need high-frequency sinusoids, because only they change fast enough to represent a fast change. The artifacts present at high time points of the impulse responses are the time-domain result of the high-frequency sinusoids necessary to represent (in the frequency domain) the sudden discontinuity in the imaginary part.

Since the artifacts do not belong to the behavior of the continuous-time model that we are attempting to implement, and since they would interfere with the convolution operation, I removed most of them by cutting the 1024 point impulse responses in half, to 512 points each. By the 512th sample, the impulse response amplitudes get very small relative to the maximum amplitudes, so it did not seem that truncation would remove much useful information from the impulse responses.

To test the degree to which truncation affects the frequency response, I took the \texttt{fft} (fast Fourier transform) of the truncated impulse responses (zero-padded to be 1024 points long) and compared the results to the original linear plots taken directly from the continuous-time model. See Figures 3.9 and 3.10 for linear plot comparisons of $G_{CAV}$ before and after truncation for the bobcat and cheetah. In magnitude, the truncated version closely follows the behavior of the original filter until above 7000 Hz, at which point it rises to roughly 1.5 dB above the original curve. The phase is only affected substantially within the deepest part of the dip, a small interval. Given that most vocalizations contain very little content in the affected range above 7000 Hz (see Part I), that a change of 1.5 dB is rather small, and that the phase is only affected substantially for a very narrow frequency-
range (which is being attenuated heavily anyhow), I decided truncation was an acceptable way of removing the unwanted artifacts from the impulse responses.

**Figure 3.9: Left Bobcat ME: Comparison Between Original and Truncated Versions**

Linear plots of left bobcat G\textsubscript{CAV} implementations with and without truncation of the impulse response. The impulse response was truncated in order to remove artifacts that occurred at high time points as a result of a discontinuity in the imaginary part between the maximum positive and maximum negative frequencies. Truncation affects the magnitude primarily above 7000 Hz, above which the largest effect is around 1.5 dB. Truncation affects the phase primarily in the close vicinity of the dip. Both of these effects were deemed acceptable, and consequently truncated versions of the impulse responses were used for filtering vocalizations.
Figure 3.10: Left Cheetah ME: Comparison Between Original and Truncated Versions

Linear plots of left cheetah $G_{CAV}$ implementations with and without truncation of the impulse response. See the caption to Figure 3.9 for further discussion of truncation’s effects on these plots.

3.4 Examination of Filters and Comparisons to Theory

This section looks at the properties of the implemented filters in comparison to the theoretical calculations of Section 3.2.

Predictions from theory and measurements of the $G_{CAV}$ implementations are shown and compared in Table 3.4. DC gain, dip center frequency, dip bandwidth, dip $Q$, and high frequency asymptotic behavior are presented for both the theoretical predictions and the implementations. The % difference between theory and implementation is given as well, calculated as:
\[ \% \text{ Difference} = 100 \times \frac{\text{Implementation} - \text{Theory}}{\text{Theory}}. \]  
(3.9)

<table>
<thead>
<tr>
<th></th>
<th>BCL</th>
<th>BCR</th>
<th>CHL</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC Gain (dB), THEORY:</td>
<td>-3.194</td>
<td>-3.119</td>
<td>-0.912</td>
<td>-0.991</td>
</tr>
<tr>
<td>IMPLEMENTATION:</td>
<td>-3.192</td>
<td>-3.118</td>
<td>-0.908</td>
<td>-0.990</td>
</tr>
<tr>
<td>% Difference:</td>
<td>-0.06</td>
<td>-0.03</td>
<td>-0.44</td>
<td>-0.10</td>
</tr>
<tr>
<td>Dip Center Freq. (Hz), THEORY:</td>
<td>2783</td>
<td>2900</td>
<td>2049</td>
<td>2103</td>
</tr>
<tr>
<td>IMPLEMENTATION:</td>
<td>2914</td>
<td>3040</td>
<td>2126</td>
<td>2189</td>
</tr>
<tr>
<td>% Difference:</td>
<td>4.7</td>
<td>4.8</td>
<td>3.8</td>
<td>4.1</td>
</tr>
<tr>
<td>Dip Bandwidth (Hz), THEORY:</td>
<td>259</td>
<td>264</td>
<td>157</td>
<td>159</td>
</tr>
<tr>
<td>IMPLEMENTATION:</td>
<td>266</td>
<td>261</td>
<td>169</td>
<td>171</td>
</tr>
<tr>
<td>% Difference:</td>
<td>2.7</td>
<td>-1.1</td>
<td>7.6</td>
<td>7.5</td>
</tr>
<tr>
<td>Dip Q, THEORY:</td>
<td>10.7</td>
<td>11.0</td>
<td>13.1</td>
<td>13.2</td>
</tr>
<tr>
<td>IMPLEMENTATION:</td>
<td>11.0</td>
<td>11.6</td>
<td>12.6</td>
<td>12.8</td>
</tr>
<tr>
<td>% Difference:</td>
<td>2.8</td>
<td>5.5</td>
<td>-3.8</td>
<td>-3.0</td>
</tr>
<tr>
<td>High Freq. Asymptote, THEORY:</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IMPLEMENTATION:</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.4: Comparisons of \( G_{CAV} \) Implementation to Theory

Comparison of DC Gain, Dip Center Frequency, Dip Bandwidth, Dip Quality Factor (\( Q = \text{center frequency} / \text{bandwidth} \)), and High Frequency Asymptotic Behavior between the theoretical \( G_{CAV} \) model and the discrete-time implementation. The \% Difference values are computed as:

\[ \% \text{ Difference} = 100 \times \frac{\text{Implementation} - \text{Theory}}{\text{Theory}}. \]

Of the implementations, DC Gain is always within 0.5% of the theory, and the Dip Center Frequencies are within 5% of theory. For the bobcat, the Dip Bandwidths are within 3% of theory, but for cheetah they are within 8% of theory. Bandwidth measurements were estimated from the Bode plots, so some of the error could be due to imprecise measurements. Dip Q was always within 6% of theory. Theory predicted that the magnitude of \( G_{CAV} \) approaches 1 at high frequencies, a behavior that occurs in the implementations as well. Overall the implementations are reasonably consistent with the theoretical predictions. One source of error is the frequency-dependence of \( R_F \) which was not accounted for in the theoretical predictions.
Based on Figures 3.3-3.6 the left and right ears of each species have similar Cavity Gains. It can be seen from Table 3.4 that the theoretical dip frequencies for left and right bobcat ears differ by only 117 Hz (≈4%), 54 Hz for cheetah ears (≈3%). Theoretical low-frequency gain is within 0.1 dB for the left and right ears of both species, and the other structural features are also similar between left and right ears. The implementations exhibit comparable similarities between left and right ears. The fact that the left and right are so similar for the same individual is the reason that I chose only to use the left $G_{CAV}$ filters for filtering the digitized vocalizations (see Chapter 5).

The % Difference rows of Table 3.4 show that, for DC Gain and Center Frequency the implementations are within 5% of the theoretical predictions. The bobcat Dip Bandwidths are within 3% of theory, but for cheetah they are only within 8% of theory. The bandwidth measurements for the implementations were estimated from the Bode plots, so some of the error could be due to lack of precision. Except for BCR, all $Q$ values are within 4% of theory (6% for BCR). The magnitudes of the implemented filters approach 1 at high frequencies, as predicted by theory. All in all, the implemented filters resemble the theoretical models within a tolerance that is acceptable for the purposes of this thesis.

3.5 Modifications of Models
So far we have confined ourselves to the normal, intact middle-ear model shown in Figure 3.1, with element values shown in Table 3.1. The resulting $G_{CAV}$ functions for the four ears are displayed as linear and log-log (Bode) plots in Figures 3.3-3.6.

In this section we explore how the frequency-domain behavior of $G_{CAV}$ is affected by three types of modification to the middle-ear structure. By understanding the effects of such modifications we will attain a better understanding of the relationships between structure and function for felid ears, as well as attain insight that will prove useful to us in
Chapter 6. The modifications that are explored in this section include 1) removal of the bullar cavity and foramen by plugging the foramen, 2) effective enlargement of the tympanic cavity by removing the septum separating it from the bullar cavity, and 3) perturbing the dip center frequency by modifying the cavity and foramen dimensions.

3.5.1 Plugging the Foramen

This modification, which effectively involves removing the foramen and bullar cavity from the middle ear, was implemented simply by replacing $Z_{BC}$ with an open circuit (see Figure 3.11). Figures 3.13-3.16 show linear and Bode plots of the frequency domain effects (dashed line). Note that the dip is absent and that the low-frequency gain is reduced by around 7 dB for bobcat and 3 or 4 dB for cheetah. High-frequency behavior of the lumped parameter model appears not to be affected.

![Figure 3.11: Middle-Ear Model with a Plugged Foramen](image)

Circuit model of a middle ear with a plugged foramen. The foramen and bullar cavity are effectively closed off from the tympanic cavity, signified by the absence of $Z_{BC}$. 

85
3.5.2 Removing the Septum

This modification, which effectively involves replacing the foramen impedance with an short circuit (joining the tympanic and bullar cavities together) was implemented by removing $R_F$ and $M_F$ from the original model (see Figure 3.12). $Z_{BC}$ and $Z_{TC}$ combine to be one capacitor with a larger value. See Figures 3.13-3.16 for the effects of this on the frequency domain (dash-dotted line). Note again that the dips are absent, but that the low-frequency transmission is not reduced as it was with the plugged foramen (a consequence of the larger total cavity volume). High-frequency behavior appears to be unchanged (for the lumped element model), but this actually disagrees with experimental results. Taberner, 1999, showed that removing the septum introduces two sizable dips in the cavity admittance around 10 kHz. This point will be revisited in Chapter 6.

![Circuit model of a middle ear with a removed septum. The foramen is effectively removed and the tympanic and bullar cavities become one, signified by parallel capacitors.](image)

**Figure 3.12:** Middle-Ear Model with Removed Septum
Figure 3.13: Left Bobcat Ear: Linear Plot Comparison of Original to Plugged Foramen and Removed Septum

Linear plot comparing $G_{CAV}$ for the left bobcat ear when: the ear is intact (see Figure 3.1), the foramen is plugged (thus removing the foramen and bullar cavity from the model: see Figure 3.11), and the septum is removed (thus combining the tympanic and bullar cavities into one: see Figure 3.12). Both modifications remove the dip from $G_{CAV}$, but plugging the foramen has the additional effect of reducing the low frequency gain. The high frequency gain approaches 1 in both cases, though experiments have shown that removing the septum introduces two dips close to 10 kHz (see Chapter 6).
Figure 3.14: Left Bobcat Ear: Bode Plot Comparison of Original to Plugged Foramen and Removed Septum

Bode plot comparing $G_{CAV}$ for the left bobcat ear under the same circumstances described in the caption to Figure 3.13.
Figure 3.15: Left Cheetah Ear: Linear Plot Comparison of Original to Plugged Foramen and Removed Septum

Linear plot comparing \( G_{CAV} \) for the left cheetah ear under the same circumstances described in the caption to Figure 3.13.
Figure 3.16: Left Cheetah Ear: Linear Plot Comparison of Original to Plugged Foramen and Removed Septum

Bode plot comparing $G_{CAV}$ for the left cheetah ear under the same circumstances described in the caption to Figure 3.13. Notice low frequency attenuation around 3-4 dB for the plugged foramen configuration.

3.5.3 Sensitivity of $G_{CAV}$ Dip to Variations of Element Values

In the equation for computing dip center frequency (Equation 3.7: $\omega_0 = 1 / \sqrt{[M_F*(C_{TC}*C_{BC}]/(C_{TC} + C_{BC})]}$), it is clear that the foramen mass and cavity compliances are responsible for the dip frequency. The values of the compliances and mass depend on the dimensions of the middle-ear air spaces, so insofar as these dimensions vary, the center frequency will vary. It is possible to examine how much the center fre-
quency is likely to vary given an estimate of how much the dimensions are likely to vary among individuals. Center frequency variability among individuals is an important consideration for the communication hypothesis that will be examined in Chapters 4, 5, and 6.

\( M_F \), the value of an acoustical mass, can be computed as \( \rho_0 l/A \), where \( \rho_0 = 1.18 \text{ kg/m}^3 \), \( l \) is the effective length of the foramen and \( A \) is its cross-sectional area. Changing the length and radius by \((+/-) 10\%\) can cause \( M_F \) to vary between 74\% and 136\% of its original value.

\( C_{TC} \) and \( C_{BC} \), the values of two acoustic compliances, can be computed as \( V_0/\rho_0 c^2 \), where \( V_0 \) is the volume of the given cavity, \( \rho_0 \) is given in the previous paragraph, and \( c \) is 345 m/s. Changing the radius of the volume by \((+/-) 10\%\) can cause each of the compliances to vary between 73\% and 133\% of their original values.

Assuming that middle-ear dimensions only vary within a 10\% range, the variability ranges computed in the previous two paragraphs imply that in a worst case scenario the center frequency can vary between 74\% and 136\% of its original value. Judging from the element values listed in Table 3.1 for the left and right ears of two individuals, most element values vary within 10\% for the left and right ears of a given individual and the center frequencies vary within 5\% of each other. If each relevant dimension varied by 10\%, the variability of the element values could be larger (as shown for the worst case above), but the variations do not always affect the center frequency in the same direction.

Since only one individual from each species is represented in Table 3.1, it is conceivable that larger variation in element values could be measured on the ears of different individuals. For example, in Huang et al., 1997a, a plot is shown for \( G_{CAV} \) of a different bobcat with a dip frequency at 3.5 kHz, 26\% higher than our center frequency for the left bobcat ear.
Basically, since the dip center frequency depends crucially on the physical dimensions of the middle ears of individual cats, one should expect its value to vary from individual to individual, at least within a 25% range.

3.6 Summary
In this chapter the felid middle-ear model was presented, along with a transfer function $G_{CAV}$ (the Cavity Gain) that describes the transmission of sound through the middle ear (Section 3.1). $G_{CAV}$ for each ear was first analyzed from the standpoint of a circuit description (Section 3.2) and later implemented in Matlab as a discrete-time impulse response (Section 3.3). The resulting impulse responses behave in a manner that is sufficiently consistent with theoretical predictions made from the circuit models (Section 3.4). Modifications to the middle-ear structure were examined for their relevance to developing an understanding of the middle-ear models as well as for their relevance to issues that will be discussed in Chapter 6. With no more than 10% variation of middle-ear dimensions, in the worst case the $G_{CAV}$ dip center frequency can vary between 74% and 136% of an original value, so any application of the dip for enhancing communication between individuals of the same species should be able to withstand variations on this order (Section 3.5)
Chapter 4

Hypothesized Role of the Middle Ear in Vocal Communication

In Chapters 1 and 2 bobcat and cheetah vocalizations were studied from the standpoint of the spectral content of their recordings. Several structural features were identified and discussed, such as pitch contour, amount of vocalized noise, duration, harmonic amplitudes, etc. These and other structural features probably contribute in some way to the kinds of percepts that are experienced for these vocalizations.

In Chapter 3 the Felid middle ear was modeled as a filter affecting the transmission of a vocalization signal from the ear canal to the cochlea. Both species’ middle-ear Cavity Gains are notch filters with notch center frequencies around 2900 and 2100 Hz for bobcat and cheetah respectively (see Table 3.4). Consequently, vocalization spectral content in the vicinity of these dips is attenuated by the middle ear.

In this chapter an hypothesis is developed concerning a possible role of the middle ear in Felid communication.

4.1 A Model for the Structure, Perception, and Meaning of Vocalizations

Before we can proceed with the development of an hypothesis involving the relationship between vocalization structure and perception, we need to define what we mean by perception.

In the Vocal Communication Chain depicted in Figure 0.1, a vocalization passes through the outer ear, middle ear, inner ear, and nervous system of the receiver, eventually triggering a “perception” in the receiver’s brain.
At the beginning of this sequence the vocalization signal consists of a travelling sound wave with a well-defined physical structure that can be recorded and analyzed spectrally. Well-defined measurements such as duration, maximum F0 frequency, pitch contour, etc. can then be made and organized into a space of structural measurements (see Figure 4.1).

Once the vocalization signal (represented as a sound wave) enters the ear canal, it causes mechanical motion of the tympanic membrane and ossicles, which is impeded by the acoustic structure of the middle ear and the mechanical properties of the tympanic membrane, ossicles, and cochlea. The ossicles, then, mechanically transfer a filtered version of the original signal to the cochlea.

The filtering properties of the Felid middle ear are also well-defined (see Chapter 3), and can be simulated computationally. After the vocalization signal enters the cochlea, it is transformed into a spatial array of neural signals (Kiang, 1965) that lead to a perception in the brain of the receiver. This final transformation is not well-defined for cats since our knowledge of the perceptual world of cats is nowhere near complete. However, one can assume that there is a space of perceptions that a cat experiences given particular sounds entering its ear (see Figure 4.1). In some cases different points in the structural measurement space may map to different points in the perception space, indicating that the structural differences between the two points trigger distinct perceptual experiences. (see ‘a’ or ‘b’ as compared to ‘c’ in Figure 4.1). An example of this might be two vocalizations that differ in duration by 1 second. Such a structural difference could easily lead to a perceptual difference, e.g. one vocalization might fall into a “short mew” perceptual category and the other might fall into a “long mew” category. In another case a structural difference may not trigger a perceptual difference (see ‘a’ and ‘b’ in Figure 4.1). An example of this might be two vocalizations that contain different amounts of noise located directly above a prominent harmonic in the spectrograms. If the noise lies within a critical band with the
prominent harmonic, the two vocalizations could produce be perceived as being identical in humans (for a discussion of masking and the critical band in humans, see Chapter 3 of Moore, 1997).

It is possible that the middle-ear transfer function may cause some structural differences in the vocalizations to map to different perceptions (and for other structural differences to map to the same perception). This idea will be the basis of the hypothesis developed in Section 4.2.

To study the transformation taking place inside a cat’s brain during the act of perception, I will assume that from the cochlea onward simple perceptual processes for humans are sufficiently similar to cats (below 8000 Hz) that it is adequate to use human perceptual abilities to study the possible perceptual influence of the middle ear (see Chapter 5).

Finally, once a vocalization is perceived, i.e. processed neurally to produce a response in perception space, an assortment of factors may map that perception to a point in a “meaning” space. Such a mapping could presumably be studied with behavioral experiments. The mapping would probably depend on factors such as the physical and mental state of the receiver, the receiver’s past experiences, instincts, relationship to the sender (e.g. a call by a kitten’s mother might lead the kitten to come closer, whereas a call by an unknown adult might lead the kitten to move further away), etc. One could imagine that different perceptions could in some cases map, to first order, to the same point in the meaning space (see ‘d’ and ‘e’ in Figure 4.1). For instance, vocalizations of the same type may be perceived as different vocalizations, yet may convey the same message. In other cases one could imagine that vocalizations that are perceived as different may map to different points in the meaning space (see ‘d’ or ‘e’ and ‘f’ in Figure 4.1). For instance, perhaps a short growl in a given context may mean something different to the receiver than a prolonged tonal mew in the same context.
Figure 4.1: A Model of Felid Perception of Vocalizations

Model of the mapping of Felid vocalizations with different structures (points in the “Structure Space”) to different percepts as experienced by a given cat (points in the “Percept Space”), to different possible meanings understood by a given cat in a given context (points in the “Meaning Space”). It is possible for vocalizations differing by some structural feature to nonetheless map to the same percept (a and b). Or, a vocalization differing from a and b by other structural features could map to a different percept (c). Similarly for the mapping of percepts to meanings, it is possible for vocalizations that map to different percepts to nonetheless be interpreted by a given cat in a given context as having the same meaning (d and e). Or, it is also possible for vocalizations with different percepts to map to different meanings (compare f to d or e).

4.2 Hypothesis of the Middle Ear’s Role in Felid Vocal Communication

We will restrict our attention to the role of the middle ear in affecting the mapping between structure and perception. Since the $G_{CAV}$ filters are notch filters, the most salient feature of the middle-ear transfer function is that it attenuates a narrow range of frequencies. There are a few ways that one might imagine such a filter could affect the perception of a vocalization, two of which are outlined below.

4.2.1 Modifying the Fundamental Frequency for Vocalizations with Flat Pitch Contours

If a vocalization has harmonic structure and the fundamental frequency (F0) does not change much with time (over a sizeable interval), the harmonics would remain at essentially the same frequencies over the course of the “constant” interval of the vocalization.
and would occur at integer multiples of the fundamental frequency. If F0 were different for another vocalization, all of the harmonic frequencies would differ.

The notch frequency for a given ear remains fixed and the notch width is generally small relative to F0, so one could pick two vocalizations with two F0 values such that in one case a harmonic is attenuated by the dip because it occurs at the dip frequency, whereas in another case with no harmonic in the dip none is significantly attenuated.

It is conceivable that attenuation of a harmonic due to the middle-ear notch may have a perceptible effect. Consider, for instance, the fact that humans categorize perceived vowel sounds based on the frequencies of formant peaks that are visible in spectral representations of vowels (see Chapter 6 of Stevens, 1998). Alteration of a formant frequency often causes listeners to perceive a different vowel, a definite change in percept category. Since the fundamental frequencies for bobcat and cheetah vocalizations are typically much higher than those of humans, the harmonic spacing is considerably wider (in frequency). Consequently, the envelope of harmonic amplitudes does not clearly indicate the “formants” in the spectra. In such a case, attenuating a harmonic might be likened to shifting a formant in frequency. If this modification has the consequence of changing the perceived vowel in humans, then one could imagine that cats might also perceive a clear difference (see Darwin, 1984 and Hienz et al., 1996).

The hypothesis thus states that, by changing F0 one could conceivably hear definite perceptual differences after passing through the middle-ear transfer function, even if the vocalizations were originally identical except for F0. Figure 4.2 contains an illustration of the hypothesis for two vocalizations.

4.2.2 Changing F0 with Time

Another case in which one might expect the middle-ear filter to cause perceptual differences in vocalizations occurs when F0 varies with time so that the amplitude of a har-
monic changes by a large amount (relative to the dip width) as it passes in and out of the dip. Such amplitude changes could be perceptible, especially if the attenuated harmonic is initially of high amplitude relative to other harmonics. The dip would effectively act like a frequency-modulation (FM) demodulator, producing large amplitude variations in a harmonic as a result of relatively small frequency variations. As in Section 4.2.1, then, the hypothesis states that the middle ear may be capable of transforming vocalizations that do not sound appreciably different to begin with (but contain different types of F0 variation) into vocalizations that do sound appreciably different. See Figure 4.2 for a diagrammatic representation of the hypothesis.

