Exploring the Interior Structure of White-handed Gibbon and Rat Vocal Communication

By

Angela Marie Dassow

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Zoology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2014

Date of final oral examination: 11 June 2014

The dissertation is approved by the following members of the Final Oral Committee:
  Michael H. Coen, Assistant Professor, Biostatistics and Medical Informatics, Computer Sciences and Zoology
  Michelle R. Ciucci, Assistant Professor, Surgery and Communication Sciences and Disorders
  Eric S. Raimy, Professor, Linguistics and English
  Lauren V. Riters, Professor, Zoology
  Charles T. Snowdon, Professor Emeritus, Psychology

Acknowledgments

The completion of this doctoral thesis would not have been possible without the guidance and expertise of several individuals. First and foremost, I would like to thank my academic advisor, Dr. Michael Coen, for his continued support and encouragement. I would also like to thank the members of my dissertation committee: Dr. Michelle Ciucci, Dr. Eric Raimy, Dr. Lauren Riters, and Dr. Charles Snowdon. Their valuable expertise and guidance helped me to develop my research. It has been a privilege to work with these scientists.

I would also like to thank the members of the Coen lab, M. Hidayath Ansari, Tim Chang, and Marissa Phillips, who offered advice throughout the development of my research. There are many other professors and staff who were also influential in the development of this research. I am grateful for the guidance and contributions of Dr. Sharon Thoma, Dr. Robert Bleiweiss, Dr. Antony Stretton, Dr. Jeff Hardin, Dr. Karen Strier, Dr. Nazan Gillie, Dr. Paul Rathouz, Cathy Whitford and the staff in the Biostatistics and Medical Informatics and Zoology Departments. I would also like to thank our collaborator, Dr. Esther Clarke, who generously shared her wild gibbon recordings.

I am also grateful for the Writing Center staff, especially: Brad Hughes, Nancy Karls, Eileen Callahan, Virginia Piper, Kevin Mullen, and Kristiane Stapleton. Their advice and guidance was influential in improving the clarity and conciseness of my writing. I would also like to thank my dissertation writing group members: Dr. Hannah Gaines, Dr. J. Ashleigh Ross, and Dr. Natalie Hunt for their weekly encouragement in finishing each chapter.

There are several individuals from the Henry Vilas Zoo and Racine Zoological Society that I would like to thank for their contributions: Erin Flynn, Jeff Stafford, Rhonda Schwetz, Jennifer Zuehlke, Jay Christie, Theresa Donarski, Gianna O'Keefe, Anne Sylvester and Crystal Champeau-Williams, and the gibbons.

Finally, I would like to express my heartfelt gratitude to my family, especial my parents, Don and Lois Dassow, and my boyfriend, Richard Statz. Their love, encouragement, patience, and support throughout this process is greatly appreciated. I would also like to thank my great grandparents, Vincent and Clara Freund, to whom this work is dedicated, for encouraging me to pursue my scientific curiosity.

Table of Contents

Exploring the Interior Structure of White-handed Gibbon and Rat Vocal Communication

Angela M Dassow

Under the supervision of Professor Michael H. Coen

At the University of Wisconsin-Madison

Abstract

What is the difference between communication and language? In non-human animals, communication is widely viewed as a behavior; a reflexive activity designed to produce behavioral responses in conspecifics or across species. In contrast, language is a human affair. It transfers conceptual knowledge from speaker to listener and has extraordinarily generalizable descriptive powers. We reevaluate this distinction in the context of vocalizations of white-handed gibbons (*Hylobates lar*), demonstrating previously unrecognized complexity and structure in their vocalizations.

We begin by employing novel machine learning and statistical methods to identify acoustic *primitives* in gibbon vocal productions that are unexpectedly governed by rules and statistical distributions common in human speech. Having identified these as the smallest acoustic units in gibbon vocalizations, we present evidence these are also the smallest "atom" that can change the *meaning* of a gibbon vocalization as determined by ethological observation. This leads to a generalized concept of a *phoneme* not confined to human language.

Our approach, called *Cepstral Self-Similarity Matrices*, enables automatic sequencing of gibbon vocalizations into their constituent phonemes. We analyze these sequences using basic Ergodic theory to segment them into distinct subsequences that appear consistently and repeatedly across our gibbon populations in specific referential contexts. For example, predator alarm calls share basic properties, statistical phonemic distributions and overall structure but each displays unique sequences associated with a particular predator. We view these as semantic units within the calls identifying the predator, as opposed to behavioral exhortations intended to trigger responses within the social group.

To further explore the value of our computational approach, the vocal repertoire of laboratory rats is examined. Distinct acoustic units are found and explored for signs of vocal degradation after the rats receive a neurotoxic brain lesion to mimic the signs associated with Parkinson’s disease. Doing so revealed distinct, differential vocal loss in acoustic units. Examining the Cepstral Self-Similarity Matrices of each acoustic unit also revealed degradation of acoustic unit structure that previously could not be detected using spectral images. In all, this dissertation has provided new insights into how gibbons and rats communicate as well as provide a new analytical tool to researchers.

"*I cannot doubt that language owes its origin to the imitation and modification, aided by signs and gestures, of various natural sounds, the voices of other animals… some early progenitor of man, probably first used his voice in producing true musical cadences, that is in singing, as do some of the gibbon-apes at the present day.*" (Darwin, 1871)

Chapter 1: Introduction

This dissertation presents a novel approach to understanding vocal communication systems in non-human animals. Drawing parallels between humans and animals has historically been problematic for many reasons. Understanding how another organism mentally represents sounds is a major roadblock in interpreting animal vocalizations. This has led to complications in using descriptive human terms from linguistics, such as *phonemes*, when describing the vocal repertoire of another species. We address this issue by developing an automated approach to segmenting and clustering acoustic units in a robust manner. Rather than relying on human interpretations of similarities between acoustic units, we additionally utilize an empirical approach of measuring acoustic variables and cluster sounds together based upon these measured values. The framework involved correctly segments and clusters English phonemes (with minor boundary differences) and produces similar divisions to the manually proposed boundaries of white-handed gibbon (*Hylobates lar*) and black rat (*Rattus norvegicus*) vocalizations. The assumption of meaning in a non-human animal brain still remains; however this assumption is defined in a more formal way by drawing upon theories in evolution and quantifying structural acoustic similarities in a mathematical way. By doing so, we move from a philosophical debate to an empirical one.

An additional issue complicating interpretation of meaning in animal communication is the viewpoint that humans share little intellectually with other animals. This has led to the notion that humans should be placed upon a pedestal with respect to their cognitive capabilities. In this respect, language can be viewed as something people *use*, whereas communication is what animals *do*. While there has been little evidence to support an alternate viewpoint, a major issue with this perspective is that there is also little evidence to support this distinction between humans and animals. To move beyond this conundrum, we adopt a less biased viewpoint and re-evaluate what connections may exist between humans and animals by examining formal acoustic properties that play semantic roles in their communication systems (Trubetzkoy, 1939; Raimy & Cairns, 2009). In this sense, we can be viewed as an alien observer trying to decipher meaning from a set of complex systems. These properties can then be compared with what we know about human languages and we can introduce linguistic terminology into systems where they fit. **We demonstrate that by looking for human language correlates in animal vocalizations, some remarkable similarities between humans and animals emerge.** As a starting point, we examine the vocalizations of white-handed gibbons. White-handed gibbons are known to have remarkable diversity in the types of sounds they make and thus present themselves as a good species to examine in more detail.

In addition to exploring the diversity of sounds in white-handed gibbons, the order with which these sounds are uttered is also examined. Sequences of acoustic units in gibbons are explored under various experimental conditions to determine whether meaningful differences exist. Previous research has made claims of syntax in gibbon vocalizations (Clarke et al., 2006);

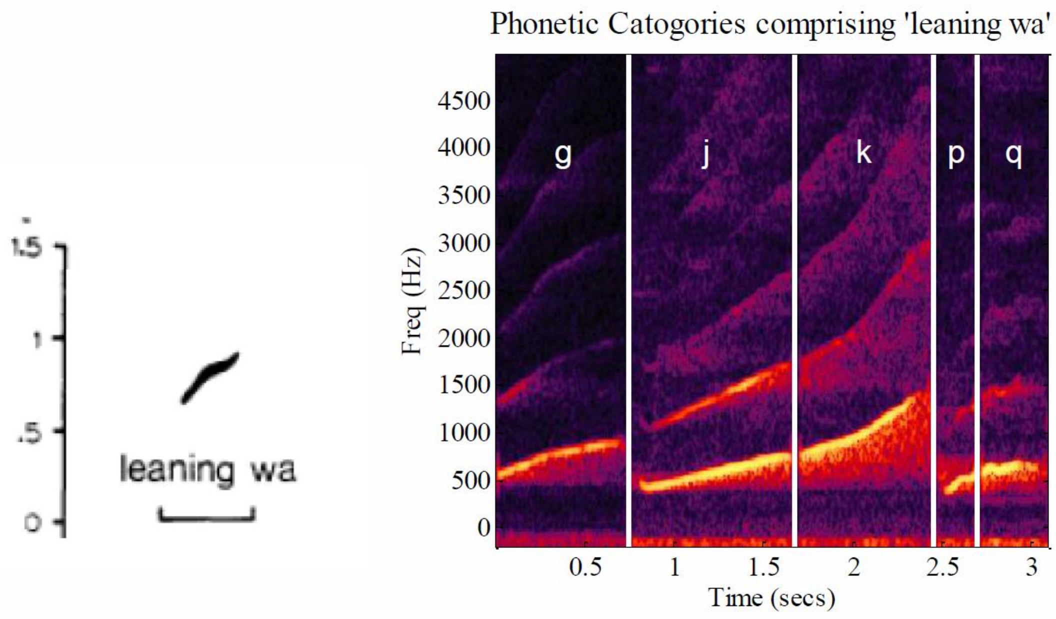


Figure 1.1: The 'leaning wa' category represents a previously described gross category of sound in the vocal repertoire of the white-handed gibbon. This category is composed of five distinct categories of acoustic units.

however the evidence has not supported this claim. This is in part due to how the vocalizations were examined via gross categories of sound, instead of acoustic units. For example, Figure 1.1 shows how a single gross category of sounds in gibbon can be represented by five distinct sounds using work in this thesis. Additionally, only a small subset of the calls was examined instead of exploring entire sequences. We argue that when looking for meaning in an unknown data set, the entire recording should be used rather than a small subset of sounds. This has been historically difficult since these data sets have been coarsely labeled by inspection; however with an automated approach more sophisticated views can be taken.

Examining gibbon vocalizations for phonemic and morphemic equivalents can answer interesting behavioral questions; however this approach can also be valuable from a translational medical standpoint. A third component of this dissertation explores the use of this computational framework in addressing questions involved in Parkinson's disease (PD) using a rat model. One of the many symptoms of PD is loss of vocal strength and quality. This often leads to the inability to effectively communicate with others. In rats, vocal loss can be measured by recording baseline and degraded vocalizations. Determining whether vocalizations degrade differentially is important for addressing future questions of vocal loss recovery via different treatment methods. While many speech therapy methods exist, treatment is often directed towards improving all phonemes regardless of idiosyncratic degradation at the individual phonemic level. Targeted speech therapy which directly addresses only the phonemes which have degraded can be time and cost effective.

Overall this dissertation moves our understanding of animal communication systems towards a unified understanding of vocal evolution. We draw upon the decades of research conducted in the disciplines of Zoology, Computer Sciences, Linguistics, Statistics and Medicine to approach controversial and complex matters in a rigorous way. Questions such as "Do animals have mental representations of their acoustic units?" or "Do any non-human species possess syntax?" are controversial questions that require the input of all of the above disciplines.

By drawing upon the knowledge developed from all of these fields we can address these questions in an open-minded manner. While some of this work challenges some long standing beliefs in complexity of animal vocal systems, it also supports many others as shown below. This work addresses several distinct, but related gaps in knowledge by making a variety of contributions to the above disciplines.

1.1 Biological Contributions

The notion of whether animals have “words” in the sense that humans do has been posed by several researchers; however, this type of question is controversial due to not knowing whether the animals are employing the same cognitive processes as humans. To address this issue in a more formal manner, we reword the question in a way which can be empirically addressed. Posing the question as “Do any repetitive patterns of acoustic units correlate to a given behavioral context?” can provide us with a testable hypothesis. We can also use this framework to ask more complex questions about meaning such as whether or not subcategories of patterns can be found within a given behavioral context.

To address these types of questions about meaning, we focus on white-handed gibbon vocalizations. A dataset from the wild is compared with a dataset from captivity and warning calls from each set are examined using our novel CSSM approach. By using CSSMs, we can rule out differences due simply to minute changes in vocalizing. Variation in vocal production was demonstrated by Peterson and Barney (1952) in their examination of vowel formants. Figure 1.2 shows an example of what kind of variation exists within vowels of the English language. Similar variation in vocal output can be found in animals and could cloud the outcome of a semantic analysis. Using Markov chains can reduce that possibility while also providing a clearer view of how to repeat a given call. Repetition of a given call correlated to a specific behavioral context has historically been used to presume meaning. We have continued to work

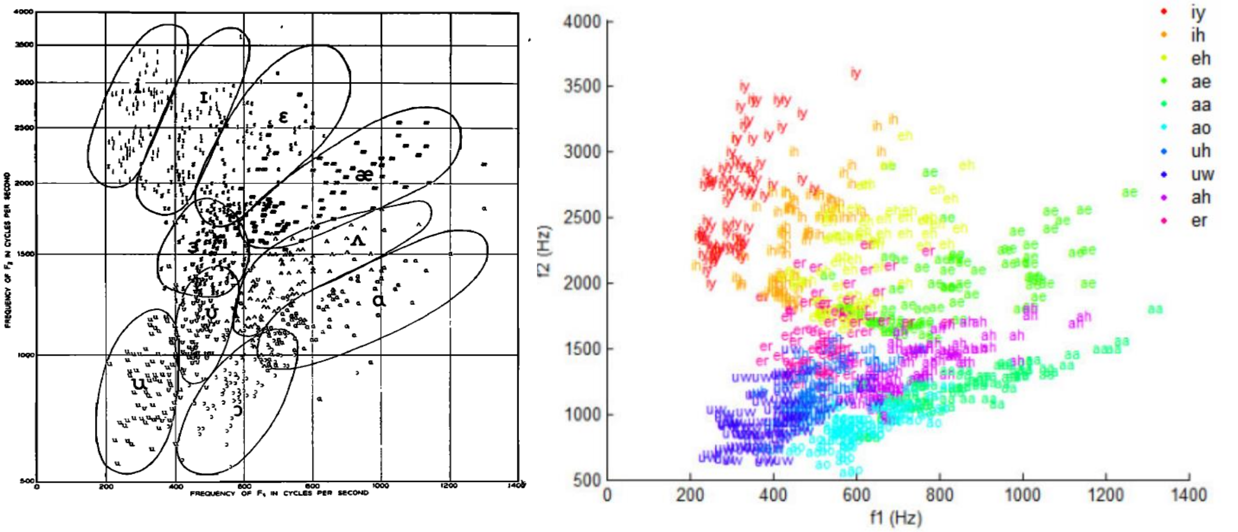


Figure 1.2: The left figure shows the original experiment conducted by Peterson and Barney (1952). The first and second formant frequencies of 10 American English vowels circa 1939 were measured and compared for their acoustical distinctions. This data was reproduced by Coen (2006) on the right. Each of the ovals or color patterns depicts a distinct vowel. Variation within each of these categories is present due to articulatory deviation.

with this assumption when examining gibbon calls.

Our work examines the meaning of gibbon vocalizations by identifying strings of acoustic units and examining what is unique in each subset of data rather than claiming a specific meaning for a general set of calls. In other words, instead of saying sequence *A* means clouded leopard, we are demonstrating that clouded leopard calls elicit any variant of call sequence *A*. The latter leaves open the possibility that additional meaning can be coded for in each of these calls. Further work examining whether any visual or olfactory cues can help to decipher other contexts will still be needed.

1.2 Computational Contributions

This dissertation employs a new computational approach to examining vocalizations. The approach used here has drawn upon several other methods which have previously been proposed and has combined these advancements into a single, universal approach to understanding vocalizations. One of the main tenets behind why we believe this is possible is evolutionary theory has taught us organisms that share a common ancestor often share ancestral traits. It stands to reason that this holds true for communication systems as well. Instead of examining vocalizations differently for each species, finding a unified approach to this problem could help address how closely related different communication systems are to one another.

To be able to address these questions in a universal manner, we must move beyond the simple analyses involving measures of start and stop frequency, etc., to signal properties that can only be computationally measured. This approach was inspired by work conducted by Thomson (1982), Fee et al. (1998), Tchernichovski et al. (2000, 2004), Coen (2006, 2007) and Saar & Mitra (2008). These papers develop more sophisticated approaches to examining vocalizations by characterizing acoustic structures of calls using properties such as entropy and spectral continuity. This research led to a remarkable advancement in how we view vocalizations and in particular, what we know about zebra finch song. **Specifically, this kind of computational approach has moved away from what a sound looks like to how it is formed.** A minor drawback to some of these approaches is that the selected parameters used in the algorithms were specific to zebra finch calls and did not generalize to the vocalizations of other species.

We address the issue of overfitting the analysis by using vocalizations of multiple species during the development of a robust tool which can segment and cluster a variety of vocalization types. In doing so, mel frequency cepstral coefficients (MFCCs) are used to measure the acoustic properties of the calls and then a *Cepstral Self-Similarity Matrix (CSSM)* is formed. This similarity matrix is what allows us to examine the relatedness of the different sounds and unlike previous approaches, it is not species specific. As we describe in Chapter 4, the parameters used here can be applied to a variety of other vocal species and therefore represent a more universal approach to studying vocal communication. The scientific value of this is illustrated by answering behavioral questions from populations of white-handed gibbons and a translational medical question using laboratory rats.

1.3 Translational Medical Contributions

In addition to answering some computationally and behaviorally interesting questions, our automated approach to segmenting vocalizations also provides a tool that has potential for addressing translational medical questions using a Long-Evans rat (Charles River Laboratories International, Inc.) model. A common sign in some neurodegenerative diseases is a loss of speech sensorimotor control, collectively termed dysarthria. How dysarthria manifests varies with each disease, but with Parkinson's Disease (PD), vocal loss is characterized by phoneme degradation (e.g. harsh, breathy quality of phonemes, reduced stress, and voice tremor), and diminished amplitude (Darley et al., 1969; Logemann et al., 1978; Hartelius & Svensson, 1994; Stewart et al., 1995; Ho et al., 1998; Fox et al., 2002; Harel et al., 2004; Plowmann-Prine et al., 2009). Degradation of vocal output has previously been established as a sign in humans and rats, the latter of which is a common model organism in PD research.

Parkinsonism is induced in rats by infusing a neurotoxin, typically unilaterally, to either the medial forebrain bundle or striatum, which results in the death of dopaminergic neurons. The death of these neurons is the hallmark pathology of PD (Bergman & Deuschl, 2002; Braak et al., 2003; Braak et al., 2004). These lesions in rats lead to similar signs seen in human patients (Ungerstedt, 1970). Rats in this model show decreased intensity, bandwidth and complexity of calls (Ciucci et al., 2007, Ciucci et al., 2009). The current study builds on this knowledge by taking a closer examination of specifically which categories degrade to address questions about regaining sound quality and quantity at an acoustic unit level. Determining whether all acoustic units degrade equally or differentially is an important basis for future studies examining treatment options. For example, if there are ten distinct types of 50 kHz calls, but only two degrade, then future studies could focus on why only a subset of these calls are affected. Identifying the acoustic units in rats requires an automated approach that had previously been unavailable to researchers. By developing such an approach, we can also address the additional problems of tracking acoustic units as they degrade and accurately matching acoustic units between healthy and lesioned rats. Addressing whether or not the difference in acoustic unit degradation is meaningful to rats requires additional tools and will be left for future work.

1.4 Theoretical Contributions

Deciphering meaning from the unknown poses many challenges, especially when the unknowns are based in the mind. Studies of human cognition have led to the development of various Theories of Mind; while that may be one feature that separates us from other animals, understanding what an animal comprehends can help us to understand the complexity of their communication systems. As mentioned previously, the viewpoint that humans and animals share little in common when it comes to vocalizations is problematic and not based on comprehensive evidence. Determining where animals fall on a language-communication continuum helps us elucidate questions concerning complexity in a more empirical manner. By doing so we can examine more specifically where animals, such as white-handed gibbons, overlap with us.

Previous attempts at formalizing these notions have been problematic due to poorly designed questions and lack of necessary tools. A few papers have methodically tackled the question of whether we can break animal vocalizations into phonemic categories instead of gross sound categories (Cleveland & Snowdon, 1982; Masataka, 1982); however when these studies were conducted, only manual measurements of these call types were available and thus they are subject to possible human bias or error for even the most meticulous researcher. Other work which has made unsupported claims of syntax (Clarke et al., 2006; Bohn et al. 2009) has divided the scientific community. While this may not have been intentional, a bold claim of something as sophisticated as syntax in an animal communication system requires more evidence for support before gaining acceptance in a broader research community. The different viewpoints between ethologists, bioacousticians and linguists have led to many disagreements between communities. This dissertation addresses some of these disagreements by incorporating important aspects of each field into a common goal of understanding vocal communication in white-handed gibbons and rats.

1.4 Hypotheses

As mentioned throughout this chapter, most prior connections between animal vocalizations and human language have struggled to find common ground due to rigid viewpoints and lack of empirical data. In developing a way to formalize this argument, I hypothesize that phonemic equivalents do exist in white-handed gibbons and that these acoustic units are combined to convey messages which possess motivational and referential content. I also hypothesize that phonemic equivalents are found in laboratory rats and that unilateral 6-OHDA lesions leads to differential loss of these acoustic units.

1.5 Roadmap

Chapter 2 sets the stage for the rest of the dissertation by revisiting a well-studied example of examining the interior structure of birdsong, specifically zebra finches. It clarifies the need for more sophisticated tools when examining the vocalizations of certain species of animals. It also highlights how an automated approach can be used to address a variety of interesting questions, including determining how similar animal vocalizations are to human language.

Chapter 3 connects the work in this dissertation with the wide array of disciplines from which it was drawn. The main motivations for the approaches taken in this dissertation are highlighted. Additionally the major contributions from the respective fields of linguistics, signal processing, neurodegenerative disorders, and bioacoustics are examined. We also examine why the advances in each of these respective fields benefit from one another and propose a way to do so in a productive and complementary manner.

Chapter 4 develops the computational tools needed to address an acoustic unit analysis of a non-human animal and subsequently uses that approach to re-examine the vocal repertoire of white-handed gibbons. This work builds upon a previous acoustic unit description of white-handed gibbons by adding data from two pairs of captive gibbons as well as a different approach for describing the acoustic unit categories. By doing so, we provide a solid foundation for addressing questions regarding meaning which are interesting from a behavioral standpoint.

Chapter 5 uses the acoustic unit categories in gibbons which were established in Chapter 4 to address a behaviorally interesting question by re-examining the *meaning* behind calls gathered from predator presentation models in the wild. We compare the original conclusions which were based on a short segment of gross sound categories to a more inclusive, acoustic unit approach to determine what potential meaning was overlooked. We also compare the wild gibbon predator calls to warning calls gathered from captive animals to determine whether captive gibbons develop novel warning calls for their perceived predators (e.g. veterinarians, cockatoos, etc.) or whether they reassign unused wild predator warning calls to captive situations (e.g. clouded leopard call becomes vet warning call). In doing so, we further our understanding of how meaning is correlated to calls as well as provide useful information for wildlife managers interested in the captive release of gibbons into the wild.

Chapter 6 uses the automated analysis developed in Chapter 4 to elucidate a finer level analysis of vocal loss in PD using a rat model. This chapter begins by describing the acoustic units found in normal, healthy male rats. We then describe the vocal repertoire of lesioned rats and correlate which categories show significant loss to acoustic structure.

Chapter 7 gives a brief summary of the contributions of this dissertation and describes future work which can build upon this framework.

Chapter References

Bergman, H., & Deuschl, G. (2002). Pathophysiology of Parkinson's disease: From clinical neurology to basic neuroscience and back. *Movement Disorders, 17,* S28-S40.

Bohn, K. M., Schmidt-French, B., Schwartz, C., Smotherman, M., & Pollak, G. D. (2009). Versatility and stereotypy of free-tailed bat songs. *Plos One, 4*(8).

Braak, H., Del Tredici, K., Rub, U., de Vos, R. A. I., Steur, E., & Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging, 24*(2), 197-211.

Braak, H., Ghebremedhin, E., Rub, U., Bratzke, H., & Del Tredici, K. (2004). Stages in the development of Parkinson's disease-related pathology. *Cell and Tissue Research, 318*(1), 121-134.

Ciucci, M. R., Ahrens, A. M., Ma, S. T., Kane, J. R., Windham, E. B., Woodlee, M. T., et al. (2009). Reduction of dopamine synaptic activity: degradation of 50-kHz ultrasonic vocalization in rats. *Behavioral neuroscience, 123*(2).

Ciucci, M. R., Ma, S. T., Fox, C., Kane, J. R., Ramig, L. O., & Schallert, T. (2007). Qualitative changes in ultrasonic vocalization in rats after unilateral dopamine depletion or haloperidol: A preliminary study. *Behavioural Brain Research, 182*(2), 284-289.

Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *Plos One, 1*(1).

Cleveland, J., & Snowdon, C. T. (1982). The complex vocal repertoire of the adult cotton-top tamarin (*Saguinus oedipus oedipus*). *Zeitschrift Fur Tierpsychologie-Journal of Comparative Ethology, 58*(3), 231-270.

Coen, M. (2006). *Self-supervised acquisition of vowels in American English*. Paper presented at the Proceedings of the Twenty First National Conference on Artificial Intelligence (AAAI), Boston, MA.

Coen, M. (2007). *Learning to sing like a bird: Self-supervised acquisition of birdsong*. Paper presented at the Proceedings of the Twenty Second National Conference on Artificial Intelligence (AAAI), Vancouver, British Columbia.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12*(2), 246-270.

Darwin, C. R. (1871). *The descent of man, and selection in relation to sex*: London.

Fee, M. S., Shraiman, B., Pesaran, B., & Mitra, P. P. (1998). The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature, 395*(6697), 67-71.

Fox, C. M., Morrison, C. E., Ramig, L. O., & Sapir, S. (2002). Current perspectives on the Lee Silverman Voice Treatment (LSVT) for individuals with idiopathic Parkinson disease. *American Journal of Speech-Language Pathology, 11*(2), 111-123.

Harel, B., Cannizzaro, M., & Snyder, P. J. (2004). Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study. *Brain and Cognition, 56*(1), 24-29.

Hartelius, L., & Svensson, P. (1994). Speech and Swallowing Symptoms Associated with Parkinsons-disease and Multiple-sclerosis - a Survey. *Folia Phoniatrica Et Logopaedica, 46*(1), 9-17.

Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1998). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurology, 11*(3), 131-137.

Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, E. R. (1978). Frequency and Co-occurrence of Vocal-tract Dysfunctions in Speech of a Large Sample of Parkinson Patients. *Journal of Speech and Hearing Disorders, 43*(1), 47-57.

Masataka, N. (1982). A field study on the vocalizations of Goeldi's monkeys (*Callimico goeldii*). *Primates, 23*(2), 206-219.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 24*(2), 10.

Plowman-Prine, E. K., Sapienza, C. M., Okun, M. S., Pollock, S. L., Jacobson, C., Wu, S. S., et al. (2009). The Relationship Between Quality of Life and Swallowing in Parkinson's Disease. *Movement Disorders, 24*(9), 1352-1358.

Raimy, E., and Charles E. Cairns, eds. (2009). *Contemporary views on architecture and representations in phonology* (Vol. 48): MIT Press.

Saar, S., & Mitra, P. P. (2008). A technique for characterizing the development of rhythms in bird song. *Plos One, 3*(1).

Stewart, C., Winfield, L., Hunt, A., Bressman, S. B., Fahn, S., Blitzer, A., et al. (1995). Speech Dysfunction in Early Parkinsons disease. *Movement Disorders, 10*(5), 562-565.

Tchernichovski, O., Lints, T. J., Deregnaucourt, S., Cimenser, A., & Mitra, P. P. (2004). Studying the song development process rationale and methods. *Behavioral Neurobiology of Birdsong, 1016*, 348-363.

Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour, 59*, 1167-1176.

Thomson, D. J. (1982). *Spectrum estimation and harmonic analysis*. Paper presented at the *Proceedings of the IEEE.*

Trubetzkoy, N. S. (1939). *Grundzüge der Phonologie* (C. A. M. Baltaxe, Trans. Principles of Phonology ed.).

Ungerstedt, U. (1970). Striatal dopamine release after amphetamine or nerve degeneration revealed by rotational behaviour. *Acta Physiologica Scandinavica, Supplementum 367*, 49-68.

"Now by these two means one can also know the difference between men and beasts. … for one sees that magpies and parrots can utter words just as we can, and yet they cannot speak as we do, that is to say, by testifying to the fact that they are thinking about what they are saying." (Descartes, 1637)

Chapter 2: Setting the Stage

We begin with an example which illustrates two fundamental problems in correlating linguistic terms to bioacoustics research:

1) Interior structure – What constitutes a unit of sound in a non-human animal?
2) Clustering data[1] – How do we know how many distinct acoustic units a given species has if we don't have ground truth from that species?

The example given below is specific to zebra finch song, but the techniques presented in later chapters can be applied to a wide variety of vertebrates including mammals, other bird species, reptiles, amphibians and fish.

2.1 Interior Structure of Zebra Finch Song

Our example begins with a spectrogram and a multitaper spectral derivative of a zebra finch song. Traditional approaches to bioacoustics research characterize song structures by measuring start and stop frequencies for each call separated by silence. For organisms with simple, tonal vocalizations, this is a relatively easy process; however for more sophisticated singers, the complexity of the calls demands a more detailed analysis. The spectral image at the top of Figure 2.1 shows acoustic complexity in a zebra finch call. In this case, start and stop frequencies cannot accurately characterize all of the variation within the sounds being produced. Additionally, it is difficult to measure the amount of complexity present because of the lack of definition in the spectral image. These issues led to the development of a spectral derivative,

---

[1] Our approach uses the unsupervised learning definition of clustering which is mathematically based (Margolis & Laurence, 1999; Xu & Wunsch II, 2005).

shown on the bottom half of Figure 2.1. A spectral derivative is a time derivative of power that clarifies the boundaries between sounds by focusing on power transitions which marks the shift between high energy and low energy signals. This improves the definition of frequency transitions by focusing on the gradient of power changes.

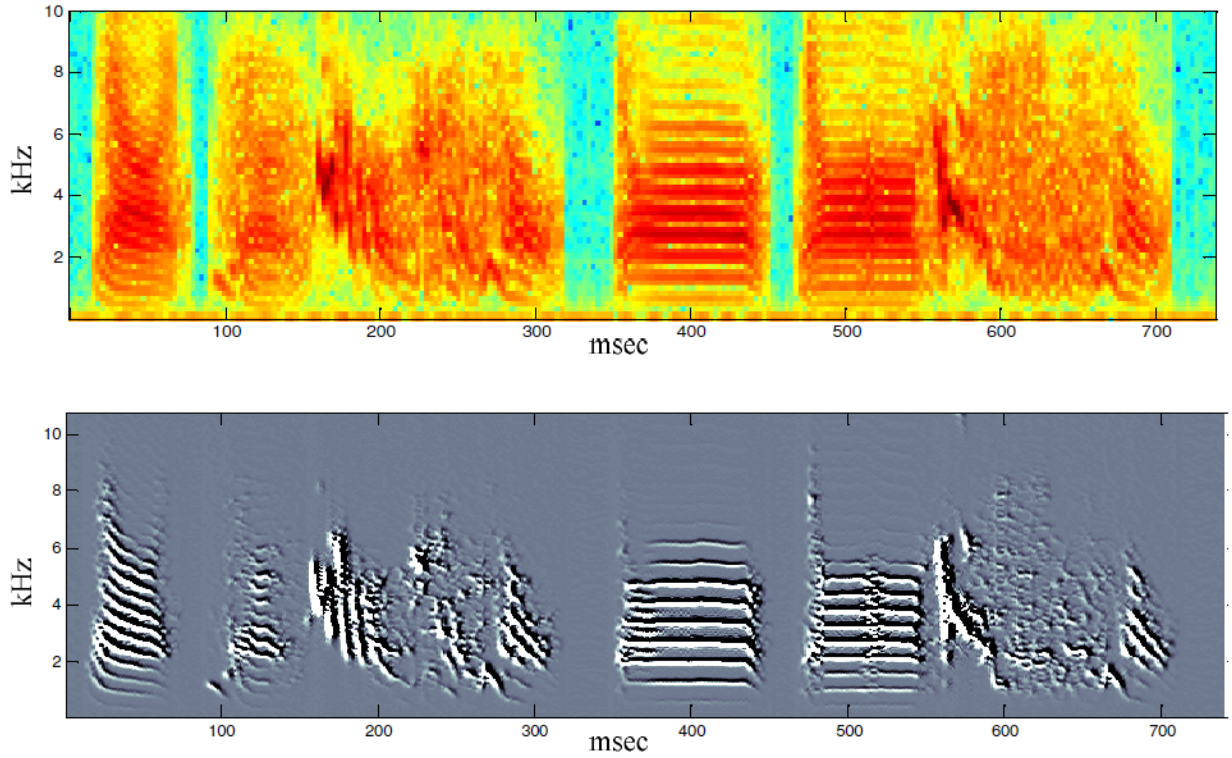


Figure 2.1 – Comparing a spectrogram to a spectral derivative display for the song of a zebra finch. A spectrogram for a zebra finch song is displayed on top. The spectral derivative of that song is shown on the bottom and provides much clearer visual detail. It also provides a framework for subsequent harmonic analysis. This figure is from Coen, 2006.

The use of multitaper spectral derivatives in song analysis was first described by Thomson (1982) and Fee et al. (1998) and later implemented by Tchernichovski et al. (2000) Coen (2006, 2007), and Saar & Mitra (2008) for use in an automated segmentation of birdsong. These computational techniques greatly advanced prior sound analysis programs by incorporating automated feature measures capable of processing even the most complex sounds. Prior programs such as Sound Ruler (Gridi-Papp, 2003 – 2007) were limited to analyzing simple, repetitive sounds and they were only capable of producing basic spectral images with low levels of clarity. Popular programs currently in use are species or group specific such as SpectraPLUS/PRO/LAB (Pioneer Hill Software, 1993), which are fit to cetacean research, or Batsound (Pettersson Elektronic, 1983) and Sonobat (Sonobat, 2013), which are tailored to bat calls and can only provide relatively simple acoustic measurements. Syrinx (Burt, 2005) initially showed much promise, but it is no longer in development and appears to be a combination of Adobe Audition (Adobe Systems Inc., 2014), Avisoft (Avisoft Bioacoustics, 2014), Raven (Bioacoustics Research Program, 2014) and SIGNAL/RTS (Engineering Design, 1980). These programs are far more advanced and versatile than most, but they are still limited to rougher scale measurements such as start and stop frequency, peak and minimum frequency. One advantage of these programs is the possibility of synthesizing new song which can be quite useful for certain types of research. Sound Analysis Pro, developed by Ofer Tchernichovski and colleagues (2000), is another program currently available; however the parameters used in the algorithms are tailored to zebra finch calls. Despite so many freely available platforms for bioacoustics research, there remains a need for an advanced collective approach that is not species biased.

To address this need, we turn our attention to Coen's work using peak detection in Figure 2.2. As shown here, this work builds off of Tchernichovski et al. (2000) by first converting a spectrogram into a spectral derivative. Again this helps to clarify the boundary definitions or transitions between sounds. The top half of this figure shows traditional boundaries of syllables which are placed between well-defined periods of silence. To determine boundary placement within a syllable, the smoothed log power of the calls (yellow lines) are calculated. Additional boundaries are determined based on the local adjacent minima of these smoothed power lines.

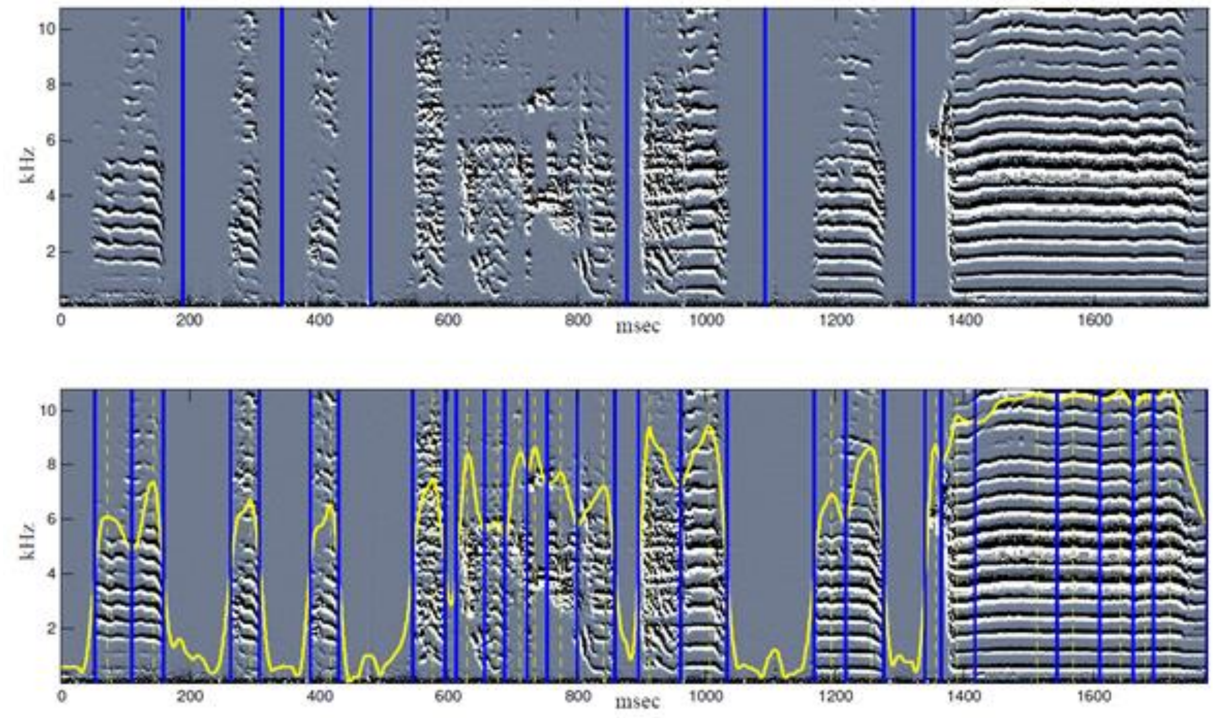


Figure 2.2 – Breaking a birdsong down into constituent songemes. On the top, the song is displayed divided into seven syllables. The 22 derived songemes, defined via peaks of the song's smoothed log(power), are shown on the bottom. The peaks are indicated by the dotted vertical yellow lines. The blue lines indicate songeme boundaries, determined by locally adjacent minima. We note the long vocalization at the end of the song corresponds to a distance call. This figure is from Coen, 2006.

If we examine a single syllable in more detail (See Figure 2.3), we can visualize how much complexity can be found in a given sound. In this sense, spectral derivatives are more precisely characterizing the complexity of the calls. While spectral derivatives can clarify these sound boundaries, we also need new tools to measure the changes which are now apparent. Simple frequency measurements can't capture this complexity. Tools which can capture these more complex measurements are now available for use (Coen, 2006, 2007). This is an important advancement for addressing questions regarding vocal complexity.

These more advanced segmenting approaches are great for examining complexity in a given call; however this does not address the issue of determining which of these segmented sounds is similar enough to one another that they can be categorized as being the same acoustic

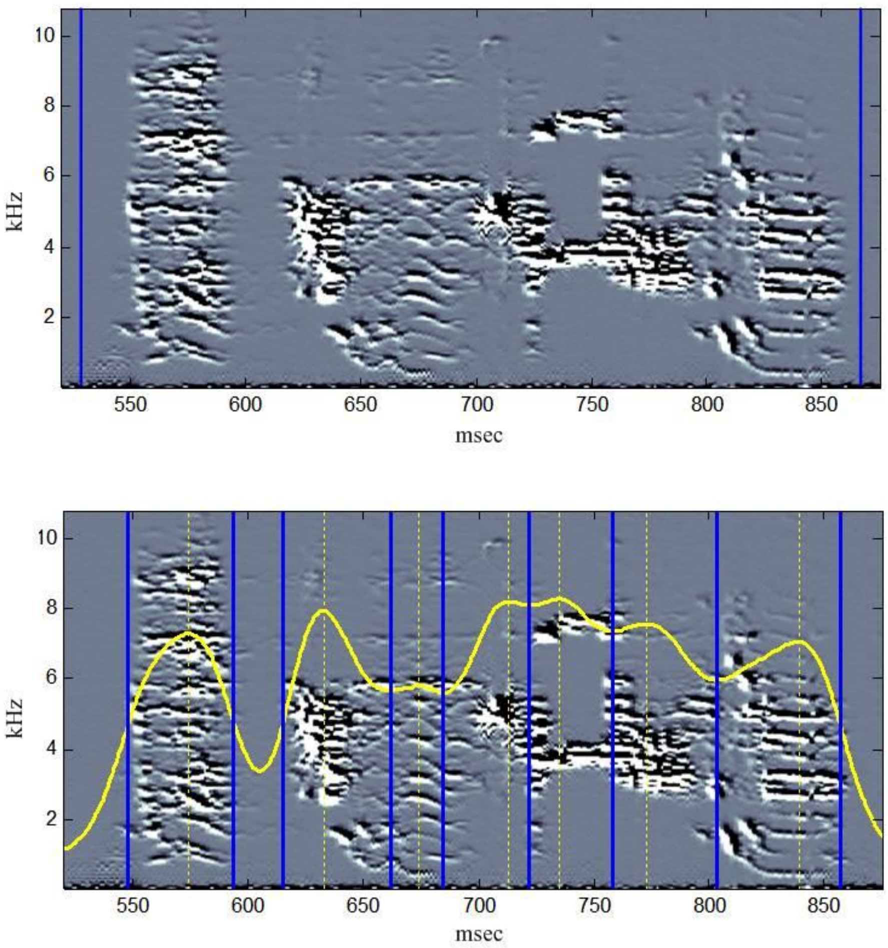


Figure 2.3 – Partitioning a single syllable into songemes. This figure displays the segment of birdsong in Figure 2.2 between 520 and 875 msecs. The single syllable shown on top has been automatically partitioned into seven songemes on the bottom, which correspond more closely to the changes in vocalization during this interval. This example supports our belief that the widespread syllabic approach to studying birdsong is a poor model for capturing its internal complexity. This figure is from Coen, 2006.

unit. To tackle this issue, we could now take the traditional frequency measurements and run a statistical analysis, but that will in some sense be reverting back to old habits. Instead, we can move forward with developing more computational tools such as feature extraction. Shown in Figure 2.4 are a few of the features used on these diverse sounds to determine which are similar to one another. This is a demonstration of clustering based upon well-defined and measureable parameters, which adds empirical validity to the subsequent claims of category assignments.

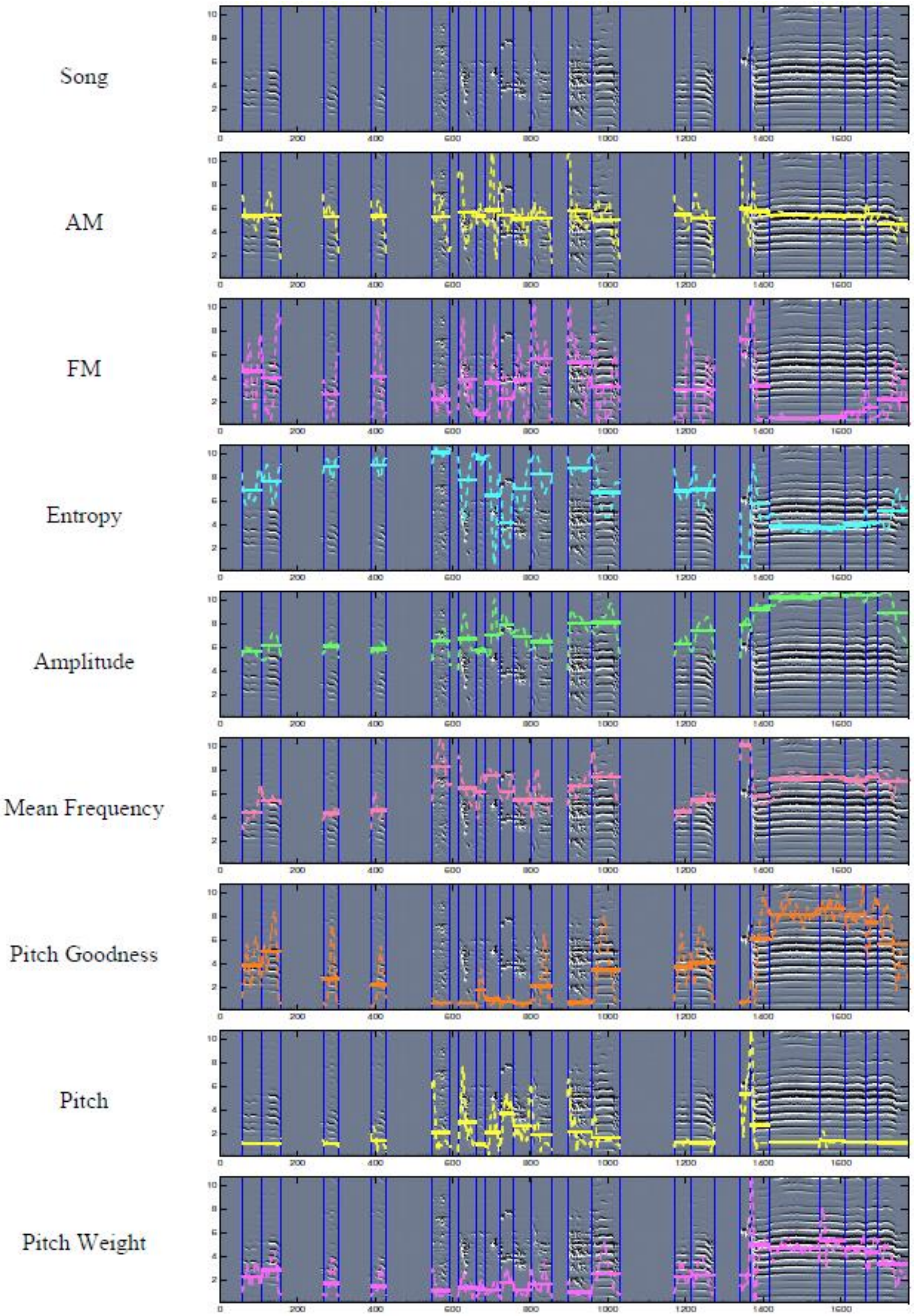


Figure 2.4 – Feature extraction for a zebra finch song partitioned into songemes. The solid line within each songeme shows the mean value for the corresponding feature within it. These values are cross-modally clustered to learn the structure of the birdsong. The dotted line within each songeme shows the actual feature data, which is smoothed with a low-pass Savitzky-Golay filter. The feature values have been normalized to fit within each plot. We note this is the same song as shown in Figure 2.2 (Coen, 2006).

2.2 What Other Species Have Complex Vocalizations?

Zebra finches are a well-known model organism for neuroscientists. As such, the scientific community knows a great deal about them including the complexity of their vocalizations. Though zebra finches produce a very diverse set of vocalizations, they are not the only species with complex vocalizations. For example, the Madagascar tree frog has a rich repertoire of 28 discrete calls classified (Narins et al., 2000). This diversity in many ways defies what one would speculate for a frog and highlights the importance of gathering empirical data prior to forming assumptions regarding complexity. **Given the diversity of vocal species alive today, why is it that more species haven't been examined using the computational approaches developed for studying the zebra finch?** As we mentioned earlier, much of this has to do with the amount of time and effort involved in developing these approaches, which in part was due to the lack of technological tools available. Older microphones and field recording media such as tape recorders do not provide the amount of clarity that digital recorders do. Figure 2.5 shows a comparison of a tape recorded call (left) and a digital recorded call (right) of a white-handed gibbon from the Khao Yai National forest. The acoustic units are the same in each recording; however the clarity of the calls in visual and auditory domains is significantly clearer in the digital recording. Likewise, computational power and storage has greatly increased over the years making it possible to carry out more

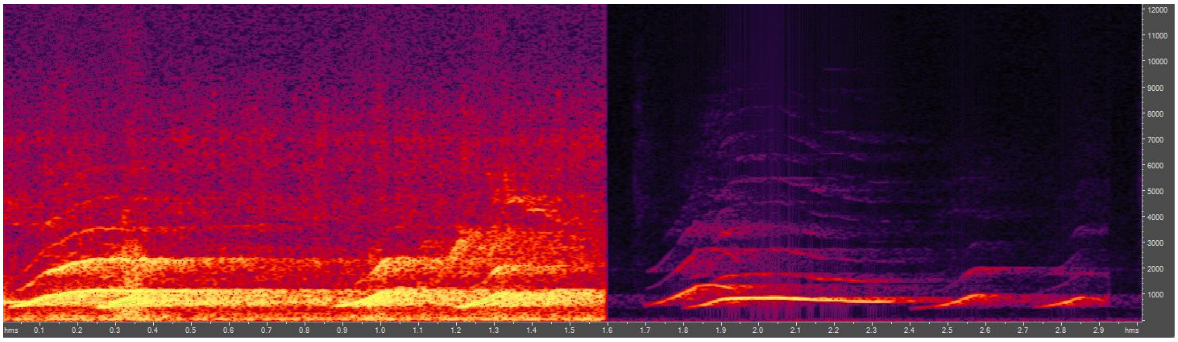


Figure 2.5: Comparison between a tape recorded call (left) and a digital recorded call (right) of a gibbon from the Khao Yai National forest. Each call consists of the same acoustic units; however the tape recorded call lacks the clarity present in the digital recorded call.

sophisticated types of analyses. Despite the limits in the tools available to researchers in the past, a few species have had their vocal repertoires examined at an acoustic unit level using a variety of frequency measurements.

Searching for similarities between human language and animal communication has been a long standing debate in research. The earliest attempts at understanding a species' vocalizations at an acoustic unit level were conducted on cotton-top tamarins (*Saguinus oedipus*) by Cleveland and Snowdon (1982) and Goeldi's Monkeys (*Callimico goeldii*) by Masataka (1982). At this time, one of the major drawbacks for these researchers was the lack of technology that can clearly detect subtle, but important variation in calls. Tape recorders were used and vocalizations were analyzed using KAY analyzers which produced low resolution spectrograms. Despite this, 38 distinct vocalizations in cotton-top tamarins and 40 distinct vocalizations in Goeldi's monkeys were identified and described. These studies achieved an important milestone for moving bioacoustics research towards a union with linguistics. One drawback to these studies was verifying the proposed categories of sound. Aside from inter-observer agreement on category designation, there was no clear way to formally defend the existence of each category.

The issue of category designation persisted and nearly two decades passed before another attempt was made to examine the vocal repertoire of a species. By this time, more advanced types of tape recording were available which helped provide a clearer spectral image. These tape recordings were utilized for recording Barbary Macaques (*Macaca sylvanus*) by Hammerschmidt & Fisher (1998), Madagascar treefrogs (*Boophis madagascariensis*) by Narins et al. (2000) and Garnett's greater bushbaby (*Otolemur garnetti*) by Becker et al. (2003). The calls of the Barbary macaque and bushbaby appeared graded and therefore clustered into only 7 categories and 4 categories (with 11 potential variants) respectively. Graded vocalizations have no clear endpoints where one category of sound ends and another begins. This makes it difficult to detect additional categories, even if they do exist in the animal's vocal repertoire.