![Diagram](image-url)

**Figure 4.2: Hypothetical Role of the Middle Ear in Vocal Communication**

Diagram illustrating how filtering through the middle ear could affect vocalizations differently. Two unfiltered vocalizations, $A_u$ and $B_u$ initially sound similar, but not identical. Perhaps they differ moderately in some structural feature such as fundamental frequency (4.2.1) or pitch contour (4.2.2). The filtered versions of the vocalizations, $A_f$ and $B_f$, sound different from one another, due to the notch in the transfer function. Vocalization $B$ contains a prominent harmonic in the vicinity of the notch and is affected significantly by the filtering, but vocalization $A$ does not have a harmonic in the vicinity of the dip and consequently is not affected significantly by the dip. In this manner, the hypothesis predicts that the two vocalizations can be made to sound more distinguishable as a result of filtering.
4.2.3 Extension of the Hypotheses to the Full Communication Chain

In the case that the sender in Figure 0.1 has a middle-ear transfer function very similar to the receiver's, both cats would hear the same effect on the vocalizations due to the middle ear. Therefore, it would be possible for the sender to use its own perception to create vocalizations that sound different as a result of the properties of the middle ear. It is also conceivable that these middle-ear properties could be exploited to create vocalizations with different mappings to the meaning space.

The possible utility of the middle ear for enhancing vocal communication in the manner described above depends both on the variability of the center frequency of the middle-ear notch among individuals of the same species, and the salience of its effects on the perception of actual vocalizations. The next section outlines a methodology that will be used in Chapter 5 to test the hypothesis that the middle-ear transfer function enhances vocal communication in felids.

4.3 Proposed Methodology for Testing the Hypothesis

The hypotheses of Section 4.2 concern the effects of the middle-ear transfer function on the mapping of vocalization structure to perception. Felid perception is difficult to study directly, so I use human subjects in the perceptual experiments described below. See Chapters 5 and 6 for a detailed application of the following methodology.

1. Generate two groups of vocalizations: filtered and unfiltered (referred to as “pre” and “post” respectively). Generate the filtered group by using the middle-ear impulse responses of the appropriate species (see Chapter 3) to filter the digitized vocalizations described in Chapters 1 and 2.

2. Listen to the vocalizations in the two groups, comparing the pre and post versions of each individual vocalization and assessing the salience of the perceived differences.
3. Conduct formal listening tests with multiple human subjects to test the preliminary assessments of step 2 above that compare vocalizations in the pre and post groups.

4. If the results of step 3 indicate that filtering affects perception in a significant way, search for structural correlates to the results of step 3 and conduct additional listening tests to determine how filtering affects the perceived contrast between vocalizations in the same group.
Chapter 5

Tests of the Perceptual Hypotheses
In this chapter I attempt to implement the testing strategy described in Section 4.3 as a method of testing the hypotheses in Section 4.2. I begin by filtering each collection of vocalizations with the appropriate impulse responses generated in Chapter 3 and listening to the vocalizations before and after filtering to gauge the feasibility and usefulness of conducting quantitative listening tests. I follow the informal preliminary listening with a quantitative listening test comparing vocalizations from the CH-III collection before and after filtering, and attempt to find structural features correlated with the results. Finally, I conduct a test comparing filtered and unfiltered background noise to filtered and unfiltered vocalizations to test whether vocalization structure is likely to be more responsible for listening test results than the background noise present in the recordings.

5.1 Filtering and Preliminary Listening
In this section I describe the methods I used to filter and listen to the vocalizations, the observations and conclusions I made after comparing filtered (post) and unfiltered (pre) versions of each vocalization, and the role that my preliminary listening observations played on the design of the formal listening tests described in Section 5.2.

5.1.1 Filtering
Since the $G_{CAV}$ impulse responses were implemented in Matlab (see Chapter 3) and the vocalizations were stored in the “Klatt .wav” format (see Chapters 1 and 2), I wrote a Matlab script (m-file) to import the digitized vocalizations into Matlab, convolve them with the appropriate impulse responses (from the same species), export them to the Klatt .wav format, and save them as files with new names. The m-file is called “filtercats.m” and can be viewed on pages 228-229 of Appendix 4. This file makes use of two scripts,
kl2mat.m and mat2kl.m, which are available and operational at the RLE Speech Communication Lab (see pp. 229-230 of Appendix 4). Mat2kl.m requires a program called raw2kl (also available at the Speech Communication Lab) in order to work. To filter the vocalizations I simply ran the filtercats.m script in Matlab, specifying inside the script the names of the files to be filtered. Since both the unfiltered vocalizations (referred to as “pre”) and filtered vocalizations (“post”) are saved as Klatt .wav files, both can be analyzed with xkl, a useful program for analyzing and listening to vocalizations (described briefly in Chapter 1).

5.1.2 Listening
After filtering the vocalizations I opened two xkl windows, one for a “post” version and one for a “pre” version of a given collection. I used xkl to listen to the pre and post versions of each vocalization several times, and noted to the best of my ability how different they sounded. I also tried modifying the playback volume between pre and post to assess the effects of absolute amplitude on my perception of pre versus post differences. I made these comparisons for all vocalizations and repeated the whole procedure once more on a different day to see how well my judgments agreed from one day to another. Additionally, I listened to the pre and post versions of each collection in full from start to finish and tried to see how vocalizations compared to one another within the pre version versus how they compared to one another in the post version.

5.1.3 Classification Scheme for Perceptual Judgments in Preliminary Listening
On the whole the differences I perceived between pre and post vocalizations were subtle. In many cases the pre and post versions were practically indistinguishable. Low pre vs. post distinguishability was evident for the majority of vocalizations in BC-IV and V, as well as for CH-I. I use the “I” symbol to denote vocalizations for which I found the pre and post versions to be essentially “Indistinguishable” from each another.
For some vocalizations I was able to hear a slight difference between the pre and post versions. I denote these vocalizations with an “M” symbol to indicate that I perceived them as sounding “Mildly different”.

I use a third symbol, “N” (for “Noticeably different”) to denote vocalizations for which the difference before and after filtering was considerably more noticeable than vocalizations denoted “M”, but still not enough to be called salient. Very few vocalizations fit this category, and even when they did, the differences were generally still subtle, just somewhat more noticeable to my ear than the “mild” differences.

A final symbol, “S” (for “Salient”) was intended for use in labeling vocalizations yielding prominent, obvious differences between the filtered and unfiltered versions. As will be shown in the next section, none of the vocalizations in the 8 collections fit this category.

In the cases of “M” and “N” ratings (“mildly” and “noticeably” different respectively) I also tested whether or not changing the playback volume between listening to the “pre” and “post” versions caused changes in my judgments. The purpose of making these modifications was to test whether the perceived pre-post differences were due to changes in the absolute amplitude caused by filtering or whether the perceived differences were due to relative amplitude changes among the different frequency bands of the vocalizations. When the difference was still noticeable after random amplitude variations I wrote “L+”, meaning the difference withstood changes in level, and “L-” meaning it did not withstand level changes, and instead became a less prominent rating.

See Table 5.1 for a summary of the classification symbols used to qualitatively summarize the effects of filtering on vocalizations.
Symbols | Meanings
--- | ---
I | Pre and post versions are Indistinguishable.
M | Pre and post versions are Mildly distinguishable.
N | Pre and post versions are Noticeably distinguishable.
S | Pre and post versions are Saliently distinguishable.
L+ | Random changes in playback level do not affect the given perceptual judgment (M, N, or S).
L- | Random changes in playback level affect the given perceptual judgment, reducing it to a subtler judgment (M -> I; N -> M or I; and S -> N, M, or I).

Table 5.1: Summary of Symbols used to Classify Perceived Pre-Post Differences

The symbols I, M, N, and S represent qualitative judgments of the perceived difference between unfiltered and filtered (pre and post) versions of a vocalization, increasing in prominence from “I” to “S”.

For a vocalization to rank as Indistinguishable (“I”), the pre vs. post difference needs to be so subtle that, unless special efforts are made (such as extra careful attention and repeated listenings), the vocalizations sound the same before and after filtering.

To rank as Mildly distinguishable (“M”), the pre vs. post difference of a vocalization needs also to be subtle, but to be more prominent than the Indistinguishable vocalizations, such that special effort may still be necessary to hear the difference, but that it is considerably easier to hear nonetheless.

To rank as Noticeably distinguishable (“N”), the pre vs. post difference should be considerably more prominent than than the Mildly distinguishable vocalizations such that the difference can be heard without special effort.

To rank as Saliently different (“S”), the pre vs. post difference should be considerably more apparent and obvious than Noticeably distinguishable vocalizations, and be possible to hear without any special effort.

The additional symbols “L+” and “L-” refer to how a given perceptual judgment (I, M, N, or S) is affected by randomly changing the playback level of the pre and/or post versions when listening to one and then the other. “L+” states that the perceptual judgment remains the same even with such changes in playback level. “L-” states that the perceptual judgment changes to a less prominent one as a result of changes in playback level.
5.1.4 Results: Perceptual Judgments for Preliminary Listening

Preliminary perceptual judgments were made and recorded on two occasions separated in time by more than one week. I was the only person to make these judgments and I used high quality headphones for listening. I listened to pre and post versions of each vocalization alternately, and chose to assign to the pair the most applicable rating of those described in Table 5.1 (I, M, N, S) according to my own judgment. Once out of the two listening sessions, for ratings of M or higher I additionally listened to the pre and post versions of the vocalizations while making random changes in playback volume as I alternated between the two (I turned the volume dial on the amplifier by a random amount). When my judgments did not change as a result of these changes, I appended an “L+” to the end of the original judgment, and when my judgments changed to a less prominent judgment I appended an “L-”.

Tables 5.2-5.9 contain a complete set of the perceptual judgments that I made for each vocalization in each collection. Two judgments were made for each vocalization, and the second judgment (made more than a week after the first) is shown in parentheses. Table 5.10 contains a summary of Tables 5.2-5.9, displaying the percentage of vocalizations in each collection that were given each type of rating. See Section 5.1.5 for a discussion of these results.
### Table 5.2: Preliminary Perceptual Judgments for the BC-I Collection

Judgments of the perceived difference between filtered and unfiltered versions of each vocalization in the BC-I collection. Two judgments were made for each vocalization, the second of which is in parentheses. For judgments of M, N, or S, an additional rating was used to indicate whether or not the pre vs. post differences could still be heard or whether they became harder to tell apart when the loudness was varied randomly on playback (L+ and L- respectively). Only one such rating was given for each vocalization.

None of the 13 BC-I vocalizations exhibited “salient” pre vs. post differences, and only 1 consistently exhibited “noticeable” differences (number 8). 8 vocalizations exhibited a mild difference at least part of the time, and the remaining 3 vocalizations consistently exhibited “indistinguishable” differences. Only the difference heard for vocalization number 8 could be heard with random changes in playback volume.
Table 5.3: Preliminary Perceptual Judgments for the BC-II Collection

See the first paragraph in the caption to Table 5.2 for an explanation that is also applicable to this table.

None of the 9 BC-II vocalizations exhibited “salient” or “noticeable” pre vs. post differences. 8 vocalizations consistently exhibited a mild difference, and the remaining vocalization consistently exhibited an “indistinguishable” difference. None of the 8 vocalizations with “mild” pre vs. post differences could still be heard after the playback volume was changed randomly.

Table 5.4: Preliminary Perceptual Judgments for the BC-III Collection

See the first paragraph in the caption to Table 5.2 for an explanation that is also applicable to this table.

None of the 8 BC-III vocalizations exhibited “salient” or “noticeable” pre vs. post differences. 6 vocalizations consistently exhibited a “mild” difference and the remaining 2 vocalizations consistently exhibited “indistinguishable” differences. Only the “mild” difference heard for vocalization 6 could be heard with random changes in playback volume.
Table 5.5: Preliminary Perceptual Judgments for the BC-IV Collection

See the first paragraph in the caption to Table 5.2 for an explanation that is also applicable to this table.

All of the 7 BC-IV vocalizations consistently exhibited “indistinguishable” pre vs. post differences.

<table>
<thead>
<tr>
<th>BC-IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voc #</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

Table 5.6: Preliminary Perceptual Judgments for the BC-V Collection

See the first paragraph in the caption to Table 5.2 for an explanation that is also applicable to this table.

None of the 7 BC-V vocalizations exhibited “salient” or “noticeable” pre vs. post differences. 2 vocalizations consistently exhibited a “mild” difference, and the remaining vocalization consistently exhibited an “indistinguishable” difference. Neither of the two “mild” vocalizations featured pre vs. post differences that could still be heard after the playback volume was changed randomly.

<table>
<thead>
<tr>
<th>BC-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voc #</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>
None of the 17 CH-I vocalizations exhibited “salient” or “noticeable” pre vs. post differences. 7 vocalizations consistently exhibited a “mild” difference, and 9 vocalization consistently exhibited an “indistinguishable” difference. 5 of the “mild” vocalizations featured pre vs. post differences that could still be heard after the playback volume was changed randomly.

### Table 5.7: Preliminary Perceptual Judgments for the CH-I Collection

See the first paragraph in the caption to Table 5.2 for an explanation that is also applicable to this table.

<table>
<thead>
<tr>
<th>Voc #</th>
<th>Ratings</th>
<th>Voc #</th>
<th>Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M L- (M)</td>
<td>10</td>
<td>I (I)</td>
</tr>
<tr>
<td>2</td>
<td>I (I)</td>
<td>11</td>
<td>M L+ (M)</td>
</tr>
<tr>
<td>3</td>
<td>I (I)</td>
<td>12</td>
<td>I (I)</td>
</tr>
<tr>
<td>4</td>
<td>I (M L-)</td>
<td>13</td>
<td>M L- (M)</td>
</tr>
<tr>
<td>5</td>
<td>I (I)</td>
<td>14</td>
<td>I (I)</td>
</tr>
<tr>
<td>6</td>
<td>I (I)</td>
<td>15</td>
<td>M L+ (M)</td>
</tr>
<tr>
<td>7</td>
<td>I (I)</td>
<td>16</td>
<td>M L+ (M)</td>
</tr>
<tr>
<td>8</td>
<td>I (I)</td>
<td>17</td>
<td>M L+ (M)</td>
</tr>
<tr>
<td>9</td>
<td>M L+ (M)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.8: Preliminary Perceptual Judgments for the CH-II Collection

See the first paragraph in the caption to Table 5.2 for an explanation that is also applicable to this table.

None of the 12 CH-II vocalizations exhibited “salient” pre vs. post differences. 6 vocalizations consistently exhibited a “mild” difference, and 5 vocalizations consistently exhibited an “indistinguishable” difference. 1 vocalization bordered between “noticeable” and “mild”. All of the “mild” differences could still be heard after randomly changing the playback volume.
<table>
<thead>
<tr>
<th>CH-III</th>
<th></th>
<th>Voc #</th>
<th>Ratings</th>
<th>Voc #</th>
<th>Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>M L- (M)</td>
<td>16</td>
<td>M L- (M)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>M L- (M)</td>
<td>17</td>
<td>I (N L+)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>I (M L-)</td>
<td>18</td>
<td>I (M L+)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>M L- (M)</td>
<td>19</td>
<td>M L- (M)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>M L- (M)</td>
<td>20</td>
<td>M L- (M)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>I (I)</td>
<td>21</td>
<td>I (M L-)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>I (I)</td>
<td>22</td>
<td>I (M L-)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>I (I)</td>
<td>23</td>
<td>I (M L-)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>M L+ (I)</td>
<td>24</td>
<td>N L+ (N)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>I (M L-)</td>
<td>25</td>
<td>N L+ (N)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>M L+ (M)</td>
<td>26</td>
<td>I (I)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>M L+ (M)</td>
<td>27</td>
<td>M L- (M)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13</td>
<td>M L+ (N)</td>
<td>28</td>
<td>M L+ (N)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14</td>
<td>M L+ (N)</td>
<td>29</td>
<td>I (M L-)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>M L- (M)</td>
<td>30</td>
<td>I (M L-)</td>
</tr>
</tbody>
</table>

Table 5.9: Preliminary Perceptual Judgments for the CH-III Collection

See the first paragraph in the caption to Table 5.2 for an explanation that is also applicable to this table.

None of the 30 CH-III vocalizations exhibited “salient” pre vs. post differences. 2 vocalizations were consistently rated “noticeable” and 3 were rated between “mild” and “noticeable”. 11 vocalizations consistently exhibited a “mild” difference, 10 were rated between “indistinguishable” and “mild”, and 4 vocalizations consistently exhibited an “indistinguishable” difference. 1 vocalization was rated as “noticeable” the first time and “indistinguishable” the second time. 10 vocalizations rated as “mild” or “noticeable” featured differences that could still be heard after randomly changing the playback volume.
Table 5.10: Summary of Perceptual Judgments for all 8 Vocalization Collections

<table>
<thead>
<tr>
<th>Judgment</th>
<th>BC-I</th>
<th>BC-II</th>
<th>BC-III</th>
<th>BC-IV</th>
<th>BC-V</th>
<th>CH-I</th>
<th>CH-II</th>
<th>CH-III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistently ‘I’</td>
<td>31%</td>
<td>11%</td>
<td>25%</td>
<td>100%</td>
<td>71%</td>
<td>53%</td>
<td>42%</td>
<td>13%</td>
</tr>
<tr>
<td>‘I’ / ‘M’</td>
<td>15%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>6%</td>
<td>0%</td>
<td>30%</td>
</tr>
<tr>
<td>Consistently ‘M’</td>
<td>31%</td>
<td>89%</td>
<td>75%</td>
<td>0%</td>
<td>29%</td>
<td>41%</td>
<td>50%</td>
<td>37%</td>
</tr>
<tr>
<td>(‘I’ or ‘M’) / ‘N’</td>
<td>15%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>8%</td>
<td>13%</td>
<td></td>
</tr>
<tr>
<td>Consistently ‘N’</td>
<td>8%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>7%</td>
<td></td>
</tr>
<tr>
<td>‘N’ / ‘S’</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Consistently ‘S’</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>‘L+’</td>
<td>8%</td>
<td>0%</td>
<td>13%</td>
<td>0%</td>
<td>0%</td>
<td>29%</td>
<td>58%</td>
<td>33%</td>
</tr>
<tr>
<td>‘L-’</td>
<td>62%</td>
<td>89%</td>
<td>87%</td>
<td>0%</td>
<td>29%</td>
<td>18%</td>
<td>0%</td>
<td>53%</td>
</tr>
</tbody>
</table>

Perceptual judgments are summarized for the 8 vocalization collections in terms of the percentage of vocalizations within a given collection that were assigned a particular rating. Each vocalization was rated twice, on two separate occasions, using the symbols ‘I’, ‘M’, ‘N’, ‘S’, ‘L+’, and ‘L-’ (see Table 5.1). Consistently ‘I’ means both of the two judgments were ‘I’, ‘I’ / ‘M’ means one of the two judgments was ‘I’ and the other was ‘M’, Consistently ‘M’ means that both of the two judgments were ‘M’, (‘I’ or ‘M’) / ‘N’ means that one of the two judgments was either ‘I’ or ‘M’ and the other was ‘N’, Consistently ‘N’ means that both of the judgments were ‘N’, ‘N’ / ‘S’ means that one of the two judgments was ‘N’ and the other was ‘S’, and Consistently ‘S’ means that both of the judgments were ‘S’. The ‘L+’ and ‘L-’ ratings are included in the rows with the headings of the same name.

None of the vocalizations was given a “salient” rating, and only BC-I and CH-III contained any vocalizations consistently rated as “noticeable”. The vast majority of vocalizations, then, were rated as having “mild” and/or “indistinguishable” pre vs. post differences. Aside from CH-II, most of time it was harder to hear pre vs. post differences when random playback volume changes were made. All percentages are out of the total number of vocalizations in a collection, and nonzero numbers are shown in boldface for emphasis.
5.1.5 Discussion of Preliminary Listening Results

1. Lack of “Salient” Judgments

The most notable finding from the informal preliminary listening tests was that none of the pre vs. post differences was “saliently” noticeable. All of the differences I heard during the tests were subtle, and much of the time they were very subtle. It was necessary for me to pay careful attention and alternately listen to the pre and post versions many times in order to hear the “mild” differences, and the few “noticeable” differences I heard still required effort to hear. To contrast with this state of affairs, I imagine a “salient” pre vs. post difference to be: two sounds that are readily and recognizably different sounding, in a manner that does not require any special effort or circumstances to hear. A simple example of this is two different vowel sounds: it is easy to readily recognize a vowel at first hearing in a variety of different circumstances. The bobcat and cheetah vocalizations exhibit pre vs. post differences which rank far below “salient” in the above sense of the word.

2. Possible Structural Reason for Lack of “Salient” Judgments

One potential reason that none of the vocalizations exhibited “salient” pre vs. post differences is that the center frequencies of the filter dips occur at frequencies that are higher than most of the prominent harmonic content of the vocalizations. In bobcat the dip center frequency occurs at nearly 2900 Hz (see Table 3.4), whereas the frequency of the lowest harmonic of the bobcat vocalizations never exceeds 1450 Hz (see Table 1.2). Therefore the first and second harmonics, which often feature the highest amplitudes (see Chapters 1 and 2), never pass through the dip and are consequently never substantially affected by filtering. In cheetah the dip center frequency occurs around 2100 Hz (see Table 3.4), whereas the frequency of the lowest harmonic of the cheetah vocalizations never exceeds
Therefore, as in bobcat, the first and second harmonics never pass through the dip, and so are never substantially affected by filtering.

For both bobcat and cheetah only the third or higher harmonics would generally occur in the vicinity of the dip, leaving the prominent first and second harmonics basically unaffected. To test whether or not attenuating the first or second harmonics could lead to “salient” pre vs. post differences I generated filters with the dip center frequencies halved (created by quadrupling $M_F$) and compared a few vocalizations before and after filtering. The differences I heard were at the least easily “noticeable” and often what I would call “salient”, indicating that salient pre vs. post differences could occur if the dip frequencies were lower.

3. Compensating for the Perception of Absolute Amplitude Changes Due to Filtering

While I was listening it was not clear whether I was primarily hearing the effects of the filter on the vocalization or the effects of the filter on the background noise. In some collections (such as BC-II and III) the background noise was loudly audible when listening to vocalizations, and its sound quality often changed in a noticeable way after filtering. In many cases the perceived loudness of the background noise decreased slightly after filtering, and I thought this might account for some of the mild differences I heard. In a simple attempt to reduce my perception of background noise loudness changes, I varied the playback volume as described in Section 5.1.4, using the symbols “L+” and “L-” to summarize the results.

Among the bobcat vocalizations most of the ratings were ‘L-’, whereas for cheetah vocalizations there was a higher percentage of ‘L+’ ratings, especially in the case of CH-II. Given that very little vocalization content appeared near 2 kHz for CH-II, perhaps either the heavy background noise was shaped by the 2 kHz notch in a manner that was more noticeable than the shaping taking place for the bobcat vocalizations, or perhaps the
peak in the cheetah filter around 1700 Hz (see Figure 3.5) affected my perception of these vocalizations in a manner that was not eliminated by changing the playback volume.

4. Comparisons Between Vocalizations Within Each Collection

The diagram in Figure 4.3 shows a scenario in which two unfiltered vocalizations sound nearly indistinguishable, but for which their filtered counterparts sound saliently different. The middle ear, then, might be viewed as a transformation that makes similar-sounding vocalizations in the pre collection sound saliently different in the post collection. During my preliminary listening I attempted to estimate whether the vocalizations in the pre group sound substantially different upon direct comparison, especially when they belong to different collections. Even in cases where vocalizations belong to the same collection they differ enough in duration, pitch contour, harmonic amplitudes, etc. that they are easily distinguishable when listened to side by side. Given that filtering does not produce “salient” pre vs. post differences for any of the bobcat or cheetah vocalizations, the filtered versions of the vocalizations are not likely to sound any more different after filtering than before. Consequently, my preliminary listening suggests that the scenario of Figure 4.3 does not hold true for these vocalizations and filters, and consequently that formal listening tests are not likely to generate results in support of the hypotheses developed in Chapter 4.