Though there were many drawbacks to using tape recorders, one of the distinct advantages of using them was that it provided an avenue for pursuing more empirical approaches to studying sound. A major breakthrough for bioacoustics was the advancement from tape recorders to digital recorders which provided a clearer representation of sound. Some of the first researchers to utilize digital recording techniques were Van Parijs and colleagues (2000) to record Australian Irrawaddy dolphins (*Orcaella brevirostris*). These dolphins use 6 distinct types of calls during a variety of activities including foraging and socializing. While this was relatively advanced in terms of recording procedures, the analysis is still very basic with a major focus on frequency descriptions. As the types of digital recorders improved, so did the software to analyze the recordings. Software, such as PRAAT (Boersma & Weenink, 2013), was specifically designed by linguists to study human language; however it can also be used for examining animal vocalizations. Gamba & Giacoma (2007) used PRAAT to identify formants in crowned lemur (*Eulemur coronatus*) vocalizations for use in classification of calls. This allowed them to use a Discriminant Function Analysis (DFA) to identify 10 acoustic categories. This was an important paper for a few reasons: 1) by using a linguistic analysis package, the bioacoustics researchers were making an effort to formalize the analysis of sounds in a manner conducive with making comparisons to human language and 2) by using a more quantitative approach, some of the criticism surrounding observer bias could be diffused. While error could still occur during the PRAAT measurements due to manual placement of boundaries, this error was reduced.

In more recent years, a variety of acoustical software packages became available for use in bioacoustics research. Gros-Louis et al. (2008) used Cool Edit and SASLab Pro to identify 27 call types in white-faced capuchins (*Cebus capucinus*). This was the first study to examine a species of monkey using digital recordings and specially designed acoustical software. Also using SASLab Pro was a study conducted with Mongolian gerbils (Kobayasi & Riquimaroux, 2012) which identified 13 categories of calls. Similar to the study with the Madagascar frogs, this study provided support for examining species outside of primates and oscine songbirds. Further support for examining a wide array of species for phonemic potential was offered by Giles et al. (2009). Giles and colleagues (2009) used SpectraPLUS to identify 17 categories in the vocal repertoire of the long-necked turtle *(Chelodina oblonga)*. One major drawback to these studies however is despite the available technology, they are still using simple frequency measurements in the analyses which have been a continued criticism for drawing parallels to human language.

What we can learn from these earlier attempts is that future studies of animal vocalizations would benefit from more rigorous analytical techniques. To examine the vast array of complexity in different species' vocalizations, we should employ knowledge from the various disciplines involved in these arguments. By combining the technical skills of the signal processing and computer science communities with the ethological knowledge of the zoological communities, we can begin to find some common ground. As a brief motivating example, white-handed gibbon vocalizations have long been considered to be categorized as 6-7 gross categories of sound (Raemaekers et al. 1984). Amongst others, the "oo" call was considered to be one category; however upon listening to a short vocal bout, it is readily apparent that the "oo" category contains a wide variety of subtypes. With more advanced technology, we can see (and hear) some remarkable diversity within this one category. Figure 2.6 shows 10 subcategories of

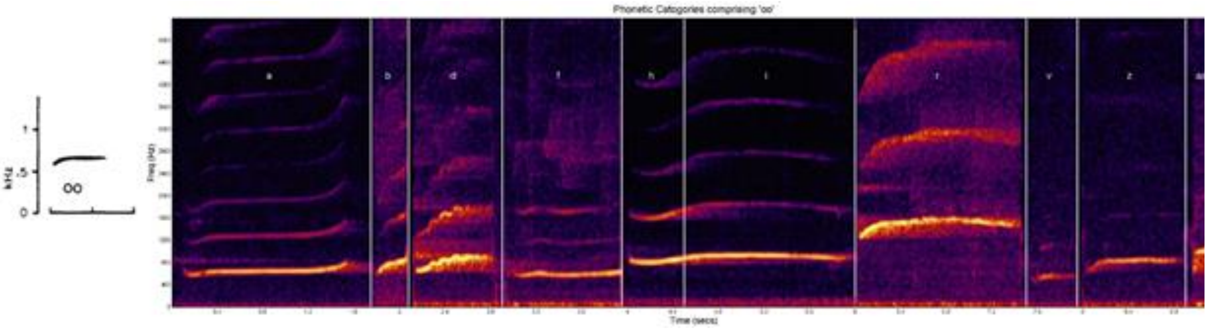


Figure 2.6: The left side of this figure shows a line drawing of the 'oo' category of calls from Raemaekers et al., 1984. After further examination of the vocal repertoire of white-handed gibbons, the 'oo' category appears to be made up of 10 distinct categories (Dassow, 2010).

"oo's". These divisions were originally proposed by Dassow (2010) using a standard manual analysis and independently corroborated using peak detection and power smoothing methods developed by Coen (2007). In Chapter 4, we will demonstrate an additional computational process which independently confirms this division.

2.3 Why are linguistic claims in animal research so controversial?

Much of the debate surrounding the application of linguistic terminology to animal vocalizations has stemmed from a disagreement regarding a few important distinctions. These primarily have to do with the differences between allophones (phones) and phonemes and between prosody and syntax. Table 2.1 contains a comparison between the linguistics and bioacoustics communities on how each term is used. Terms like allophone and prosody would be less controversial to use than terms like phoneme and syntax; however they are rarely used. In part, this may be due to the technical distinction between each category.

The major distinction between allophone and phoneme is that allophones are acoustical properties of sounds which can be measured and categorized, whereas phonemes are realized versions of allophones which require a mental representation on behalf of an individual. An example the Oxford dictionary gives describes the difference between the letters in the word "catch" and the phonemes. In this example, the phonemes in a word do not correspond to the letters with which we write it (e.g. the word *catch* contains five letters: *c – a – t – c – h,* but only three sounds: *c – a – tch).* From a bioacoustics viewpoint, this level of comprehension in an animals' mind is currently unrealistic to verify and therefore the term phoneme could never pass the standards of a linguist. On the other hand, non-human animals could have mental representations of their acoustic units and we would have no way of demonstrating the validity of such an argument. Recent work in human language acquisition however has demonstrated that we need not view these two terms as separately as they once were, which may open the door for future work examining this delicate issue.

	Linguistic	Bioacoustic
Allophone	One of two or more variants of the same phoneme <the aspirated \p\ of *pin* and the unaspirated \p\ of *spin* are *allophones* of the phoneme \p\>	Used as an equivalent to phoneme
Phoneme	Any of the perceptually distinct units of sound in a specified language that distinguish one word from another, for example *p*, *b*, *d*, and *t* in the English words *pad*, *pat*, *bad*, and *bat*	Any measureable difference in sound category. Differences need not be confirmed by the focal species
Prosody[2]	The systematic study of metrical structure; the rhythmic and intonational aspect of language	Duration of silent intervals, fundamental frequency shifts and entropic shifts
Syntax	The way in which linguistic elements (as words) are put together to form constituents (as phrases or clauses) A set of rules for the formal properties of languages	Any predictable combination of calls. Can be a repeated pattern and does not need to be recursively embedded structure.

Table 2.1: A variety of definitions exist for each of the linguistic terms in the above table. This variation is compounded by the discrepancy in usage when making comparison between language and animal vocalizations. This table summarizes a few of these notable distinctions.

Recent work in language acquisition of the Inuktitut language suggests that a two stage acquisition of allophones and phonemes may not be the most realistic and in fact presents multiple flaws (Dillon et al, 2013). One of the flaws involves how errors in allophone acquisition are mapped onto subsequent phonemes. Errors in a 2-stage model would accumulate and therefore create an issue for proper mapping of phonemes. This issue is resolved if a 1-stage model of acquisition were used by the learner. This would also lead to questions of how this evolved and leave the potential for some or all animals to possess similar frameworks. For the purposes of this work, we are moving forward with the assumption that gibbons and other animals do possess phoneme-like representations, though the exact amount of mental representation will still be unknown.

---

[2] Merriam-Webster defines the basis of "traditional" prosody in English as the classification of verse according to the syllable stress of its lines. This can include the study of the elements of language, such as metre, that contribute to rhythmic and acoustic effects in poetry. A poem's "sound meaning" can also be influenced by effects such as rhyme scheme, alliteration, and assonance. Nonmetrical prosody can apply to modern poetry, and visual prosody is used when verse is "shaped" by its typographical arrangement.

Another major misunderstanding is between prosody and syntax. Current work in bioacoustics has not provided sufficient data to support claims of syntax. Take for instance the description of free-tailed bat (*Tadarieda brasiliensis*) in Figure 2.7. While the authors are using the term syntax, this does not fit the modern linguistic definition of the arrangement of words and phrases as there is no indication that chirps, trills and buzzes are words or display nested structure. Further, there are no described rules which would move this beyond a display of prosodic rhythm which the top half of the figure shows. While some order exists, there is no evidence that the meaning will change if a buzz and a trill were switched. As an English example in Merriam-Webster describes, “The boy loves his dog” follows standard subject-verb-object word order, and switching the order of such a sentence would change the meaning or make the sentence meaningless (Merriam-Webster, 2014).

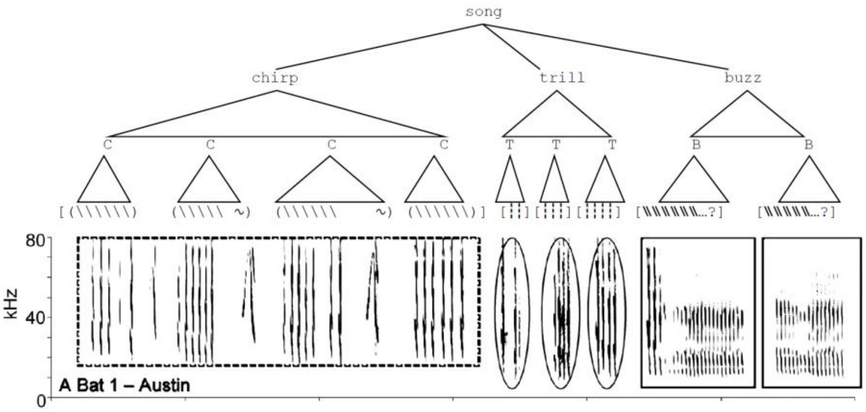


Figure 2.7: The top half of this figure shows how the chirps, trills and buzzes of free-tailed bat vocalizations can be viewed as an example of prosodic organization (Figure provided by Eric Raimy). The bottom half of this figure appears in Bohn et al. (2009) and is part of the claim for syntax. While there is clear structure in these calls, the sequences do not pass the modern linguistic definition for syntax.

How can we move forward from these discrepancies in term usage? Acknowledging the drawbacks to such rigid definitions and finding a mathematical way of drawing comparisons between human language and animal communication is a good way to start. The following chapters will outline a proposed method for doing this and give examples from two species of mammals, white-handed gibbons and black rats.

Chapter References

Becker, M. L., Buder, E. H., & Ward, J. P. (2003). Spectrographic description of vocalizations in captive Otolemur garnettii. *International Journal of Primatology, 24*(2), 415-446.

Bioacoustics, A. (2014). Avisoft. Retrieved from http://www.avisoft.com/

Bohn, K. M., Schmidt-French, B., Schwartz, C., Smotherman, M., & Pollak, G. D. (2009). Versatility and stereotypy of free-tailed bat songs. *Plos One, 4*(8).

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer. Retrieved from http://www.praat.org/

Burt, J. (2005). Syrinx. Retrieved from http://www.syrinxpc.com/

Cleveland, J., & Snowdon, C. T. (1982). The complex vocal repertoire of the adult cotton-top tamarin (*Saguinus oedipus oedipus*). *Zeitschrift Fur Tierpsychologie-Journal of Comparative Ethology, 58*(3), 231-270.

Coen, M. (2006). *Self-supervised acquisition of vowels in American English*. Paper presented at the Proceedings of the Twenty First National Conference on Artificial Intelligence(AAAI), Boston, MA.

Coen, M. (2007). *Learning to sing like a bird: Self-supervised acquisition of birdsong*. Paper presented at the Proceedings of the Twenty Second National Conference on Artificial Intelligence(AAAI), Vancouver, British Columbia.

Dassow, A. M. (2010). *Vocal repertoire of the white-handed gibbon (Hylobates lar)*. Masters of Zoology, University of Wisconsin-Madison.

Descartes, R. (1637). *Discourse on method and meditations on first philosophy* (D. A. Cress, Trans.). Indianapolis: Hackett Pub. Co.

Design, E. (1980). SIGNAL/RTS. Retrieved from http://www.engdes.com/

Dillon, B., Dunbar, E., & Idsardi, W. (2013). A single-stage approach to learning phonological categories: Insights from Inuktitut. *Cognitive Science, 37*(2), 344-377.

Elektronik, P. (1983). Batsound. Retrieved from http://www.batsound.com

Fee, M. S., Shraiman, B., Pesaran, B., & Mitra, P. P. (1998). The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature, 395*(6697), 67-71.

Gamba, M., & Giacoma, C. (2007). Quantitative acoustic analysis of the vocal repertoire of the crowned lemur. *Ethology Ecology & Evolution, 19*(4), 323-343.

Giles, J. C., Davis, J. A., McCauley, R. D., & Kuchling, G. (2009). Voice of the turtle: The underwater acoustic repertoire of the long-necked freshwater turtle, Chelodina oblonga. *Journal of the Acoustical Society of America, 126*(1), 434-443.

Gridi-Papp, M. (2003-2007). SoundRuler: Acoustic Analysis for Research and Teaching. Retrieved from http://soundruler.sourceforge.net

Gros-Louis, J. J., Perry, S. E., Fichtel, C., Wikberg, E., Gilkenson, H., Wofsy, S., et al. (2008). Vocal repertoire of Cebus capucinus: Acoustic structure, context, and usage. *International Journal of Primatology, 29*(3), 641-670.

Hammerschmidt, K., & Fischer, J. (1998). The vocal repertoire of Barbary macaques: A quantitative analysis of a graded signal system. *Ethology, 104*(3), 203-216.

Incorporated, A. S. (2014). Adobe Audition. Retrieved from http://www.adobe.com/products/audition.html

Kobayasi, K. I., & Riquimaroux, H. (2012). Classification of vocalizations in the Mongolian gerbil, Meriones unguiculatus. *Journal of the Acoustical Society of America, 131*(2), 1622-1631.

Margolis, E., and Stephen Laurence, eds. (1999). *Concepts: core readings*: MIT Press.

Masataka, N. (1982). A field study on the vocalizations of Goeldi's monkeys (*Callimico goeldii*). *Primates, 23*(2), 206-219.

Merriam-Webster(2014). Available from http://www.merriam-webster.com/dictionary/syntax

Narins, P. M., Lewis, E. R., & McClelland, B. E. (2000). Hyperextended call note repertoire of the endemic Madagascar treefrog Boophis madagascariensis (Rhacophoridae). *Journal of Zoology, 250*, 283-298.

Program, B. R. (2014). Raven Interactive Software. Retrieved from http://www.birds.cornell.edu/raven

Raemaekers, J. J., Raemaekers, P. M., & Haimoff, E. H. (1984). Loud calls of the gibbon (*Hylobates lar*) - Repertoire, organization and context. *Behaviour, 91*, 146-189.

Raimy, E. (2011). Bat Prosody.

Saar, S., & Mitra, P. P. (2008). A technique for characterizing the development of rhythms in bird song. *Plos One, 3*(1).

Software, P. H. (1993). SpectraPLUS. Retrieved from http://www.spectraplus.com/

SonoBat. (2013). SonoBat. Retrieved from http://www.sonobat.com/index.html

Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour, 59*, 1167-1176.

Thomson, D. J. (1982). *Spectrum estimation and harmonic analysis*. Paper presented at the *Proceedings of the IEEE*.

Van Parijs, S. M., Parra, G. J., & Corkeron, P. J. (2000). Sounds produced by Australian Irrawaddy dolphins, Orcaella brevirostris. *Journal of the Acoustical Society of America, 108*(4), 1938-1940.

Xu, R., & Wunsch, D., II. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks, 16*(3), 645-678.

"By far the simplest assumption regarding the social behavior of the chimpanzee, for example, is that if this species' behavior resembles that of ourselves then the underlying psychological and mental processes must be similar too. To propose otherwise requires that we assume the evolution of divergent processes for the production of similar behavior." (de Waal, 1991)

Chapter 3: Literature Review

Significant progress has been made over the past century regarding the understanding of animal communication systems and human language. Though these lines of research have predominantly been separate from one another, it has become increasingly common in recent years to search for areas of overlap. Although spoken language is widely accepted as the most complex form of vocal communication, what if any overlap exists with animal vocalizations remains largely unexamined. Fueling this interest is constant advancements in technology that continue to provide more empirical methods to studying and comparing vocal systems between different organisms. **Given the distinct histories in each of these fields, this literature review examines the most influential work which has contributed to the current, combined approach used in this dissertation to examine white-handed gibbon and rat vocalizations.**

Section 3.1 of this review surveys the literature which has contributed to our current evolutionary understanding of vocal communication systems. The anatomical constraints in vocal production systems across taxa are examined as well as commonalities which exist between animals and humans. Additionally, proposed approaches to examining animal vocalizations from the linguistics community are discussed. Section 3.2 considers how the zoological community has approached the problem of finding commonalities between language and communication. The reasoning that has been used for isolating subunits of calls in other studies is included in this section. Section 3.3 covers the signal processing literature as it pertains to examining animal vocalizations. This section includes previously used approaches to segmenting and classifying the vocalizations of various species. Section 3.4 reviews how neurobiological and ethological studies have contributed to our current understanding of the meaning of animal vocalizations and how these studies have influenced bioacoustics research.

3.1 Evolutionary Commonalities and Linguistic Viewpoints

Evolutionary viewpoint

A long standing question in research is what if any commonalities exist between humans and other animals. Tenets from evolutionary theory suggest that there should be at least some overlap in derived features. The approaches used to examine potential commonalities in vocal communication have been divided. A predominant viewpoint from the linguistics community is that animals must understand human language before any claims of vocal complexity are accepted. Alternatively, the bioacoustics community asserts a rationale for language-like complexity in animals based upon the evolutionary unlikelihood of language spontaneously developing without any precursors. This review evaluates the relevant literature from linguistics and bioacoustics which has sought to join each view.

The linguistics community has a lengthy tradition of separating humans from other animals without formal justification for doing so. Numerous studies have overturned this division. One key approach has been examining genetic similarities based upon evolutionary theories. For instance, one can theorize that vocalizing is a common trait and therefore its conservation and modification should be found as phylogenetic commonalities across species. When examining genetic evidence, the development of vocalizations have been traced back to an evolutionary origin (Bass et al., 2008), thus supporting a theory of vocal commonality. Novelties, such as vocal learning (Marler, 1970; Nottebohm, 1972; Scharff, 1991; Adret, 1993; Marler, 1997; Deecke, 2000; Hahnloser, 2010; Goodale, 2010; Hoeschele, 2010; Lahti, 2011) and sophisticated mating duets (Raemaekers, 1984; Levin, 1996a; Levin, 1996b; Nietsch, 1999; Hall, 2000; Grafe, 2004; Murphy, 2006; Benedict, 2009; Bradley, 2009) have been shown to arise as the ancestrally shared vocal basis evolved in different groups. Humans are not separate in this evolution, as phonemic diversity has been traced to an inferred origin in Africa (Atkinson, 2011) and variation in ossicle structure has been discovered in the fossil hominin *Australopithecus africanus* (Quam et al., 2013). That these fossil hominins possessed an extra ossicle suggests that our distant relatives perceived sound in a different manner than we do, thus highlighting the importance of anatomical structures.

Structural similarity in neural regions has been studied extensively within a species and between species. Functional differences within brain regions have been described by mapping brain regions within a given species (Zilles & Amunts, 2010). In humans, this effort began with a cytoarchitectural map of the brain better known as Brodmann's map (Brodmann, 1909). While many of these regions have been debated and described in more detail, this work greatly influenced neuroscientific research. An important contribution to this field was the development of understanding the functional significance of these different brain regions. In human brains, early neurobiological studies learned that areas like Broca's area (Broca, 1861) is important for speech production, while Wernicke's (Wernicke, 1874) area is important for comprehending spoken and written language. These areas are found in the pars opercularis, Brodmann's area 44, and pars triangularis, Brodmann's area 45, of the inferior frontal gyrus and the posterior section of the superior temporal gyrus, Brodmann's area 22 &39, in the dominant cerebral hemisphere (Brodmann, 1909). For one interested in studying speech production or comprehension, these areas are reliably traced from one person to another with a small amount of variation due to intersubject variability or problems of observer-dependent parcellation techniques. Discovering general regions which can be reliably traced was an important step in furthering our understanding of how to formally study vocal development.

Finer-grained analyses of brain regions helped to establish further justification for viewing vocal development as a heavily intertwined part of neural development. Additional language related regions of Wernicke's area have been discovered and include the superior temporal gyrus located in the caudal part of area 22, angular gyrus located in area 39 of the parietal lobe and supramarginal gyrus located in area 40 of the parietal lobe (Zilles & Amunts, 2010; Gannon et al., 2001). The prefrontal cortex, areas 9, 10 and 46, and the neocerebellum were confirmed to play an important role in semantic processing (Petersen et al., 1998; Petersen and Fiez, 1993). Further discoveries include the primary auditory cortex, area 41 & 42, which is important for auditory perception and processing semantic information, the sylvian fissure, and the planum temporale, located behind areas 41 & 42 within the Sylvian fissure, which interprets language sounds (Gannon et al., 2001). The discovery and understanding of these areas has been important for understanding human languages and it has facilitated an approach to drawing insight into other species' vocalizations by looking for similarities in neural structures and inferring similarities in function.

A common assumption in cross species studies is that similarities in structural form match similarities in function. This assumption has formed the basis for justifying the use of model organisms in several lines of research, including neurobiology. For studies exploring vocal evolution, songbirds (e.g. zebra finches, canaries, white-crowed sparrows and starlings) have been the preferred organisms to examine (Thorpe, 1961; Margoliash, 1983; Simpson & Vicario, 1990; Nordeen & Nordeen, 1992; Vu et al., 1994; Yu & Margoliash, 1996; Riters & Ball, 1999; Slater, 2003; Haesler et al., 2004; Jarvis, 2004; Nottebohm & Liu, 2010). Research with these organisms has led to the discovery of important neural pathways involved in song production and learning, as well as what motivates singing (Riters et al., 1999; Riters et al., 2000; Ball et al., 2004; Heimovics & Riters, 2005; Alger & Riters, 2006; Alger et al., 2009; Riters, 2012; Riters & Stevenson, 2012; Kelm-Nelson & Riters, 2013). While it is slightly less common, primates have also served as important model organisms for evolutionary studies on vocal development (Hopkins et al., 1998; Eliades & Wang, 2008; Losin et al., 2008; Jürgens, 2009). As Gannon et al. (2001) have shown, the planum temporale in white-handed gibbons plays an important role in the perception of vocalizations. Additional studies have been important for furthering our understanding of other areas related to communication systems as well as offer support for the view that vocal systems have evolved and as such we should expect similarities between species. One such example is a study conducted by Martin-Ordas et al. (2013). In this study, chimpanzees and orangutans displayed evidence of the ability to recall past events. Demonstrating this ability outside of humans is important for establishing more credibility for evolved communication systems, as it suggests that these animals use an active thought process when assessing situations rather than simply reacting to events.

Numerous studies have discovered that vocal tract anatomy plays an important role in predicting what constraints are found in sound production between species (Fitch and Hauser, 1995; Lieberman and McCarthy, 1999; Lieberman, 2002; Nishimura, 2003; Nishimura et al., 2003; Nishimura, 2006; Nishimura et al., 2006). Nishimura and colleagues (2003a; 2003b; 2006a; 2006b) have provided significant empirical evidence demonstrating the importance of a specific shared feature, the lateral thyrohyoid ligament, between humans and non-human hominoids. This piece of ligament joins the hyoid bone to the larynx and provides a greater level of flexibility within the supralaryngeal vocal tract. The flexibility in the supralaryngeal vocal tract is known to be important for generating vowels in human speech and was previously viewed as a major limiting factor in vocal production of non-human animals (Lieberman, 1968; Lieberman et al., 1969). More recently, however, there has been support for a descended vocal tract in other species such as red deer (Fitch & Reby, 2001) as well as formant production, which is an important component of vowels, in Diana monkeys (Riede et al., 2005). These lines of evidence open up the possibility that vocal features are not necessarily exclusive to humans.

The similarities in neural, vocal tract and auditory anatomy provide more support for the evolution of communication systems, including those of humans. A considerable amount of work will be needed to test the evolutionary relationship between humans and non-humans in other aspects of vocal communication, such as memory storage and additional acoustic feature production. Moreover, future work should continue to strive towards developing empirical assessments of comparative studies, while remaining cautious of overstating the significance of findings which have multiple factors influencing their outcome (e.g. neural substrates and vocal tract restrictions on sound production). Taken together, the current findings suggest we may proceed cautiously with drawing parallels between human and non-human animal vocal communication systems.

Linguistic viewpoint

Human languages have been studied extensively with a major focus on defining the linguistic rules which describe a particular language. Common ways to describe a particular language may include a description of allophones, phonemes, prosody, syntax, vowel harmony, syllables, transitional probabilities, tone, etc. (Hockett, 1947; Mandelbrot, 1953; Harris, 1955; Chomsky, 1957; Harris, 1955; Shipman & Zue, 1982; Cutler & Carter, 1987; Cutler & Norris, 1988; Goodsitt et al., 1993; Saffran et al., 1996; Kessler & Treiman, 1997; Jusczyk et al., 1999; Krakow, 1999; Johnson & Jusczyk, 2001; Pelucchi et al., 2009a; Pelucchi et al., 2009b; Saffran et al., 1999; Saffran, 2003). **Given the vast amount of work describing human languages, it should come as no surprise that there are significantly more properties described in human language than there are in animal vocalizations.** This is in part due to the unlikelihood of certain properties showing up in animal vocalizations and in part due to the lack of studies examining linguistic correlates. While the number of properties in human languages exceeds what can be described here, a few candidates that are likely to show up in animal vocalizations will be discussed. Allophones and phonemes are the basic building blocks of language and as such they are a good place to start with analyzing sounds. Identification of allophones and phonemes in human languages has been thoroughly described by a set of rules specific to each language that loosely follow describing the place features, manner features and laryngeal features required for sound production (Ladefoged & Halle, 1988). The number of phonemes in human languages varies greatly (Evans & Levinson, 2009), yet they all appear to follow the same foundational principle of passing the minimal pair test (Pike, 1947). Pike (1947) defined the minimal pair test as a comparison between two basic words that only vary in one sound. For example, the English words ‘cat’ and ‘bat’. In this example the /k/ sound in ‘cat’ is different than the /b/ sound in ‘bat’. These differences also correspond to different meanings in these two words. This test works well for human language because we *know* what these words mean; however as we will point out later, applying these same procedures to non-human animals is often times problematic.

Allophones have widely been accepted as fundamental units of sound because each allophone is found in a place or environment which is described by predictable statistical distributions (Twaddell, 1952). While the environments are not necessarily found as normal distributions, each environment is reliably found by a set of specific rules[1]. The distinction between allophones and phonemes was discovered by linguists several decades ago and relies on the difference between them (Trubetzkoy, 1939). Phonemes are abstract units, while allophones are acoustic characterizations which are dependably found. To describe the differences between allophones and phonemes, comprehension of the rules guiding the language must be present; however classifying the statistical distribution of the allophones merely requires a large enough data set which encompasses all variation in acoustical unit distribution. Since allophones are described by statistical rules, they are widely accepted as fundamental units in non-human animal species. How they overlap with phonemic representations or broader range categories is highly debatable (Tesar & Smolensky, 1998; Boersma & Hayes, 2001; Hayes, 2004; Peperkamp et al., 2006; Goldsmith & Xanthos, 2009; Dillon et al., 2012).

In the literature there is a debate about whether any non-human animals are capable of comprehending signals which contain the linguistic phenomena present in human languages

---

[1] In examining the distribution of laterals or /l/ sounds in the word set [blend, cling, file, wealth], the [ɫ̪] is always found after a vowel and before a voiceless interdental fricative, [θ]. The [ɫ] is always found at the end of the word and is preceded by a vowel or a glide. The [l̥] is always found after an aspirated sound and before a vowel. The [l] is in a slightly more complex environment so it is defined as the elsewhere case. Since these environments slightly overlap, they are often simplified. To summarize our rules in a more concise manner, our phoneme is /l/ with [l̥] being found after a voiceless sound, [ɫ] being found at the end of a word, [ɫ̪] being found before a [θ] and [l] being found in all other cases.

(Hauser et al., 2002; Fitch et al., 2005; Pinker & Jackendoff, 2005; Jackendoff & Pinker, 2005; Berwick et al., 2011). **The dismissal that animals cannot comprehend complex linguistic phenomena has often times been made without any formal analysis of data**, but it has also been made by demonstrating the lack of certain linguistic phenomena (Yip, 2006). A common argument for dismissing an animal's ability to comprehend these types of signals is to test their ability in identifying linguistic properties which are highly unlikely to occur in their natural repertoire (Hurford, 2004; Yip, 2006). Given the vast number of theoretical properties identified in human languages, it is no surprise that other communication systems will lack at least some of the exact same properties. For example, vowel harmony is a property that is found in several languages. An argument made by Yip (2006) asserts that animals must be sensitive to vowel harmony if we are to classify them as sophisticated vocalizers. This argument is illogical for the vast majority of animals who do not possess the required anatomical structures for producing vowels. If no mechanism for production exists, then there would not have been any selective evolutionary pressure for comprehension. In this sense, the question of linguistic analogs in animal vocalizations is often tautologically dismissed, by posing it in terms of phenomena that are unique to humans. This seems unreasonable and would be better presented as a generalizable notion of phonetic structure that captures properties of phonemes that are observed in a wide variety of species (Samuels, 2009a, b; Samuels et al., 2010). Doing this neither devalues unique properties of human phonology nor does it attribute unrealistic capabilities, such as an ability to distinguish trochaic vs. iambic rhythm (Yip, 2006), to other animals. It does however allow us to address questions about vocal complexity in ways open to evaluation.

Methods for examining phonemic categories, such as minimal pair testing, have been proposed by linguists, but have not been formally studied in many species of non-human animals (Pike, 1947). This gap is in part due to the constraints in experimental design. Yip (2006) provides a classic example of how a category error in experimental design often times yields misleading results. Yip argues a need to test an animal's ability in understanding human language before making claims regarding comprehension. Her proposed experimental design, which tests a tamarin's ability to comprehend Turkish, is based on previous research which tested cotton-top tamarin's comprehension of Dutch and Japanese (Ramus et al., 2000). Yip (2006) asserts that a failure to comprehend Turkish demonstrates a lack of sophisticated linguistic ability in tamarins. This interpretation does not take into account the fact that Turkish is not the tamarins natural communication system and therefore their comprehension of Turkish is as irrelevant to their vocal communication system as a comprehension of ultrasonic calls are to our communication system. While the concept of a minimal pair test has the potential to be a valuable tool to researchers, the units being tested must be part of the animal's natural repertoire to have any value in bioacoustics studies.

Habituation to experimental procedures has also been identified by many researchers (Thompson & Spencer, 1966; Groves & Thompson, 1970; Rankin et al., 2009) as an additional obstacle in applying minimal pair testing methods to animal vocalizations. The issue of habituation has led to problems with interpreting and validating experimental results. Interpreting the results of a minimal pair test in animals is challenging because the animals often times become habituated to the experiments after the first test and they fail to perform reliably during subsequent attempts (Thompson & Spencer, 1966; Groves & Thompson, 1970; Rankin et al., 2009). Even without habituation, the ground truth to decipher meaning is still only available to members of each species. The effort required to circumvent these issues is often far greater than the benefits received from doing so. In spite of this, Zuberbühler and colleagues (1999) and Weary (1989) attained some success in examining the vocalizations of a few species via elaborate experiments to test an individuals' ability in perceiving an isolated phoneme. For many other species, an assumption of meaningful differences in phonemes is made. This assumption is supported by examining the distinctiveness of acoustic features from one phoneme to another. In some species, it is also supported by similarities in auditory perception between the focal species and human observers (Kuhl & Miller, 1975).

Whether or not referential signaling, an association between call sequence and environmental event, occurs in non-human animals has also been debated in the literature (Marler et al., 1992). Questions regarding whether any linguistic properties exist in animal vocalizations have been discussed in great detail using a well-known bioacoustics example of vervet monkey alarm calls (Seyfarth & Cheney, 1980a, b). Hurford (2004) argues that predator warning signals from animals such as vervet monkey barks are "not inherently leopard-like" and thus not referential. Hurford employs the analogy to a human jogging to assert vervet monkey barks cannot be symbolic of leopards. In his example, jogging could indicate someone is being chased or catching the bus. This is a poor comparison however because the human example is indicating an action, running, whereas the vervet example is likely indicating an object, leopard. While actions often have multiple meanings, objects typically have a single meaning, though there are exceptions to each. Despite this discrepancy, Hurford is likely correct in the assumption that humans have significantly more referential signals than any other species. This should not however lead to the conclusion that animals are incapable of having referential signals.

Despite some of the above disagreements in term usage, there is much to be learned from the linguistic community regarding the applications of methods and rules to studying vocal communication in animals. Future work which keeps a strong focus on the animal's primary mode of communication is needed and this will require an assumption of phonemic referential signals being present. Also, it is important to acknowledge the fact that humans and animals possess different perceptual features such as vowel harmony vs. phase. Animal vocalizations need not rely on vowel harmony any more than human language needs to rely on phase. By moving past these discrepancies, we can move forward with experiments which examine animal vocalizations in more detail and start to describe their capabilities on a spectrum based view rather than a discrete view. This research assumes that the phonemes being described have a meaningful difference in the given species vocal repertoire. By doing so, we will address the unanswered question of whether interior structure is present in animal vocalizations.

3.2 Phonetics in Non-human Animals

Given the highly controversial nature of this topic, we believe that the subset of bioacoustics papers covering possible correlations to human linguistic terms deserved its own section in this literature review. Despite the flaws in many of these studies, they present an important step towards reaching out to other scientific communities to unify ideas on how vocal communications systems evolved. This section will review the most controversial and influential work to date and discuss some ways to move beyond some of the rejected claims.

Issues with examining phonetic correlates in animal vocalizations

There are surprisingly few attempts made at examining the vocal repertoires of other species, especially given the vast amount of knowledge we have concerning human phonetics. This discrepancy has arisen for a few reasons. First, gathering large data sets of animal vocalizations is significantly more difficult than gathering data sets from humans. To acquire a rich and diverse set of data from an animal, a researcher must spend months to years in the wild recording daily bouts of calling. These efforts are both time consuming and costly. To reduce costs, some researchers have opted for recording captive animals (Shepherdson et al., 1989; Geissmann, 2002; Ruppell, 2007; Frey et al., 2007); however caution must be used when relying solely on captive vocalizations, as they may not be representative of natural repertoires. A second reason examining vocal repertoires have been problematic is due to lack of technology. The majority of the papers published on the matter employed various forms of tape recorders which produce low resolution or low quality representations of the sounds (Masataka, 1982; Cleveland & Snowdon, 1982; Hammerschmidt & Fisher, 1998; Van Parijs et al., 2000; Narins et al., 2000; Becker et al., 2003; Gamba & Giacoma, 2007; Gros-Louis, 2008). This is an issue because without higher resolution, the subtleties within the general sound categories may not be able to be distinguished from one another. An additional reason for the lack of vocal repertoire descriptions is due to the lack of automated analysis techniques (See section 3.3 for further discussion). In the bioacoustics field, animal vocalizations have traditionally been segmented and labeled by hand. This is an incredibly labor intensive process which requires months or years to complete. To ensure there isn't observer bias, these sounds typically have to be labeled by multiple observers and then compared with one another. This is often times far more time consuming than what a researcher can reasonably put into a single project. Also, determining whether to divide a sound category into subcategories is an issue. For example, Shapiro et al. (2011) demonstrate that both types of classifications are possible for Norwegian whales, but further testing is needed to determine what the whales perceive. Despite these issues, there are a few examples of how segmentation and classification of vocal repertoires have been approached in the past.

It has previously been noted that several species of animals appear to have graded signals rather than discrete signals. Graded signals are particularly difficult to classify because of the lack of a distinct separation point between categories (Newman & Goedeking, 1992). For these signals, methods such as k-means clustering have been employed (Hammerschmidt & Fisher, 1998). A major drawback to this type of approach is that k-means clustering relies on the observer to choose the number of categories (k) that are expected in a data set. In other words, it would be looking for *k* number of categories because one expects to find k number of categories. This is not the most unbiased approach to the problem, but as mentioned earlier many of these early studies were constrained by the tools available at the time. Still, this leaves an open-ended question of how to best approach the problem of identifying categories, especially when they are not discrete.

Another method that has been employed, albeit controversially, by researchers is the use of Zipf's law (Zipf, 1949; Mandelbrot, 1953; Hailman et al., 1985; McCowan et al., 1999; Suzuki et al., 2004; McCowan et al., 2005). Zipf's law is an empirical observation which predicts the occurrence of an entity be exponentially proportional to its rank. It has been demonstrated that the distribution of sounds and words we use in speech follows Zipf's law, which has both helped and hurt its use when applied to animal vocalizations. McCowan et al. (1999) suggested Zipf's law as a good tool for comparing the internal information structure of bottlenose dolphin whistles, but received a strong rebuttal against its use by Suzuki et al. (2004). In this rebuttal, Suzuki et al. (2004) argues against the value of this tool by illustrating Zipf's law also applies to a die-rolling experiment. In this example, the authors demonstrate that words generated by the experiment occur at a frequency which follows Zipf's law, but they do not contain semantic information and therefore there is no use for its' application to communicative repertoires. The major drawback to this line of thought is that the experimental design does not observe the application of the Zipf statistic properly. Data sets which follow Zipf's law may either indicate that there *is* or *might be* communicative value present. In other words, Zipf's law is not a definitive tool for claiming communicative value, but it is a useful statistic for providing support for the existence of such value. If a data set of animal vocalizations does follow Zipf's law, then there is reason to pursue further testing of the data set; however if it does not follow Zipf's law then there is no communicative value in it and it is best to move on to a different species.

What does the number of sound categories in a given species mean?

Diversity of sounds within a vocal repertoire has been studied in several species, but what this diversity means has received little attention. Of the work currently published, cotton-top tamarins have 38 acoustic units (Cleveland & Snowdon, 1982), Goeldi's monkeys have 40 acoustic units (Masataka, 1982), Barbary macaques have 7 acoustic units (Hammerschmidt & Fisher, 1998), Madagascar treefrogs have 28 acoustic units (Narins et al., 2000), Australian

Irrawaddy dolphins have 6 acoustic units (Van Parijs et al., 2000), Garnett's greater bush babies have 11 acoustic units (Becker et al., 2003), crowned lemurs have 10 acoustic units (Gamba & Giacoma, 2007), white-faced capuchins have 27 acoustic units (Gros-Louis et al., 2008), Australian long-necked turtles have 17 acoustic units (Giles et al., 2009), and Mongolian gerbils have 8 acoustic units (Kobayasi & Riquimaroux 2012). While it is interesting to note relatively simple organisms such as a long-necked turtle appear to possess a more diverse vocal repertoire than a relatively sophisticated organism such as a dolphin, there is little to infer from this knowledge regarding how this relates to communicative complexity. This is in part due to how vocalizations are examined. Describing repertoires based upon frequency characteristics does not indicate anything about how these sounds are used to pass messages along to a receiver and it does not indicate anything about how vocalizations of one species are related to another species. The first issue concerning meaning is something that behavioral studies often times address. The second issue concerning relatedness has received less attention; however it is an important issue for those interested in learning about how vocalizations have evolved.

Describing vocal repertoires using frequency range limits what can be inferred about relationships to the relative size of an organism. For example, a set of ultrasonic vocalizations implies the animal producing the calls must be a small organism such as a mouse. In a set of infrasonic vocalizations, the animal producing them must be a larger animal such as an elephant or the animal must possess an atypical supralaryngeal vocal tract such as that found in koalas (Charlton et al., 2013). Examining the acoustical properties of sound and tracing their occurrences throughout time provides a more inclusive approach to studying the evolutionary relationships of calls. For example, it is well known that owls use phase to help them hunt for their prey (Knudsen, 1982). Interesting questions to pursue would be what other species use phase and are there any other functions of this sensitivity. While these questions have yet to be addressed formally, the methods necessary to address these questions will be described and future research will be devoted to discovering these types of connections.

Claims of syntax

Many attempts to draw parallels between human language and animal communication systems have been made. In some cases, the claim of syntax has been presented. These studies have been criticized by the linguistic community. Some of this criticism is due to variations in how the meaning of *syntax* was intended (See Table 2.1). It most frequently appears in papers which are showing some regular expression of order, but no embedded structure (Clarke et al., 2006; Bohn et al. 2009; Abe & Watanabe, 2011). The differences between these meanings are significant. Nested embedded structure is required for recursive grammar and has only been conclusively demonstrated in humans. Regular expression or prosody explains the stresses or patterns of vocal data sets and is found in some non-human animals.

The conflict in term usage arises in papers in several ways. One of the major drawbacks to Clarke et al. (2006) was how the predator categories were classified and examined. When examining predator call categories, gross categories of sounds were used and the labeling and analysis was limited to a sequence of 10 sounds with all individuals being analyzed together. Though this may not seem like an issue, the gibbons in these recordings are not always calling at the same time. This introduces the possibility that they may not have all been calling about the same thing which in turn introduces the problem of suggesting an overly complex system of vocal warnings. In Chapter 5, we will explore this issue further and offer suggestions for ways to improve future analyses to avoid this possible mischaracterization. Bohn et al.'s (2009) paper concerning free-tailed bats has shown prosody in bat vocalizations, which is interesting in its own respect; however with the data at hand, there is no evidence that these 3 call types are used in more complex ways. Finally, in the paper by Abe & Watanabe (2011), the claim of learning artificial grammar rules by Bengalese finches has elicited objections. One of the main arguments against this claim is that the songs are k-reversible so they could still be learned (Becker et al., 2012). Becker et al. (2012) offer a plausible alternative to Abe & Watanabe's

(2011) conclusion that suggests the birds are recognizing and discriminating the sound patterns based on acoustical memory representations, not context-free grammar rules.

Attempts to draw parallels between human and animal vocalizations have previously failed due to a misunderstanding of linguistic terms and not enough computational advancement to illustrate similarities. Given the collaborative nature of current research and the technological advancements in recent years, these obstacles have been reduced. Resolving the issues of call classification is still needed before future questions regarding vocal usage are addressed.

3.3 Signal Processing in Bioacoustics

The field of signal processing addresses a vast array of applications, including the processing of images, sensors, electromagnetic radiation and sound. For bioacoustics researchers, it is the advancements in sound processing or analysis that has been the most beneficial. These advancements have helped to address a persistent issue in bioacoustics research, which is the problem of how to analyze data sets consistently and efficiently.

One subfield of sound processing, automated speech recognition (ASR), has been key to developing algorithms to identify lexical sequences in spoken human language (Rabiner & Juang, 1986; Juang & Rabiner & 1986; Juang et al., 1986; Rabiner, 1989; Huang et al., 2001) and some have believed these algorithms have the potential to be used in bioacoustics. ASR algorithms rely on models of language that enable phonetic identification, lexical categories, phrasal groupings and sentence assembly. **However, modifying human language algorithms to fit the communication systems of animals is problematic because we lack formal theories of what rules govern their communication systems.** The most we have are statistical models of acoustic elements. One way of addressing this issue, which we elaborate on in chapter 4, is to test whether an automated approach is sufficiently robust to detect phoneme boundaries in human languages as well as similar proposed boundaries in other species vocalizations. By doing so, some of the bias associated with boundary placement is diminished.

Other approaches to automating analysis of animal vocalizations and human language have been developed with varying levels of success. These are roughly broken down into attempts to segment sequences of calls and attempts to classify calls. Within these, more effort has gone into classifying calls which have been segmented manually. Attempts to automatically segment a continuous stream of vocalizations has been limited (see below for more discussion), though the work that has been done has provided some extremely useful tools for our current work.

Segmenting vocalizations

Segmenting a continuous set of animal vocalizations into its subunits is a long-standing interest to those studying vocal communication. The traditional approach to segmenting calls has been to identify boundaries of notes by the silences that precede and follow each sound. Boundaries are typically identified by a researcher manually, which is both extremely time consuming and subjective. To address these issues, several methods for automating, or at least partially automating, the segmentation of calls has been developed.

Two of the earliest semi-automated methods tested on a dataset of animal vocalizations were the use of Dynamic Time Warping (DTW) and Hidden Markov Models (HMM) on a data set of bird song by Kogan & Margoliash (1998). In DTW, a series of song templates are used to label unknown sequences. This was first introduced in human speech recognition by Vintsyuk (1971), and later modified by Bridle et al. (1982) and Ney (1984). One of the major drawbacks to its use with animal vocalizations is that it is only reliable for species whose vocalizations are not modified during production. For species with variable calls, templates either fail to match all of the variability within a set of calls or the templates fail to discriminate between differences in calls. This trade-off between template effectiveness and vocalization separability has greatly limited the use of this approach in animal communication studies. The other approach tested in this study was the use of HMMs. HMMs were developed by Rabiner and Huang (1986) and work by modeling temporal data using a series of hidden states. States are often separated by Markov chains using graphical Ergodic theory. HMMs differ from template matching by forming statistical boundaries for a given vocalization. These boundaries help to establish models that generalize better and account for the subtle variability in vocal output. Despite HMMs being better than DTW at consistently separating calls, they still have the drawback of not performing well in certain conditions such as low signal to noise environments (a common problem with animal vocalization files) and also lack the utilization of discriminative features to separate calls.

The drawbacks of HMMs have led to the development of hybrid analysis systems. These hybrid systems use HMMs and Artificial Neural Networks (ANNs), which rely more on transitional probabilities rather than observer probabilities (Bourlard et al., 1994). ANNs use a learning/training phase and a classification phase to segment calls. The learning phase takes in inputs to create different categories by setting weights or values to the data set and the classification phase takes new data and determines which of the predefined categories they most closely match. Potter et al. (1994) presented this technique as an improved way of detecting bowhead whale (*Balaena mysticetus*) song notes. The success of this approach led to its use by other researchers interested in call discrimination (Mercado & Kuh, 1998; Murray et al., 1998; Mellinger & Clark, 2000; Schon et al., 2001), individual identification (Reby et al., 1997; Deecke et al., 1999; Campbell et al., 2002), call structure analysis (Dawson et al., 2006; Nickerson et al., 2006), and species identification (Houser et al., 1999; Parsons & Jones, 2000; Parsons, 2001; Deregnaucourt et al., 2001; Chesmore, 2001, 2004; Chesmore & Ohya, 2004; Tian & Shang, 2006). A drawback to this approach and a general issue in machine learning is that each of these studies has been heavily influenced by the training stage and requires predefined categories which limit their ability to detect new categories.

Moving in a slightly different direction, Tchernichovski et al. (2004) developed an approach to map out the development of song in zebra finches. This approach builds Dynamic Vocal Development (DVD) maps as the birds song begins to crystallize with age. The maps utilize spectral derivatives rather than a traditional sound spectrogram (Thomson, 1982; Fee et al., 1998; Tchernichovski et al., 2000) to enhance harmonic clarity. The advantage of the spectral derivative is that it is utilizing the change of power rather than the power spectrum versus time. This sharpens the frequency contours which make them easier to detect. These derivatives are then applied to all of the songs gathered by the developing bird and acoustic features, such as frequency modulation, are used to trace the development of each syllable. This representation is useful for researchers interested in song learning and development; however it also requires a large data set which can only reasonably be gathered in a lab and therefore is limited in its usefulness in understanding wild populations.

Subsequent work with this data set developed a promising approach to segmenting adult vocalizations by introducing self-supervised machine learning and cross-modal clustering (Coen, 2006a; Coen, 2006b; Coen, 2007). Unlike other work, this approach segments calls by detecting peaks in each signal after the smoothed log power is determined. After these peaks are found, the local adjacent minima are then used to establish the boundaries of each sound. This was an important innovation because it established a way of examining the interior structure of sound rather than relying solely on general sound categories. After all of the boundaries are placed, each sound has several different acoustic features extracted and compared via a clustering algorithm. A drawback to this approach is that it is parametrically tuned to zebra finch song and requires modification for use with other species; however it achieves a far more important goal of providing a way to move from syllable recognition to acoustic unit recognition.

As highlighted by the above approaches, there is a great deal of potential for developing an automated approach to segmenting animal vocalizations. Difficulties such as signal to noise ratios have been dealt with by filtering out ambient noise to a certain extent. By working with a more diverse data set, the issue of overfitting models to a single species may be addressed. Given the vast array of bioacoustics libraries of sound, finding samples to work with is possible. It is important to address a concern such as overfitting since there is a wide array of vocal species and to be able to make any evolutionary comparisons, it is helpful to maintain some consistency between analyses. Additionally, a more universal approach to analysis is beneficial for connecting previously disconnected communities within the bioacoustics field. Finally, an inclusive automated approach for sound segmentation will help to reduce any bias in observer error or unusual call identification. While these approaches may over segment calls by dividing categories that shouldn't be separated, this is more of an issue for classifying calls which will be discussed next.

Classifying calls

Determining what constitutes a sound category for a given species is a fundamental, yet challenging aspect of bioacoustics research (Margolis & Laurence, 1999). Varying methods for how to identify perceptual boundaries of calls has been addressed by several researchers. Many of these approaches are good attempts at solving the problem of call classification; however; there are several drawbacks to each process. Some of the major drawbacks to the various approaches include having limits to the robustness of the approach (Tooze et al 1990, Fristrup & Watkins 1992, Collier et al. 2010; Han et al 2011; Kobayasi & Riquimaroux 2012), using only a subset of the vocal repertoire (Tooze et al., 1990; Clemins et al, 2005) or being overfit to a single group or species (Chesmore 2001, Campbell et al. 2002, Mellinger & Bradbury 2006, Pozzi et al. 2010, Adi et al 2010). Despite these drawbacks, these studies are all moving towards a more formalized approach to sound classification.

A common approach to classifying calls is to use a variety of frequency measurements to describe each sound. This is either done by hand using sound processing programs like Raven or SASLab Pro (Tooze et al 1990; Kobayasi & Riquimaroux, 2012) or it is automated (Fristrup &

Watkins, 1993; Mellinger & Bradbury, 2006). One of the major drawbacks to using programs like Raven or SASLab Pro is that the researcher has to identify the beginning and end of each sound. This is extremely subjective and is difficult to maintain consistency. To counter this drawback, Mellinger & Bradbury (2006) developed an automated approach which uses an annotation box to identify where a sound is and a feature box within the annotation box which measures values such as duration, bandwidth, and energy threshold to identify time and frequency boundaries. This has been quite successful with marine mammal vocalizations but needs to be explored further for its potential use with other species.

Another approach to classifying calls is to use acoustic features that are extracted from the sounds. This has an advantage over using frequency measurements by moving the analysis towards the possibility of a more universal approach to understanding vocal communication. By looking at different acoustic features, it is possible to examine similarities and differences in feature usage (e.g. phase) across species which draws upon evolutionary commonalities. To do this, Fristrup & Watkins (1993) have developed their own analysis package, AcouStat, which automatically extracts 26 different acoustic features. This initially seemed promising; however further development of it appears to have been stalled. In other efforts, Han et al. (2011) tested the value of using spectral centroid, Shannon entropy and Rényi entropy to classify Australian anuran vocalizations. These features appeared to work well for simple calls (e.g. frequency unmodulated, short duration calls), but are less useful for complex calls such as in bird song. Though these initial attempts have been limited in their success, they move us closer to finding an automated approach which is useful across lineages.