5.1.6 Conclusions and Further Work

Based on my preliminary listening I come to the following tentative conclusions, subject to formal testing:

1. There are no “salient” differences in vocalization sound quality due to filtering with the middle-ear transfer function.
2. The “noticeable” differences are subtle, and require careful listening to hear. Of the 8 collections, the most consistently “noticeable” pre vs. post differences occur in the CH-III collection.

3. Introducing a simple distraction to listening, such as randomly varying the playback level, can make it significantly harder to hear pre vs. post differences, especially for bobcat vocalizations.

4. The likelihood is small that formal listening tests using the available filters and vocalizations will generate support for the hypothesis developed in Chapter 4, given the lack of “salient” pre vs. post differences and the lack of nearly identical sounding unfiltered vocalizations.

5. To formally test the suitability of the available vocalizations and filters to the hypothesis in Chapter 4 I will begin by quantitatively testing the perception of pre vs. post differences for multiple listeners under semi-controlled conditions (see Section 5.2).

5.2 Quantitative Perceptual Tests
The preliminary listening tests described in Section 5.1 show that a single individual who is familiar with the vocalizations and their sound (myself) found that the filtered and unfiltered versions of the vocalizations are usually only mildly different sounding. The preliminary tests, however, were neither formal nor quantitative, two shortcomings that are addressed by the formal, quantitative listening tests of this section.

5.2.1 Methods
Two types of listening tests were performed on four subjects, an “identification” test and a “discrimination”, or “ABX” test. Both test a subject’s ability to tell apart pre and post versions of vocalizations, but differ by the manner in which the individual trials are presented. Six vocalizations were selected for the tests (their selection is described below), and are referred to as “tokens” 1-6.
The identification test was conducted on one subject at a time, for each of the 6 tokens. Beginning with token 1 the subject was allowed to hear pre and post versions of the token upon request as many times as he or she wished. This practice period typically lasted less than 5 minutes. The subject was not allowed to view spectrograms of the tokens. After the practice period the subject was given ten presentations of either pre or post, selected randomly on each presentation, and asked to identify each as either pre or post. Each presentation was given only once and no feedback was given during the test until all six tokens were completed. After the first token, an identical procedure was repeated for each additional token, and identical procedures were applied to the remaining three subjects.

The other type of test, the “discrimination” test (also known as an “ABX” test) is conducted in the same manner as the identification test except for the way in which each of the 10 trials is presented. Instead of presenting a pre or post version alone and expecting the listener to correctly identify it, each trial is presented as: “pre, post, X”, where X is randomly chosen to be either pre or post (the “ABX” name refers to this structure in each trial). This pattern of presentation enables the subject to hear the two choices immediately before hearing the unknown vocalization. Such a test does not require a subject to maintain any memory of the two choices, unlike the identification test, but instead requiring the subject to directly discriminate the two choices. The discrimination test tends to be considerably easier than the identification test. After testing the 6 tokens, the tests were repeated for the remaining three subjects. The four subjects who participated in the tests are referred to as D, J, P, and A. P and J are male, and D and A are female.

All six tokens, listed in Table 5.11, were taken from the CH-III collection. The CH-III collection was chosen because it had the greatest selection of vocalizations with different preliminary ratings. I chose vocalizations that were given different ratings, in an effort to
see whether my preliminary judgments bear any resemblance to the quantitative results of listening tests.

<table>
<thead>
<tr>
<th>Token #</th>
<th>Vocalization</th>
<th>Initial Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CH-III-3</td>
<td>I (M L-)</td>
</tr>
<tr>
<td>2</td>
<td>CH-III-7</td>
<td>I (I)</td>
</tr>
<tr>
<td>3</td>
<td>CH-III-12</td>
<td>M L+ (M)</td>
</tr>
<tr>
<td>4</td>
<td>CH-III-14</td>
<td>M L+ (N)</td>
</tr>
<tr>
<td>5</td>
<td>CH-III-21</td>
<td>I (M L-)</td>
</tr>
<tr>
<td>6</td>
<td>CH-III-24</td>
<td>N L+ (N)</td>
</tr>
</tbody>
</table>

Table 5.11: Listening Test Tokens

Listing of the 6 tokens used in the identification and discrimination (ABX) listening tests. All were taken from the CH-III collection in an effort to use tokens with a variety of different ratings. Three tokens were each rated consistently as I, M, and N. Two were rated as I and M, and one was rated as M and N. See Appendix 2 for spectrograms of the tokens prior to filtering.

5.2.2 Results

The results of the identification test are shown in Table 5.12 and the results of the discrimination test are shown in Table 5.13. In both the identification and discrimination tests, each token was presented in a total of 40 trials (10 per subject) and each subject participated in 60 trials (10 per token). Raw scores are given for the number of correct responses out of 10 for each token and subject, as well as cumulative results of individual tokens and subjects. The scores for the discrimination test, in Table 5.13, are displayed in the same format as Table 5.12.

The results in Tables 5.12 and 5.13 are plotted in several ways on the following pages. Figure 5.1 shows the average percent correct scores for each token (averaged over the four subjects), with error bars representing the standard deviation of the four scores that were combined for the total mean % correct. The overall mean, averaged over all 6 tokens, was 56.7% for the identification test and 74.2% for the discrimination test.
Figure 5.2 contains bar charts for each token, showing how each subject scored on a given token, as well as the average and standard deviation of the scores of the four subjects (both of which were shown in Figure 5.1).

Figure 5.3 contains a line graph for each subject, showing his or her performance on the 6 tokens in both tests.

<table>
<thead>
<tr>
<th>Subject / Token #</th>
<th># of Correct Responses (out of 10 per trial)</th>
<th>Overall Scores (by Token)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D</td>
<td>J</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Mean % Correct</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>9.6</td>
</tr>
</tbody>
</table>

**Table 5.12: Identification Test Results**

Table of identification test results. The six rows beneath the subject headings (D, J, P, and A) show the raw scores (out of 10) for each subject and token. The total number correct for each token (out of 40) is shown in the 6th column, with the mean % correct and standard deviations (of the % correct numbers) in the 7th and 8th columns. The bottom 3 rows contain totals, mean % correct, and standard deviations for each subject (each based on the 60 trials). The overall mean % correct for all tokens and all subjects is shown in the bottom right corner.
<table>
<thead>
<tr>
<th>Subject / Token #</th>
<th>D</th>
<th>J</th>
<th>P</th>
<th>A</th>
<th>Total (of 40)</th>
<th>Mean % Correct</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>9</td>
<td>4</td>
<td>9</td>
<td>28</td>
<td>70</td>
<td>21.2</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>32</td>
<td>80</td>
<td>7.1</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>10</td>
<td>2</td>
<td>10</td>
<td>29</td>
<td>72.5</td>
<td>32.7</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>10</td>
<td>6</td>
<td>9</td>
<td>32</td>
<td>80</td>
<td>15.8</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>10</td>
<td>26</td>
<td>65</td>
<td>22.9</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>9</td>
<td>10</td>
<td>8</td>
<td>31</td>
<td>77.5</td>
<td>22.8</td>
</tr>
<tr>
<td>Total (of 60)</td>
<td>36</td>
<td>54</td>
<td>34</td>
<td>54</td>
<td>Overall Mean:</td>
<td>74.2</td>
<td></td>
</tr>
<tr>
<td>Mean % Correct</td>
<td>60</td>
<td>90</td>
<td>56.7</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>15.3</td>
<td>10</td>
<td>24.9</td>
<td>8.2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.13: Discrimination Test Results**

Table of discrimination test results. See the caption to Table 5.12 for an explanation of the table format.
Figure 5.1: Overall Scores of Identification and Discrimination Listening Tests

Average performance of four test subjects (D, J, P, A) on each of six tokens for identification and discrimination tests. The vertical axis shows the percent correct, based on 40 trials for each data point. The error bars represent the standard deviation of the four individual scores for a given token. The overall mean is the mean value of the six points in a given panel. By chance alone one would expect each point to be a small distance above or below 50%.
Scores of individual subjects on each of the tokens (out of 10), as well as mean scores for each token. The vertical axis represents the number of correct responses out of 10 given for each test. The right-most bars in each panel show the mean of the other bars in the same panel, with error bars representing the standard deviation. By chance alone one would expect each score to fall close to 5 out of 10.
Figure 5.3: Scores of Individual Test Subjects

Test results broken down by subject. Scores for the 6 tokens are shown as the number correct out of the 10 presentations in each test. By chance alone one would expect to have most data points lie close to the 5 out of 10 mark.

Efforts were made to compute the statistical significance of these results. Since each trial involved making a choice between two options, the listening tests can be modeled as a Bernoulli process with a certain probability P of correctly answering and probability 1-P of incorrectly answering. If we assume that the probabilities of correctly or incorrectly
answering are identical, P would be equal to 0.5 and one could perform the test by flipping an unbiased coin to respond to each presentation.

The measure of statistical significance that I used, commonly known as “alpha”, estimates the probability that a given test outcome (number of correct responses out of 10, 40, or 60 trials) could be achieved by chance alone, or by flipping an unbiased coin to choose each response. The actual results would be considered more significant if the probability of obtaining equal or better results by flipping an unbiased coin is small.

To compute these significance levels I generated probability mass functions (PMFs) to describe the probability of obtaining k correct responses in n Bernoulli trials. The formula for obtaining k correct responses out of n Bernoulli trials, with a probability P of correctly answering is given in Equation 5.1.

\[
p(k) = \frac{n!}{k!(n-k)!} \times P^k (1-P)^{n-k}
\] (5.1)

For the PMFs used to measure statistical significance, P is set to 0.5 and N is set to 10, 40, or 60 according to which result’s significance is being measured. Significance levels, alpha, were found for each test result by summing p(k) for k greater than or equal to the number of correct responses in the given result (a one-sided test). Alpha, then, indicates the probability that a result could be equalled or exceeded by chance alone. Larger values of alpha, then, mean lower significance. Alpha values for the identification and discrimination tests are given in Tables 5.14 and 5.15 respectively.

Alpha values are given for the 10 trial tests of a given subject and token. These numbers have a tendency to be fairly large because with only 10 trials it is not exceedingly rare to perform above chance level (5 out of 10) by guessing blindly. As the number of trials increases the probability of performing better than chance (20 correct out of 40 or 30 correct out of 60) becomes considerably smaller. Thus, the cumulative significance measures
in the right-most column and bottom-most row (of Tables 5.14 and 5.15) have the opportunity of becoming very high (small percentages). Alpha levels less than or equal to 0.05 are considered “statistically significant” by me and are marked with a “∗∗” in the tables.

<table>
<thead>
<tr>
<th>Subject / Token #</th>
<th>D</th>
<th>J</th>
<th>P</th>
<th>A</th>
<th>Using all 40 trials/token</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.83</td>
<td>0.99</td>
<td>0.38</td>
<td>0.17</td>
<td>0.32</td>
</tr>
<tr>
<td>2</td>
<td>0.83</td>
<td>0.17</td>
<td>0.05*</td>
<td>0.38</td>
<td>0.008*</td>
</tr>
<tr>
<td>3</td>
<td>0.62</td>
<td>0.17</td>
<td>0.83</td>
<td>0.17</td>
<td>0.04*</td>
</tr>
<tr>
<td>4</td>
<td>0.95</td>
<td>0.83</td>
<td>1.0</td>
<td>0.38</td>
<td>0.87</td>
</tr>
<tr>
<td>5</td>
<td>0.95</td>
<td>0.83</td>
<td>0.95</td>
<td>0.01*</td>
<td>0.32</td>
</tr>
<tr>
<td>6</td>
<td>0.99</td>
<td>0.62</td>
<td>0.99</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>Using all 60 trials/person</td>
<td>0.88</td>
<td>0.18</td>
<td>0.65</td>
<td>0.0005*</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.14: Alpha Values for Identification Scores (the Probability that Scores Could be Equalled or Exceeded by Chance Alone)

Alpha significance levels for identification listening tests, computed by summing over appropriately chosen binomial probability mass functions described by Equation 5.1. The numbers found in the 24 cells in the middle of the table (for individual tokens and subjects) correspond to alpha levels for the number of correct responses out of 10 listed for the same subject and token pair in Table 5.12. The right-most column gives alpha levels for the total number of correct responses (out of 40) given for a particular token. The bottom row gives alpha levels for the responses of a particular subject, based on the total number correct out of 60. Alpha values less than or equal to 0.05 are marked with an asterisk to indicate that they are significant at the 0.05 level (smaller alpha means higher significance). Only 5 scores meet this criterion for the identification test.
Alpha Values for Discrimination Test Trials

<table>
<thead>
<tr>
<th>Token #</th>
<th>D</th>
<th>J</th>
<th>P</th>
<th>A</th>
<th>Using all 40 trials/token</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.62</td>
<td>0.05*</td>
<td>0.95</td>
<td>0.05*</td>
<td>0.02*</td>
</tr>
<tr>
<td>2</td>
<td>0.17</td>
<td>0.05*</td>
<td>0.38</td>
<td>0.17</td>
<td>0.0003*</td>
</tr>
<tr>
<td>3</td>
<td>0.38</td>
<td>0.01*</td>
<td>0.01*</td>
<td>0.01*</td>
<td>0.0008*</td>
</tr>
<tr>
<td>4</td>
<td>0.38</td>
<td>0.01*</td>
<td>0.62</td>
<td>0.05*</td>
<td>0.0003*</td>
</tr>
<tr>
<td>5</td>
<td>0.95</td>
<td>0.38</td>
<td>0.83</td>
<td>0.01*</td>
<td>0.08</td>
</tr>
<tr>
<td>6</td>
<td>0.95</td>
<td>0.05*</td>
<td>0.01*</td>
<td>0.17</td>
<td>0.001*</td>
</tr>
</tbody>
</table>

Using all 60 trials/person 0.12 < 0.0001* 0.26 < 0.0001*

Table 5.15: Alpha Values for Discrimination (ABX) Test Scores (the Probability that Scores Could be Equalled or Exceeded by Chance Alone)

Alpha significance levels for discrimination (ABX) listening tests. See the caption to Table 5.14 for an explanation of the table format. 17 discrimination test scores had alpha levels less than or equal to 0.05, compared to only 5 in the identification test. The higher number of significant results is consistent with the fact that the discrimination test is much easier to do well on than the identification test. Subject A did well overall on both tests (identification and discrimination), and the overall scores for tokens 2 and 3 were also significant for both tests. Subjects J and A performed well on the discrimination test.

5.2.3 Discussion

Overall scores for the identification listening test (see Figure 5.1) show that the pre and post versions of tokens 2 and 3 were more often correctly identified than the other four tokens. Only tokens 2 and 3 are significant at the 0.05 level (see Table 5.14) with alphas for the other four exceeding 0.3. These results imply that people had a hard time telling apart the pre and post versions of each token based on memory, which is consistent with
my preliminary listening observation that the pre vs. post difference was always subtle for vocalizations, including those marked as “N”. Despite the subtlety, under the testing conditions it was possible to correctly identify pre and post more than half the time, with an overall mean of 56.7%.

Subject A performed considerably better on the identification test than the other subjects, which may be partially due to the fact that she chose to practice for considerably longer than the other 3 subjects.

Alpha levels for the overall scores in the discrimination (ABX) test (see the right-most column of Table 5.15) show that, for all tokens other than number 5, overall scores received alpha level of 0.05 or less. Subjects J and A both received overall alpha levels below 0.0001. Several tests on individual tokens and subjects also received alpha levels less than or equal to 0.05. Since the same tokens were used in the identification and discrimination tests, these results reinforce the notion that a discrimination test is considerably easier to score well on when pre and post versions are only subtly different. Basing a response on direct comparisons of similar-sounding vocalizations is more likely to yield correct answers than trying to remember a subtle difference and mentally comparing it to future presentations.

For the preliminary judgments of Section 5.1 to be consistent with the quantitative listening test scores of this section, one would expect token 2, for which I consistently gave an “I” rating, to receive low scores, token 6, which I consistently rated as “N”, to receive high scores, and the other tokens to receive scores in between those of token 6 and 2 according to the prominence of their preliminary ratings. In actuality the listening test scores were in disagreement with my preliminary listening ratings. For example, token 2 received the highest overall score and token 6 the lowest overall score, the opposite of what would be needed to agree with my preliminary ratings. For the discrimination test
there is not a very large spread of average scores for the tokens, so no token received particularly higher scores than any other token, and consequently there is no clear relationship of these scores to my preliminary judgments.

Even though separate symbols were used in the preliminary listening section to denote “indistinguishable”, “mildly” distinguishable, and “noticeably” distinguishable pre vs. post differences, all three categories were subtle and described differences that usually required careful listening on my part to hear. Additionally, it was rather difficult to compare the noticeability of differences, and there is a good chance that another person listening to the vocalizations would judge them differently.

The results of these tests indicate that there is not a clear relationship between my own preliminary judgments and the cumulative listening test scores, and when comparing plots of individual tokens and individual subjects there also doesn’t appear to be a clear relationship between one subject’s performance and that of another, e.g. high scores and low scores do not usually occur on the same tokens from one subject to another.

As the discrimination test results indicate, when pre and post versions can be compared directly, scores improve (over identification), but only up to a certain point. The overall scores had a mean of 74.2%. Since discrimination (ABX) is probably the easiest test that could be given without offering repeated presentations or feedback, and there is only a small difference in the averages of different tokens, it seems as though correct comparisons are difficult to make, and do not have much to do with which vocalization is listened to. Anecdotally, the subjects taking the tests agreed that it was difficult to tell the difference between pre and post versions.

In a case that pre and post versions of a token are saliently different sounding, e.g. in a manner similar to the way two vowel sounds differ, one would expect subjects to require
one or two listens during the practice period (compared to the 2-5 minutes needed by most subjects) to recognize and remember the difference.

For such a token, one would expect all subjects to score near 100% in both listening tests. Based on the scores resulting from the actual tests and the difficulty subjects had in taking the tests, it is apparent that none of the pre vs. post differences were “salient” in the manner just described, which agrees with my finding in the preliminary listening section for all vocalizations.

5.2.4 Conclusions

I draw the following conclusions from the quantitative listening tests described in this section:

1. Pre and post versions of all tokens are hard to correctly tell apart in an “identification” test, but for larger numbers of trials, people nonetheless can answer correctly more than half of the time.

2. Pre and post versions of all tokens are also hard to correctly tell apart in a “discrimination” test, but the ability to directly compare the vocalizations allows people to perform considerably better than chance in the identification test.

3. In the discrimination test there is no substantial difference in the subjects’ mean performance on the different tokens.

4. In the identification test there does appear to be a difference in peoples’ performance on the tokens, but it disagrees with my preliminary classifications.

5. To test the relationship between these listening tests and the hypotheses presented in Chapter 4, the structure of the tokens should be examined before and after filtering, focusing on the effects of filtering on harmonic amplitudes, and an attempt should be made to find structural correlates to the listening test scores of this section.
5.3 Correlations Between Test Results and Vocalization Structure

Subjects were, on the average, able to answer correctly more than half of the time in the listening tests of Section 5.2. In this section I attempt to find a structural explanation for the subjects’ performance based on the predictions of the hypotheses presented in Section 4.2.

The first version of the hypothesis in Section 4.2 predicts that salient pre vs. post differences might result from the static attenuation of a harmonic throughout a vocalization as a result of filtering. Such attenuation would alter the relative amplitudes of the harmonics in a manner that could conceivably be heard by people as a change in vowel quality. The second version of the hypothesis predicts that salient pre vs. post differences might result from a harmonic varying in amplitude by a large amount over the course of a vocalization, caused by the harmonic’s frequency changing such that the harmonic passes in and out of the dip in a characteristic manner.

In this section I measure the frequencies and amplitudes of the two harmonics that are most affected by the filter, the second and third, and define quantities based on these measurements that should vary in proportion to the applicability of the two hypotheses. Finally, I compute the correlations of these quantities to the listening test results of (Section 5.2) as a way of testing of the hypotheses’ ability to account for the listening test results.

5.3.1 Methods

My first task was, for pre and post versions of each token, to measure how the harmonic frequencies and amplitudes vary with time. I initially found frequency and amplitude measures for the first four harmonics, but found that only the second and third varied to any substantial degree as a result of filtering, due primarily to the effects of the filter peak and dip (respectively).
I made measurements within each token, spaced every 50 ms, starting 50 ms past the beginning of each vocalization. Of the 6 tokens, vocalization durations varied such that the number of measurements varied from 6 to 11. Measurements were made on windowed fft plots, computed from a 25.6 ms hamming-windowed portion of the vocalization, centered at each measurement point. Sometimes it was difficult to find the harmonic peaks in the fft plots, so some frequency measurements may appear to be out of place. Amplitude measurements were made in decibels. Plots of the measurements just described can be found in Figures 5.4 to 5.9 (see Section 5.3.2). From these measurements I wanted to define quantities that would reflect: a) how much the third harmonic was attenuated relative to the second harmonic over the course of a vocalization (to test the hypothesis in Section 4.2.1) and b) how much the third harmonic changed in amplitude relative to the second harmonic over the course of a vocalization (to test the hypothesis in Section 4.2.2).

For a) above I wanted a measure that would account for the largest changes in the relative amplitudes of the second and third harmonics as a result of filtering. For this, I took the average of the three maximum values of the changes in relative harmonic amplitudes as a result of filtering. More precisely, I found:

\[
\text{Amplitude Change} = |H_2\text{amp, pre} - H_3\text{amp, pre}| - |H_2\text{amp, post} - H_3\text{amp, post}|
\]  

(5.2)

for every measuring time and took the average of the three largest values of this quantity, yielding the Maximum Amplitude Change. The values of this quantity are shown for each token in Table 5.16 under the column labeled “Max |H2-H3| Change”. Since harmonic amplitudes changed with time for most vocalizations, H3 is affected by the dip to varying degrees so taking an average of three maximum changes partially corrects for especially large values that occur when a measuring time happens to correspond to a moment that a harmonic enters the center of the dip (as in the 250 ms point of token 4, in Figure 5.7).
For b) above I wanted a measure that also accounted for the smallest changes in relative harmonic amplitudes that occurred as a result of filtering, such that the measure would be larger if the relative amplitudes varied more over the course of the vocalization than if they stayed at a fixed value. I computed the average of the three smallest changes in \(|H2_{amp} - H3_{amp}|\) (see “Min. \(|H2-H3|\) Change” in Table 5.16) that occurred as a result of filtering and subtracted this from the Maximum Amplitude Change for this measure. The resulting measure is referred to as “Max. - Min.” in Table 5.16, and is intended to be correlated with the applicability of the hypothesis in Section 4.2.2. See Table 5.16 for a listing of these measures for the six tokens, as well as Figure 5.10 for plots of these measures placed on the same axes as the listening test results.