As mentioned earlier, Hidden Markov Models (HMMs), and Artificial Neural Networks (ANNs) have been used for segmenting vocalizations; however they have also been used for classifying calls. Additionally, Gaussian Mixture Models (GMMs) have been used for classifying calls (Dempster et al., 1977; Bilmes, 1998). Static GMMs differ from generative HMMs by not utilizing the temporal progression of the call and therefore a HMM is a series of GMMs which are used to explain the temporal progression. The use of HMMs (Anderson, 1996; Clemins et al., 2005; Herbst et al., 2008; Adi et al., 2010), Gaussian Mixture Models (GMMs) (Adi et al., 2010) and ANNs (Chesmore, 2001; Campbell et al., 2002; Pozzi et al., 2010) in call classification have been used by many researchers with varying levels of success. Anderson (1996) also tested a Dynamic Time Warping (DTW) system (Vintsyuk, 1971; Bridle et al., 1982; Ney, 1984) for classifying the calls of 2 song bird species and concluded that DTW works better for small data sets, but HMM is more robust to noise and more variable or complex vocalizations. Given the fact that most bioacoustics research takes place outside of anechoic chambers, the ability for an automated process to handle ambient noise is extremely useful and as such HMMs show the most promise for use in future classification tasks.

Perhaps the most advanced work in this field has come from Tchernichovski et al. (2000) and Saar & Mitra (2008). Tchernichovski et al (2000) used zebra finch song to develop an automated measurement to song similarity using spectral derivatives and a similarity matrix. This removed much of the bias of measurement error by being an automated approach. Saar & Mitra (2008) developed an important contribution of features by developing a method for determining rhythmic structure in zebra finch song development. Both of these advancements helped to influence the approach developed by Coen (2006a, 2006b, 2007) and the work in this dissertation.

There are many reasons why so much effort has gone into developing automated approaches to sound segmentation and classification. Automated processing is much faster which leaves more time for conducting field or lab experiments which test behaviorally inspired questions. A more empirical approach to data processing also reduces criticism from skeptics by moving the analysis from a heavily biased human process to a less biased computer process, though it is important to note that humans still program computers so bias is never completely removed. Overall, there remains a gap in the development of a fully automated, robust approach to segment and cluster vocalizations across taxa.

3.4 Bioacoustics

A desire to comprehend the diverse array of sounds produced by different species has led to the field of bioacoustics. Observations of animal interactions generated an early understanding of communication requirements, which included a signaler and receiver. A signaler must produce a sound and a receiver must comprehend this sound for communication between the two individuals to be effective (Bradbury & Vehrencamp, 1998). Sound production and comprehension require a variety of neural and muscular inputs and outputs. This is necessary for all species, regardless of the type of mechanism employed for sound production (e.g. stridulation of two body parts, swim bladders inflation, or air passage across a single syrinx, a dual syrinx, vocal folds or vocal cords). Examination of the similarities and differences in sound production and usage has led to much input from other disciplines. Ethology has had a particularly strong influence on the bioacoustics field. Some of the contributions from this discipline are discussed as well as potential applications of our research to questions in this field.

Contributions from Ethology

Understanding animal behavior, including various signaling behaviors, has been the focus of many ethological studies. These studies range in their invasiveness, but they have contributed some important information in understanding the evolutionary significance of different communication systems. Many organisms use multiple modes of signaling to communicate with conspecifics. Gestural (Tomasello et al., 1985; Goodall, 1986; de Waal and Johanowicz, 1993; Tomasello et al., 1993; Tanner & Byrne, 1993, 1996; Pika et al., 2003), olfactory (Bossert & Wilson, 1963; Eisenberg & Kleiman, 1972; Charpentier et al, 2010) and vocal (Sutton et al., 1973; Seyfarth et al., 1980a, b; Hodun et al., 1981; Geissmann, 1984; Seyfarth & Cheney, 1986, 1997; Mitani et al., 1992; Owren et al., 1993; Snowdon & Elowson, 1999; Roush & Snowdon, 1999, 2000) communication are all regularly used in many of the species studied. Given the difficulty in studying each individual signal modality, many of the studies have focused solely on the most dominant form of signaling used by the species of interest.

For primates that rely heavily on acoustic signaling, understanding whether calls are learned or innate has been the focus of many studies. More studies have focused on prosimians and monkeys rather than apes, in part because of how difficult and invasive vocal learning studies are to carry out. To examine vocal learning, one often times must manipulate the conditions with which an infant is raised. For any species of ape, the benefits of these studies would not outweigh the costs. For the studies carried out with monkeys, interesting results regarding vocal development have been obtained (Masataka & Fujita, 1989; Owren et al., 1992, 1993). Masataka & Fujita (1989) studied food calls in cross fostered rhesus and Japanese macaques. They presented their findings as evidence of vocal learning by the cross fostered infants. A drawback to this study is that some of the vocalizations produced by each species are not acoustically distinct in the adult population which obscures the researcher's ability to accurately determine whether the production is due to learning or innateness. An additional flaw in their experimental design was that the rhesus and Japanese macaques were housed in the same room, which means there is the possibility for the infants to still hear the calls from each species. A flawed statistical analysis also led to improper results. These potential problems were later confirmed by Owren et al. (1992, 1993) who conducted a follow up study which contradicted these earlier results. Despite these conflicting results, these studies provide evidence that some plasticity exists within the vocal system of macaques. This helped to support further examination of other primates for vocal plasticity (de la Torre & Snowdon, 2009), as well as studies examining the meaning of calls.

Applications of Bioacoustics

As mentioned in previous sections, characterizing the vocal repertoire of a species is becoming more realistic due to the advances in computational techniques for analyzing sounds. Many interesting applications of advanced vocal analyses exist. For instance, it is becoming more common to use a species' vocalizations as a method of population censusing (Harrington & Mech, 1982; Estrada et al., 2004; Collier et al., 2010; Adi et al., 2010). This approach is exceptionally useful in the case of rare species, which are less suitable for riskier trapping methods. It is also useful for species that are difficult to trap due to their evasive behavior or if they only live in a remote region where it is difficult to access traps on a daily basis. In other cases identifying vocal repertoires is helpful for identifying new species (Nietsch, 1999). While this is less common and not as definitive as genetic analyses, it does provide an important first step for protecting a species and raising awareness of a need to allocate more resources to understating the new organisms. Still other studies have focused on using vocalizations to identify known species (Masters, 1991; Gerhardt, 1994; Zimmermann et al., 2000; Anderson et al., 2000; Ambrose, 2003; Yovel et al., 2009; Goodale & Podos, 2010), individuals within a population (Marler & Hobbett, 1975; Waser, 1977; Tooze et al., 1990; Zimmermann & Lerch, 1993; Prestwich, 1994; Rendall, 1996; Zimmermann & Hafen, 2001; Reby, 2003; Palacios et al., 2007; Spillmann et al., 2010) or dialects in a population (Tyack, 1986; Whitehead et al., 1998; Risch et al., 2007; Podos, 2007; Garland et al., 2011; Kershenbaum et al., 2012). Overall, these studies illustrate a need to continually search for more accurate and cost effective methods of analysis. Additionally, these studies have provided valuable insight and support for studying the development of vocal communication.

Vocal development has historically been divided between sounds that are learned and sounds that are innate. This relationship became slightly more complex in Marler (1997) where species of songbirds were grouped into one of the following three categories: 1) animals that learn by instruction (Sensorimotor model), 2) animals that learn by selection, and 3) animals that memorize by instruction followed by selection in production. This study highlighted the problem of separating species into only 2, rigid categories. Though this work focused solely on songbirds, it has wide-ranging implications for other organisms including primates. Less invasive measures are often necessary for studying vocal production in primates; however several innovative studies have found ways to produce informative results. An option that has been employed for studying primate vocal production is to incorporate zoo animals into a study as a replacement to isolation rearing or brain lesioning (Sheperdson et al., 1989; Geissmann, 2000, 2002; Hosey, 2005; Ruppell, 2007). The contrastive social conditions between the zoo and the wild environments provide a platform for testing whether learning plays a role in vocal development. Differences observed in acoustical unit production or context of sounds used have been demonstrated in several species (Sheperdson et al., 1989; Geissmann, 2000, 2002; Ruppell, 2007). Although zoo studies are less definitive than invasive studies, they still help us to understand more about communication in primates.

Another application of bioacoustics research is to examine the order of sounds and the rules which govern this order in animal vocalizations. As complexity of social systems increases within species, one might suppose a greater need for linguistic structure evolves simultaneously. Thus, there is little surprise that pressure toward complex phenomena, as seen in vocal learning (Marler, 1970, 1997; Nottebohm, 1972; Scharff & Nottebohm, 1991; Adret, 1992; Deecke, 2000; Goodale, 2010; Hoeschele, 2010; Hahnloser, 2010; Lahti, 2011) and in sophisticated mating duets (Raemaekers et al., 1984; Levin, 1996a, b; Nietsch, 1999; Hall, 2000; Grafe, 2004; Murphy, 2006; Benedict, 2009; Bradley, 2009) are commonly observed. For example, increased complexity of vocalizations in mating rituals can display an individual's reproductive fitness, which is strongly believed with birdsong (Alvarez, 1993; Murphy, 2006; Goodale, 2010; Anderson et al., 2010; Hoeschele, 2010; Hetrick, 2012). In humans, rules which govern complexity in sound order are represented as syntax (Chomsky, 1957). In animals, it is still unclear what level of organization exists in many species; however the context in which calls are given has been examined by a variety of approaches.

Advancements in understanding the basic units of sound in primates has led to a large body of research focused on understanding what, if any meaning exists in calls (Marler et al., 1992). Previous assumptions were that humans possess referential signals and animals use motivational signals. This does not take into account the referential nature of the alarm calls of vervets (Struhsaker, 1967; Seyfarth et al., 1980 a,b) or the food calls of toque macaques and chimpanzees (Dittus, 1984; Goodall, 1986). These studies suggest that motivational signals may also have referential information. Describing calls as categorical or non-categorical signals has also been explored (Newman & Goedeking, 1992; Owren et al., 1992). The existence of non-categorical or graded signals has been proposed as evidence for greater variability in the internal motor control of signal production (Campos et al., 1989). Though the exact meaning of these types of calls has yet to be confirmed, acknowledgment of these differences has led to a greater focus on understanding the underlying mechanisms controlling the vocal production. Another major focus of vocalization studies examines differences in species based upon whether they are production learners, usage learners or comprehension learners. Seyfarth & Cheney (2010) outline and describe the differences between these categories and where various species fall. While production learning has been demonstrated in relatively few species, an overwhelming number of species have not been examined yet. This is also true for usage and comprehension learning. All of these components are important parts of human language learning, but the extent to which each of these categories plays a role in other species is part of an ongoing effort to understand vocal communication systems.

Summary

The understanding of behavior in other species has opened up the door for correlating acoustical signals with more specific meaning. The computational approach developed in this thesis will help future studies carried out on referential signaling by provided a more sophisticated way of describing sound properties and unifying the way in which we can compare results. This will help us to analyze longer segments of vocalizations which in turn will reduce the potential error associated with not looking at the "important" part of the calls. Since meanings are unknown, we can't be certain that previous studies did not overlook segments which were referential to the animals. The goal of this research is to discover and elucidate some of the observable similarities in how sound waves are utilized by different species and their implications for linguistic structure.

Several factors have influenced the current line of research in this dissertation. Contributions to the previously discussed fields will include 1) development of an automated computational approach which segments and clusters acoustic unit categories in white-handed gibbons and rats; 2) implement linguistically inspired methods to examine white-handed gibbon vocalizations for semantic meaning; and 3) elucidate vocal degradation and recovery of acoustic unit categories in a translational medical study with rats. Completion of this work contributes to previous studies by narrowing the gap in several areas of research.

Chapter References

Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience, 14*(8), 1067-U1173.

Adi, K., Johnson, M. T., & Osiejuk, T. S. (2010). Acoustic censusing using automatic vocalization classification and identity recognition. *Journal of the Acoustical Society of America, 127*(2), 874-883.

Adret, P. (1992). Vocal learning induced with operant techniques - An overview. *Netherlands Journal of Zoology, 43*(1-2), 125-142.

Alger, S. J., Maasch, S. N., & Riters, L. V. (2009). Lesions to the medial preoptic nucleus affect immediate early gene immunolabeling in brain regions involved in song control and social behavior in male European starlings. *European Journal of Neuroscience, 29*(5), 970-982.

Alger, S. J., & Riters, L. V. (2006). Lesions to the medial preoptic nucleus differentially affect singing and nest box-directed behaviors within and outside of the breeding season in European starlings (Sturnus vulgaris). *Behavioral Neuroscience, 120*(6), 1326-1336.

Alvarez, F. (1993). Alertness signaling in 2 rail species. *Animal Behaviour, 46*(6), 1229-1231.

Ambrose, L. (2003). Three acoustic forms of Allen's galagos (Primates; Galagonidae) in the Central African region. *Primates, 44*(1), 25-39.

Anderson, M. G., Brunton, D. H., & Hauber, M. E. (2010). Species specificity of grey warbler begging solicitation and alarm calls revealed by nestling responses to playbacks. *Animal Behaviour, 79*(2), 401-409.

Anderson, M. J., Ambrose, L., Bearder, S. K., Dixson, A. F., & Pullen, S. (2000). Intraspecific variation in the vocalizations and hand pad morphology of southern lesser bush babies (Galago moholi): A comparison with G. senegalensis. *International Journal of Primatology, 21*(3), 537-555.

Anderson, S. E., Dave, A. S., & Margoliash, D. (1996). Template-based automatic recognition of birdsong syllables from continuous recordings. *Journal of the Acoustical Society of America, 100*(2), 1209-1219.

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science, 332*(6027), 346-349.

Ball, G. F., Auger, C. J., Bernard, D. J., Charlier, T. D., Sartor, J. J., Riters, L. V., et al. (2004). Seasonal plasticity in the song control system - Multiple brain sites of steroid hormone action and the importance of variation in song behavior. In H. P. Zeigler & P. Marler (Eds.), *Behavioral Neurobiology of Birdsong* (Vol. 1016, pp. 586-610).

Bass, A. H., Gilland, E. H., & Baker, R. (2008). Evolutionary origins for social vocalization in a vertebrate hindbrain-spinal compartment. *Science, 321*(5887), 417-421.

Becker, M. L., Buder, E. H., & Ward, J. P. (2003). Spectrographic description of vocalizations in captive Otolemur garnettii. *International Journal of Primatology, 24*(2), 415-446.

Beckers, G. J. L., Bolhuis, J. J., Okanoya, K., & Berwick, R. C. (2012). Birdsong neurolinguistics: songbird context-free grammar claim is premature. *Neuroreport, 23*(3), 139-145.

Benedict, L., & McEntee, J. P. (2009). Context, structural variability and distinctiveness of California Towhee (Pipilo crissalis) vocal duets. *Ethology, 115*(1), 77-86.

Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Sciences, 15*(3), 113-121.

Bilmes, J. A. (1998). *A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models* (No. Technical Report ICSI-TR-97-02). University of Berkeley: International Computer Science Institute.

Boersma, P., & Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry, 32*(1), 45-86.

Bohn, K. M., Schmidt-French, B., Schwartz, C., Smotherman, M., & Pollak, G. D. (2009). Versatility and stereotypy of free-tailed bat songs. *Plos One, 4*(8).

Bossert, W. H., & Wilson, E. O. (1963). The analysis of olfactory communication among animals. *Journal of theoretical biology, 5*(3), 443-469.

Bourlard, H., Konig, Y., & Morgan, N. (1995). REMAP: recursive estimation and maximization of a posteriori probabilities in connectionist speech recognition. *Eurospeech*.

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication* (Second ed.). Sunderland, Massachusetts: Sinauer Associates, Inc.

Bradley, D. W., & Mennill, D. J. (2009). Strong ungraded responses to playback of solos, duets and choruses in a cooperatively breeding Neotropical songbird. *Animal Behaviour, 77*(5), 1321-1327.

Bridle, J. S., Brown, M. D., & Chamberlain, R. M. (1982). An algorithm for connected word recognition *Automatic Speech Analysis and Recognition* (Vol. 88, pp. 191-204): Springer Netherlands.

Broca, P. (1861). Remarques Sur Le Siége De La Faculté Du Langage Articulé, Suivies D'une Observation D'aphémie (Perte De La Parole). *Bulletin de la Société Anatomique, 6*, 330-357.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Groβhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth, Leipzig.

Campbell, G. S., Gisiner, R. C., Helweg, D. A., & Milette, L. L. (2002). Acoustic identification of female Steller sea lions (Eumetopias jubatus). *Journal of the Acoustical Society of America, 111*(6), 2920-2928.

Campos, J. J., Campos, R. G., & Barrett, K. C. (1989). Emergent themes in the study of emotional development and emotion regulation. *Developmental Psychology, 25*(3), 394-402.

Charlton, B. D., Frey, R., McKinnon, A. J., Fritsch, G., Fitch, W. T., & Reby, D. (2013). Koalas use a novel vocal organ to produce unusually low-pitched mating calls. *Current Biology, 23*(23).

Charpentier, M. J. E., Crawford, J. C., Boulet, M., & Drea, C. M. (2010). Message 'scent': lemurs detect the genetic relatedness and quality of conspecifics via olfactory cues. *Animal Behaviour, 80*(1), 101-108.

Chesmore, D. (2004). Automated bioacoustic identification of species. *Anais Da Academia Brasileira De Ciencias, 76*(2), 435-440.

Chesmore, E. D. (2001). Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals. *Applied Acoustics, 62*(12), 1359-1374.

Chesmore, E. D., & Ohya, E. (2004). Automated identification of field-recorded songs of four British grasshoppers using bioacoustic signal recognition. *Bulletin of Entomological Research, 94*(4), 319-330.

Chomsky, N. (1957). *Syntactic Structures*. Netherlands: Mouton & Co., The Hague.

Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *Plos One, 1*(1).

Clemins, P. J., Johnson, M. T., Leong, K. M., & Savage, A. (2005). Automatic classification and speaker identification of African elephant (Loxodonta africana) vocalizations. *Journal of the Acoustical Society of America, 117*(2), 956-963.

Cleveland, J., & Snowdon, C. T. (1982). The complex vocal repertoire of the adult cotton-top tamarin (*Saguinus oedipus oedipus*). *Zeitschrift Fur Tierpsychologie-Journal of Comparative Ethology, 58*(3), 231-270.

Coen, M. (2006). *Multimodal Dynamics: Self-supervised learning in perceptual and motor systems*. Unpublished Doctoral dissertation, Massachusetts Institute of Technology.

Coen, M. (2006). *Self-supervised acquisition of vowels in American English*. Paper presented at the Proceedings of the Twenty First National Conference on Artificial Intelligence(AAAI), Boston, MA.

Coen, M. (2007). *Learning to sing like a bird: Self-supervised acquisition of birdsong*. Paper presented at the Proceedings of the Twenty Second National Conference on Artificial Intelligence(AAAI), Vancouver, British Columbia.

Collier, T. C., Kirschel, A. N. G., & Taylor, C. E. (2010). Acoustic localization of antbirds in a Mexican rainforest using a wireless sensor network. *Journal of the Acoustical Society of America, 128*(1), 182-189.

Cutler, A., & Carter, D. (1987). Metrical structure of initial syllables in English. *Journal of the Acoustical Society of America, 81*, S67.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology-Human Perception and Performance, 14*(1), 113-121.

Dawson, M. R. W., Charrier, I., & Sturdya, C. B. (2006). Using an artificial neural network to classify black-capped chickadee (Poecile atricapillus) call note types. *Journal of the Acoustical Society of America, 119*(5), 3161-3172.

de la Torre, S., & Snowdon, C. T. (2009). Dialects in pygmy marmosets? Population variation in call structure. *American Journal of Primatology, 71*(4), 333-342.

de Waal, F. B. (1991). Complementary methods and convergent evidence in the study of primate social cognition. *Behaviour*, 297-320.

de Waal, F. B. M., & Johanowicz, D. L. (1993). Modification of reconciliation behavior through social experience - an experiment with 2 macaque species. *Child Development, 64*(3), 897-908.

Deecke, V. B., Ford, J. K. B., & Spong, P. (1999). Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (Orcinus orca) dialects. *Journal of the Acoustical Society of America, 105*(4), 2499-2507.

Deecke, V. B., Ford, J. K. B., & Spong, P. (2000). Dialect change in resident killer whales: implications for vocal learning and cultural transmission. *Animal Behaviour, 60*, 629-638.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society Series B-Methodological, 39*(1), 1-38.

Deregnaucourt, S., Guyomarc'h, J. C., & Richard, V. (2001). Classification of hybrid crows in quail using artificial neural networks. *Behavioural Processes, 56*(2), 103-112.

Dillon, B., Dunbar, E., & Idsardi, W. (2013). A single-stage approach to learning phonological categories: Insights from Inuktitut. *Cognitive Science, 37*(2), 344-377.

Dittus, W. P. J. (1984). Toque macaque food calls - Semantic communication concerning food distribution in the environment. *Animal Behaviour, 32*(2), 470-477.

Eisenberg, J. F., & Kleiman, D. G. (1972). Olfactory communication in mammals. *Annual Review of Ecology and Systematics, 3*, 1-32.

Eliades, S. J., & Wang, X. Q. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature, 453*(7198), 1102-U1108.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*(5), 429-448.

Fee, M. S., Shraiman, B., Pesaran, B., & Mitra, P. P. (1998). The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature, 395*(6697), 67-71.

Fitch, W. T., & Hauser, M. D. (1995). Vocal production in nonhuman-primates - Acoustics, physiology, and functional constraints on honest advertisement. *American Journal of Primatology, 37*(3), 191-219.

Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition, 97*(2), 179-210.

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society of London Series B-Biological Sciences, 268*(1477), 1669-1675.

Frey, R., Gebler, A., Fritsch, G., Nygren, K., & Weissengruber, G. E. (2007). Nordic rattle: the hoarse vocalization and the inflatable laryngeal air sac of reindeer (Rangifer tarandus). *Journal of Anatomy, 210*(2), 131-159.

Fristrup, K. M., & Watkins, W. A. (1992). *Characterizing acoustic features of marine animal sounds*. Massachusetts: Woods Hole Oceanographic Institution.

Fristrup, K. M., & Watkins, W. A. (1993). *Marine animal sound classification*. Massachusetts: Woods Hole Oceanographic Institution.

Gamba, M., & Giacoma, C. (2007). Quantitative acoustic analysis of the vocal repertoire of the crowned lemur. *Ethology Ecology & Evolution, 19*(4), 323-343.

Gannon, P. J., Kheck, N. M., & Hof, P. R. (2001). Language areas of the hominoid brain: a dynamic communicative shift on the upper east side planum *Evolutionary anatomy of the primate cerebral cortex*. Cambridge: Cambridge University Press.

Garland, E. C., Goldizen, A. W., Rekdahl, M. L., Constantine, R., Garrigue, C., Hauser, N. D., et al. (2011). Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale. *Current biology, 21*(8), 687-691.

Geissmann, T. (1984). Inheritance of song parameters in the gibbon song, analyzed in 2 hybrid gibbons (Hylobates pileatus X Hylobates lar). *Folia Primatologica, 42*(3-4), 216-235.

Geissmann, T. (2000). Gibbon songs and human music from an evolutionary perspective. In N. L. Wallin, B. Merker & S. Brown (Eds.), *The origins of music* (pp. 103-123): Massachusetts Institute of Technology.

Geissmann, T. (2002). Duet-splitting and the evolution of gibbon songs. *Biological Reviews, 77*(1), 57-76.

Gerhardt, H. C. (1994). The evolution of vocalization in frogs and toads. *Annual Review of Ecology and Systematics, 25*, 293-324.

Giles, J. C., Davis, J. A., McCauley, R. D., & Kuchling, G. (2009). Voice of the turtle: The underwater acoustic repertoire of the long-necked freshwater turtle, Chelodina oblonga. *Journal of the Acoustical Society of America, 126*(1), 434-443.

Goldsmith, J., & Xanthos, A. (2009). Learning phonological categories. *Language, 85*(1), 4-38.

Goodale, E., & Podos, J. (2010). Persistence of song types in Darwin's finches, Geospiza fortis, over four decades. *Biology Letters, 6*(5), 589-592.

Goodall, J. (1986). *The chimpanzees of Gombe: patterns of behavior*. Cambridge: Harvard University Press.

Goodsitt, J. V., Morgan, J. L., & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language, 20*(2), 229-252.

Grafe, T. U., Bitz, J. H., & Wink, M. (2004). Song repertoire and duetting behaviour of the tropical boubou, Laniarius aethiopicus. *Animal Behaviour, 68*, 181-191.

Gros-Louis, J. J., Perry, S. E., Fichtel, C., Wikberg, E., Gilkenson, H., Wofsy, S., et al. (2008). Vocal repertoire of Cebus capucinus: Acoustic structure, context, and usage. *International Journal of Primatology, 29*(3), 641-670.

Groves, P. M., & Thompson, R. F. (1970). Habituation: a dual-process theory. *Psychological review, 77*(5), 419.

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E. E., Lints, T., Jarvis, E. D., et al. (2004). FoxP2 expression in avian vocal learners and non-learners. *Journal of Neuroscience, 24*(13), 3164-3175.

Hahnloser, R. H. R., & Kotowicz, A. (2010). Auditory representations and memory in birdsong learning. *Current Opinion in Neurobiology, 20*(3), 332-339.

Hailman, J. P., Ficken, M. S., & Ficken, R. W. (1985). The Chick-a-dee calls of Parus atricapillus - a recombinant system of animal communication compared with written English. *Semiotica, 56*(3-4), 191-224.