5.3.2 Results

The harmonic frequencies and amplitude measures are displayed in Figures 5.4 to 5.9. The structural measures described in the previous section are given in Table 5.16 and plotted on the same axes as the identification and ABX scores in Figure 5.10. Here, the vertical axis has units of “% Correct” for the upper 2 lines (the listening test results shown in Figure 5.1), but “dB” for the bottom line, which corresponds to each of the amplitude change measurements listed in Table 5.16. The upper left panel contains the maximum \(|H2-H3|\) amplitude change, which should test for the applicability of the hypothesis in Section 4.2.1, and the lower panel is the maximum - minimum \(|H2-H3|\) amplitude change, which is intended to test for the applicability of the hypothesis in Section 4.2.2.

To see if the scores rose and fell in proportion to any of these measures, I found the “Pearson’s product-moment correlation coefficient” for the scores and measures paired by token number. Correlation coefficients can have values between -1 and 1, with 0 indicating no correlation, -1 indicating total negative correlation, and 1 indicating total positive cor-
relation. The correlation coefficients for each of the three measures are shown at the bottom of Table 5.16.

![Token 1 Harmonic Frequencies](image)

**Figure 5.4:** Token 1 H2 and H3 Frequencies and Amplitudes Before and After Filtering

First of a series of six plot pairs showing measurements of frequency and amplitude of the 2nd and 3rd harmonics in each of the tokens. Measurements were made beginning 50 ms from the beginning of each token and were spaced 50 ms apart thereafter. The top panel shows the frequencies of the 2nd and 3rd harmonics as well as could be determined by examining windowed Fourier transform plots. Horizontal lines marked PEAK (between 1700 and 1800 Hz) and DIP (between 2100 and 2200 Hz) represent the center-frequency of the peak and dip found for the left cheetah ear, as shown in Figure 3.5. Amplitude changes due to filtering, as shown in the bottom panel, can be viewed in terms of how close the harmonics get to the DIP or PEAK. In the bottom panel the harmonic amplitude changes due to filtering can be seen by comparing the appropriate lines. For token 1 the 3rd harmonic is attenuated between 5 and 10 dB for much of the vocalization, with the 2nd not substantially affected.
Figure 5.5: Token 2 H2 and H3 Frequencies and Amplitudes Before and After Filtering

Plot pair showing frequencies and amplitudes of the second and third harmonics over the course of token 2 (see description in Figure 5.4). The third harmonic lies in the vicinity of the dip from 200 to 300 ms, reflected by the drop in H3 amplitude after filtering. The 2nd harmonic is not substantially affected by filtering.
Figure 5.6: Token 3 H2 and H3 Frequencies and Amplitudes Before and After Filtering

Plot pair showing frequencies and amplitudes of the second and third harmonics over the course of token 3 (see description in Figure 5.4). None of the harmonics get very close to the dip frequency, though the 3rd harmonic is generally attenuated after filtering by around 5 or more dB. The 2nd harmonic is not substantially affected by filtering.
**Figure 5.7:** Token 4 H2 and H3 Frequencies and Amplitudes Before and After Filtering

Plot pair showing frequencies and amplitudes of the second and third harmonics over the course of token 4 (see description in Figure 5.4). At 250 ms the 3rd harmonic falls right over the dip, causing a large attenuation. Besides that point the third harmonic is attenuated by between 5 and 10 dB for much of the vocalization. The 2nd harmonic is not substantially affected by filtering.
Figure 5.8: Token 5 H2 and H3 Frequencies and Amplitudes Before and After Filtering

Plot pair showing frequencies and amplitudes of the second and third harmonics over the course of token 5 (see description in Figure 5.4). The 3rd harmonic lies close to the dip from 100 to 450 ms, accounting for the large amplitude difference before and after filtering. The 2nd harmonic is not substantially affected by filtering.
Figure 5.9: Token 6 H2 and H3 Frequencies and Amplitudes Before and After Filtering

Plot pair showing frequencies and amplitudes of the second and third harmonics over the course of token 6 (see description in Figure 5.4). The third harmonic is attenuated between 2 and 10 dB over the measured times. The 2nd harmonic is not substantially affected by filtering.
Table 5.16: Structural Measurements and Correlations to Listening Test Results

| Token Number | Max. $|H_2-H_3| \text{change (avg. of top 3)}$ (dB) | Min. $|H_2-H_3| \text{change (avg. of bottom 3)}$ (dB) | Max. - Min. (dB) |
|--------------|------------------------------------------|-----------------------------------------------|-----------------|
| 1            | 14.1                                     | 9                                             | 5.1             |
| 2            | 18.3                                     | 0.5                                           | 17.8            |
| 3            | 11.7                                     | 9.4                                           | 2.3             |
| 4            | 23.6                                     | 5.3                                           | 18.3            |
| 5            | 25.6                                     | 6                                             | 19.6            |
| 6            | 12.4                                     | 7.4                                           | 5               |
| Identification Test Correlation | -0.088                                   | -0.31                                         | 0.061           |
| Discrimination Test Correlation   | -0.15                                    | -0.5                                          | 0.093           |

Table 5.16: Structural Measurements and Correlations to Listening Test Results

Table of three structural measurements made on the six tokens, as well as their correlations to the listening test scores of Section 5.2. The first measurement is, for each token, the average of the three highest changes in the relative amplitudes of $H_2$ and $H_3$ as a result of filtering (see Equation 5.2). These values are shown in the second column, with correlations to the listening tests given at the bottom. This measure was intended to be correlated to the structural features proposed in the first hypothesis of Chapter 4 (see Section 4.2.1). The second measure, in the third column, is the average of the three smallest changes in the relative amplitudes of $H_2$ and $H_3$ as a result of filtering. This measure was not intended to directly indicate the applicability of the hypotheses of Chapter 4, but was rather used in the third measure, which is simply the difference between the maximum and minimum measures (columns 2 and 3 respectively). This final measure, in the fourth column, is intended to correlate with the structural features applicable to the second hypothesis given in Chapter 4 (see Section 4.2.2).
Figure 5.10: Plots of Test Scores and Structural Measurements on Same Axes

Three panels showing Identification and Discrimination test scores on the same axes as each of the structural measurements made on the vocalizations (see Table 5.16). The first and last structural measurements were intended to test for the applicability of the hypothesis in Sections 4.2.1 and 4.2.2 respectively.

Visually comparing the bottom curves to each of the listening test results, it appears to be the case that the two do not consistently vary together for any of the measures.

5.3.3 Discussion

Referring to the frequency and amplitude plots of Figures 5.4-5.9, it can be seen that the amplitude of the second harmonic varies only by a small amount for all of the vocalizations, whereas the third harmonic tends to change more dramatically as a result of filtering. By carefully comparing the frequency plots to the amplitude plots one can see that
as the third harmonic gets closer to the dip line its amplitude strays from the pre amplitude by, at times, over 30 dB.

If the harmonic amplitudes constitute structural features that affect the way vocalizations are perceived, then one might expect to see a correlation between measures of filter-induced amplitude changes and the perceptual experiment results. By plotting the measures of Table 5.16 on the same axes as the listening test results it is possible by visual inspection to see if the measures and scores increase and decrease together from token to token, implying a positive correlation if a high value in one curve corresponded to a high value in the other, and a low in one corresponded to a low value in the other (see Figure 5.10). The plots would be negatively correlated if a high value in one corresponded to a low value in the other and a low value in one corresponded to a high value in the other. To be uncorrelated, there would be no consistent relationship between the values of the corresponding points on the two plots.

The correlation coefficients computed for the maximum |H2-H3| change are within 0.15 of 0 (uncorrelated) for both the identification and discrimination tests, implying no strong correlation between this measure and either set of listening test results (see Table 5.16 for all correlation coefficients). The correlation coefficients for the maximum - minimum |H2-H3| change are within 0.1 of 0 for both tests, also implying no strong correlation between this measure and the listening test results.

It is evident from the plots of harmonic amplitudes for all 6 tokens that, at practically any given time point the 2nd harmonic is higher in amplitude than the 3rd harmonic, often by more than 20 dB. The prominent 2nd harmonic, in contrast to the weaker 3rd harmonic, never changes amplitude by much more than 5 dB as a result of filtering. The 3rd harmonic, however, sometimes changes by over 20 dB as a result of filtering.
Given that the 2nd harmonic is so much more prominent than the 3rd harmonic, it would not be surprising if the percept of a vocalization were more easily influenced by modifying the 2nd harmonic than the third. This prediction is reinforced by the informal experiments I conducted in which I constructed filters with notches at half their original frequencies and filtered some of the vocalizations with them (see Section 5.1.5, number 2). The pre vs. post differences that resulted were far more noticeable than those with the original filters, to the extent that I would have easily given many of them a “salient” rating by the notation of Table 5.1. These observations reinforce the notion that the existing filters, which contain notches too high in frequency to attenuate either the 1st or 2nd harmonics, are not well-suited for creating salient pre vs. post differences.

If one were to design a notch filter capable of creating salient pre vs. post differences of the type required by the hypothesis of Chapter 4 for the given vocalizations, one would certainly place the notch at a low-enough frequency that the prominent 2nd harmonic could be selectively attenuated by varying the fundamental frequency or pitch contour within the normal range for the given species. Such a filter could conceivably do a far better job at affecting percepts of the vocalizations than the actual filters. Similarly, if the filters were given, the vocalizations would need a higher fundamental frequency to be affected saliently.

5.3.4 Conclusions

I draw the following conclusions from the structural measurements, observations, and correlation measures of this section:

1. Of all the tokens, and indeed most of the vocalizations in the other collections, the 2nd harmonic tends to have higher amplitudes than those of the 3rd harmonic.
2. Of all the tokens and other vocalizations, the center frequencies of the middle-ear filters are too high to significantly affect the amplitudes of anything but the 3rd (or higher) harmonics.

3. Measurements on the changes in the relative amplitudes of H2 and H3 before and after filtering do not explain the variability in the listening test results from token to token.

4. The lack of correlation to harmonic amplitudes and test results is incompatible with the hypotheses presented in Chapter 4.

5. The next task is to search for another way of explaining the fact that, in spite of the subtlety of pre vs. post differences, people were often able to perform better than chance at identifying and especially distinguishing the pre and post versions of the vocalizations. Given that the measures made on the vocalizations do not seem to account for listening test results, I will search for an element of the recordings that is independent of the vocalizations that could account for better than chance results on the listening tests. In particular, I will test the ability of the background noise to support pre vs. post differences of the sort described in this chapter.

5.4 Perception of Filtered or Unfiltered Noise versus Vocalizations
In this section I go a step beyond the earlier findings in this chapter that the vocalizations, filters, listening test results, and structural measurements are not well-suited to the hypotheses presented in Chapter 4, by searching for an explanation of the listening test results that is independent of the vocalizations themselves.

Since all of the vocalization collections contain background noise that lasts for the full length of the recording, the noise is present both during vocalizations and between them. Since the noise tends to have a wider spectrum than harmonics of the vocalizations, it is conceivable that filtering the noise could consistently change its spectral shape to a mildly
perceivable extent. In this section I select tokens of noise and vocalizations (also containing background noise) with which to conduct a listening test to determine whether filtering the noise alone produces an identifiable difference in the pre and post versions comparable to that found by testing the vocalizations.

5.4.1 Methods

My task in this section is to perform listening tests with tokens selected as vocalizations and background noise, to determine whether filtered versions of the vocalizations were any easier to identify than filtered versions of the noise. Since background noise varies over the course of the recordings I selected one noise token to be in the immediate proximity of each vocalization token. All noise tokens were selected to be 300 ms in duration.

Vocalization and noise tokens were taken from three collections: CH-III, CH-II, and BC-III. CH-III was selected because the tokens used in the listening tests of Section 5.2 came from CH-III, so using some of them again enables the results to be compared. CH-II and BC-III were chosen because the preliminary listening judgments I made on the vocalizations in these collections (see Tables 5.4 and 5.8) contained a considerable number of "mildly" distinguishable ratings even though the vocalizations themselves generally did not appear to have much frequency content near the dip center frequency (see spectrograms in Appendices 1 and 2).

The vocalizations and background noise tokens are listed in Table 5.17, along with descriptions of where they came from. The two vocalization tokens taken from the CH-III collection correspond to tokens 2 and 6 of the listening test of Section 5.2. These two were selected because they received the highest and lowest scores (respectively) in the identification test, and I wanted to test if these results could be explained by vocalization structure. Noise tokens are given the same numbers as the vocalization tokens closest to them.
in the recordings. The three vocalizations from the CH-II collection were all consistently given “mildly” distinguishable ratings in Section 5.1, with differences heard even when the playback volume was changed (“L+”). The three vocalizations from BC-III were all consistently rated as “mildly” distinguishable and two of them were also rated as “L-”, suggesting that changes in playback volume made it hard to hear the pre vs. post differences.

For each token an identification test was performed, requiring the subject to correctly identify each of 10 presentations as either the pre or the post version as the case may be. I was the test subject for all of the tokens, with another person giving me the test. A practice period preceded each listening test, which typically lasted under 2 minutes for a given token, in which I was able to hear either the pre or the post version upon request. When I was ready, the test was given in the same manner as the identification tests of Section 5.2.
Table 5.17: Noise and Vocalizations Selected for Listening Test

<table>
<thead>
<tr>
<th>Collection</th>
<th>Noise Trial Number</th>
<th>Noise Description</th>
<th>Vocalization Trial Number</th>
<th>Vocalization Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH-III</td>
<td>1</td>
<td>300 ms before CH-III-7</td>
<td>1</td>
<td>CH-III-7, “token 2”</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>300 ms before CH-III-24</td>
<td>2</td>
<td>CH-III-24, “token 6”</td>
</tr>
<tr>
<td>CH-II</td>
<td>1</td>
<td>100 ms after CH-II-5</td>
<td>1</td>
<td>CH-II-5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>500 ms before CH-II-7</td>
<td>2</td>
<td>CH-II-7</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>400 ms before CH-II-9</td>
<td>3</td>
<td>CH-II-9</td>
</tr>
<tr>
<td>BC-III</td>
<td>1</td>
<td>100 ms after BC-III-2</td>
<td>1</td>
<td>BC-III-2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>300 ms before BC-III-6</td>
<td>2</td>
<td>BC-III-6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>600 ms after BC-III-8</td>
<td>3</td>
<td>BC-III-8</td>
</tr>
</tbody>
</table>

Listing of the noise and vocalization tokens used in the listening test of this section. Tokens were selected from the CH-III, CH-II, and BC-III collections, with noise tokens selected for their temporal proximity to the vocalization token with the same number. All noise tokens were selected to be 300 ms in duration, and consist of the background noise present in the original recordings. The Noise Description column lists how far apart the noise token was from its respective vocalization.

5.4.2 Results

The results of the noise and vocalization identification tests are presented in Table 5.18, and graphically presented in Figure 5.11. In the graphical presentation the plot is divided vertically into three regions, one for each collection from which vocalizations were taken (CH-III, CH-II, and BC-III). Within each region the trials are enumerated, with numbers corresponding to the vocalizations listed in Table 5.17. Noise and vocalization
tokens with the same numbers are plotted with the same horizontal coordinate because they came from nearly the same region of the original recording and consequently have comparable types of background noise.

<table>
<thead>
<tr>
<th>Collection</th>
<th>Trial #</th>
<th>Noise Correct (out of 10)</th>
<th>Vocalization Correct (out of 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH-III</td>
<td>1</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>CH-II</td>
<td>2</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>BC-III</td>
<td>2</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

**Table 5.18: Noise vs. Vocalization Listening Test Results**

Results of the listening test of Section 5.4. Each of the noise and vocalization scores is the result of an identification test of the pre and post versions of each token.

Scores for noise tokens exceeded or equalled the scores for vocalization tokens in all but two trials (trial 1 of CH-II and trial 1 of BC-III), implying that the background noise in the vocalizations could fully account for the better than chance listening test scores in Section 5.2 as well as the “mildly distinguishable” and “noticeably distinguishable” preliminary ratings of Section 5.1.
Figure 5.11: Listening Test Results for Vocalization and Noise Tokens

Plot of the listening test results of Section 5.4, made by comparing the identification of filtered noise versus vocalizations. Noise tokens 300 ms in length were selected that were close to the vocalization tokens in the original recordings, such that the noise in both cases would be comparable in structure. Two noise and vocalization pairs were tested for the CH-III collection and three pairs each for the CH-II and BC-III collections. Scores are out of 10 and a score of 5 would be expected by chance.

It is noteworthy that, in 6 of 8 trials, performance for the noise tokens equalled or exceeded performance for the vocalization tokens. In the first BC-III trial the noise was tested first, without an initial presentation of the pre and post pair. For all subsequent BC-III tests the pre and post pair was presented once immediately before the test, offering a cue that made it much easier to perform well on the test.

5.4.3 Discussion

The listening test results clearly indicate that identifying filtered background noise is as easy or easier than identifying filtered vocalizations with the same type of background noise. With the exception of two token pairs, trial 1 in CH-II and trial 1 in BC-III, the test
results for background noise were equal or better to the test results for vocalizations. This implies that the vocalization itself is not necessary to hear the types of pre vs. post differences that were described in Sections 5.1 and 5.2.

In the case of the first trial in BC-III, where the vocalization received a perfect 10 but the noise received a 7, there was a difference in presentation that is worth noting. The first noise trial was given before the first vocalization trial in this collection, but after the first noise trial I requested that the pre and post versions be presented once before the test began so I could attune my ears to the difference between them immediately before the test. This proved very helpful for the BC-III collection because the background noise was so heavy that there was a noticeable change in its absolute amplitude before and after filtering. Once the pre and post versions were presented side by side before the test I could gauge the absolute amplitudes in my mind and consequently was able to achieve perfect scores on the rest of the BC-III trials. The lack of this additional cue is the likely reason that the first noise trial had a lower score than the remaining noise and vocalization trials for BC-III.

In the preliminary listening section I attempted to account for the absolute amplitude change in comparing pre and post versions by changing the playback volume when I listened to the pre and post versions. In cases where I could no longer identify a perceptual difference between the two versions, I gave an L- rating. For the BC-III collection all but one of the “mildly” distinguishable vocalizations received an L- rating, which is consistent with my experience in primarily noticing only the amplitude changes in the background noise before and after filtering. In the case of CH-II all of the mildly distinguishable vocalizations were additionally given “L+” ratings. This implies that something other than an absolute change in noise amplitude was responsible for the perceived pre vs. post differences. Nonetheless, the results of the listening test in this section indicate that comparable
scores are achievable with or without vocalizations, suggesting that the background noise could be responsible for the perceived pre vs. post differences.

Both trials of the CH-III collection featured the same scores for the two vocalization tokens as well as the same scores for the two noise tokens. The first and second vocalization tokens in this test correspond to the second and sixth tokens in the listening tests of Section 5.2. The identification test scores in Section 5.2 were considerably higher for the second token compared to the sixth token, which disagrees both with the test scores of this section as well as my preliminary judgments of Section 5.1. Such inconsistencies in listening test scores and preliminary judgments reinforce the observation that it is difficult to hear the differences between pre and post versions of vocalizations, such that listening at a different time in different circumstances can lead to different scores.

5.4.4 Conclusions

The following conclusions were drawn from the listening tests of this section:

1. The listening test results imply that the background noise of the recordings can explain the ability of people to identify filtered versions of vocalizations or distinguish filtered and unfiltered versions.

2. The subtle pre vs. post differences noted in Section 5.1 as well as the better than chance results of the listening tests in Section 5.2 are likely the result of hearing the effects of filtering on the background noise rather than on the vocalizations.

3. The hypotheses of Chapter 4, which are based on the effects of vocalization structure on perception, are inconsistent with the results of this section and chapter.
Chapter 6

Overall Conclusions

Based on the work done in this thesis, I conclude that the available vocalizations and middle-ear models do not support the communication hypotheses presented in Chapter 4.

In this section I summarize the results as related to the hypotheses, suggest additional experiments that might be worthwhile, and propose an alternative hypothesis that appears to be more consistent with the results.

6.1 Summary of Results

6.1.1 Preliminary Listening

When I compared versions of the vocalizations before and after filtering (pre and post respectively) I found their differences to be subtle or inaudible on all occasions. A small number of vocalizations were somewhat more distinguishable before and after filtering than others (denoted by the “N” rating in Section 5.1), but those differences were still subtle. When I listened to the vocalizations through a filter with a dip at half the original frequency, I heard pre vs. post differences that were salient. For the actual vocalizations and filters to support the Chapter 4 hypotheses, some vocalizations should sound genuinely different after filtering while others should sound largely indistinguishable before and after filtering. As shown in the preliminary listening section, however, all vocalizations sounded at best subtly different before and after filtering.

6.1.2 Listening Tests

In the formal listening tests conducted on 6 vocalizations from the CH-III collection, on average the 4 subjects performed significantly better than chance on some vocalizations, suggesting that under the testing conditions it was possible for people to perceive some differences between pre and post versions. Overall, however, the test results confirm the preliminary assessment that all pre vs. post differences are subtle, considering that the
listening task would be extremely easy if pre and post versions sounded saliently different, and the scores would be higher than those observed. Additionally, the variable performance of individuals on the different tokens of the identification test does not appear to be caused by differences in the harmonic structure of the vocalizations. In the discrimination test, scores improved overall (as expected given that it is supposed to be easier than the identification test), but the results were still lower than one would expect if salient pre vs. post differences were present. All of the tokens received similar scores in the discrimination test, which also implies that differences in vocalization harmonic structure did not significantly impact the results.

To be consistent with the Chapter 4 hypotheses, the identification test results for vocalizations with structures leading to salient pre vs. post differences should be close to 100% on average, and the results for those lacking these structural features should be close to 50% on average. Additionally, salient pre vs. post differences should be easy to hear in the presence of distractions such as varying the playback volume. In the listening tests none of the tokens exhibited pre vs. post differences that were easy enough to hear that they would support the Chapter 4 hypotheses.

6.1.3 Structural Investigation

During attempts to find structural features in the vocalizations consistent with the Chapter 4 hypotheses, I found that the first two harmonics, which were often of the highest amplitude in a vocalization, were never significantly affected by filtering. The dip was only able to significantly affect higher harmonics, such as the third, which were usually considerably weaker than the first and second harmonics. To be consistent with the hypotheses, for a vocalization to be saliently affected by the filter, a strong harmonic should be attenuated by the dip, either statically or in a time-varying manner. For vocalizations not significantly affected by filtering, strong harmonics should avoid the dip. Over-
all, it appeared that strong harmonics nearly always avoided the dip, especially since the
dips were too high in frequency to affect the first or second harmonic. For the hypotheses
to be more likely, then, it appears that either the filter dips would need to be lower in fre-
quency, the high-amplitude harmonics would need to be higher in frequency, or the lower
harmonics would need to be smaller in magnitude than the harmonic passing through the
dip.

6.1.4 Background Noise Shaping

In a test to find a basis for better than chance listening test scores, I found that the
background noise in the recordings had pre vs. post differences that were usually at least
as noticeable as those of the vocalizations. This suggests that both the subtle differences
heard in the preliminary listening as well as the scores on listening tests could result from
the noticeability of changes in the percept of background noise rather than changes in the
percept of vocalizations. Such a finding is inconsistent with the Chapter 4 hypotheses,
which state that perceived pre vs. post differences should result from the effects of filtering
on the vocalizations themselves.