Hall, M. L. (2000). The function of duetting in magpie-larks: conflict, cooperation, or commitment? *Animal Behaviour, 60*, 667-677.

Hammerschmidt, K., & Fischer, J. (1998). The vocal repertoire of Barbary macaques: A quantitative analysis of a graded signal system. *Ethology, 104*(3), 203-216.

Han, N. C., Muniandy, S. V., & Dayou, J. (2011). Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. *Applied Acoustics, 72*(9), 639-645.

Harrington, F. H., & Mech, L. D. (1982). An analysis of howling response parameters useful for wolf pack censusing. *Journal of Wildlife Management, 46*(3), 686-693.

Harris, Z. S. (1955). From phoneme to morpheme. *Language, 31*(2), 190-222.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*(5598), 1569-1579.

Hayes, B. (2004). Phonological acquisition in Optimality Theory: the early stages. In R. Kager, J. Pater & W. Zonneveld (Eds.), *Constraints in phonological acquisition* (pp. 158-203): Cambridge University Press.

Heimovics, S. A., & Riters, L. V. (2005). Immediate early gene activity in song control nuclei and brain areas regulating motivation relates positively to singing behavior during, but not outside of, a breeding context. *Journal of Neurobiology, 65*(3), 207-224.

Herbst, J. A., Gammeter, S., Ferrero, D., & Hahnloser, R. H. R. (2008). Spike sorting with hidden Markov models. *Journal of Neuroscience Methods, 174*(1), 126-134.

Hetrick, S. A., & Sieving, K. E. (2012). Antipredator calls of tufted titmice and interspecific transfer of encoded threat information. *Behavioral Ecology, 23*(1), 83-92.

Hockett, C. F. (1947). Morphology: the descriptive analysis of words. *Language, 23*(3), 273-285.

Hodun, A., Snowdon, C. T., & Soini, P. (1981). Subspecific variation in the long calls of the tamarin, Saguinus fuscicollis. *Zeitschrift Fur Tierpsychologie-Journal of Comparative Ethology, 57*(2), 97-110.

Hoeschele, M., Moscicki, M. K., Otter, K. A., van Oort, H., Fort, K. T., Farrell, T. M., et al. (2010). Dominance signalled in an acoustic ornament. *Animal Behaviour, 79*(3), 657-664.

Hopkins, W. D., Marino, L., Rilling, J. K., & MacGregor, L. A. (1998). Planum temporale asymmetries in great apes as revealed by magnetic resonance imaging (MRI). *Neuroreport, 9*(12), 2913-2918.

Hosey, G. R. (2005). How does the zoo environment affect the behaviour of captive primates? *Applied Animal Behaviour Science, 90*(2), 107-129.

Houser, D. S., Helweg, D. A., & Moore, P. W. (1999). Classification of dolphin echolocation clicks by energy and frequency distributions. *Journal of the Acoustical Society of America, 106*(3), 1579-1585.

Huang, X., Acero, A., & Hon, H.-W. (2001). *Spoken Language Processing: A guide to theory, algorithm and system development*: Prentice Hall PTR.

Hurford, J. R. (2004). Human uniqueness, learned symbols and recursive thought. *European Review, 12*(4), 551-565.

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language - (Reply to Fitch, Hauser, and Chomsky). *Cognition, 97*(2), 211-225.

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Behavioral Neurobiology of Birdsong, 1016*, 749-777.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*(4), 548-567.

Juang, B., & Rabiner, L. (1986). *Mixture autoregressive Hidden Markov Models for speaker independent isolated word recognition*. Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing.

Juang, B., Rabiner, L., & Wilpon, J. (1986). *On the use of bandpass liftering in speech recognition*. Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing.

Jurgens, U. (2009). The neural control of vocalization in mammals: A review. *Journal of Voice, 23*(1), 1-10.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics, 61*(8), 1465-1476.

Kelm-Nelson, C. A., & Riters, L. V. (2013). Curvilinear relationships between mu-opioid receptor labeling and undirected song in male European starlings (Sturnus vulgaris). *Brain Research, 1527*, 29-39.

Kershenbaum, A., Ilany, A., Blaustein, L., & Geffen, E. (2012). Syntactic structure and geographical dialects in the songs of male rock hyraxes. *Proceedings of the Royal Society B-Biological Sciences, 279*(1740), 2974-2981.

Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language, 37*(3), 295-311.

Knudsen, E. I. (1982). Auditory and visual maps of space in the optic tectum of the Owl. *Journal of Neuroscience, 2*(9), 1177-1194.

Kobayasi, K. I., & Riquimaroux, H. (2012). Classification of vocalizations in the Mongolian gerbil, Meriones unguiculatus. *Journal of the Acoustical Society of America, 131*(2), 1622-1631.

Kogan, J. A., & Margoliash, D. (1998). Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. *Journal of the Acoustical Society of America, 103*(4), 2185-2196.

Krakow, R. A. (1999). Physiological organization of syllables: a review. *Journal of Phonetics, 27*(1), 23-54.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *190*(4209), 69-72.

Ladefoged, P., & Halle, M. (1988). Some major features of the International Phonetic Alphabet. *Language, 64*(3), 577-582.

Lahti, D. C., Moseley, D. L., & Podos, J. (2011). A tradeoff between performance and accuracy in bird song learning. *Ethology, 117*(9), 802-811.

Levin, R. N. (1996a). Song behaviour and reproductive strategies in a duetting wren, Thryothorus nigricapillus .1. Removal experiments. *Animal Behaviour, 52*, 1093-1106.

Levin, R. N. (1996b). Song behaviour and reproductive strategies in a duetting wren, Thryothorus nigricapillus .2. Playback experiments. *Animal Behaviour, 52*, 1107-1117.

Lieberman, D. E., & McCarthy, R. C. (1999). The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions. *Journal of Human Evolution, 36*(5), 487-517.

Lieberman, P. (1968). Primate vocalizations and human linguistic ability. *Journal of the Acoustical Society of America, 44*(6), 1574-1584.

Lieberman, P., Klatt, D. H., & Wilson, W. H. (1969). Vocal tract limitations on vowel repertoires of Rhesus monkey and other nonhuman primates. *Science, 164*(3884), 1185-1187.

Losin, E. A. R., Russell, J. L., Freeman, H., Meguerditchian, A., & Hopkins, W. D. (2008). Left hemisphere specialization for oro-facial movements of learned vocal signals by captive chimpanzees. *Plos One, 3*(6).

Mandelbrot, B. (1953). An informational theory of the statistical structure of language. In W. Jackson (Ed.), *Communication theory* (pp. 486-502): Academic Press Inc.

Margoliash, D. (1983). Acoustic parameters underlying the responses of song-specific neurons in the White-crowned sparrow. *Journal of Neuroscience, 3*(5), 1039-1057.

Margolis, E., and Stephen Laurence, eds. (1999). *Concepts: core readings*: MIT Press.

Marler, P. (1970). A comparative approach to vocal learning: Song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology, 71*(2, Pt. 2), 1-25.

Marler, P. (1997). Three models of song learning: Evidence from behavior. *Journal of Neurobiology, 33*(5), 501-516.

Marler, P., Evans, C. S., & Hauser, M. D. (1992). Animal signals: motivational, referential, or both *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 66-86). Cambridge: Cambridge University Press.

Marler, P., & Hobbett, L. (1975). Individuality in a long-range vocalization of wild chimpanzees. *Zeitschrift für Tierpsychologie, 38*(1), 97-109.

Martin-Ordas, G., Berntsen, D., & Call, J. (2013). Memory for distant past events in chimpanzees and orangutans. *Current Biology, 23*(15), 1438-1441.

Masataka, N. (1982). A field study on the vocalizations of Goeldi's monkeys (*Callimico goeldii*). *Primates, 23*(2), 206-219.

Masataka, N., & Fujita, K. (1989). Vocal learning of Japanese and Rhesus monkeys. *Behaviour, 109*(No. 3/4), 191-199.

Masters, J. C. (1991). Loud calls of Galago crassicaudatus and G. garnettii and their relation to habitat structure. *Primates, 32*(2), 153-167.

McCowan, B., Doyle, L. R., Jenkins, J. M., & Hanser, S. F. (2005). The appropriate use of Zipf's law in animal communication studies. *Animal Behaviour, 69*(1), F1-F7.

McCowan, B., Hanser, S. F., & Doyle, L. R. (1999). Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Animal Behaviour, 57*, 409-419.

Mellinger, D. K., & Clark, C. W. (2000). Recognizing transient low-frequency whale sounds by spectrogram correlation. *Journal of the Acoustical Society of America, 107*(6), 3518-3529.

Mellinger, D. K., & Clark, C. W. (2006). MobySound: A reference archive for studying automatic recognition of marine mammal sounds. *Applied Acoustics, 67*(11-12), 1226-1242.

Mercado III, E., & Kuh, A. (1998). *Classification of humpback whale vocalizations using a self-organizing neural network*. Paper presented at the IEEE World Congress on Computational Intelligence.

Mitani, J. C., Hasegawa, T., GrosLouis, J., Marler, P., & Byrne, R. (1992). Dialects in wild chimpanzees. *American Journal of Primatology, 27*(4), 233-243.

Murphy, T. G. (2006). Predator-elicited visual signal: why the turquoise-browed motmot wag-displays its racketed tail. *Behavioral Ecology, 17*(4), 547-553.

Murray, S. O., Mercado, E., & Roitblat, H. L. (1998). The neural network classification of false killer whale (Pseudorca crassidens) vocalizations. *Journal of the Acoustical Society of America, 104*(6), 3626-3633.

Narins, P. M., Lewis, E. R., & McClelland, B. E. (2000). Hyperextended call note repertoire of the endemic Madagascar treefrog Boophis madagascariensis (Rhacophoridae). *Journal of Zoology, 250*, 283-298.

Newman, J. D., & Goedeking, P. (1992). Noncategorical vocal communication in primates: the example of common marmoset phee calls *Nonverbal vocal communication* (pp. 87-101). Cambridge: Cambridge University Press.

Ney, H., Mergel, D., Noll, A., & Paeseler, A. (1992). *Data driven search organization for continuous speech recognition*. Paper presented at the IEEE Transactions on Signal Processing.

Nickerson, C. M., Bloomfield, L. L., Dawson, M. R. W., & Sturdy, C. B. (2006). Artificial neural network discrimination of black-capped chickadee (Poecile atricapillus) call notes. *Journal of the Acoustical Society of America, 120*(2), 1111-1117.

Nietsch, A. (1999). Duet vocalizations among different populations of Sulawesi tarsiers. *International Journal of Primatology, 20*(4), 567-583.

Nishimura, T. (2003). Comparative morphology of the hyo-laryngeal complex in anthropoids: two steps in the evolution of the descent of the larynx. *Primates, 44*(1), 41-49.

Nishimura, T. (2006). Descent of the larynx in chimpanzees: Mosaic and multiple-step evolution of the foundations for human speech. In T. Matsuzawa, M. Tomonaga & M. Tanaka (Eds.), *Cognitive Development in Chimpanzees* (pp. 75-95). Japan: Springer.

Nishimura, T., Mikami, A., Suzuki, J., & Matsuzawa, T. (2003). Descent of the larynx in chimpanzee infants. *Proceedings of the National Academy of Sciences of the United States of America, 100*(12), 6930-6933.

Nishimura, T., Mikami, A., Suzuki, J., & Matsuzawa, T. (2006). Descent of the hyoid in chimpanzees: evolution of face flattening and speech. *Journal of Human Evolution, 51*(3), 244-254.

Nordeen, K. W., & Nordeen, E. J. (1992). Auditory-feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behavioral and Neural Biology, 57*(1), 58-66.

Nottebohm, F. (1972). Origins of vocal learning. *American Naturalist, 106*(947).

Nottebohm, F., & Liu, W. C. (2010). The origins of vocal learning: New sounds, new circuits, new cells. *Brain and Language, 115*(1), 3-17.

Owren, M. J., Dieter, J. A., Seyfarth, R. M., & Cheney, D. L. (1992). Food calls produced by adult female Rhesus (Macaca mulatta) and Japanese (M. fuscata) macaques, their normally-raised offspring, and offspring cross-fostered between species. *Behaviour, 120*(No. 3/4), 218-231.

Owren, M. J., Dieter, J. A., Seyfarth, R. M., & Cheney, D. L. (1993). Vocalizations of Rhesus (Macaca mulatta) and Japanese (M. fuscata) macaques cross-fostered between species show evidence of only limited modification. *Developmental Psychobiology, 26*(7), 389-406.

Owren, M. J., Seyfarth, R. M., & Hopp, S. L. (1992). Categorical vocal signaling in nonhuman primates *Nonverbal vocal communication* (pp. 102-122). Cambridge: Cambridge University Press.

Palacios, V., Font, E., & Marquez, R. (2007). Iberian wolf howls: Acoustic structure, individual variation, and a comparison with North American populations. *Journal of Mammalogy, 88*(3), 606-613.

Parsons, S. (2001). Identification of New Zealand bats (Chalinolobus tuberculatus and Mystacina tuberculata) in flight from analysis of echolocation calls by artificial neural networks. *Journal of Zoology, 253*, 447-456.

Parsons, S., & Jones, G. (2000). Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. *Journal of Experimental Biology, 203*(17), 2641-2656.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009a). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition, 113*(2), 244-247.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009b). Statistical learning in a natural language by 8-month-old infants. *Child Development, 80*(3), 674-685.

Peperkamp, S., Le Calvez, R., Nadal, J. P., & Dupoux, E. (2006). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition, 101*(3), B31-B41.

Petersen, S. E., & Fiez, J. A. (1993). The processing of single words studied with positron emission tomography. *Annual Review of Neuroscience, 16*, 509-530.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331*(6157), 585-589.

Pika, S., Liebal, K., & Tomasello, M. (2003). Gestural communication in young gorillas (Gorilla gorilla): Gestural repertoire, learning, and use. *American Journal of Primatology, 60*(3), 95-111.

Pike, K. (1947). *Phonemics: A technique for reducing languages to writing*. Ann Arbor: University of Michigan: University of Michigan Publications Linguistics 3.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition, 95*(2), 201-236.

Podos, J. (2007). Discrimination of geographical song variants by Darwin's finches. *Animal Behaviour, 73*, 833-844.

Potter, J. R., Mellinger, D. K., & Clark, C. W. (1994). Marine mammal call discrimination using artificial neural networks. *Journal of the Acoustical Society of America, 96*(3), 1255-1262.

Pozzi, L., Gamba, M., & Giacoma, C. (2010). The use of artificial neural networks to classify primate vocalizations: A pilot study on black lemurs. *American Journal of Primatology, 72*(4), 337-348.

Prestwich, K. N. (1994). The energetics of acoustic signaling in anurans and insects. *American Zoologist, 34*(6), 625-643.

Quam, R. M., de Ruiter, D. J., Masali, M., Arsuaga, J. L., Martinez, I., & Moggi-Cecchi, J. (2013). Early hominin auditory ossicles from South Africa. *Proceedings of the National Academy of Sciences of the United States of America, 110*(22), 8847-8851.

Rabiner, L. (1989). *A tutorial on Hidden Markov Models and selected applications in speech recognition*. Paper presented at the IEEE.

Rabiner, L., & Juang, B. (1986). *An introduction to Hidden Markov Models*. Paper presented at the *IEEE Acoustics Speech and Signal Processing Magazine*.

Raemaekers, J. J., & Raemaekers, P. M. (1984). The ooaa duet of the gibbon (*Hylobates lar*) - A group call which triggers other groups to respond in kind. *Folia Primatologica, 42*(3-4), 209-215.

Raemaekers, J. J., Raemaekers, P. M., & Haimoff, E. H. (1984). Loud calls of the gibbon (*Hylobates lar*) - Repertoire, organization and context. *Behaviour, 91*, 146-189.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*(5464), 349-351.

Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D. F., Colombo, J., et al. (2009). Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiology of Learning and Memory, 92*(2), 135-138.

Reby, D., Lek, S., Dimopoulos, I., Joachim, J., Lauga, J., & Aulagnier, S. (1997). Artificial neural networks as a classification method in the behavioural sciences. *Behavioural Processes, 40*(1), 35-43.

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour, 65*, 519-530.

Rendall, D., Rodman, P. S., & Emond, R. E. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Animal Behaviour, 51*, 1007-1015.

Riede, T., Bronson, E., Hatzikirou, H., & Zuberbühler, K. (2005). Vocal production mechanisms in a non-human primate: morphological data and a model. *Journal of Human Evolution, 48*(1), 85-96.

Risch, D., Clark, C. W., Corkeron, P. J., Elepfandt, A., Kovacs, K. M., Lydersen, C., et al. (2007). Vocalizations of male bearded seals, Erignathus barbatus: classification and geographical variation. *Animal Behaviour, 73*, 747-762.

Riters, L. V. (2012). The role of motivation and reward neural systems in vocal communication in songbirds. *Frontiers in Neuroendocrinology, 33*(2), 194-209.

Riters, L. V., & Ball, G. F. (1999). Lesions to the medial preoptic area affect singing in the male European starling (Sturnus vulgaris). *Hormones and Behavior, 36*(3), 276-286.

Riters, L. V., Eens, M., Pinxten, R., Duffy, D. L., Balthazart, J., & Ball, G. F. (2000). Seasonal changes in courtship song and the medial preoptic area in male European starlings (Sturnus vulgaris). *Hormones and Behavior, 38*(4), 250-261.

Riters, L. V., & Stevenson, S. A. (2012). Reward and vocal production: Song-associated place preference in songbirds. *Physiology & Behavior, 106*(2), 87-94.

Roush, R. S., & Snowdon, C. T. (1999). The effects of social status on food-associated calling behaviour in captive cotton-top tamarins. *Animal Behaviour, 58*, 1299-1305.

Roush, R. S., & Snowdon, C. T. (2000). Quality, quantity, distribution and audience effects on food calling in cotton-top tamarins. *Ethology, 106*(8), 673-690.

Ruppell, J. (2007). *Environmental enrichment for captive animals: A playback experiment involving captive gibbons at the Oregon Zoo*. The Gibbon's Voice newsletter of the Gibbon Conservation Center.

Saar, S., & Mitra, P. P. (2008). A technique for characterizing the development of rhythms in bird song. *Plos One, 3*(1).

Saffran, J. R. (2003). Absolute pitch in infancy and adulthood: the role of tonal structure. *Developmental Science, 6*(1), 35-43.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27-52.

Samuels, B. (2009). The third factor in phonology *Biolinguistics, 3*, 355-382.

Samuels, B., Hauser, M. D., & Boeckx, C. (2010). *Do animals have Universal Grammar? A case study in phonology*: The Oxford Handbook of Universal Grammar.

Samuels, B. D. (2009). *The structure of phonological theory*. Unpublished Doctoral dissertation, Harvard University, Cambridge, Massachusetts.

Scharff, C., & Nottebohm, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system - Implications for vocal learning. *Journal of Neuroscience, 11*(9), 2896-2913.

Schon, P. C., Puppe, B., & Manteuffel, G. (2001). Linear prediction coding analysis and self-organizing feature map as tools to classify stress calls of domestic pigs (Sus scrofa). *Journal of the Acoustical Society of America, 110*(3), 1425-1431.

Seyfarth, R. M., & Cheney, D. L. (1986). Vocal development in vervet monkeys. *Animal Behaviour, 34*, 1640-1658.

Seyfarth, R. M., & Cheney, D. L. (1997). Behavioral mechanisms underlying vocal communication in nonhuman primates. *Animal Learning & Behavior, 25*(3), 249-267.

Seyfarth, R. M., & Cheney, D. L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain and Language, 115*(1), 92-100.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980a). Monkey responses to 3 different alarm calls - Evidence of predator classification and semantic communication. *Science, 210*(4471), 801-803.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980b). Vervet monkey alarm calls - semantic communication in a free-ranging primate. *Animal Behaviour, 28*(4), 1070-1094.

Shapiro, A. D., Tyack, P. L., & Seneff, S. (2011). Comparing call-based versus subunit-based methods for categorizing Norwegian killer whale, Orcinus orca, vocalizations. *Animal Behaviour, 81*(2), 377-386.

Shepherdson, D., Bemment, N., Carman, M., & Reynolds, S. (1989). Auditory enrichment for Lar gibbons (*Hylobates lar*) at London Zoo. *International Zoo Yearbook, 28*, 256–260.

Shipman, D., & Zue, V. (1982). *Properties of large lexicons: Implications for advanced isolated word recognition systems*. Paper presented at the *IEEE International Conference on ICASSP*. IEEE.

Simpson, H. B., & Vicario, D. S. (1990). Brain pathways for learned and unlearned vocalizations differ in zebra finches. *Journal of Neuroscience, 10*(5), 1541-1556.

Slater, P. J. B. (2003). Fifty years of bird song research: a case study in animal behaviour. *Animal Behaviour, 65*, 633-639.

Snowdon, C. T., & Elowson, A. M. (1999). Pygmy marmosets modify call structure when paired. *Ethology, 105*(10), 893-908.

Spillmann, B., Dunkel, L. P., van Noordwijk, M. A., Amda, R. N. A., Lameira, A. R., Wich, S. A., et al. (2010). Acoustic properties of long calls given by flanged male orang-utans (Pongo pygmaeus wurmbii) reflect both individual identity and context. *Ethology, 116*(5), 385-395.

Struhsaker, T. T. (1967). Auditory communication among vervet monkeys (Cercopithecus aethiops) *Social Communication Among Primates* (pp. 281-324). Chicago, IL: University of Chicago Press.

Sutton, D., Larson, C., Taylor, E. M., & Lindeman, R. C. (1973). Vocalization in Rhesus-monkeys - Conditionability. *Brain Research, 52*, 225-231.

Suzuki, R., Buck, J. R., & Tyack, P. L. (2005). The use of Zipf's law in animal communication analysis. *Animal Behaviour, 69*(1), F9-F17.

Tanner, J. E., & Byrne, R. W. (1993). Concealing facial evidence of mood - Perspective-taking in a captive gorilla. *Primates, 34*(4), 451-457.

Tanner, J. E., & Byrne, R. W. (1996). Representation of action through iconic gesture in a captive lowland gorilla. *Current Anthropology, 37*(1), 162-173.

Tchernichovski, O., Lints, T. J., Deregnaucourt, S., Cimenser, A., & Mitra, P. P. (2004). Studying the song development process rationale and methods. *Behavioral Neurobiology of Birdsong, 1016*, 348-363.

Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour, 59*, 1167-1176.

Tesar, B., & Smolensky, P. (1998). Learnability in optimality theory. *Linguistic Inquiry, 29*(2), 229-268.

Thompson, R. F., & Spencer, W. A. (1966). Habituation: a model phenomenon for the study of neuronal substrates of behavior. *Psychological review, 73*(1), 16.

Thomson, D. J. (1982). *Spectrum estimation and harmonic analysis*. Paper presented at the *Proceedings of the IEEE*.

Thorpe, W. H. (1961). *Bird-song: the biology of vocal communication and expression in birds*. Cambridge, England: University Press.

Tian, H., & Shang, Z. (2006). Artificial neural network as a classification method of mice by their calls. *Ultrasonics, 44*, E275-E278.

Tomasello, M., George, B. L., Kruger, A. C., Farrar, M. J., & Evans, A. (1985). The development of gestural communication in young chimpanzees. *Journal of Human Evolution, 14*(2), 175-186.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences, 16*(3), 495-511.

Tooze, Z. J., Harrington, F. H., & Fentress, J. C. (1990). Individually distinct vocalizations in timber wolves, Canis-lupus. *Animal Behaviour, 40*, 723-730.

Trubetzkoy, N. S. (1939). *Grundzüge der Phonologie* (C. A. M. Baltaxe, Trans. Principles of Phonology ed.).

Twaddell, W. F. (1952). Phonemes and allophones in speech analysis. *The Journal of the Acoustical Society of America, 24*(607).

Tyack, P. (1986). Population biology, social-behavior and communication in whales and dolphins. *Trends in Ecology & Evolution, 1*(6), 144-150.

Van Parijs, S. M., Parra, G. J., & Corkeron, P. J. (2000). Sounds produced by Australian Irrawaddy dolphins, Orcaella brevirostris. *Journal of the Acoustical Society of America, 108*(4), 1938-1940.

Vintsyuk, T. K. (1971). Element-wise recognition of continuous speech composed of words from a specified dictionary. *Cybernetics and Systems Analysis, 7.2*, 361-372.

Vu, E. T., Mazurek, M. E., & Kuo, Y. C. (1994). Identification of a forebrain motor programming network for the learned song of zebra finches. *Journal of Neuroscience, 14*(11), 6924-6934.

Waser, P. M. (1977). Individual recognition, intragroup cohesion and intergroup spacing - Evidence from sound playback to forest monkeys. *Behaviour, 60*, 28-74.

Weary, D. M. (1989). Categorical perception of bird song - How do great tits (*Parus major*) perceive temporal variation in their song. *Journal of Comparative Psychology, 103*(4), 320-325.

Wernicke, C. (1874). *Der aphasische Symptomencomplex. Eine psychologische Studie auf anatomischer Basis*.

Whitehead, H., Dillon, M., Dufault, S., Weilgart, L., & Wright, J. (1998). Non-geographically based population structure of south Pacific sperm whales: dialects, fluke-markings and genetics. *Journal of Animal Ecology, 67*(2), 253-262.

Yip, M. J. (2006). The search for phonology in other species. *Trends in Cognitive Sciences, 10*(10), 442-446.

Yovel, Y., Melcon, M. L., Franz, M. O., Denzinger, A., & Schnitzler, H. U. (2009). The voice of bats: How greater mouse-eared bats recognize individuals based on their echolocation calls. *Plos Computational Biology, 5*(6).

Yu, A. C., & Margoliash, D. (1996). Temporal hierarchical control of singing in birds. *Science, 273*(5283), 1871-1875.

Zilles, K., & Amunts, K. (2010). TIMELINE Centenary of Brodmann's map - conception and fate. *Nature Reviews Neuroscience, 11*(2), 139-145.

Zimmermann, E., & Hafen, T. G. (2001). Colony specificity in a social call of mouse lemurs (Microcebus ssp.). *American Journal of Primatology, 54*(3), 129-141.

Zimmermann, E., & Lerch, C. (1993). The complex acoustic design of an advertisement call in male-mouse lemurs (Microcebus-murinus, Prosimii, Primates) and sources of its variation. *Ethology, 93*(3), 211-224.

Zimmermann, E., Vorobieva, E., Wrogemann, D., & Hafen, T. (2000). Use of vocal fingerprinting for specific discrimination of gray (Microcebus murinus) and rufous mouse lemurs (Microcebus rufus). *International Journal of Primatology, 21*(5), 837-852.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge, Mass: Addison-Wesley Press.

Zuberbühler, K., Cheney, D. L., & Seyfarth, R. M. (1999). Conceptual semantics in a nonhuman primate. *Journal of Comparative Psychology, 113*(1), 33-42.

"When we treat of sexual selection we shall see that primeval man, or rather some early progenitor of man, probably first used his voice in producing true musical cadences, that is in singing, as do some of the gibbon-apes at the present day; and we may conclude from a widely-spread analogy, that this power would have been especially exerted during the courtship of the sexes,--would have expressed various emotions, such as love, jealousy, triumph,--and would have served as a challenge to rivals. It is, therefore, probable that the imitation of musical cries by articulate sounds may have given rise to words expressive of various complex emotions." (Darwin, 1871)

Chapter 4: Phonologically inspired segmentation of white-handed gibbon vocalizations

Abstract

Classification of animal vocalizations has increasingly employed the use of automated methods, despite the major influence manual analyses has on the field of bioacoustics. Though many automated platforms currently exist, most rely on measured properties tied to frequency descriptions. The drawback to this approach is that it overlooks any interior structure present in a given vocal bout. In this chapter, we describe a novel approach to segmenting and clustering acoustic units using a Cepstral Self-Similarity Matrix to define borders and classes of sounds. By doing so, a more empirical approach to segmenting calls is developed. This is used to illustrate that white-handed gibbons possess far more structure within their vocal repertoire than previously established. It is also used to define a novel vocalization gathered from two captive gibbons.

Introduction

The *Hylobatidae* family is uncommonly vocal among apes. This observation easily dates to Darwin (Darwin, 1871), who proposed human language shared a common evolutionary pathway with the singing ability of white-handed gibbons (*Hylobates lar*). Gibbons have subsequently been the subject of numerous studies which have focused on understanding the structure of their vocalizations, as well as the function of their vocalizations in mediating their behavioral dynamics (Marshall & Marshall, 1976; Raemaekers & Raemaekers, 1984, 1985;

Raemaekers et al., 1984; Mitani & Marler, 1989; Cowlishaw, 1992; Geissmann & Nijman, 2006; Bartlett, 2003). Much of this work has employed ethological observations and field studies correlating gibbon vocal communication with ongoing behaviors. These have revealed vocal patterns specific to mating, bonding, predator detection, and territorial defense. We note that analyses of the acoustic structure of gibbon song have been based upon manual inspection of spectral waveforms and spectrograms—particularly formant envelope structures in the latter—dating to the seminal works of Marshall and Marshall (1976) and Raemaekers et al. (1984).

Though commonly employed, a side effect of visual and manual classification of many acoustic structures, such as frequency envelopes, is they may be incapable of distinguishing significant differences in recorded data. This can be due to descriptional inadequacies in acoustic representations (e.g., spectrograms), subjectivity on the part of human researchers, lack of incorporation of ethological specificity (as discussed below), and perceptual differences in acoustic processing between target species and human observers.

Let us consider an example to make this clear. A widely used label for describing a particular gibbon vocalization is referred to as an "OO" (Raemaekers et al., 1984), one of six primary "types" of produced sounds described therein (see Figure 4.1). However, examining this classification in more detail, we have found (as detailed in the results section) that there are ten distinct types of "OO" sounds. While they are superficially similar from many views, they are extremely well distinguished by: their behavioral specificity; their positions in sequences; acoustic properties that do not reveal themselves using common analytic methods, and finally, theoretical mathematical properties to be discussed below.

Figure 4.1 shows three sub-categories of the "OO" sound illustrated in panel (A). These are labeled OO_1, OO_2, and OO_3 in panel (B). While all three of these acoustic units have clear characteristics of the archetype "OO" waveform, they are vocally employed in highly distinct ways. The methods introduced in this chapter demonstrate that they are far more acoustically distinct than many common representations reveal, as shown in panel (C). Even though the spectrograms of these three sounds are quite similar, from the perspective of the Cepstral Self-Similarity Matrix (CSSM) approach, their differences are readily apparent both to computational classification algorithms and to human observers.

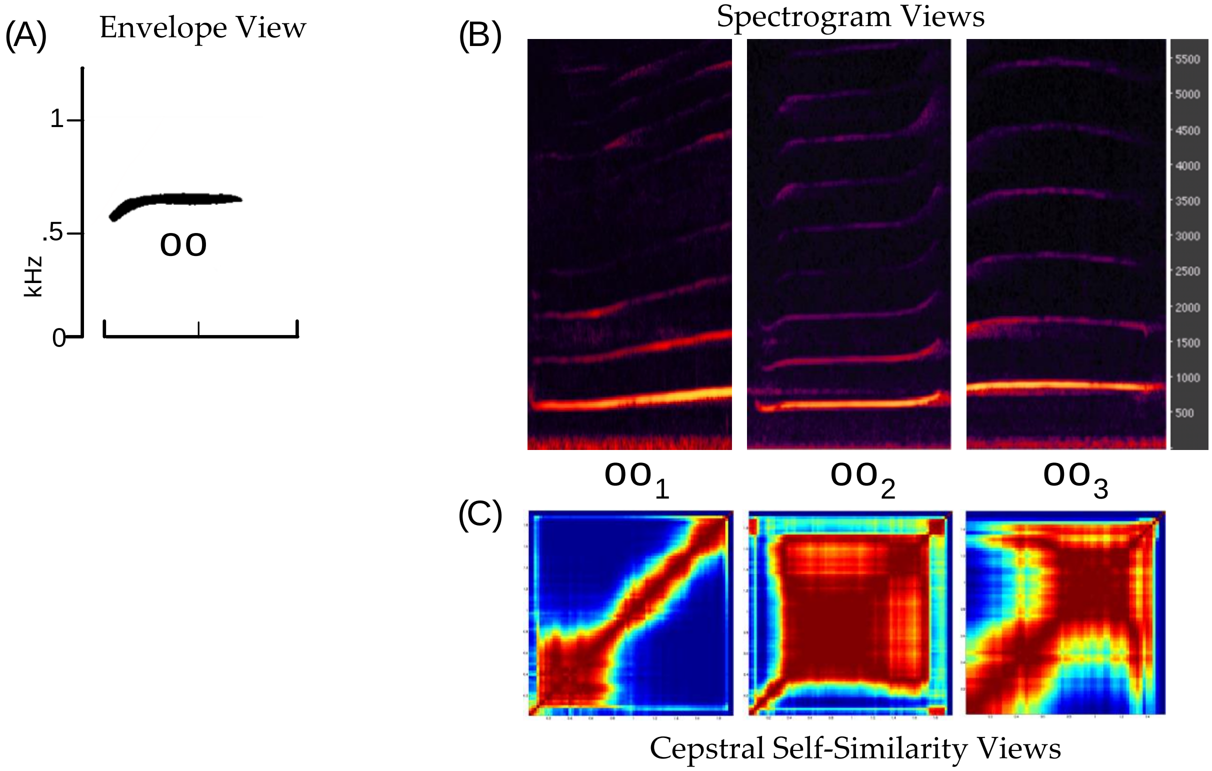


Figure 4.1 – Panel (A) shows the frequency envelope of the "OO" sound according to the classifications proposed in Raemaekers et al. (1984). Panel (B) shows three acoustic units, all classified as "OO"s by this view. However, we have found that each of these units, here labeled OO_1, OO_2, and OO_3, are employed in extremely distinct behavioral and contextual ways in gibbon communication. We have developed a more sophisticated acoustic representation, called a Cepstral Self-Similarity Matrix (CSSMs), which accurately classify these primitive sounds. They reveal far more detailed acoustic structure, as shown in Panel (C), which enables automated computational analysis of gibbon vocalizations.

Thus, this work expands upon earlier studies classifying gibbon vocalizations (Raemaekers et al., 1984, Clarke et al., 2006) by examining them in far greater depth to identify structure that has previously been unrevealed. That these structures have strong analogues to ones in spoken human language cannot yet be established. However, this work suggests that some human linguistic phenomena may not be uniquely restricted to humans. The approach here may generalize with many types of animal vocalizations and could provide an avenue for investigating other species more formally and objectively. Our aim is to make these analyses less subjective and more rigorous to better understand vocal generative processes as well as the behavioral and cognitive aspects underlying them.

In this chapter, we present the CSSM framework for elucidating behavioral and communicative distinctions in gibbon vocalizations that have hitherto been unknown. We demonstrate that the original six categories of gibbon vocalizations are comprised of 27 well-defined acoustic units. We may view these units as refinements of the earliest proposed categories in gibbon song, but we demonstrate they are not interchangeable. Each of the 27 categories appears in particular behavioral and sequential contexts, has unique mathematical properties, and can be computationally identified with extremely high accuracy. Our goals in this chapter are to: (1) elucidate the *interior* structures of calls in white-handed gibbons; and (2) demonstrate that previous call characterizations in *H. lar* grouped visually similar sounds that have extremely different communicative and behavioral roles.

Furthermore, using this framework, we present evidence of a previously undescribed vocalization. We have discovered captive gibbons that vocalize behaviorally specific calls that have not been previously observed (personal communication with T. Geissmann, and gibbon caretakers at 6 zoos and sanctuaries), as far as we have been able to determine. The timbre of these sounds is extraordinarily distinct from other gibbon vocalizations, as are their formal acoustic properties. Though there is insufficient evidence to claim vocal learning, there is sufficient evidence to merit further examination of this highly atypical call.

Prior Work

Many studies have characterized gibbon sounds perceptually and phenomenologically. An early, common approach—that has displayed remarkable longevity—is based upon visually characterizing the fundamental frequency envelopes of spectral sound representations, often

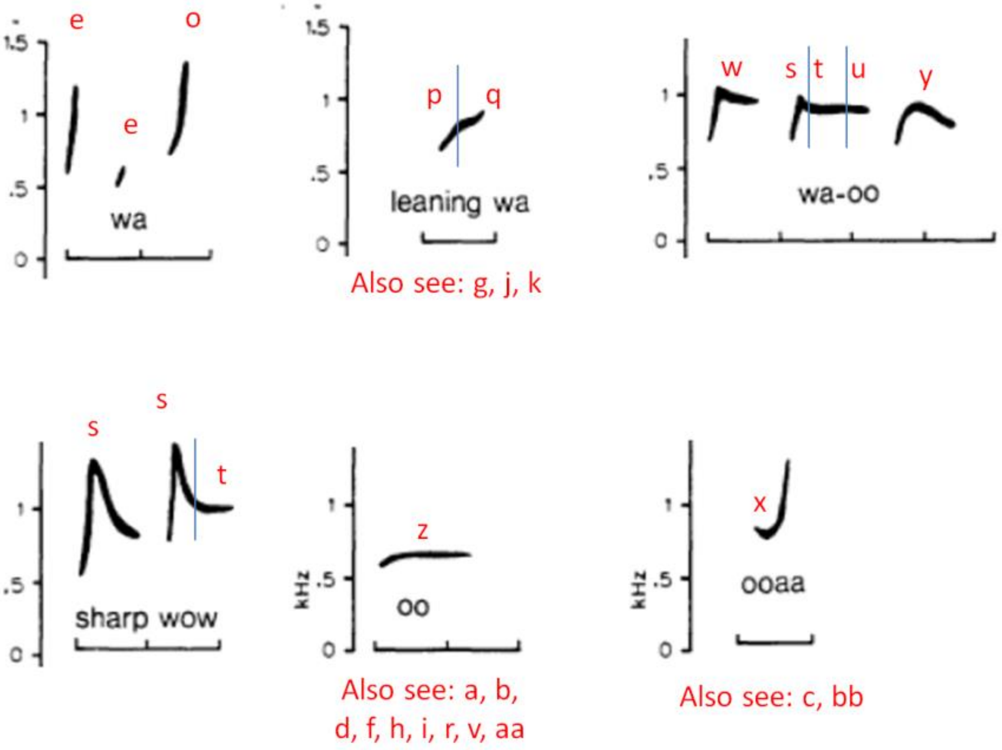


Figure 4.2: This figure shows a comparison between the 1984 classification of white-handed gibbon vocalizations conducted by Jeremy & Patricia Raemaekers (black labels) and the current classification of acoustic units proposed in this chapter (red labels and blue boundaries). Note how some of the classifications have little to no change (e.g. wa notes are the same as e and o) whereas some classification have been greatly expanded (e.g. wa-oo may be w, y, or a combination of s, t, u). There are also some additional acoustic units described in this current classification that do not fit in any of the above categories and include acoustic unit's l, m, and n.

coupled with onomatopoeic descriptions of sound groupings. For example, sustained vocalizations with minimal frequency modulation might be referred to as "OO" sounds (See Figure 4.1). We see this in the work of (Raemaekers & Raemaekers 1984, 1985), which partitioned gibbon vocalizations into six general categories as seen in Figure 4.2. This conceptual framework has been utilized in subsequent bioacoustic and communicative analyses of gibbons (Clarke et al. 2006, Geissmann and Nijman 2006, Wanelik et al. 2013) and other species (Bohn, 2009).

Visual inspection of spectrograms has a long history in bioacoustics (Masataka 1982; Cleveland & Snowdon 1982; Hammerschmidt & Fischer 1998; Van Parijs et al. 2000; Narins et al. 2000; Becker et al. 2003; Gamba & Giacoma 2007; Gros-Louis 2008). This approach is clearly useful for elucidating the basic acoustic repertoire of species vocal productions and its articulatory capabilities; however it is highly lacking for analyzing intraspecies communication because it "lumps" together sounds that are differentially and sometimes exclusively used in specific behavioral and communicative contexts. While more sophisticated signal processing tools are increasingly employed to elucidate these differences (Thomson, 1982; Fee et al., 1998; Kogan & Margoliash, 1998; Tchernichovski et al., 2000, 2004; Coen, 2006a; Coen, 2006b; Coen, 2007; Saar & Mitra, 2008), one cannot ignore the continued influence of early approaches. Specifically, what is lost in the abstract visual representation of a complex sound, e.g., through its spectral envelope? This concern is further exacerbated by the tendency to use these "lumped" categories when analyzing sequences of vocalizations. For example, Clarke et al. (2006) uses these six basic categories to analyze sequences of behaviorally-correlated gibbon vocalizations for syntactic structure. However, given there are many distinct sounds that fall into the "OO" category—most of which are not used interchangeably—one suspects identifying the basic constituent sounds is a prerequisite for sequence analysis. Although approximate characterizations of gibbon vocalizations are descriptively helpful, they do not necessarily invite more careful investigation of differences between sounds within the same general category. Furthermore, while it is widely accepted that gibbons make "OO"-type sounds, we suggest it is equally important to examine the constituent—yet highly distinct vocalizations—that comprise this group. As such, one may view our approach as a natural extension of the early characterization of gibbon vocalizations, using modern statistical, signal processing, and machine learning methodology to "dig deeper" into the early results of the field.

This view is perhaps most straightforward to appreciate via appeal to human language. Figure 4.3 shows spectrograms for five distinct English vowels. While their harmonic structures clearly differ, the envelopes of their fundamental frequencies (shown in purple in the figure) are sufficiently similar that envelope-based analyses would be tempted to "lump" them together were one sufficiently unfamiliar with human language. The folly of doing so is readily apparent here only because we are so familiar with the importance of distinguishing these sounds in human communication. However in examining non-perspicacious communicative acts in other species, it is worthwhile to preserve distinctions that are mathematically and ethologically significant.

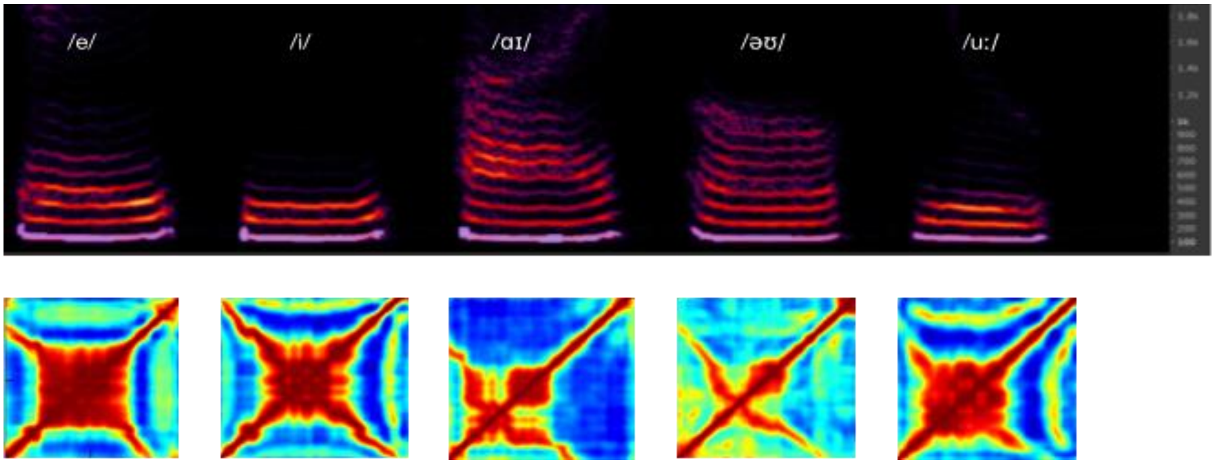


Figure 4.3—An illustration of a basic difficulty with envelope-based classification. This figure displays a spectrogram of English vowels, identified by their IPA labels. (Each of these represents in order the vowels in the words: “hay,” “eat,” “pie,” “mow,” and “you.”) For each vowel, the envelope of the fundamental frequency is highlighted in purple. Even though these envelopes are in fact not identical, gross acoustic classification based on visual inspection hides distinctions among these sounds that are structurally and semantically essential. Evidence suggests abstracting gibbon communication this way obfuscates rather than illuminates the content and meaning of their vocalizations. Were the displayed sounds all produced by gibbons, they would be grouped together as “OO”s without further distinction.

In recent years, more sophisticated approaches have been introduced into bioacoustics. As examined above, spectral representations (spectrograms) consisting of discrete Fourier transforms of short overlapped, segments of auditory signals are ubiquitous and characterize well acoustic features related to spectral profile and frequency modulation, many of which are relevant in animal acoustic communication. More formal analyses of these may exploit features such as minimum and maximum fundamental frequency, slope of the fundamental frequency, number of inflection points, and the presence of harmonics (Oswald et al., 2007) that vary between individuals (Buck & Tyack, 1993; Blumstein & Munos, 2005; Koren & Geffen, 2011; Ji et al., 2013; Root-Gutteridge et al., 2014), and the effects of different environmental contexts (Matthews et al., 1999; Taylor, Reby & McComb, 2008; Henderson, Hildebrand & Smith, 2011).

Extensions to spectral methods include incorporation of multi-taper methods (Thomson, 1982; Tchernichovski et al., 2000; Baker & Logue, 2003), Wigner-Ville spectra (Martin & Flandrin, 1985), and wavelet analysis (Mallat, 1999). These methods often provide improved frequency resolution at low frequencies and finer temporal resolution at higher frequencies. Cepstral analysis (Oppenheim & Schafer, 2004), which conceptually may be viewed itself as a (recursive) spectrum of a power spectrum, is widely used in human speech recognition (Remez et al., 1994). It provides a low-dimensional representation of a spectrum that is computationally efficient to process. In addition to analytic features derived from frequency-related features, other auditory characteristics such as prosody (Saar and Mitra, 2008) that are temporally sensitive to power can be incorporated into what are otherwise primarily frequency-based analyses.

Often one finds these approaches applied to tracing articulatory development during the period corresponding to vocal acquisition. This is particularly common in the birdsong community. Alternatively, these methods may be used to provide more precise characterizations of vocal acoustic properties, e.g., the degree of frequency modulation in a specific call. Other popular areas of investigation include species and even individual identification based on acoustic signatures.

In contrast, the approach taken here is to use novel statistical machine learning tools to analyze the structure of gibbon vocalizations themselves. Our goal is to “deconstruct” utterances into a small number of reused building blocks, out of which we propose all gibbon vocalizations are constructed. In doing so, we believe there is sufficient evidence to support the existence of acoustic units in gibbon vocalizations.

Methods

In an effort to remain consistent with previous bioacoustics research (Marshall & Marshall 1976; Cleveland & Snowdon 1982; Raemaekers et al. 1984; Raemaekers & Raemaekers

1984; Raemaekers & Raemaekers 1985; Mitani & Marler 1989; Hammerschmidt & Fischer 1998; Hsu et al. 2005; Gamba & Giacoma 2007; Gros-Louis et al. 2008) and to contribute new tools to analysis methods, the species being examined have had their vocalizations hand-labeled as well as computationally labeled. The hand-labeled analysis will provide the assumed ground truth for the more precise computational analysis.

Data Collection

Wild gibbon vocalizations were gathered by Esther Clarke from April 2004 to August 2005 at the Khao Yai National Park in Thailand and include 47 individuals across 12 groups. Recording methods can be found in Clarke et al. (2006). *Ad libitum* recordings were gathered as well as predator-induced (e.g. clouded leopard, tiger, and reticulated python) vocalizations. A total of 24.5 hours of vocalizations were obtained.

Zoo gibbon vocalizations were gathered by Angela Dassow and Michael Coen from August 2012 to September 2013 at the Racine Zoological Society and Henry Vilas Zoo. Each zoo houses a pair of adult white-handed gibbons. The gibbons at the Racine Zoo are a father-daughter pair. The father was born in the wild in Thailand. His daughter was born in captivity and housed with him for most of her life. The gibbons at the Henry Vilas Zoo are captive born and they are unrelated to each other or the Racine pair. The animals were habituated to the presence of the observers and recording equipment for a minimum of one hour prior to recording each day. A total of 28 hours of vocalizations were obtained.

For data collected at the zoos, recording equipment was set up outside of the gibbon enclosures to record their natural vocalizations. The recording equipment included two Vidpro XM-55 Condenser Shotgun Microphones connected to a four-track Zoom H4n digital recorder; a Blue Yeti Pro USB Condenser Microphone, all recording at 44.1 kHz, and a shock mount to reduce ambient noise interference. Encounters were filmed in high-definition with a Canon VIXIA HF M500 HD image stabilized camcorder equipped with a DM-100 shotgun microphone.

The microphones were distributed equidistantly around the circumferences of the gibbon enclosures. Recordings were taken 2-3 times per week for approximately 7 hours per day. During this time the gibbons were free to move around their enclosure and were not exposed to any toxic agents or restrained for any invasive medical procedures. These recordings were used to establish the vocal repertoires of the captive gibbons. The work at the zoo was approved by the University of Wisconsin-Madison's Research Animal Resource Center (RARC) and Institutional Animal Care and Use Committee (IACUC protocol number is L00452-0-08-12), Gibbon Species Survival Plan (SSP) Coordinator and Directors of the Henry Vilas Zoo and Racine Zoological Society which are both Association of Zoos and Aquariums (AZA) accredited zoos.

Analysis

This research made extensive use of computational tools and state-of-the-art machine learning methods. These included several off-the-shelf software platforms, including Matlab, R, Adobe Audition, Logic Pro, PRAAT, and Audacity. Data analysis made use of substantial mathematical and software tools written by Michael Coen. The University of Wisconsin-Madison's HTCondor high-throughput, massively distributed computing infrastructure was also utilized.

Approach

Our aim is to analyze gibbon vocalizations in a more objective and rigorous fashion. Specifically, previous analyses have provided useful frameworks for characterizing the gross structures of utterances largely based on visual inspection of non-acoustic features. While not an "exact" or even easily repeatable process, these high-level characterizations provide a useful landscape for classifying utterances in ways humans find easy to employ. However, as we show, there are far more subtle distinctions within these classes that demonstrate they provide a very abstract view of observed communications. By way of analogy, it is not entirely dissimilar from observing that basic sounds in English can be described as composed of consonants and vowels. While an important first step in characterizing the language, it reveals little about the structures of the many sounds comprising vocalizations. Even were we to "break" these sounds down further into (what are common accepted) groups such as *dentals, palatals, fricatives, etc.*, each of these subgroups in turn is composed of its own repertoire of sounds with distinct communicative importance.

We hypothesize that gibbon's use a finite number of distinct, repeated sounds out of which all utterances are generated. We further postulate these sound "units" are highly distinct. By this, we mean they: are not exchangeable; appear in behaviorally distinct contexts; co-occur in structured and predicable ways with other distinct "units" in vocal sequences; and finally, have unique mathematical properties upon which we elaborate.

Our approach takes two fronts. The first has a human expert familiarizing herself with gibbon vocalizations. This highly involved and labor-intensive process is described immediately below. The second utilizes state-of-the-art machine learning and digital signal processing tools to identify these repeated sounds, which we have hypothesized, are the building blocks of gibbon vocalization. Evaluating these approaches is essential for validating our theses. The most basic, but perhaps most complex test is whether the human and computational analyses—done independently—agree with one another. That is, do both approaches find the "same" building blocks for gibbon vocalizations? If so, one must ask if this is an artifact of the analysis or an emergent phenomenon worthy of further investigation. Assuming significant agreement, we must turn to ethological and contextual circumstances in which utterances were made, asking if particular acoustic units preferentially or exclusive appear only during particular behavioral displays. If this is the case, a sequence analysis is called for to elucidate structure and sometimes even attribute semantic value to particular orderings of these sounds. Finally, we address basic and theoretical statistical properties of these units that would be expected in a communicative system and establish independently that these categories are not "random" from the perspective of communication. We detail this validation in the results section.