6.1.5 Variability of Dip Frequencies

Another requirement of the Chapter 4 hypotheses is that the dip center frequency for a
given species should not vary by much from one individual to another. With high variabil-
ity of the dip center frequency the filter would not have the same effect on the vocalization
percept for different individuals. Consequently, the communication hypotheses could not
apply to different individuals in the same manner. In Section 3.5 I estimated that the dip
center frequency could vary by as much as 25% if the middle-ear dimensions only varied
within 10% for different individuals. Such variation could move the bobcat dip center fre-
quency up or down by more than 700 Hz, or more than 500 Hz for cheetah. These changes
are on the order of the fundamental frequency of the vocalizations. This situation is
incompatible with the Chapter 4 hypotheses because harmonic spacing would be compa-
rable in size to the variability of the dip frequency among individuals, preventing the middle-ear transformation from affecting the percepts of different individuals in a consistent manner.

6.2 Additional Work

Despite the lack of evidence in support of the Chapter 4 hypotheses, it may be worth testing the hypotheses further. Here I suggest two areas that may be worth investigating.

6.2.1 Collecting More Data

The listening tests were performed using recordings of vocalizations from a small number of individuals, in a small number of uncontrolled behavioral situations. Consequently, the available recordings probably do not fully represent the possible vocalizations of the two species and it is conceivable that there may exist vocalizations with structures that are more consistent with the Chapter 4 hypotheses. If so, the only way to find out would be to obtain a larger collection of recordings, preferably from multiple individuals in multiple well-controlled behavioral situations. Obtaining more middle-ear data would allow one to see how much the dip center frequency varies among individuals.

For new recordings of vocalizations to be consistent with the hypotheses they would probably need to have considerably higher fundamental frequencies, and to place prominent harmonics in the range of the dip. For additional middle-ear data to be consistent with the hypotheses they would have to show small variability in dip center frequency (relative to the fundamental frequency of the vocalizations) among individuals of the same species.

6.2.2 Synthesis Experiments

Another approach to conducting additional tests that does not require the difficult task of obtaining additional recordings and middle-ear measurements is to synthesize vocalizations that are similar in character to the available recordings. Synthesis could offer full control of vocalization structural features that are supposedly necessary to support the
Chapter 4 hypotheses. With appropriately designed vocalizations, experiments could be conducted to determine whether or not the hypotheses could even be supported theoretically. If such experiments yield results that support an hypothesis under a given set of structural features, there would be a strong motivation to look for additional field recordings to see if those features occur in nature. If the results are unsupportive, however, they would constitute an additional reason to reject that hypothesis.

6.3 An Alternate Hypothesis

The work done in this thesis involved testing two closely-related hypotheses concerning a function of the middle-ear structure found in felids. As my findings are not supportive of those hypotheses, I outline here a different reason that the felid middle-ear structure may be the way it is, which has been proposed in Huang et al., 2000.

6.3.1 Low-Frequency Hearing

As described in Section 3.5 and shown in the Figures 3.13-3.16, low-frequency hearing improves when total cavity volume increases. Ears with and without the septum both had an increase in low frequency gain of between 3 and 7 dB over an ear with no bullar cavity (plugged foramen). This could explain the advantage of enlarging the total cavity volume with the addition of the bullar cavity, but it does not explain why the two cavities are separated by a septum.

6.3.2 Experimental High-Frequency Behavior

Contrary to the high-frequency behavior shown in Figures 3.13-3.16 for the lumped models, it has been found experimentally that removing the septum introduces two notches in the cavity gain around 10 kHz (as implied by the admittance plots in Taberner, 1999). With the septum intact the high-frequency dips disappear, and instead we see the familiar dip of our intact middle-ear filter models.
It is believed that identification of spectral minima in the 10 kHz region is important for sound localization (see Huang, 1996), so for that reason it may be advantageous to ensure that no notches appear near 10 kHz. Introducing the septum removes the high-frequency notches as desired, but has the cost of introducing a notch at a lower frequency (seen in the $G_{CAV}$ plots of Section 3.3). As the listening experiments indicate, however, this dip does not appear to have a significant influence on the perception of vocalizations. Perhaps the dips naturally occur at a high-enough frequency to avoid interfering with the prominent first two harmonics in the vocalizations of the given species. Thus, the results reported here are consistent with this alternative hypothesis.

6.3.3 Summary

The hypotheses presented in Chapter 4 concerning a possible function of the middle-ear notch in communication are not supported by the tests performed in this thesis.

An alternate hypothesis to that tested in this thesis is as follows: the two middle-ear cavities are present to improve low-frequency hearing and the septum is present to eliminate high frequency notches that could adversely affect sound localization. The dip introduced by the septum is high enough in frequency that it does not significantly affect the way that species’ vocalizations are heard.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

Contents

BC-I Collection
Summary .......................................................................................................................... 159
Timeline ......................................................................................................................... 160
Spectrograms.................................................................................................................. 161

BC-II Collection
Summary .......................................................................................................................... 165
Timeline ......................................................................................................................... 166
Spectrograms.................................................................................................................. 167

BC-III Collection
Summary .......................................................................................................................... 170
Timeline ......................................................................................................................... 171
Spectrograms.................................................................................................................. 172

BC-IV Collection
Summary .......................................................................................................................... 175
Timeline ......................................................................................................................... 176
Spectrograms.................................................................................................................. 177

BC-V Collection
Summary .......................................................................................................................... 182
Timeline ......................................................................................................................... 183
Spectrograms.................................................................................................................. 184
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

BC-I Summary

Source of Recordings: Dr. Gustav Peters (Zoologisches Forschungsinstitut und Museum Alexander Koenig, Bonn, Germany).

Background Information: 13 vocalizations (in a 26 second recording) by one adult female bobcat (all within a 15 second section). Described as “mews” in Peters (1987). Four spectrograms of vocalizations from this collection (9-12) appear in Figure 2 on page 321 of Peters (1987), with the following captions: (9 & 10) “Medium intensity mews by a female”, and (11 & 12) “Mews of low and medium intensity by a female. The calls show rhythmical amplitude modulation.” Voice recording on the tape specifies that it was recorded at a high level, with heavy background noise.

Frequency Range of Vocalizations: Between 0.6 and 7 kHz. Darkest content occurs between 1 and 3 kHz.

Duration Stats (milliseconds):

<table>
<thead>
<tr>
<th>Stat</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>453</td>
</tr>
<tr>
<td>Std. Dev</td>
<td>90</td>
</tr>
<tr>
<td>Min</td>
<td>246</td>
</tr>
<tr>
<td>Max</td>
<td>560</td>
</tr>
</tbody>
</table>

Interval Stats (milliseconds):

<table>
<thead>
<tr>
<th>Stat</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>706</td>
</tr>
<tr>
<td>Std. Dev</td>
<td>285</td>
</tr>
<tr>
<td>Min</td>
<td>494</td>
</tr>
<tr>
<td>Max</td>
<td>1519</td>
</tr>
</tbody>
</table>

Pitch Contours: “UCD” or some variant. Often peaks in the frequency of the lowest harmonic occur earlier in the vocalization (1, 4, 8, 9, 11, 12). Sometimes the peaks seem to occur closer to the center, and the contour appears to be flat on top (2, 5, 6, 7, 10).

H1 Peak Stats (Hz): (lowest harmonic does not appear to be at the fundamental frequency)

<table>
<thead>
<tr>
<th>Stat</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1294</td>
</tr>
<tr>
<td>Std. Dev</td>
<td>105</td>
</tr>
<tr>
<td>Min</td>
<td>1040</td>
</tr>
<tr>
<td>Max</td>
<td>1449</td>
</tr>
</tbody>
</table>

Characteristics of Harmonics: Harmonic amplitudes typically appear noisy and irregular. It is rare to see them form smooth, continuous lines. The lowest harmonic is sometimes relatively smooth and continuous (1, 2, 5, 6, 7, 8, 9). The lowest harmonic is not always the darkest, and it is common for weaker noisy bands to occur between the darker bands.

Additional Features: Pulsed amplitude modulation occurs at the beginning of 10, 11, and 12. In 10 only two pulses are clearly visible, in 11 around 6 are visible, and in 12 around 9 are visible. The pulse rate is roughly 90 per second.
BC-I Timeline
(durations and locations are approximate)
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

BC-I Collection

1. 6097-6558 ms: 461 ms
   “UCDCD”, Freq. of lowest harmonic (H1) peaks early in the vocalization. Noisy harmonics at beginning and end, smoother near middle. Darkest between 1 & 3 kHz. Lowest harmonic referred to as H1 instead of F0 since it doesn’t appear to be at the fund. freq.
   
   H1 @ 6197 ms = 1418 Hz (peak)
   H1 @ 6391 ms = 1229 Hz

2. 7089-7335 ms: 246 ms
   “UCD”, H1 peaks near the midpoint. Weak overall; darkest between 1 and 3 kHz.
   
   H1 @ 7206 ms = 1292 Hz

3. 7895-8237 ms: 342 ms
   “UCD”, H1 seems to peak to the left of the midpoint. Weak except between 1.5 and 3 kHz. H1 is weak.
   
   H1 @ 7970 ms = 1040 Hz

4. 8731-9229 ms: 498 ms
   “UCD”, H1 peaks early then tapers off gradually. Noisy content between 1 and 7 kHz. Harmonic amplitudes vary in a rough manner judging by their splotchy patterning.

   H1 @ 8899 ms = 1229 Hz
5. 9767-10173 ms: 406 ms  
“UCD”, H1 contour appears to lean to the left but is fairly flat on top. Noisy content from 1 to 7 kHz with a few prominent harmonics, including H1.

H1 @ 9936 ms = 1386 Hz (peak).

6. 10758-11141 ms: 383 ms  
“UCDCD”, H1 contour leans slightly to the left. Noisy content, darkest between 1.5 and 3.5 kHz.

H1 @ 10898 ms = 1229 Hz (peak).  
H1 @ 11000 ms = 1103 Hz.

7. 11793-12214 ms: 421 ms  
“UCD”, H1 contour peaks, slightly, early in the vocalization. Noisy content, darkest between 1 and 3 kHz.

H1 @ 11958 = 1197 Hz.

8. 13244-13677 ms: 433 ms  
“UCD”, H1 contour peaks early then falls off gradually. Noisy content, with dark bands occurring between 1 and 4.3 kHz.

H1 @ 13358 ms = 1386 Hz (peak)  
H1 @ 13550 ms = 1166 Hz.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

9. 15196-15741 ms: 545 ms
“UCD”, H1 contour peaks near the left then falls off gradually. Noisy content with some darker harmonics. H1 is particularly dark.

H1 @ 15301 ms = 1449 Hz (peak)
H1 @ 15588 ms = 1134 Hz.

10. 16571-17086 ms: 515 ms
“UCD”, H1 contour peaks, slightly, closer to the beginning. Content is noisy not as full as the previous six vocalizations. Some harmonics appear prominent. Pulsed amplitude modulation occurs at the beginning (~2 pulses).

H1 @ 16740 ms = 1260 Hz.

11. 17684-18244 ms: 560 ms
“UCD”, peaks near the left. Pulsed amplitude modulation occurs at the beginning (~6 pulses).

H1 @ 17819 ms = 1292 Hz.

12. 18789-19348 ms: 559 ms
“UCD”, H1 peaks near the beginning. Similar structure to 10 and 11, all three of which sound somewhat different from the other vocalizations in BC-I. Pulsed amplitude modulation occurs at the beginning (~9 pulses).

H1 @ 18889 ms = 1355 Hz (peak)
H1 @ 19042 ms = 1197 Hz.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

13. 19932-20445 ms: 513 ms
“UCDCD”, H1 peaks a little to the left of center. Compared to the other vocalizations in BC-I, less vocalized noise is present. Most of the noise that appears is background noise.

H1 @ 20100 ms = 1292 Hz (peak)
H1 @ 20256 ms = 1040 Hz.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

BC-II Summary

Source of Recordings: Gustav Peters

Background Information: 9 vocalizations (in a 31 second recording) by the same adult female bobcat as in BC-I. Described as “mews” by Peters. No spectrograms from this collection were published in Peters (1987), though they were most likely part of the data set he used for that paper. Recorded at a high level, with heavy background noise.

Frequency Range of Vocalizations: Between 0.6 and 7 kHz. Darkest bands occur between 1.2 and 3 kHz.

Duration Stats (milliseconds):
- Mean: 530
- Std. Dev: 149
- Min: 333
- Max: 791

Interval Stats (milliseconds):
- Mean: 1482
- Std. Dev: 1002
- Min: 631
- Max: 3611

Pitch Contours: “UCD” or some variant. They differ from BC-I in that many of the peaks occur either near the center (4, 5, 8, 9) or closer to the end (2, 3). Even in the cases where the actual peak occurs closer to the beginning (1, 6, 7), the pitch appears to rise more gradually at the beginning fall off more rapidly at the end.

H1 Peak Stats (Hz): (Lowest harmonic does not appear to be at the fundamental frequency.)
- Mean: 1078 (this mean is a little more than a minor third below the BC-I mean)
- Std. Dev: 78
- Min: 945
- Max: 1166

Characteristics of Harmonics: Harmonic amplitudes tend to be irregular and noisy in a similar way to the BC-I collection. The amplitude of the lowest harmonic is never the strongest: the highest harmonic amplitudes typically occur between 1.2 and 3 kHz. It is common for some harmonics to vary in amplitude by a large amount and for others to change their amplitudes only a small amount.

Additional Features: Amplitude modulation is present at the beginning of 7. Around 6 pulses are visible, and the rate is roughly 85 pulses per second. Vocalization 9 contains what appears to be amplitude modulation in its middle section, with very closely spaced pulses. Roughly 17 pulses are visible, for a rate of approximately 140 pulses per second.
BC-II Timeline
(durations and locations are approximate)
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

BC-II Collection

1. 4417-5124 ms: 707 ms
   “UCDCUCD”. H1 is weak, H3 and H5 are fairly smooth and dark. H2 and H4 cut in and out, but become fairly dark near the center.
   
   H1 @ 4661 ms = 945 Hz.

2. 6110-6564 ms: 454 ms
   “UCD”, H1 contour leans toward the right. Noisy content, H1 is weak, H2 appears the darkest.
   
   H1 @ 6382 ms = 1134 Hz (peak)

3. 7745-8395 ms: 650 ms
   “UCD”, with a prolonged “C” section. Noisy content, darkest between 1 and 3 kHz. Some harmonic amplitudes appear to cut in and out.
   
   H1 @ 8134 ms = 977 Hz.

4. 9382-9820 ms: 438 ms
   “UCD”, pitch seems to peak near the center. Harmonic content is smoother, but still somewhat noisy.
   
   H1 @ 9599 ms = 1103 Hz.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

5. 10766-11144 ms: 378 ms
   “UCD”, pitch seems to peak close to the center. Odd harmonics receive more emphasis, with the exception of H2. Content is a little less noisy than the others in BC-II.
   
   H1 @ 10933 ms = 1166 Hz.

6. 11775-12566 ms: 791 ms
   “UCDCUCDCD”. Noisier content at beginning than middle. Even harmonics disappear just before the center.
   
   H1 @ 11940 ms = 1166 Hz
   H1 @ 12099 ms = 1040 Hz
   H1 @ 12147 ms = 1071 Hz
   H1 @ 12399 ms = 914 Hz.

7. 13386-13965 ms: 579 ms
   “UCUCDCD”. Noisy from the middle to the end. Pulsed amplitude modulation occurs near the beginning (~6 pulses). A dark H2 appears from the onset of the noise until the end.
   
   H1 @ 13610 ms = 1008 Hz
   H1 @ 13808 ms = 945 Hz.

8. 16656-17099 ms: 443 ms
   “UCD”, peak occurs near the center. Noisy content, harmonics cut in and out.
   
   H1 @ 16882 ms = 1134 Hz.
9. 20710-21043 ms: 333 ms
“UCD”, peaks near the center. Different appearance than others in BC-II. Vertical striations appear in the middle. Content is noisy in the middle.

H1 @ 20859 ms = 1071 Hz.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

BC-III Summary

Source of Recordings: Gustav Peters

Background Information: 8 vocalizations (in a 16 second recording) by an adult male bobcat. Also described as “mews” by Peters, spectrograms of two vocalizations from this collection (1 & 2) appear in Figure 2 on page 321 of Peters (1987). They are described as “Mews of low intensity by a male.” Recorded at a high level, with very heavy background noise.

Frequency Range of Vocalizations: Between 0.5 and 7 kHz, though most content is below 4 kHz.

Duration Stats (milliseconds):
- Mean: 766
- Std. Dev: 407
- Min: 359
- Max: 1752

Interval Stats (milliseconds):
- Mean: 838
- Std. Dev: 193
- Min: 527
- Max: 1143

Pitch Contours: Typically “UCD” or some variant. This collection of vocalizations is generally less intense (and with heavier background noise) than either of the previous bobcat collections. Consequently it is sometimes difficult to get a clear view of the fundamental harmonic across its length. As well as can be told, peaks seem to occur closer to the beginning (1, 4, 5), and near the center (2, 3, 6, 7, 8) of the vocalizations.

F0 Peak Stats (Hz): (Note: the lowest harmonic is at the fundamental frequency.)
(Vocalization 3 was so faint that no measurement of its peak fundamental frequency was made.)
- Mean: 869 (mean is a little less than a perfect fifth below the BC-I mean)
- Std. Dev: 29
- Min: 819
- Max: 914

Characteristics of Harmonics: Often the harmonics become darker near the end of the vocalizations (2, 3, 5, 6, 7, 8), at which point sometimes as many as seven can become visible. Insofar as they are visible, the harmonics are smooth curves and do not seem to contain much “vocalized” noise. With the exception of 7, the darkest bands occur between 1 and 3 kHz.

Additional Features:
Vocalization 4 is significantly longer and more intense than the others in the collection.
BC-III Timeline
(durations and locations are approximate)
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

BC-III Collection

1. 3606-4526 ms: 920 ms
   "UCDCD". Aside from heavy background noise, vocal content is not very noisy. Lowest harmonic is the fundamental, and harmonics form relatively smooth lines. Sounds qualitatively calmer than the BC-I & II collections.
   
   F0 @ 3885 ms = 882 Hz
   F0 @ 4260 ms = 662 Hz.

2. 5053-5598 ms: 545 ms
   "UCD". Finding the start and end is difficult with the background noise. Harmonics are fairly smooth, but almost too weak to see until the end.
   
   F0 @ 5304 ms = 882 Hz.

3. 6521-6965 ms: 444 ms
   Too much background noise to see clearly or measure F0.

4a. 8108-9860 ms: 1752 ms
    ~"UCDCD", H1 peak occurs closer to the beginning. Near the beginning seven or more harmonics are visible, afterward only around three can be seen prominently.
    
    F0 @ 8314 ms = 851 Hz
    F0 @ 8740 ms = 662 Hz
    F0 @ 9058 ms = 567 Hz
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

4b.

5. 10521-11207 ms: 686 ms “UCD”, H1 contour appears to lean somewhat to the left. Sounds like something hit the microphone in the middle (~10900 ms).

   F0 @ 10819 ms = 851 Hz.

6. 12087-12776 ms: 689 ms “UCD”, difficult to see vocalization within heavy background noise.

   F0 @ 12430 ms = 819 Hz.

7. 13432-14119 ms: 687 ms “UCD”, H1 seems to peak near the center. Aside from the first and third, most harmonics are not visible until close to the end.

   F0 @ 13776 ms = 914 Hz.
8. 14868-15603 ms: 735 ms
“UCD”, H1 appears to peak slightly before the center of the vocalization. Toward the end several harmonics are visible.

F0 @ 15179 ms = 882 Hz.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

BC-IV Summary

Source of Recordings: Gustav Peters

Background Information: 7 vocalizations (in a 35 second recording) by an adult male and adult female bobcat. Peters was unable to attribute sounds to particular individuals. Several different types of sound are present, including growls, hisses, and prolonged calls with varied, complex structures. Peters refers to this and the following collection as “diverse sounds”, and he mentions that the prolonged, mew-like sounds have a “growly superposition of varying degrees; for some it is relatively strong.” Mild background noise.

Frequency Range of Vocalizations: Varied

Duration Stats (milliseconds):
- Mean: 2760
- Std. Dev: 2055
- Min: 441
- Max: 5212

Interval Stats (milliseconds):
- Mean: 2088
- Std. Dev: 2068
- Min: 223
- Max: 6494

Pitch Contours: Complex, but often flat.

H1 Peak Stats (Hz): Some measurements were made, but they are usually not intended to be H1/F0 peak measurements.

Characteristics of Harmonics: Complex and varied: Sometimes smooth, sometimes rough and noisy, sometimes pulsed, etc.

Additional Features: Vocalizations 1, 2, 3, 4, and 7 are heterogeneous mew-like sounds with growly parts. 5 sounds like a breathing sound and 6 sounds like a short growl followed by a “huff.”
BC-IV Timeline
(durations and locations are approximate)
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

BC-IV Collection

1. 1438-2613 ms: 1175 ms  
   Not very loud, and not harsh sounding.  
   F0 variations produce a "wavery"  
   sound. Darkest content occurs between  
   1 and 2kHz.  
   
   F0 @ 2102 ms = 347 Hz.

2a. 3943-5856 ms: 1913 ms  
   Somewhat prolonged, with "growly"  
   parts interspersed with more "tonal"  
   sounding parts.  
   
   F0 @ 5454 ms = 504 Hz.  
   (F0 was hard to see)

2b. 

3a. 12350-17510 ms: 5160 ms  
   Long vocalization that sounds like a  
   varied "tonal" sound, punctuated with  
   two growly bursts at the end (~16050  
   and 16950 ms). Parts sound like a  
   motor suddenly changing speeds.  
   Three such changes occur at 13874,  
   14180 and 14800 ms.  
   
   F0 @ 13302 ms = -441 Hz  
   F0 @ 14264 ms = 347 Hz  
   F0 @ 14572 ms = 473 Hz.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

3b. 

3c. 

3d. 

4. 19812-24663 ms: 4851 ms
Long vocalization with a variety of different textures visible in the harmonics (smooth, rough, wavering, noisy, pulsed amplitude modulation, etc.)

F0 @ 20972 ms = 378 Hz
F0 @ 23379 ms = 473 Hz
F0 @ 24338 ms = 126 Hz.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

4b. 25468-25909 ms: 441 ms
Possibly a breathing sound.

4c.

4d.

5. 25468-25909 ms: 441 ms
Possibly a breathing sound.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

6. 26132-26701 ms: 569 ms short “snorty” growl sound followed by a “huff” sound.

7a. 28076-33288 ms: 5212 ms Variable harmonic textures are present.

F0 @ 28736 ms = ~410 Hz
F0 @ 31008 ms = 473 Hz
F0 @ 32942 ms = 189 Hz.

7b.

7c.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

7d.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

BC-V Summary

Source of Recordings: Gustav Peters

Background Information: 7 vocalizations (in a 50 second recording) by the same adult male and female bobcats as in BC-IV. Peters’ descriptions also apply to this collection. Mild background noise.

Frequency Range of Vocalizations: Variable

Duration Stats (milliseconds):
- Mean: 3142
- Std. Dev: 2242
- Min: 471
- Max: 6206

Interval Stats (milliseconds):
- Mean: 3271
- Std. Dev: 2342
- Min: 909
- Max: 6342

Pitch Contours: Complex, but often flat.

H1 Peak Stats (Hz): Some measurements were made, but they are often not intended to indicate H1/F0 peak measurements.

Characteristics of Harmonics: Complex and varied: sometimes smooth, sometimes rough and noisy, sometimes pulsed, etc.

Additional Features:
Vocalization 4 is a snort-like sound, 5 is a “hissy growl”-like sound, 6 sounds like two cats vocalizing at the same time, one growling and the other making a mew-like sound. 7 is a “growl”.
BC-V Timeline
(durations and locations are approximate)
Variable harmonic textures. Sounds like there’s a slowly wavering pitch change.

F0 @ 642 ms = 189 Hz
F0 @ 2743 ms = 189 Hz.

1a. 185-6391 ms: 6206 ms
1b.
1c.
1d.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

1e.

2a. 12733-18661 ms: 5928 ms
Long "growly" sound with a "snorty" burst at the end (panel 2e).

\[ F0 @ 15269 \text{ ms} = 378 \text{ Hz} \]

2b.

2c.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

2d. 

3a. 19622-22868 ms: 3246 ms
A softer “growly” sound.

F0 @ 20698 ms = 599 Hz

3b.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

3c.

4. 28755-29473 ms: 718 ms
A “snort-like” sound.

5. 30382-30853 ms: 471 ms
A “hissy-growl” sound.

6a. 32004-36242 ms:
Sounds like two cats at the same time, one “growly” and the other one “mew-like”.

F0 @ 34492 ms = 284 Hz.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms

6b. BC-V-6b

6c. BC-V-6c

6d. BC-V-6d

7. 40620-41807 ms: 1187 ms
A "growl" sound.
Appendix 1: Bobcat Vocalization Summaries and Spectrograms
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

Contents

CH-I Collection
  Summary ................................................................. 191
  Timeline ............................................................... 193
  Spectrograms ....................................................... 194

CH-II Collection
  Summary ................................................................. 198
  Timeline ............................................................... 200
  Spectrograms ....................................................... 201

CH-III Collection
  Summary ................................................................. 204
  Timeline ............................................................... 205
  Spectrograms ....................................................... 206
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

CH-I Summary

Source of Recordings: Stuart Wells (National Zoological Park, Washington, D.C.)

Background Information: 17 vocalizations (in a 39 second recording) by one male cheetah. Described (with some exceptions) as “stutters” by Wells. Refer to Ruiz-Miranda et al. (1998) for a description of the behavioral context of stutters. Examples of a vocalization called “gurgling” were displayed and described in Peters (1984). The structure of the CH-I collection of vocalizations appears similar to the examples in Peters’ paper, though harmonic details visible in this collection are not visible in his paper (perhaps due to different frequency-domain resolutions). Background noise is fairly heavy and the recording quality is relatively poor relative to the bobcat recordings.

Frequency Range of Vocalizations: Between 0.45 and 7 kHz.

Duration Stats (milliseconds):
- Mean: 569
- Std. Dev: 348
- Min: 45
- Max: 1300

Interval Stats (milliseconds):
- Mean: 1521
- Std. Dev: 1709
- Min: 58
- Max: 5784

Pitch Contours: For all stutters (1, 2, 3, 4, 6, the first half of 7, 8, 9, parts of 12, 14, 16, and 17), the pitch contour is essentially flat for stutters. Vocalization 5 (a non-stutter) contains no pulse modulation and has a “CUCDCD” pitch contour. Number 10 lacks a clear harmonic structure. Number 11 contains a relatively flat pitch contour, and appears to contain pulses that are very closely spaced (~26 / .198 seconds) = 131 pulses/second. Numbers 12 and 13 are very difficult to see within the background noise, though 12 appears to have a flat contour and 13 seems to have some concave-down curvature to it. Number 15 is a very short tonal sound with a relatively flat contour (it seems to slope down slightly). Looking more closely at the contours of individual pulses in stutters, it appears that they tend to have concave-down curvature.

F0 Peak Stats (Hz): (fundamental is present)
These statistics were computed only for the stutters in the CH-I collection.
- Mean: 504
- Std. Dev: 42
- Min: 473
- Max: 630

Characteristics of Harmonics: For stutters, the structure of each pulse contains a potentially large number of visible harmonics. Given how much background noise is present, it is hard to gauge how noisy the harmonics themselves are, but from the spectrograms it appears that some “vocalized” noise is present. The amplitudes vary from pulse to pulse, though quite often the lowest harmonic is the darkest. In many cases the next darkest harmonic is the third instead of the second.
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

Stutter Rate Stats (# pulses / duration, given in pulses per second):
(Based on vocalizations 1-4, 6-9, and 16-17.)
Mean: 21.7
Std. Dev: 3.6
Min: 18.2
Max: 29.1

Additional Features:
Numbers 5, 10, 11, 13, and 15 are diverse sounds other than stutters.
CH-I Timeline
(durations and locations are approximate)

- Stutter
- Other
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

**CH-I Collection (all vocalizations are stutters unless otherwise noted)**

1. 1061-1575 ms: 514 ms
   ~10 pulses, mainly < 5kHz.
   
   F0 @ 1426 ms = 504 Hz.

2. 1773-2338 ms: 565 ms
   ~10-11 pulses, higher harmonics are weaker at spots near the middle.
   
   F0 @ 2044 ms = 504 Hz.

3. 3076-3910 ms: 834 ms
   ~17 pulses. Shortly after onset, mid-high frequencies weaken, but they return by the middle.
   
   F0 @ 3562 ms = 504 Hz.

4. 7032-8061 ms: 1029 ms
   ~21 pulses, with full mid-high frequency harmonics.
   
   F0 @ 7533 ms = 504 Hz.
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

5. 8803-9225 ms: 422 ms
   "CUCDCD". Not a stutter

   F0 @ 8854 ms = 504 Hz.

6. 9865-10570 ms: 705 ms
   ~17 pulses, though pulse structure becomes less clear for the second half.

   F0 @ 10023 ms = 536 Hz.

7. 12259-13356 ms: 1097 ms
   ~14 pulses, through 12873 ms, after which pulse structure is unclear.

   F0 @ 12702 ms = 473 Hz.

8. 14857-15504 ms: 647 ms
   ~12-13 pulses.

   F0 @ 15253 ms = 473 Hz.
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

9. 15903-16616 ms: 713 ms ~12-14 pulses.
   F0 @ 16044 ms = 504 Hz.

10. 17221-17668 ms: 447 ms
    Noisy and with an unclear structure.
    Not a stutter.

11. 17978-18176 ms: 198 ms
    Can see vertical striations in spectrogram. Not a stutter.
    F0 @ 18080 ms = 630 Hz.
    [A few weak sounds without clear beginnings and endings fall between 11 & 12.]

12. 23960?-25260 ms: 1300 ms
    A "soft tonal whine" until ~24532 ms, and pulsed amplitude modulation afterward, though the structure of the vocalization is not clear in the spectrogram. Not a stutter.
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

13. 25629?-26186 ms: 557 ms
   Noisy, “tonal” sound. Not a stutter.

14. 28845?-29237 ms: 392 ms
   ~6 pulses? Unclear structure.
   \[ F_0 @ 29078 \text{ ms} = 504 \text{ Hz}. \]

15. 29669-29714 ms: 45 ms
   Very short tonal sound. Not a stutter.
   \[ F_0 @ 29078 \text{ ms} = 504 \text{ Hz}. \]

16. 34803-34906 ms: 103 ms
   Short tonal sound, which appears to have ~3 pulses.
   \[ F_0 @ 34859 \text{ ms} = 473 \text{ Hz}. \]

17. 34964-35072 ms: 108 ms
   Another short tonal sound ~3 pulses.
   \[ F_0 @ 35030 \text{ ms} = 473 \text{ Hz}. \]
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

CH-II Summary

Source of Recordings: Stuart Wells

Background Information: 12 vocalizations (in a 22 second recording) by the same male cheetah in CH-I and possibly another cheetah and described as “eeaows” by Wells. See Ruiz-Miranda (1998) for a description of the behavioral context of eeaows. This collection also contains three stutters (CH-II-(3, 10, 11)). When listening to the recording it sounds like one cat is responsible for the eeaows and another for the stutters, partially because of how quickly CH-II-4 follows the stutter CH-II-3, and because the stutters sound like they were produced closer to the microphone than the eeaows.

Frequency Range of Vocalizations: Between 0.5 to 5.5 kHz (with the majority of content below 2 kHz)

Duration Stats (milliseconds):
- Mean: 405
- Std. Dev: 135
- Min: 240
- Max: 711

Interval Stats (milliseconds):
- Mean: 1109
- Std. Dev: 487
- Min: 388
- Max: 1857

Pitch Contours: Typically “UCDC” or some variant. The eeaows follow gentle pitch contours, for which only the last part is clearly visible in most spectrograms. There usually is a peak, and it is always closer to the beginning. Often after peaking the contour falls a little to form a flat line, during which higher harmonics become visible. The stutters follow flat contours as in CH-I.

F0 Stats (Hz): (fundamental is present)
Because the fundamental is often very weak, measurements cannot always be made on the eeaows during the peaks. Instead measurements are often made on the flat part closer to the end. As a result, these statistics do not reflect the peak of the lowest harmonic. Additionally, since there are two types of vocalization present, these numbers are not an ideal description of eeaows.

These statistics were computed only for the eeaows in the CH-II collection.
- Mean: 735
- Std. Dev: 33
- Min: 693
- Max: 788

Characteristics of Harmonics:
For eeaows the first two harmonics are most prominent, and usually only near the end do higher harmonics appear. The vocalizations are relatively weak and the background noise level is high, so it is hard to say for sure if the the harmonics contain vocalized noise content. Based on the spectrograms and their smooth sound, I would say that vocalized noise is minimal. For the stutters, the harmonic structure is less clearly visible than in CH-I, but the pulses appear similar in structure.
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

Stutter Rate Stats (# pulses / duration, given in pulses/second):
(Based on vocalizations 3, 10, and 11.)
Mean: 21.2
Std. Dev: 0.4
Min: 20.7
Max: 21.8

Cumulative Stutter Rate Stats (pulses per second):
(Combining CH-I and CH-II stutters)
Mean: 21.6
Std. Dev: 3.2
Min: 18.2
Max: 29.1

Additional Features: --
CH-II Timeline
(durations and locations are approximate)

Time (seconds)

Eeaows

Stutters

END
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

CH-II Collection (vocalizations are eeaows unless noted otherwise)

1. 3325-3919 ms: 594 ms
   Tonal, sounds “squeal-like”. By 3685 ms, mid-high frequencies enter.
   \[ F_0 @ 3745 \text{ ms} = 694 \text{ Hz}. \]

2. 4882-5288 ms: 406 ms
   “UCDC”. Only first two harmonics prominent.
   \[ F_0 @ 5136 = 725 \text{ Hz}. \]

3. 7145-7604 ms: 459 ms
   Stutter, ~10 pulses.
   \[ F_0 @ 72886 \text{ ms} = 441 \text{ Hz}. \]

4. 7992-8289 ms: 297 ms
   “UCDC”. Short tonal sound, somewhat noisy, with only first two harmonics visible.
   \[ F_0 @ 8151 \text{ ms} = 788 \text{ Hz}. \]
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

5. 10137-10424 ms: 287 ms
   “UCDCD”. Soft sounding
   F0 @ 10263 ms = 756 Hz.

6. 11919-12269 ms: 350 ms
   “UCDCD”. Most content < 2kHz.
   F0 @ 12014 ms = 1008 Hz (C1)
   F0 @ 12167 ms = 725 Hz (C2).

7. 13829-14174 ms: 345 ms
   “UCDCD”. Tonal “squeal-like” sound.
   F0 @ 14066 ms = 725 Hz.

8. 15036-15276 ms: 240 ms
   “UCDC”
   F0 @ 15201 ms = 725 Hz.
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

9. 16055-16337 ms: 282 ms “UCDC”
   
   \[ \text{F0 @ 16211 ms} = 788 \text{ Hz.} \]

10. 17480-18191 ms: 711 ms Stutter, -15 pulses.
    
    \[ \text{F0 @ 17774 ms} = 473 \text{ Hz.} \]

11. 18747-19133 ms: 386 ms Stutter, -8 pulses.
    
    \[ \text{F0 @ 18804 ms} = 410 \text{ Hz.} \]

12. 19880-20386 ms: 506 ms “UCDCD”. Tonal sound, only first two harmonics prominent, both well below 2 kHz. There appears to be some vibrato near the end.

    \[ \text{F0 @ 20194 ms} = 693 \text{ Hz.} \]
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

CH-III Summary

**Source of Recordings:** Stuart Wells

**Background Information:** 30 vocalizations (in a 47 second recording) by the same male cheetah heard in CH-I and II, and described as “chirps” by Wells. See Ruiz-Miranda (1998) for a discussion of the behavioral contexts of chirps.

**Frequency Range:** Between 0.1 and 7.5 kHz

**Duration Stats (milliseconds):**
- Mean: 494
- Std. Dev: 81
- Min: 311
- Max: 702

**Interval Stats (milliseconds):**
- Mean: 1050
- Std. Dev: 493
- Min: 362
- Max: 2047

**Pitch Contours:** Typically “UCDC” or some variant. Usually rises to a peak to the left of center, then falls off to a lower, nearly flat level that remains nearly flat until the end (1-5, 7, 8, 10, 11, 13-19, 21, 22, 24, 25, 27-30). Sometimes after peaking the contour remains nearly flat for an extended period before falling down to a lower level, if it even does (6, 9, 12, 20, 23, 26).

**F0 Peak Stats (Hz):** (fundamental is present)
- Mean: 882
- Std. Dev: 94
- Min: 662
- Max: 1040

**Characteristics of Harmonics:**
- Each vocalization contains visible evidence of three or more harmonics. In some, over ten are visible during some of the time. The harmonics almost always contain evidence of “vocalized” noise, but the amount varies from a small amount (e.g. 9) to a large amount (e.g. 1).

**Additional Features:** --
CH-III Timeline
(durations and locations are approximate)
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

CH-III Collection (all vocalizations are chirps)

1. 761-1232 ms: 471 ms
   "UCDC". Noisy.
   F0 @ 846 ms = 977 Hz (peak)
   F0 @ 1046 ms = 630 Hz.

2. 1814-2206 ms: 392 ms
   "UCDC".
   F0 @ 1897 ms = 977 Hz (peak)
   F0 @ 2089 ms = 630 Hz.

3. 2938-3488 ms: 550 ms
   "UCDC". Harmonics do not rise as high or as steeply as in 1 and 2.
   F0 @ 3030 ms = 882 Hz (peak)
   F0 @ 3287 ms = 662 Hz.

4. 4540-5032 ms: 492 ms
   "UCDC". Three visible harmonics.
   F0 @ 4651 ms = 914 Hz (peak)
   F0 @ 4831 ms = 662 Hz.
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

5. 6865-7343 ms: 470 ms
"UCDC". Three harmonics visible.

F0 @ 6935 ms = 977 Hz (peak)
F0 @ 7172 ms = 630 Hz.

6. 9360-9957 ms: 597 ms
"UCDC". Prolonged "C" section. ~9 harmonics visible. Second harmonic is darkest.

F0 @ 9444 ms = 504 Hz
F0 @ 9687 ms = 725 Hz.

7. 10352-10914 ms: 562 ms
"UCDC". Second harmonic appears most prominent.

F0 @ 10467 ms = 819 Hz (peak)
F0 @ 10617 ms = 662 Hz.

8. 11276-11769 ms: 493 ms
"UCDCDC".

F0 @ 11380 ms = 882 Hz (peak)
F0 @ 11560 ms = 725 Hz.
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

9. 12702-13404 ms: 702 ms “UCDCUCDC”.
   F0 @ 12713 ms = 441 Hz
   F0 @ 13034 ms = 662 Hz

10. 14493-15030 ms: 537 ms “UCDCDC”.
    F0 @ 14587 ms = 882 Hz (peak)
    F0 @ 14793 ms = 692 Hz.

11. 16100-16549 ms: 449 ms “UCDCDC”.
    F0 @ 16172 ms = 914 Hz (peak)
    F0 @ 16388 ms = 662 Hz.

12. 17900-18448 ms: 548 ms “UC”.
    F0 @ 18104 ms = 725 Hz.
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

13. 18936-19407 ms: 471 ms
   "UCDC".
   F0 @ 19056 ms = 819 Hz (peak)
   F0 @ 19227 ms = 662 Hz.

14. 21109-21452 ms: 343 ms
   "UCDC".
   F0 @ 21198 ms = 945 Hz (peak)
   F0 @ 21389 ms = 662 Hz.

15. 22906-23408 ms: 502 ms
   "UCDC". 3 noisy harmonics at beginning, 10 harmonics with vibrato near the end, and 2 weaker harmonics at the end.
   F0 @ 22992 ms = 851 Hz (peak)
   F0 @ 23207 ms = 693 Hz.

16. 24012-24481 ms: 469 ms
   "UCDC". Noisy.
   F0 @ 24101 ms = 945 Hz (peak)
   F0 @ 24314 ms = 693 Hz.
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

17. 25679-26133 ms: 454 ms
   “UCDC”. Noisy.
   F0 @ 25749 ms = 977 Hz (peak)
   F0 @ 25980 ms = 693 Hz.

18. 26680-27239 ms: 559 ms
   “UCDC”.
   F0 @ 26798 ms = 945 Hz (peak)
   F0 @ 27011 ms = 662 Hz.

19. 27846-28339 ms: 493 ms
   “UCDC”.
   F0 @ 27943 ms = 945 Hz (peak)
   F0 @ 28201 ms = 630 Hz.

20. 30386-30849 ms: 463 ms
   “CUC”.
   F0 @ 30491 ms = 725 Hz (peak)
   F0 @ 30648 ms = 630 Hz.
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

21. 32417-32961 ms: 544 ms
   “UCDCDC”. Noisy.
   
   F0 @ 32508 ms = 882 Hz (peak)
   F0 @ 32781 ms = 693 Hz.

22. 33857-34349 ms: 492 ms
   “UCDC”. Not as noisy.
   
   F0 @ ~33940 ms = 882 Hz (peak)
   F0 @ 34182 ms = 693 Hz.

23. 35201-35638 ms: 437 ms
   “UCDC”.
   
   F0 @ 35294 ms = 851 Hz (peak)
   F0 @ 35429 ms = 725 Hz.

24. 36855-37217 ms: 362 ms
   “UCDC”. Noisy.
   
   F0 @ 36913 ms = 1040 Hz (peak)
   F0 @ 37108 ms = 662 Hz.
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

25. 37868-38323 ms: 455 ms “UCDC”. Noisy.
   F0 @ 37933 ms = 1008 Hz (peak)
   F0 @ 38151 ms = 693 Hz.

26. 40088-40671 ms: 583 ms “UCDC”.
   F0 @ 40368 = 725 Hz.

27. 41709-42339 ms: 630 ms “UCDC”.
   F0 @ 41880 ms = 819 Hz (peak)
   F0 @ 42153 ms = 630 Hz.

28. 42878-43189 ms: 311 ms “UCDC”.
   F0 @ 42943 ms = 945 Hz (peak)
   F0 @ 43126 ms = 630 Hz.
Appendix 2: Cheetah Vocalization Summaries and Spectrograms

29. 44501-45017 ms: 516 ms
   “UCDC”.
   
   F0 @ 44617 ms = 882 Hz (peak)
   F0 @ 44833 ms = 662 Hz.

30. 45572-46036 ms: 464 ms
    “UCDC”.
    
    F0 @ 45668 ms = 945 Hz (peak)
    F0 @ 45850 ms = 693 Hz.
function frqresp = BCL(f, swch)
%function frqresp = BCL(f, swch)
% % uses six element (left) bobcat ear model to generate the frequency
% % of the transfer function of pressure from the ear canal to the tympanic
% % cavity, evaluated at pre-specified frequencies.
% % % swch = 0 means Rf is not frequency dependent, whereas
% % swch = 1 means it is proportional to sqrt(f)
% % (for comparison purposes)
% % % MODEAR allows you to choose which ear model to use.
% % 0 = use intact ear model.
% % 1 = plugged foramen (just tympanic cavity)
% % 2 = removed septum (tympanic and bullar cavities are combined)
% % 3 = modify the Rf & Mf values to alter the bandwidth & freq. of the dip.

global MODEAR% global variable needs to be set somewhere else
global RFMOD
global MFMOD
Ts = 1/16129.030273;
w = 2*pi*f;
N = length(f);

% left ear of bobcat:

%%---------------
Ctc = 2.57e-12; %Annette Taberner, 1999 (AT)
Cbc = 9.367e-12; %AT
Ctoc = 5.305e-12; %<---- approximated from Huang 1997 graph

Mf = 1622; % 'free parameter'
R = 5e4; % 'free parameter'
Rf = (1-swch)*R + swch*(R.* sqrt(f));
Rtoc = 0.055e9;%<---- approximated from Huang 1997 graph

%---------------
% impedances

Zmf = j*w*Mf;
Zbc = 1./(j*w*Cbc);
Ztc = 1./(j*w*Ctc);
Rtoc = 1./(j*w*Ctoc);

switch MODEAR
  case 0 % NO CHANGE
    Zl = Zbc + Rf + Zmf;
    Zcav = (Zl .* Ztc) ./ (Zl + Ztc);
  case 1 % PLUGGED FORAMEN: just tympanic cavity
    Zcav = Ztc;
  case 2 % REMOVED SEPTUM: mass and resistance are zero
    Zcav = (Ztc .* Zbc) ./ (Ztc + Zbc);
  case 3 % CHANGE RESISTOR AND MASS VALUES
    R = RFMOD;% 'free parameter': change w/ scale factors
    Rf = (1-swch)*R + swch*(R.* sqrt(f));
    Zmf = j*w*MFMOD;
    Zl = Zbc + Rf + Zmf;
    Zcav = (Zl .* Ztc) ./ (Zl + Ztc);
end

Z2 = Rtoc + Ztoc;
frqresp = 1 - Zcav ./ (Zcav + Z2);
BCR.m

function frqresp = BCR(f, swch)
  %function frqresp = BCR(f, swch)
  %
  % uses six element (right) bobcat ear model to generate the frequency
  % response
  % of the transfer function of pressure from the ear canal to the tympanic
  % cavity, evaluated at pre-specified frequencies.
  %
  % swch = 0 means Rf is not frequency dependent, whereas
  % swch = 1 means it is proportional to sqrt(f)
  % (for comparison purposes)
  %
  % MODEAR allows you to choose which ear model to use.
  % 0 = use intact ear model.
  % 1 = plugged foramen (just tympanic cavity)
  % 2 = removed septum (tympanic and bullar cavities are combined)
  % 3 = modify the Rf & Mf values to alter the bandwidth & freq. of the dip.

  global MODEAR% global variable needs to be set somewhere else
  global RFMOD
  global MFMOD

  Ts = 1/16129.030273;
  w = 2*pi*f;
  N = length(f);

  % right ear of bobcat:
  %
  Ctc = 2.28e-12; %Annette Taberner, 1999 (AT)
  Cbc = 1.6e-11; %AT
  Ctoc = 5.305e-12; %approximated from Huang 1997 graph

  Mf = 1622;%AT
  R = 5e4; %AT
  Rf = (1-swch)*R + swch*(R.*sqrt(f));
  Rtoc = 0.055e9;%approximated from Huang 1997 graph

  global MODEAR% global variable needs to be set somewhere else
  global RFMOD
  global MFMOD

  Ts = 1/16129.030273;
  w = 2*pi*f;
  N = length(f);

  switch MODEAR
  case 0 % NO CHANGE
    Z1 = Zbc + Rf + Zmf;
    Zcav = (Z1 .* Ztc) ./ (Z1 + Ztc);
  case 1 % PLUGGED FORAMEN: just tympanic cavity
    Zcav = Ztc;
  case 2 % REMOVED SEPTUM: mass and resistance are zero
    Zcav = (Ztc .* Zbc) ./ (Ztc + Zbc);
  case 3 % CHANGE RESISTOR VALUE
    R = RFMOD;% 'free parameter': change w/ scale factors
    Rf = (1-swch)*R + swch*(R.*sqrt(f));
    Rtoc = 0.055e9; %approximated from Huang 1997 graph
    Z2 = Rtoc + Ztoc;

    frqresp = 1 - Zcav ./ (Zcav + Z2);
  end

  % impedances
  Zmf = j*w*Mf;
  Zbc = 1./(j*w*Cbc);
  Ztc = 1./(j*w*Ctc);
  Ztoc = 1./(j*w*Ctoc);

  switch MODEAR
function frqresp = CHL(f, swch)
% function frqresp = CHL(f, swch)
% uses six element (left) cheetah ear model to generate the frequency
% response
% of the transfer function of pressure from the ear canal to the tympanic
% cavity, evaluated at pre-specified frequencies.
% swch = 0 means Rf is not frequency dependent, whereas
% swch = 1 means it is proportional to sqrt(f)
% (for comparison purposes)
% MODEAR allows you to choose which ear model to use.
% 0 = use intact ear model.
% 1 = plugged foramen (just tympanic cavity)
% 2 = removed septum (tympanic and bullar cavities are combined)
% 3 = modify the Rf & Mf values to alter the bandwidth & freq. of the dip.

global MODEAR
global RFMOD
global MFMOD

Ts = 1/16129.030273;
w = 2*pi*f;
N = length(f);

%%%%%%
% left ear of cheetah:
%%%%%%
%Annette Taberner, 1999 (AT)
Ctc = 0.7454e-6 / 1.4e5;%AT
Cbc = 3.4374e-6 / 1.4e5;%AT
Ctoc = 3.3e-12; %AT

Mf = 1380;  % 'free parameter'
R = 3e4;    % 'free parameter'
Rf = (1-swch)*R + swch*(R.*sqrt(f));
Rtoc = 2e7; %AT

% impedances
Zmf = j*w*Mf;
Zbc = 1./(j*w*Cbc);
Ztc = 1./(j*w*Ctc);
Ztoc = 1./(j*w*Ctoc);

switch MODEAR
  case 0  % NO CHANGE
    Z1 = Zbc + Rf + Zmf;
    Zcav = (Z1 .* Ztc) ./ (Z1 + Ztc);
  case 1  % PLUGGED FORAMEN: just tympanic cavity
    Zcav = Ztc;
  case 2  % REMOVED SEPTUM: mass and resistance are zero
    Zcav = (Ztc .* Zbc) ./ (Ztc + Zbc);
  case 3  % CHANGE RESISTOR VALUE
    R = RFMOD;% 'free parameter': change w/ scale factors
    Rf = (1-swch)*R + swch*(R.*sqrt(f));
    Zmf = j*w*MFMOD;
    Z1 = Zbc + Rf + Zmf;
    Zcav = (Z1 .* Ztc) ./ (Z1 + Ztc);
  end

Z2 = Rtoc + Ztoc;
frqresp = 1 - Zcav ./ (Zcav + Z2);
function frqresp = CHR(f, swch)

% uses six element (right) cheetah ear model to generate the frequency response of the transfer function of pressure from the ear canal to the tympanic cavity, evaluated at pre-specified frequencies.

% swh = 0 means Rf is not frequency dependent, whereas swh = 1 means it is proportional to sqrt(f)
% (for comparison purposes)

% MODEAR allows you to choose which ear model to use.
% 0 = use intact ear model.
% 1 = plugged foramen (just tympanic cavity)
% 2 = removed septum (tympanic and bullar cavities are combined)
% 3 = modify the Rf & Mf values to alter the bandwidth & freq. of the dip.

global MODEAR % global variable needs to be set somewhere else
global RFMOD
global MFMOD

Ts = 1/16129.030273;
w = 2*pi*f;
N = length(f);

% right ear of cheetah:
% Annette Taberner, 1999 (AT)

Ctc = 0.6974e-6 / 1.4e5;%AT(VO / (rho0*c^2))
Cbc = 3.4842e-6 / 1.4e5;%AT
Ctoc = 3.61e-12;%AT

Mf = 1380; % 'free parameter'
R = 3e4; % 'free parameter'
Rf = (1-swhc)*R + swhc*(R .* sqrt(f));
Rtoc = 2e7; %AT

%-----------------------------
% impedances

Zmf = j*w*Mf;
Zbc = 1./(j*w*Cbc);
Ztc = 1./(j*w*Ctc);
Ztoc = 1./(j*w*Ctoc);

switch MODEAR
  case 0 % NO CHANGE
    Z1 = Zbc + Rf + Zmf;

  case 1 % PLUGGED FORAMEN: just tympanic cavity
    Zcav = (Z1 .* Ztc) ./ (Z1 + Ztc);

  case 2 % REMOVED SEPTUM: mass and resistance are zero
    Zcav = (Ztc .* Zbc) ./ (Ztc + Zbc);

  case 3 % CHANGE RESISTOR VALUE
    R = RFMOD;% 'free parameter': change w/ scale factors
    Rf = (1-swhc)*R + swhc*(R .* sqrt(f));
    Zmf = j*w*MFMOD;
    Z1 = Zbc + Rf + Zmf;
    Zcav = (Z1 .* Ztc) ./ (Z1 + Ztc);
end

Z2 = Rtoc + Ztoc;
frqresp = 1 - Zcav ./ (Zcav + Z2);
BCear.m

% M-file for generating impulse responses of Bobcat left and right middle-ear % models, with options for generating various plots.

% Axis variables, for uniformity
global BodeAxesM BodeAxesP LinAxesM LinAxesP
BodeAxesM = [.1 10 -20 5]; % magnitude plot axes
BodeAxesP = [.1 10 -.3 .3]; % phase plot axes
LinAxesM = [0 8050 0 2];
LinAxesP = [0 8050 -.3 .3];

% GLOBAL VARIABLES:
%global MODEAR    MODEAR = 0;
global RFMOD    RFMOD = RF_ORIG; % only used when MODEAR = 3
% global HFMOD    HFMOD = HF_ORIG; % *
% PCT_chg = 0.2;

% The sampling period is set according to how the relevant vocalizations % were digitized. The impulse responses generated here need to be compatible % with those recordings.
Ts = 1/16129.030273;

% Number of points in FFT and impulse response
N = 1024;

% SWITCHES FOR GENERATING PLOTS
LINplots = 1;
LvsR = 1;
FrqR = 0;
TruncComp = 1;
Sept = 1;
Dip = 0;
BodeScomp = 1;
FinalBode = 1;
BodeSept = 1;
BodeDip = 0;
BodeFrqR = 0;

IMPLS = 1; % impulse response plots
IMAGsize = 0; % print out the average size of the imaginary part of the % impulse responses, to see if there's a problem with % symmetry. when it's on the order of 1e-17, it's probably % just due to roundoff error. i take the real part for the % final impulse responses.

% Switches for truncation and windowing
IMPLSlengh = N/2; % length of impulse responses (truncated to that length) % must be less than or equal to N.

% In selecting the frequency vector, evaluate it from -.5Fs to +.5Fs with % N+1 elements (an odd number) so that the DC component occurs on a sample, % and to create sample points evenly spaced around the entire unit circle.
f = linspace(-1/(2*Ts), 1/(2*Ts), N+1);

% Since the frequency vector passes through zero, it's necessary to replace % the zero element with a very small number to prevent division by zero errors.
mask = (f == 0);
mask = mask .* 1e-100;
f = f + mask;

% Get rid of the frequency at +.5Fs because we want a power of 2 elements
f = f(1:N);
w = 2*pi*f;

%%%%% % left & right bobcat ears
%%%%% % First, frequency-dependent resistor (switch = 1)
BCLlin1 = BCL2(f, 1);
BCRlin1 = BCR2(f, 1);

% Non-frequency-dependent resistor (switch = 0)
BCLlin2 = BCL2(f, 0);
BCRlin2 = BCR2(f, 0);

% For the purposes of generating odd symmetry in the phase, the imag. part % of the freq. resp. at the -pi frequency needs to be set to zero. it's % normally the site of a discontinuity in the phase, and since we're sup- % posed % to be bandlimited anyhow, this shouldn't be problematic and will allow % the resulting impulse response to be real!
With the $\sqrt{f}$ term in the model with the frequency-dependent resistor, the frequency response loses its symmetry with positive and negative frequencies. For this reason, and since we feel that this model works well for the positive frequencies, I'm going to replace the freq. resp. values at negative frequencies with their respective conjugates of the positive frequency values.

```matlab
BCLlin1(N:-1:(N/2+2)) = conj(BCLlin1(2:1:(N/2)));  
BCRlin1(N:-1:(N/2+2)) = conj(BCRlin1(2:1:(N/2)));
```

So far the vectors have spanned from a normalized frequency of $-\pi$ to just before the normalized frequency of $+\pi$. The ifft function expects to see the DC component first, so the two halves of the vectors need to be exchanged. Since the first element in the vector corresponds to $-\pi$, and there are a power of 2 elements, element $N/2$ is just before DC and $N/2 + 1$ corresponds to DC. Therefore, swapping the two halves as-is is what I want. I've used ifftshift for this, though fftshift will also work, it seems.

```matlab
BCLlin1 = ifftshift(BCLlin1);  
BCRlin1 = ifftshift(BCRlin1);  
BCLlin2 = ifftshift(BCLlin2);  
BCRlin2 = ifftshift(BCRlin2);  
```

% Take inverse transforms.

```matlab
BCLimpls1 = ifft(BCLlin1);  
BCLimpls2 = ifft(BCLlin2);  
BCRimpls1 = ifft(BCRlin1);  
BCRimpls2 = ifft(BCRlin2);
```

% Optionally display a measure of the size of the imaginary parts of the impulse responses.

```matlab
if(IMAGsize)  
  Avg_Imag_Part_Left_Earl = mean(abs(imag(BCLimpls1)));  
  Avg_Imag_Part_Left_Earl2 = mean(abs(imag(BCLimpls2)));  
  Avg_Imag_Part_Right_Earl = mean(abs(imag(BCRimpls1)));  
  Avg_Imag_Part_Right_Earl2 = mean(abs(imag(BCRimpls2)));  
end
```

% Wipe out imag parts, which should be very small, and truncate.

```matlab
BCLimpls1 = real(BCLimpls1(1:IMPLSlength));  
BCLimpls2 = real(BCLimpls2(1:IMPLSlength));  
BCRimpls1 = real(BCRimpls1(1:IMPLSlength));  
BCRimpls2 = real(BCRimpls2(1:IMPLSlength));  
```
grid on
xlabel('Frequency (Hz)')
legend('Original', 'Truncated', '4')
end

if(LvsR)
    ctr = ctr+2;
    figure(ctr)
    subplot(2,1,1), plot(findex, abs(BCLlin1(1:(N/2+1))), '-')
    grid on
    ylabel('|H(s)|')
    hold on
    subplot(2,1,1), plot(findex, abs(BCRlin1(1:(N/2+1))), '--')
    hold off
    axis(LinAxesM)
    title('Comparison of Left and Right Bobcat Middle-Ear Models (Linear Plot)')
    grid on
    subplot(2,1,2), plot(findex, (1/(2*pi))*angle(BCLlin1(1:(N/2+1))), 'hold on
    subplot(2,1,2), plot(findex, (1/(2*pi))*angle(BCRlin1(1:(N/2+1))), axis(LinAxesP)
    ylabel('Phase (periods)')
    grid on
    xlabel('Frequency (Hz)')
    legend('Left Ear', 'Right Ear', '4')
end

if(FrqR)
    ctr = ctr+2;
    figure(ctr)
    subplot(2,1,1), plot(findex, abs(BCLlin1(1:(N/2+1))), '-')
    grid on
    ylabel('|H(s)|')
    hold on
    subplot(2,1,1), plot(findex, abs(BCLlin2(1:(N/2+1))), '--')
    hold off
    axis(LinAxesM)
    title('Effects of Using a Frequency-Dependent Foramen Resistance (Left Bobcat Ear, Linear Plot)')
    grid on
    subplot(2,1,2), plot(findex, (1/(2*pi))*angle(BCLlin1(1:(N/2+1))), 'hold on
    subplot(2,1,2), plot(findex, (1/(2*pi))*angle(BCLlin2(1:(N/2+1))), 'hold off
    axis(LinAxesP)
    ylabel('Phase (periods)')
    grid on
    xlabel('Frequency (Hz)')
    legend('Original Model', '4')
end

if(Sept)
    ctr = ctr+2;
    figure(ctr)
    subplot(2,1,1), plot(findex, abs(BCLlin1(1:(N/2+1))), '-')
    grid on
    ylabel('|H(s)|')
    hold on
    subplot(2,1,1), plot(findex, abs(BCLtcLin(1:(N/2+1))), '-')
    hold off
    axis(LinAxesM)
    title('Effects of Removing the Septum and Plugging the Foramen (Left Bobcat Ear, Linear Plot)')
    grid on
    subplot(2,1,2), plot(findex, (1/(2*pi))*angle(BCLlin1(1:(N/2+1))), 'hold on
    subplot(2,1,2), plot(findex, (1/(2*pi))*angle(BCLtcLin(1:(N/2+1))), 'hold off
    axis(LinAxesP)
    ylabel('Phase (periods)')
    grid on
    xlabel('Frequency (Hz)')
    legend('Original Model', 'Plugged Foramen', 'Removed Septum', '4')
end

if(Dip)
    ctr = ctr+2;
    figure(ctr)
    subplot(2,1,1), plot(findex, abs(BCLlin1(1:(N/2+1))), '-')
    grid on
    ylabel('|H(s)|')
    hold on
    subplot(2,1,1), plot(findex, abs(BCLlmLin(1:(N/2+1))), 'hold on
    subplot(2,1,1), plot(findex, abs(BCLhmLin(1:(N/2+1))), 'hold off
    axis(LinAxesM)
    title('Effects of Increasing and Decreasing the Foramen Mass (Left Bobcat Ear, Linear Plot)')
    grid on
    subplot(2,1,2), plot(findex, (1/(2*pi))*angle(BCLlin1(1:(N/2+1))), 'hold on
    subplot(2,1,2), plot(findex, (1/(2*pi))*angle(BCLlmLin(1:(N/2+1))), 'hold off
    axis(LinAxesP)
    ylabel('Phase (H(s))')
    grid on
    xlabel('Frequency (Hz)')
    legend('Original Model', 'Mf reduced by 20%', 'Mf increased by 20%', '4')
end
if (BODEcomp)
    MODEAR = 0;
    f_log = logspace(2, 4, N);
    w_log = 2*pi*f_log;
    % With Freq-Dep. Foramen Resistance
    BCLlog1 = BCL2(f_log, 1);
    BCRlog1 = BCR2(f_log, 1);
    % With NON-Freq-Dep. Foramen Resistance
    BCLlog2 = BCL2(f_log, 0);
    BCRlog2 = BCR2(f_log, 0);
    % To use the 'bode' function, I need to structures using 'frd'.
    sysBCLlog1 = frd(BCLlog1, w_log, Ts);
    sysBCLlog2 = frd(BCLlog2, w_log, Ts);
    sysBCRlog1 = frd(BCRlog1, w_log, Ts);
    sysBCRlog2 = frd(BCRlog2, w_log, Ts);
% septum modifications:
    MODEAR = 1; % plugged foramen
    BCLtcLog1 = BCL2(f_log, 1);
    sysBCLtcLog1 = frd(BCLtcLog1, w_log, Ts);
    sysBCLnsLog1 = frd(BCLnsLog1, w_log, Ts);
% dip modifications
    MODEAR = 3;
    MFMOD = MF_ORIG*(1-PCT_CHG); % reduce MF by PCT_CHG
    BCLlmLog1 = BCL2(f_log, 1); % 'lower MF'
    BCLhmLog1 = BCL2(f_log, 1); % 'higher MF'
    sysBCLlmLog1 = frd(BCLlmLog1, w_log, Ts);
    sysBCLhmLog1 = frd(BCLhmLog1, w_log, Ts);
if (BodeSept)
    ctr = ctr+2;
    figure(ctr)
    % LEFT EAR, comparison of septum configurations
    [MAG1, PHASE1] = bode(sysBCLnsLog1, w_log);
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG1)))
    hold on
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG2)), '--')
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG3)), '-')
    title('Effects of Removing the Septum and Plugging the Foramen (Left Bobcat Ear, Bode Plot)')
    grid on
    ylabel('Magnitude (dB)')
    axis([BodeAxes])
    subplot(2,1,2), semilogx(f_log/1000, squeeze(PHASE1)/360)
    hold on
    subplot(2,1,2), semilogx(f_log/1000, squeeze(PHASE2)/360, '--')
    subplot(2,1,2), semilogx(f_log/1000, squeeze(PHASE3)/360, '-')
    legend('Original ', 'Plugged Foramen ', 'Removed Septum ', 3)
end
if (BodeDip)
    ctr = ctr+2;
    figure(ctr)
% LEFT EAR, comparison of dip locations from a change in MF.
% assuming Rf stays the same for now.
    [MAG1, PHASE1] = bode(sysBCLlog1, w_log);
    [MAG2, PHASE2] = bode(sysBCLlmLog1, w_log);
    [MAG3, PHASE3] = bode(sysBCLhmLog1, w_log);
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG1)))
    hold on
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG2)), '--')
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG3)), '-')
    title('Effects of Increasing and Decreasing the Foramen Mass (Left Bobcat Ear, Bode Plot)')
    grid on
    ylabel('Magnitude (dB)')
    axis([BodeAxes])
    subplot(2,1,2), semilogx(f_log/1000, squeeze(PHASE1)/360)
    hold on
    subplot(2,1,2), semilogx(f_log/1000, squeeze(PHASE2)/360, '--')
    subplot(2,1,2), semilogx(f_log/1000, squeeze(PHASE3)/360, '-')
    grid on
    ylabel('Phase (periods)')
    xlabel('Frequency (kHz)')
    legend('Original ', 'Mf reduced by 20% ', 'Mf increased by 20% ', 3)
end
if(FinalBode)
    ctr = ctr+2;
    figure(ctr)
    % LEFT EAR
    [MAG1, PHASE1] = bode(sysBCLlogl, w_log);
    [MAG2, PHASE2] = bode(sysBCRlogl, w_log);
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG1)));
    hold on
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG2)), '--')
    title('Comparison of Left and Right Bobcat Middle-Ear Models (Bode Plot)')
    axis(BodeAxesM)
    ylabel('Magnitude (dB)')
    subplot(2,1,2), semilogx(f_log/1000, squeeze(PHASE1)/360)
    hold on
    subplot(2,1,2), semilogx(f_log/1000, squeeze(PHASE2)/360, '--')
    ylabel('Phase (periods)')
    xlabel('Frequency (kHz)')
    legend('Left Ear ', 'Right Ear ', 3)
end

if(BodeFrqR)
    ctr = ctr+2;
    figure(ctr)
    % LEFT EAR
    [MAG1, PHASE1] = bode(sysBCLlogl, w_log);
    [MAG2, PHASE2] = bode(sysBCRlogl, w_log);
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG1)));
    hold on
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG2)), '--')
    title('Effects of Using a Frequency-Dependent Foramen Resistance (Left Bobcat Ear, Bode Plot)')
    grid on
    ylabel('Magnitude (dB)')
    subplot(2,1,2), semilogx(f_log/1000, squeeze(PHASE1)/360)
    hold on
    subplot(2,1,2), semilogx(f_log/1000, squeeze(PHASE2)/360, '--')
    ylabel('Phase (periods)')
    xlabel('Frequency (kHz)')
    legend('Frequency-Dependent R_f ', 'Constant R_f ', 3)
end
end

if(IMPLS) % run the cheetah script beforehand.
    ctr = ctr+2;
    figure(ctr)
    M = 100; IMPLSlength/8; \%/4;
    t = 0:T_s:((M-1)*T_s);
    subplot(2,1,1), stem(t*1000, BCLimplsl(1:M))
    axis([-1 6.2 -.5 1])
    title('Impulse Responses of the Cavity Gain for Bobcat (top) and Cheetah (bottom), First 100 points')
    ylabel('Amplitude')
    subplot(2,1,2), stem(t*1000, CHLimplsl(1:M))
    ylabel('Amplitude')
end
end

% Plot both bobcat and cheetah impulse responses:
CHear.m

% M-file for generating impulse responses of Cheetah left and right middle-ear
% models, with options for generating various plots.

% Axes for phase on linear plots:
chLinAxesP = [0 8050 *.4 .4];

% GLOBAL VARIABLES:
global MODEAR
MODEAR = 0;
global RFMOD
RFMOD = 5e4; % only used when MODEAR = 3
global MFMOD
MFMOD = MFMOD; %

% The sampling period is set according to how the relevant vocalizations
% were digitized. The impulse responses generated here need to be compatibl-
% with those recordings.
Ts = 1/16129.030273;

% Number of points in FFT and impulse response
N = 1024;

% SWITCHES FOR GENERATING PLOTS
LINplots = 1;
LinRs = 1;
Freq = 0;
TruncCmp = 1;
Sept = 1;
Dip = 0;
BODEcomp = 1;
FinalBode = 1;
BodeSept = 1;
BodeDip = 0;
BodeFreq = 0;

IMAGSize = 0; % print out the average size of the imaginary part of the
% impulse responses, to see if there's a problem with
% symmetry. when it's on the order of 1e-17, it's probably
% just due to roundoff error. I take the real part for the
% final impulse responses.

IMPLSlength = N/2; % length of impulse responses (truncated to that length)
% must be less than or equal to N.

% In selecting the frequency vector, evaluate it from -.5Fs to +.5Fs with
% N+1 elements (an odd number) so that the DC component occurs on a sample,
% and to create sample points evenly spaced around the entire unit circle.
f = linspace(-1/(2*Ts), 1/(2*Ts), N+1);
% Since the frequency vector passes through zero, it's necessary to replace
% the zero element with a very small number to prevent division by zero
errors.
mask = (f == 0);
mask = mask .* 1e-100;
f = f + mask;

% Get rid of the frequency at +.5Fs because we want a power of 2 elements
f = f(1:N);
w = 2*pi*f;

%%%%
% left & right Cheetah ears
%%%%

% First, frequency-dependent resistor (switch = 1)
CHLlin1 = CHL2(f, 1);
CHRlin1 = CHR2(f, 1);

% Non-frequency-dependent resistor (switch = 0)
CHLlin2 = CHL2(f, 0);
CHRlin2 = CHR2(f, 0);

% For the purposes of generating odd symmetry in the phase, the imag. part
% of the freq. resp. at the -pi frequency needs to be set to zero. It's
% normally the site of a discontinuity in the phase, and since we're sup-
% posed
% to be bandlimited anyhow, this shouldn't be problematic and will allow
% the resulting impulse response to be real!
CHLlin1(1) = 0; % real(CHLlin1(1));
CHRlin1(1) = 0; % real(CHRlin1(1));

CHLlin2(1) = 0; % real(CHLlin2(1));
CHRlin2(1) = 0; % real(CHRlin2(1));

% With the sqrt(f) term in the model with the frequency-dependent resistor,
% the frequency response loses its symmetry with positive and negative
% frequencies. For this reason, and since we feel that this model works
% well for the positive frequencies, I'm going to replace the freq. resp.
% values at negative frequencies with their respective conjugates of the
% positive frequency values.
CHLlin1(N-1; N/2+2) = conj(CHLlin1(2; N/2+2));
CHRlin1(N-1; N/2+2) = conj(CHRlin1(2; N/2+2));
So far the vectors have spanned from a normalized frequency of -π to just before the normalized frequency of +π. The ifft function expects to see the DC component first, so the two halves of the vectors need to be exchanged. Since the first element in the vector corresponds to -π, and there are a power of 2 elements, element N/2 is just before DC and N/2 + 1 corresponds to DC. Therefore, swapping the two halves as-is is what I want. I've used ifftshift for this, though fftshift will also work, it seems.

CHLlin1 = ifftshift(CHLlin1);
CHRlin1 = ifftshift(CHRlin1);
CHLlin2 = ifftshift(CHLlin2);
CHRlin2 = ifftshift(CHRlin2);

% Take inverse transforms.
CHLimplsl = ifft(CHLlinl);
CHLimpls2 = ifft(CHLlin2);
CHRimplsl = ifft(CHRlinl);
CHRimpls2 = ifft(CHRlin2);

% Optionally display a measure of the size of the imaginary parts of the impulse responses.
if(IMAGsize)
    Avg_Impart_Left_Earl = mean(abs(imag(CHLimplsl)));
    Avg_Impart_Left_Ear2 = mean(abs(imag(CHLimpls2)));
    Avg_Impart_Right_Earl = mean(abs(imag(CHRimplsl)));
    Avg_Impart_Right_Ear2 = mean(abs(imag(CHRimpls2)));
end

% Wipe out imag parts, which should be very small, and truncate.
CHLimplsl = real(CHLimplsl(1:IMPLSlength));
CHLimpls2 = real(CHLimpls2(1:IMPLSlength));
CHRimplsl = real(CHRimplsl(1:IMPLSlength));
CHRimpls2 = real(CHRimpls2(1:IMPLSlength));

% Generate and display graphs according to switches set above.
CHLlin1trunc = fft(CHLimplsl,N);
CHRlin1trunc = fft(CHRimplsl,N);

findex = linspace(0, 1/(2*Ts), (N/2+1));
findex(1) = 1e-100;

% septum modifications:
MDEAR = 2; % removed septum
CHLtcLin = CHL2(findex, 1);

% dip modifications
MDEAR = 3;
MFMOD = MFORIG*(1-PCTCHG); % reduce MF by PCTCHG
CHLlmLin = CHL2(findex, 1); % "lower MF"

MFMOD = MFORIG*(1+PCTCHG); % raise MF by PCTCHG
CHLhmLin = CHL2(findex, 1); % "higher MF"

ctr = 1;
if(LINplots)
    if(TruncCmp)
        figure(ctr)
        subplot(2,1,1), plot(findex, abs(CHLlin1(1:(N/2+1))), '-')
        grid on
        ylabel(' H(s)I')
        hold on
        subplot(2,1,1), plot(findex, abs(CHLlin1trunc(1:(N/2+1))), '--')
        hold off
        title('Effects of Impulse Response Truncation (Left Cheetah Ear, Linear Plot)')
        axis(LinAxesM)
    else
        subplot(2,1,1), plot(findex, abs(CHLlin1(1:(N/2+1))), '-')
        grid on
        ylabel(' H(s)I')
        hold on
        subplot(2,1,1), plot(findex, abs(CHLlin1trunc(1:(N/2+1))), '--')
        hold off
        title('Comparison of Left and Right Cheetah Middle-Ear Models (Linear Plot)')
    end
    if(LvsR)
        ctr = ctr + 2;
        figure(ctr)
        subplot(2,1,1), plot(findex, abs(CHLlin1(1:(N/2+1))), '-')
        grid on
        ylabel(' H(s)I')
        hold on
        subplot(2,1,1), plot(findex, abs(CHLlin1trunc(1:(N/2+1))), '--')
        hold off
        axis(LinAxesM)
        title('Comparison of Left and Right Cheetah Middle-Ear Models (Linear Plot)')
    end
end
hold on
subplot(2,1,2), plot(findex, (1/(2*pi))*angle(CHRlin1(1:(N/2+1))), 'r--')
axis(chLinAxesP)
hold off
xlabel('Frequency (Hz)')
ylabel('Phase (periods)')
grid on
legend('Left Ear ', 'Right Ear ', 4)
end

if(FrqR)
ctr = ctr + 2;
figure(ctr)
subplot(2,1,1), plot(findex, abs(CHLlin1((N/2+1))), 'r--')
hold on
subplot(2,1,1), plot(findex, abs(CHLlin2((N/2+1))), 'g-.')
title('Effects of Using a Frequency-Dependent Foramen Resistance (Left Cheetah Ear, Linear Plot)')
hold off
ylabel(' H(s)|')
grid on
subplot(2,1,2), plot(findex, (1/(2*pi))*angle(CHLlin1(1:(N/2+1))), 'r--')
hold on
subplot(2,1,2), plot(findex, (1/(2*pi))*angle(CHLlin2(1:(N/2+1))), 'g-.')
hold off
ylabel('Phase (periods)')
grid on
xlabel('Frequency (Hz)')
legend('Frequency-Dependent R_f  ', 'Constant Rf ', 4)
end

if(Sept)
ctr = ctr+2;
figure(ctr)
subplot(2,1,1), plot(findex, abs(CHLlin1(1:(N/2+1))), 'r--')
hold on
subplot(2,1,1), plot(findex, abs(CHLtcLin(1:(N/2+1))), 'g-.')
hold off
axis(LinAxesM)
title('Effects of Removing the Septum and Plugging the Foramen (Left Cheetah Ear, Linear Plot)')
hold off
ylabel(' H(s)|')
grid on
subplot(2,1,2), plot(findex, (1/(2*pi))*angle(CHLlin1(1:(N/2+1))), 'r--')
hold on
subplot(2,1,2), plot(findex, (1/(2*pi))*angle(CHLtcLin(1:(N/2+1))), 'g-.')
hold off
axis(chLinAxesP)
title('Frequency (Hz)')
legend('Frequency-Dependent R_f  ', 'Constant Rf ', 4)
end
end

if(Dip)
ctr = ctr+2;
figure(ctr)
subplot(2,1,1), plot(findex, abs(CHLlin1(1:(N/2+1))), 'r--')
hold on
subplot(2,1,1), plot(findex, abs(CHLlmLin(1:(N/2+1))), 'g-.')
hold off
axis(LinAxesM)
title('Effects of Increasing and Decreasing the Foramen Mass (Left Cheetah Ear, Linear Plot)')
hold off
ylabel(' H(s)|')
grid on
subplot(2,1,2), plot(findex, (1/(2*pi))*angle(CHLlin1(1:(N/2+1))), 'r--')
hold on
subplot(2,1,2), plot(findex, (1/(2*pi))*angle(CHLlmLin(1:(N/2+1))), 'g-.')
hold off
axis(chLinAxesP)
title('Frequency (Hz)')
legend('Original Model ', 'Mf reduced by 20% ', 'Mf increased by 20% ', 4)
end

if(BODEcomp)
MODEAR = 0;
f_log = logspace(1, 4, N);
w_log = 2*pi*f_log;

% With Freq-Dep. Foramen Resistance
CHLlog1 = CHL2(f_log, 1);
CHRlog1 = CHR2(f_log, 1);

% With NON-Freq-Dep. Foramen Resistance
CHLlog2 = CHL2(f_log, 0);
CHRlog2 = CHR2(f_log, 0);

% To use the 'bode' function, I need to create system function data % structures using 'frd'.
sysCHLlog1 = frd(CHLlog1, w_log, Ts);
sysCHLlog2 = frd(CHLlog2, w_log, Ts);
sysCHRlog1 = frd(CHRlog1, w_log, Ts);
end

end

end

end

end

end

end

end

if(BODEcomp)
end

end

end

end
sysCHRlog2 = frd(CHRlog2, w_log, Ts);

% septum modifications:
MODEAR = 1; % plugged foramen
CHLttcLog2 = CHL2(f_log, 1);
 MODEAR = 2; % removed septum
CHLnsLog2 = CHL2(f_log, 1);

sysCHLttcLog2 = frd(CHLttcLog2, w_log, Ts);
sysCHLnsLog2 = frd(CHLnsLog2, w_log, Ts);

% dip modifications
MODEAR = 3;
MFMOD = MF_ORIG*(1-PCTCHG); % reduce MF by PCTCHG
CHLlmLog2 = CHL2(f_log, 1); % "lower MF"
MFMOD = MF_ORIG*(1+PCTCHG); % raise MF by PCTCHG
CHLhmLog2 = CHL2(f_log, 1); % "higher MF"

sysCHLlmLog2 = frd(CHLlmLog2, w_log, Ts);
sysCHLhmLog2 = frd(CHLhmLog2, w_log, Ts);

if(BodeSept)
    ctr = ctr+2;
ficgure(ctr)
    % LEFT EAR, comparison of septum configurations
    [MAG1, PHASE1] = bode(sysCHLtcLog2, w_log);
    [MAG2, PHASE2] = bode(sysCHLnsLog2, w_log);
    [MAG3, PHASE3] = bode(sysCHLtcLog2, w_log);

    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG1))
    hold on
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG2))
    hold on
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG3))
    hold off
    title('Effects of Removing the Septum and Plugging the Foramen (Left Cheetah Ear, Bode Plot)')
    grid on
    ylabel('Magnitude (dB)')
    axis(BodeAxesM)
end

if(FinalBode)
    ctr = ctr+2;
    figure(ctr)
    % LEFT EAR, comparison of dip locations from a change in MF.
    % assuming MF stays the same for now.
    [MAG1, PHASE1] = bode(sysCHRlog2, w_log);
    [MAG2, PHASE2] = bode(sysCHLlmLog2, w_log);
    [MAG3, PHASE3] = bode(sysCHLhmLog2, w_log);

    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG1))
    hold on
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG2))
    hold on
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG3))
    hold off
    title('Effects of Increasing and Decreasing the Foramen Mass (Left Cheetah Ear, Bode Plot)')
    grid on
    ylabel('Magnitude (dB)')
    axis(BodeAxesM)
end

if(BodeDip)
    ctr = ctr+2;
    figure(ctr)
    % LEFT EAR, comparison of dip locations
    [MAG1, PHASE1] = bode(sysCHLlog2, w_log);
    [MAG2, PHASE2] = bode(sysCHLlmLog1, w_log);
    [MAG3, PHASE3] = bode(sysCHLhmLog1, w_log);

    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG1))
    hold on
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG2))
    hold on
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG3))
    hold off
    title('Comparison of Left and Right Cheetah Middle Ear Models (Bode Plot)')
    grid on
    ylabel('Phase (periods)')
    axis(BodeAxesP)
    legend('Left Ear ', 'Right Ear ', 3)
end
if(BodeFrqR)
    ctr = ctr+2;
    figure(ctr)
    % LEFT EAR
    [MAG1, PHASE1] = bode(sysCHLlog1, w_log);
    [MAG2, PHASE2] = bode(sysCHLlog2, w_log);
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG1)))
    hold on
    subplot(2,1,1), semilogx(f_log/1000, 20*log10(squeeze(MAG2)), '--')
    hold off
    title('Effects of Using a Frequency-Dependent Foramen Resistance (Left Cheetah Ear, Bode Plot)')
    grid on
    ylabel('Magnitude (dB)')
    subplot(2,1,2), semilogx(f_log/1000, squeeze(PHASE1)/360)
    hold on
    subplot(2,1,2), semilogx(f_log/1000, squeeze(PHASE2)/360, '--')
    hold off
    grid on
    ylabel('Phase (periods)')
    xlabel('Frequency (kHz)')
    legend('Frequency-Dependent R_f', 'Constant R_f', 3)
end
end
The files in Appendix 4 are not intended to be run without appropriate modifications. File and directory names, for instance, would need to be changed to match a different environment.

Much of the complexity in filtercats.m is devoted to string manipulations to allow file names to be listed only once at the beginning of the script, and to automatically rename the output files.

filtercats.m

% filtercats makes it easy to filter the cat sounds through various filters.
% The present implementation only works for the bobcats and cheetahs, but it could easily be extended to handle filters from other species.
% IndexBC and indexCH include the row numbers of the cat sounds to be filtered, the paths for reading and writing files can be modified as necessary.
% You must run BCear and CHear (in that order) before running this script.
% You must have properly modified versions of mat2kl.m and kl2mat.m in the same directory for this to work, and raw2kl must be available in your path.
% Created 3/12/00 -KO

INDEXBC = [1 2 3];
cheetahs = strvcat('/cheetahs/CHI.wav', ...
'/cheetahs/CHII.wav', ...
'/cheetahs/CHIII.wav');
CHroot = strvcat('/CHI', '/CHII', '/CHIII');

% modification information for end of name
% norm = filtered normally, plug = filtered with plugged septum
% middle ear, nosept = filtered with removed septum middle ear.
% RF###M$$ mean RF and MF were modified to values ### and $$ respectively.

mods = strvcat('norm', 'plug', 'nosept', ...
strcat('rf', num2str(RFMOD), 'mf', num2str(MFMOD)));
readPATH = ['/usr/users/kevinoc/recordings/filtered/norml', ...
'/usr/users/kevinoc/recordings/filtered/nosept', ...
'/usr/users/kevinoc/recordings/filtered/modRfMf'];

% SWITCHES TO SELECT bobcat FILES TO FILTER
indexBC = [1 2 3 4 5];
bobcats = strvcat('/bobcats/BCIV.wav', ...
'/bobcats/BCV.wav', ...
'/bobcats/BCI.wav', ...
'/bobcats/BCII.wav', ...
'bobcats/BCIII.wav');
BCroot = strvcat('/BCIV', '/BCV', '/BCI', ...
'/BCII', '/BCIII');

% SWITCHES TO SELECT cheetah FILES TO FILTER
indexCH = [1 2 3];
cheetahs = strvcat('/cheetahs/CHI.wav', ...
'/cheetahs/CHII.wav', ...
'/cheetahs/CHIII.wav');

% Swift to select which operations to perform.
LOADfromKLATT = 1;
FILTERfiles = 1;
SAVE2KLATT = 1;

% Switches to select which operations to perform.
LOADfromKLATT = 1;
FILTERfiles = 1;
SAVE2KLATT = 1;

% SWITCHES TO SELECT bobcat FILES TO FILTER
INDEXBC = [1 2 3];
bobcats = strvcat('/bobcats/BCIV.wav', ...
'/bobcats/BCV.wav', ...
'/bobcats/BCI.wav', ...
'/bobcats/BCII.wav', ...
'bobcats/BCIII.wav');
BCroot = strvcat('/BCIV', '/BCV', '/BCI', ...
'/BCII', '/BCIII');

% Switches to select which operations to perform.
LOADfromKLATT = 1;
FILTERfiles = 1;
SAVE2KLATT = 1;

% SWITCHES TO SELECT bobcat FILES TO FILTER
INDEXBC = [1 2 3];
bobcats = strvcat('/bobcats/BCIV.wav', ...
'/bobcats/BCV.wav', ...
'/bobcats/BCI.wav', ...
'/bobcats/BCII.wav', ...
'bobcats/BCIII.wav');
BCroot = strvcat('/BCIV', '/BCV', '/BCI', ...
'/BCII', '/BCIII');

% Switches to select which operations to perform.
LOADfromKLATT = 1;
FILTERfiles = 1;
SAVE2KLATT = 1;

% SWITCHES TO SELECT bobcat FILES TO FILTER
INDEXBC = [1 2 3];
bobcats = strvcat('/bobcats/BCIV.wav', ...
'/bobcats/BCV.wav', ...
'/bobcats/BCI.wav', ...
'/bobcats/BCII.wav', ...
'bobcats/BCIII.wav');
BCroot = strvcat('/BCIV', '/BCV', '/BCI', ...
'/BCII', '/BCIII');

% Switches to select which operations to perform.
LOADfromKLATT = 1;
FILTERfiles = 1;
SAVE2KLATT = 1;

% SWITCHES TO SELECT bobcat FILES TO FILTER
INDEXBC = [1 2 3];
bobcats = strvcat('/bobcats/BCIV.wav', ...
'/bobcats/BCV.wav', ...
'/bobcats/BCI.wav', ...
'/bobcats/BCII.wav', ...
'bobcats/BCIII.wav');
BCroot = strvcat('/BCIV', '/BCV', '/BCI', ...
'/BCII', '/BCIII');

% Switches to select which operations to perform.
LOADfromKLATT = 1;
FILTERfiles = 1;
SAVE2KLATT = 1;
Kl2mat and mat2kl are scripts that were written for use at the RLE Speech Communication Lab for importing and exporting (respectively) Klatt .wav formatted files into and out of Matlab.

**kl2mat.m**

```matlab
function [y, Fs] = kl2mat(fname)
% KL2MAT Read a .wav (Speech Group/xkl) file, specified by FNAME, and loads it into matlab via the variable Y. FS is the sampling rate.
% $Log: kl2mat.m,v$
% Revision 1.5 1998/07/22 01:13:59 krishna
% Added ability to read in .wav files that have different header
% versions (versions 0, 1, 2).
% Revision 1.4 1998/07/16 23:29:09 krishna
% Added argument checking and prettified.
% Revision 1.3 1998/06/12 03:52:36 krishna
% Hmm. Don't really know what I changed.
% Revision 1.2 1998/06/06 22:52:00 krishna
% Added RCS log.
if (nargin -= 1)
    error( 'Specify filename. Usage: kl2mat(fname)');
end
fid = fopen(fname, 'r');
if (fid -= 1)
    error(['File 'fname ' not readable.']);
end
% Get the header info
% [usec, count] = fread(fid, 1, 'int');, % sampling period in microsecs.
% [numSamples, count] = fread(fid, 1, 'int');, % number of samples
% [version, count] = fread(fid, 1, 'int');, % .wav version
Fs = 10^6/usec; % sampling rate based on sampling period in microseconds
if (version == 1)
    [junk, count] = fread(fid, 125, 'int'); % read remainder of header
elseif (version == 2)
    [Fs, count] = fread(fid, 1, 'int');
    [junk, count] = fread(fid, 124, 'int'); % read remainder of header
end
%----- CONVERT BACK TO KLATT FORMAT
if (SAVE2KLATT) % if (not(0cross))
    for nn=1:length(CHread(:,1))
        eval(['CH' num2str(nn) '_bcl = conv(CH' num2str(nn) ', BCLimplsl);']);
    end
end
% 2. through cheetah left ear
if (not(0cross)) % if (not(0cross))
    for nn=1:length(BCread(:,1))
        eval(['BC' num2str(nn) '_chl = conv(BC' num2str(nn) ', CHLimpls1);']);
    end
end
if (0cross) % if (not(0cross))
    for nn=1:length(CHread(:,1))
        eval(['CH' num2str(nn) '_chl = conv(CH' num2str(nn) ', CHLimpls1);']);
    end
end
end

end

K12mat and mat2kl are scripts that were written for use at the RLE Speech Communication Lab for importing and exporting (respectively) Klatt .wav formatted files into and out of Matlab.

**K12mat.m**

```matlab
function [y, Fs] = k12mat(fname)
% K12MAT Read a .wav (Speech Group/xkl) file, specified by FNAME, and loads it into matlab via the variable Y. FS is the sampling rate.
% $Log: k12mat.m,v$
% Revision 1.5 1998/07/22 01:13:59 krishna
% Added ability to read in .wav files that have different header
% versions (versions 0, 1, 2).
% Revision 1.4 1998/07/16 23:29:09 krishna
% Added argument checking and prettified.
% Revision 1.3 1998/06/12 03:52:36 krishna
% Hmm. Don't really know what I changed.
% Revision 1.2 1998/06/06 22:52:00 krishna
% Added RCS log.
if (nargin -= 1)
    error( 'Specify filename. Usage: k12mat(fname)');
end
fid = fopen(fname, 'r');
if (fid -= 1)
    error(['File 'fname ' not readable.']);
end
% Get the header info
% [usec, count] = fread(fid, 1, 'int');, % sampling period in microsecs.
% [numSamples, count] = fread(fid, 1, 'int');, % number of samples
% [version, count] = fread(fid, 1, 'int');, % .wav version
Fs = 10^6/usec; % sampling rate based on sampling period in microseconds
if (version == 1)
    [junk, count] = fread(fid, 125, 'int'); % read remainder of header
elseif (version == 2)
    [Fs, count] = fread(fid, 1, 'int');
    [junk, count] = fread(fid, 124, 'int'); % read remainder of header
end
```
else
    [junk, count] = fread(fid, 121, 'int'); % read remainder of header
end

[y, count] = fread(fid, inf, 'int16');
fclose(fid);

For mat2kl, note that the executable file “raw2kl” needs to be available in order for the script to work. Raw2kl is available at the RLE Speech Communication Lab.

mat2kl.m

function mat2kl(wave, fname, Fs)

%MAT2KL(WAVE, FNAME, FS)
% MAT2KLSaves wave into a .wav (MIT speech group format) as fname.wav
%   (which can be read by xkl).
% The sample values are saved as 16 bit integers, and the vector x is
% scaled to fit within the range -2^15+1 to 2^15.
% WAVE: waveform (vector of samples)
% FNAME: filename of .wav file
% FS: sampling frequency

% 3/12/00: since raw2kl requires an integer argument for the sampling
% rate, I had to put 'round()' in the below expression because it was
% somehow generating a floating point number. -KO

fsstr=sprintf('%d', round(Fs));
unix(['raw2kl -f ' fsstr ' tmp.mat2kl ' fname']);
unix(['/bin/rm tmp.mat2kl']);

if (maxx>2^15)||(minx<-2^15+1)
    wave = wave/scale;
end

 fid = fopen('tmp.mat2kl', 'w');
 fwrite(fid, wave, 'int16');
 fclose(fid);

if (nargin ~= 3)
    error(' Usage: mat2kl(wave, fname, Fs)');
end

maxx = max(wave);
minx = min(wave);
scale = max(abs(maxx),abs(minx))/(2^15-1);
Bibliography