Call classification – human observers

Classifying vocalizations of non-human species presents several challenges, including whether perceptual similarities exist between a given species and the researchers studying them. In many cases, human agreement on classification is used to determine accuracy rates for category identification. This process works well when cataloging general groups of sounds which are intuitive to segment; however categorizing acoustic units or subcategories of sounds requires a greater time commitment which limits the feasibility of multiple human classifiers. As detailed below, much of this time commitment is involved in determining whether subcategories of general groups are justifiable or not. The classification of acoustic units in gibbons has an added difficulty due to noticeable inter-individual variation.

Previous bioacoustics experience

Mastering a technical skill requires a great deal of effort. A couple of key skills are required for analyzing acoustic units in gibbon. First is the ability to consistently isolate differences between categories. Second is the ability to determine whether variation between two sounds is due to individual differences in vocal production or categorical differences. I acquired some of these abilities over a 10 year period of time working on a wolf monitoring project in Wisconsin. As part of this project, I conducted wolf howling surveys which consisted of transecting known wolf habitat, producing a series of howls for 10-15 seconds and then waiting for a reply. On successful occurrences, the number of wolves howling, the approximate age (alpha adult, non-alpha adult, immature, or pup) of each wolf, and the sex of the adult wolves was recorded. This skill was acquired over my first year of surveys and was taught to me by the state's first wolf biologist who is to this day a leading expert in wolf biology. Verification of our howling assessments was made by subsequently radio collaring and tracking each pack. While gibbons and wolves have very different vocal repertoires, the skills necessary to identify individuals as well as their approximate ages are very similar.

In addition to my experience conducting wolf howl surveys, I have had several years of experience surveying hundreds of species of song birds, raptors, and amphibians using vocal IDs. Distinguishing the difference between the wide varieties of calls which exist amongst these diverse organisms requires finely tuned perceptual abilities. These same skills are highly valuable for identifying acoustic units in gibbons.

Category designation

Prior to any attempts in labeling categories, several weeks were spent listening to each of the 52 audio files. During this time, general impressions about call types, group dynamics, and potential associations between sound categories were made. Given the diversity in the types of sounds experienced during this time, the files were subdivided into three general levels. Level 1 included all of the great calls. Great calls are well known to be part of the female gibbon's song. They are also known to be highly stereotypical in their acoustic complexity with minor individual-specific variation. Level 2 included all of the codas. The codas are well known to be part of the male's repertoire and are less variable in length and complexity than the female great calls. Level 3 included all of the *other* calls (i.e. any call that was not a great call or coda). Call classification began with the great calls and codas, in part because the stereotyped nature of these calls made boundary placement between each call category easier to identify and compare between individuals. For example, female great calls have two distinct frequency peaks at approximately 1/3 and 2/3 of the way through the bout. These peaks are easily identifiable between individuals and provided a standard reference point for comparison between individuals. This greatly assisted with determining whether variation was due to individual vocal production or the presence of two different sounds.

Classifying great calls and codas first also provided valuable impressions about individual vocal production quality. For instance, the quality of calls from the female in group C was noticeably better than the female from group D. Quality in this case is represented by amplitude of vocal production (Note: direction of recording equipment to animal was taken into consideration here), consistency of vocal production, and duration of vocal bouts. Female D often failed to produce a full great call. This lack of consistent vocalizing persisted throughout the rest of her calls as well. As such, she was not used as a *model* for comparisons. Rather her vocalizations were classified after categories were established. Female C on the other hand produced consistent, strong calls which provided a good basis for comparisons between others.

Though certain individuals were singled out as being ideal candidates to begin an in depth classification, all individuals followed the same general analysis. All recordings were initially processed in Audition and spectral images were created to provide a visual image of each vocalization. At the beginning of each day, the files being examined were listened to in their entirety several times before defining any boundaries between acoustical units. This was done to ensure that all sounds within the file could be identified and compared to each other. Similar to previous work (Raemaekers et al. 1984; Clarke et al. 2006; Dassow, 2010), boundaries were placed to exclude silence between utterances. In addition to this, boundaries were placed within an utterance if the sound was perceived to be a combination of 2 or more distinct acoustical units. This perception was based on whether or not the acoustical units could be found either in isolation or in combination with a different acoustical unit. After all of the perceived variation in acoustic units was identified, each category was compared to previous groupings. Acoustic units were assembled as sub types of previously established categories whenever possible. For example, the previously established ‘oo’ sound category described in Clarke (2006) and Raemaekers & Raemaekers (1984) contains 10 subtypes of acoustical categories. New categories were designated for three acoustic unit types which could not be classified into previously established groups. Following this procedure, 27 acoustic units were identified (Figure 4.4) and were determined to be distinct based upon visual and auditory perception.

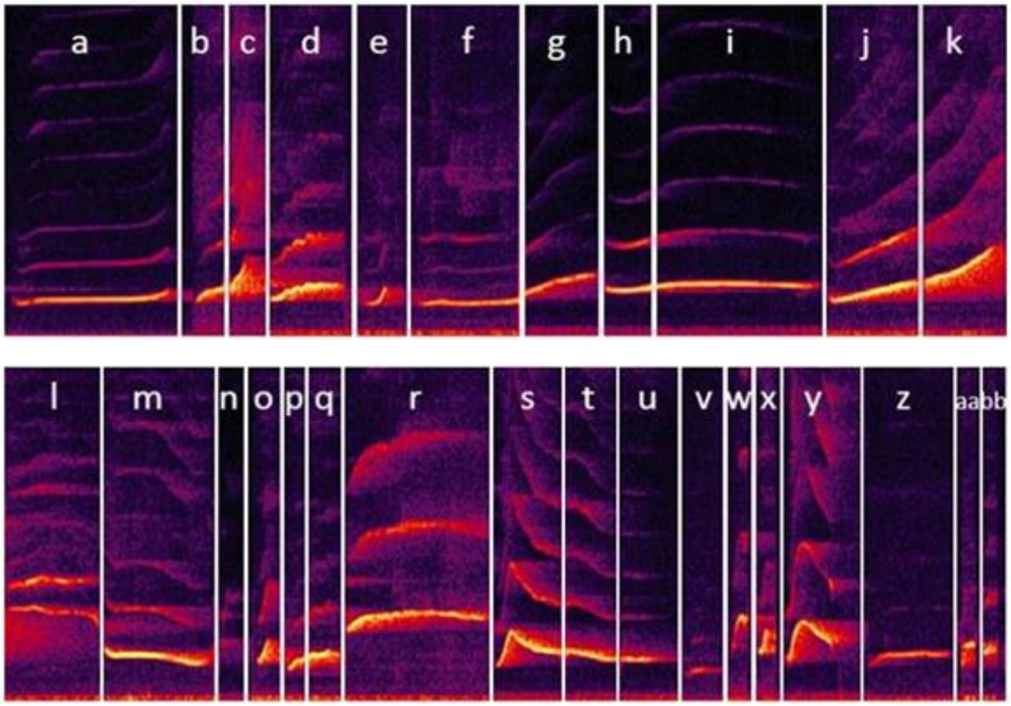


Figure 4.4: Acoustic categories of white-handed gibbon vocalizations represent a diverse range of sounds. Though some of these categories may seem like they should not be divided, closer examination demonstrates they are unique combinations of multiple acoustic units. Unit /n/ was initially identified as a possible acoustic category; however further examination of this unit suggests it is an artefact of a physiological constraint and not a communicative unit.

Call classification – computational approach

A computational analysis of the vocalizations was performed to validate the accuracy of the hand-labeling approach and also to provide a rigorous way to accurately process animal calls. Sound files from each species were analyzed in MATLAB using previously developed approaches (Coen, 2006a; Coen, 2006b; Coen 2007) and the MIR Toolbox (Lartillot & Toiviainen, 2007). Each file was filtered to remove background noise and then processed as a similarity matrix to determine acoustic unit boundaries. The cepstrum of these acoustic units was processed and Mel-Frequency Cepstral Coefficients (MFCC) from each acoustic unit was extracted. These features were used to cluster the acoustic units based upon their measured similarity. Figure 4.5 provides a flow chart of this process as well as an example of a single gibbon file.

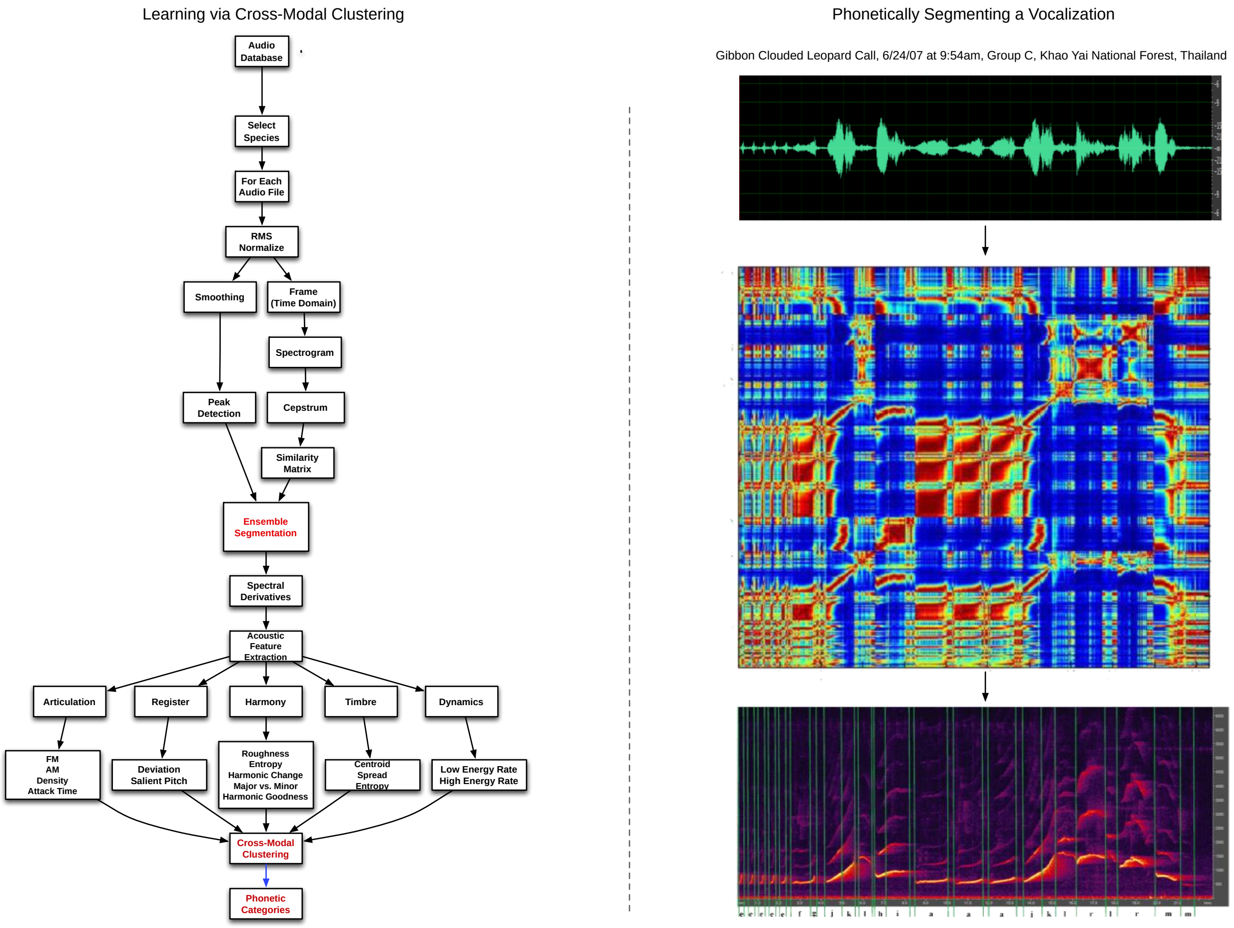


Figure 4.5: The left hand side of this figure depicts the order that a file is processed. The right hand side shows the waveform of an original file, a similarity matrix of that file and a segmented representation of acoustic units identified.

As shown in Figure 4.6, visual inspection of some spectral images does not provide a clear picture of whether multiple sounds are present. For example, acoustic unit's /j/ and /k/ may appear to be a single continuous sound which would not merit a boundary between them. However in examining these units using the CSSM approach, clear distinctions within their acoustical representations is observed. By using a CSSM, subtle differences in acoustic phenomena can be detected and more clearly illustrated. These subtle differences correspond to measureable distinctions in sounds which can be heard by a well-trained observer and supported via computational measurements.

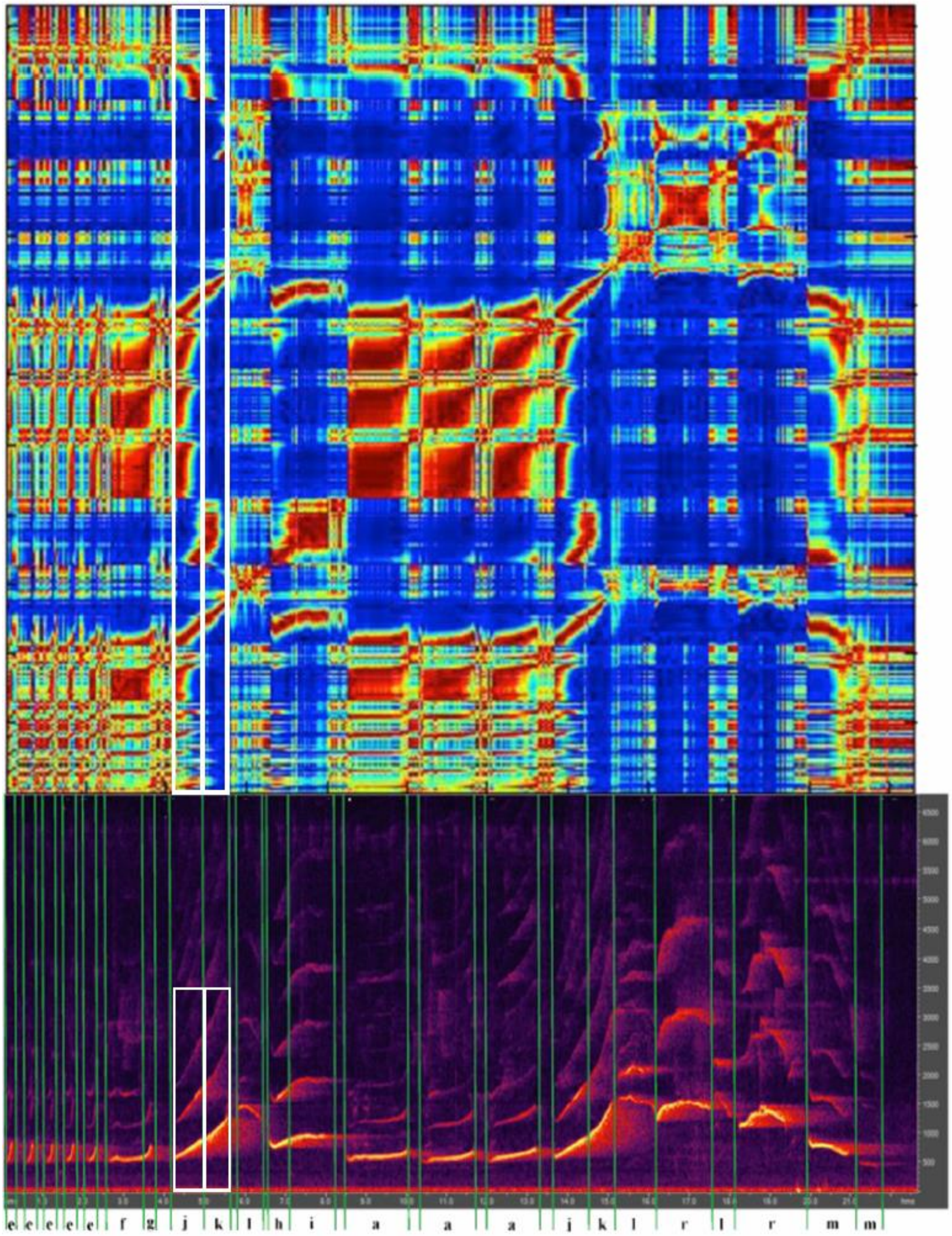


Figure 4.6: This female great call illustrates an example where viewing a spectrogram doesn't capture an obvious auditory distinction, but a CSSM makes it obvious. The white boxes highlight two acoustic units, /j/ and /k/. In viewing a spectral image, these units may appear to be a continuous sound; however in viewing their CSSM image, clear variation in their structure emerges.

Results

Categories vs. Acoustic units

The 21,000+ (12,449 wild, ~9,000 captive) calls used in this analysis belonged to 27 acoustic unit categories. As shown in Figure 4.7, the acoustic unit categories described here are primarily subsets of the previously described general categories. Our acoustic unit categories support previous work and also clarify some of the noticeable differences within each of the categories. For instance, the "oo" category from Raemaekers et al. (1984, 1985) and Clarke et al.

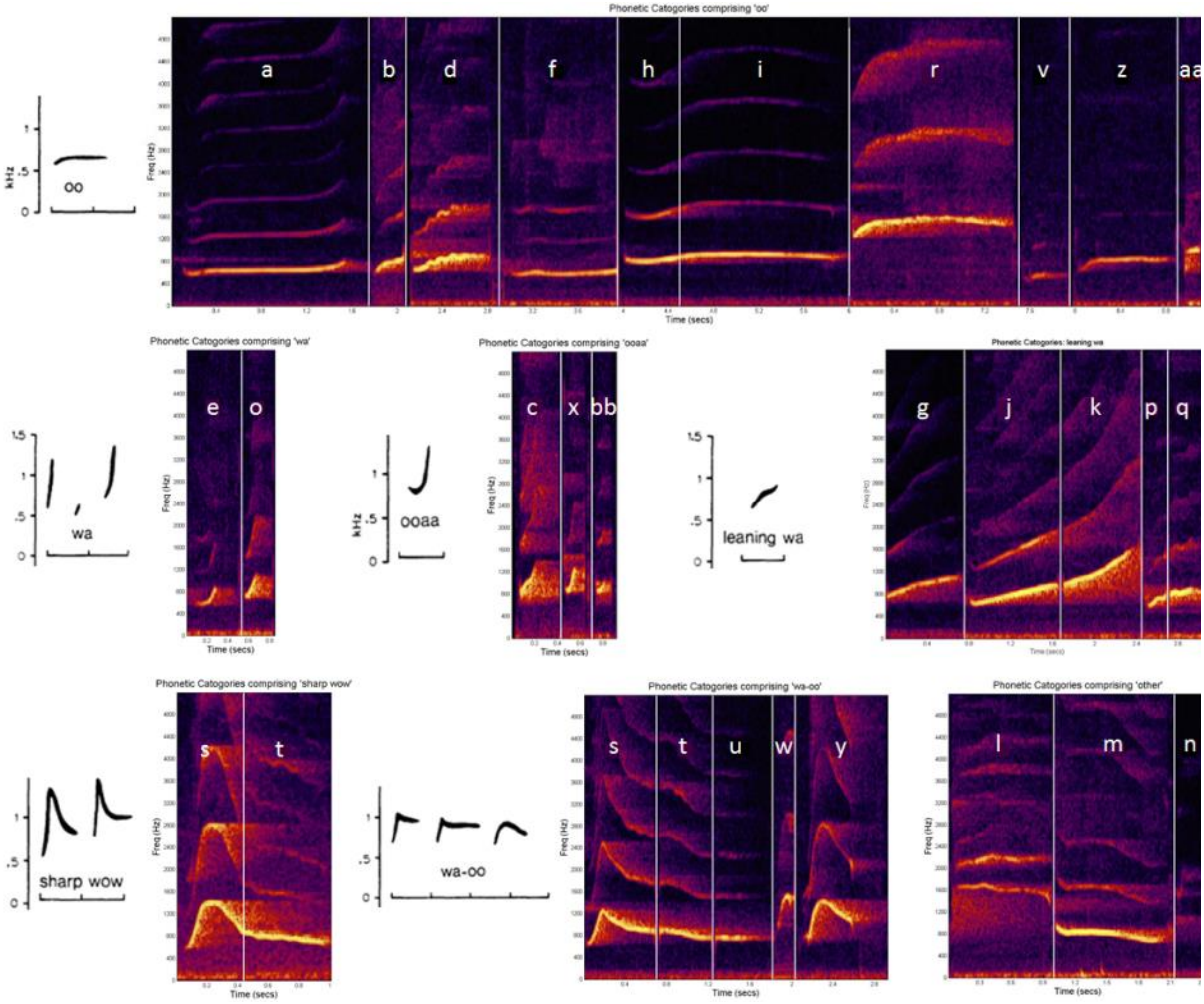


Figure 4.7: The line drawings in this figure represent the categories described in Raemaekers et al., 1984. After further examination of the vocal repertoire of white-handed gibbons, each category appears to be made up of several distinct units.

(2006) have 10 distinctive subtypes or acoustic units. The "wa" category includes acoustic units /e/ and /o/, the "ooaa" category includes acoustic units /c/, /x/, and /bb/, the "leaning wa" category includes acoustic units /g/, /j/, /k/, /p/, and /q/, the "sharp wow" category includes acoustic units /s/ and /t/, and the 'wa-oo' category includes acoustic units /s/, /t/, /u/, /w/, and /y/. The /s-t/ combination consists of a great deal of variation and overlapped two of the previously described categories. As illustrated in Figure 4.7, this combination ranged from very sharp frequency transitions to relatively moderate transitions. While these categories may be able to be separated further by visual cues, the auditory cues being used in this analysis were insufficiently different to warrant further divisions. Given the variability of these acoustic units, a best estimate for classification is given here and further sub-categorization will be described by utilizing the computational approach. Additional acoustic units which were previously unclassified include units' /l/, /m/, and /n/. Unit /n/ was only observed four times. Due to the lack of adequate samples, this unit was not treatable using our current framework. That the rest of these acoustic units are separable and repeated suggests they are deserving of their own classification. Chapter 6 describes how these acoustic unit categories are combined to elicit different meaning or behavioral responses.

Table 4.1 presents results for the manually labeled acoustic units. The duration and three frequency measurements were taken for the description of each acoustic unit. The shortest acoustic unit described was acoustic unit /x/ at 173.48 ±49.925ms. The longest acoustic unit described was acoustic unit /a/ at 1821.4 ± 356.68 ms. Frequency measurements were taken in Hz at the first quarter, middle and third quarter of each acoustic unit's duration. Acoustic unit /e/ occurred the most frequently at 2993 times and acoustic unit /u/ occurred the least frequently at 24 times. These categories were further verified by comparing each to the computational segmentation.

Acoustic unit	Length (ms)	StDev	Mass1 (Hz)	Mass2	Mass3	Sdm1	Sdm2	Sdm3	n
a	1821.40	356.68	556.55	589.54	584.70	60.14	63.66	73.44	579
aa	212.18	85.11	954.17	981.17	984.03	169.92	152.18	161.67	60
b	236.06	91.32	643.28	821.99	875.11	110.32	89.48	100.29	295
bb	186.64	62.74	913.38	992.07	1027.10	135.26	88.87	133.71	281
c	210.82	69.23	900.96	990.39	1017.40	128.49	116.93	157.11	301
d	728.52	218.59	778.66	910.42	868.23	110.08	117.65	137.02	443
e	257.73	70.69	545.14	772.69	767.33	124.77	139.09	176.66	2993
f	1548.60	418.85	541.22	619.67	662.96	64.45	76.07	98.40	265
g	772.47	265.21	728.13	780.92	811.08	105.05	112.48	143.13	188
h	595.32	116.96	620.35	767.28	799.84	114.78	81.97	92.15	100
i	1649.70	377.02	740.62	798.91	738.41	134.16	110.65	146.58	197
j	1060.10	387.29	657.95	863.71	955.93	154.32	147.26	156.17	244
k	1247.10	440.07	1042.80	1168.20	1200.10	153.45	132.02	173.73	384
l	695.58	318.52	1321.00	1449.40	1273.10	244.36	323.99	299.69	224
m	1248.60	291.81	679.76	706.81	618.17	149.98	191.01	173.08	299
n	346.00	114.18	1086.50	1222.10	1075.60	125.07	213.00	126.11	4
o	297.54	65.30	577.42	1036.60	1041.00	115.51	113.54	162.91	259
p	194.01	40.56	616.48	825.07	885.53	106.38	84.50	99.75	491
q	174.91	61.80	897.46	938.02	905.98	107.64	106.06	153.64	485
r	1362.00	190.02	1287.30	1434.10	1394.00	149.13	136.30	182.71	25
s	374.92	120.00	668.40	1042.60	1031.90	133.76	167.09	175.79	838
t	539.26	186.74	938.26	886.06	777.67	142.52	126.23	134.17	643
u	452.62	207.89	687.64	675.36	651.70	83.99	65.86	82.94	24
v	213.40	66.14	521.22	591.48	564.21	156.03	110.08	148.46	1804
w	272.06	57.29	694.59	1150.80	1158.20	170.32	117.15	144.32	87
x	173.48	49.93	927.78	998.38	1015.30	140.96	84.74	123.92	335
y	648.71	123.73	787.41	1196.10	967.14	151.20	129.51	147.88	324
z	1032.20	357.85	588.96	634.05	601.06	105.79	98.30	100.05	277

Table 4.1: A description of the measured acoustic properties of acoustic units in the gibbons' vocal repertoire includes the length or duration of each category, the sample size, and the frequency of the first quarter (Mass 1, Sdm1 is Standard deviation for this mass), half (Mass2, Sdm2) and third (Mass3, Sdm3) quarter measured in Hertz.

Novel vocalization

The zoo gibbon vocalizations consist of approximately 4,500 acoustic units from each pair for a total of 9,000 acoustic units. They contain the same acoustic units found in the vocal repertoires of the wild gibbons. Additionally, a new acoustic unit was discovered in the repertoire of the gibbons at the Racine Zoological Society. This acoustic unit is characterized by a guttural sound that does not resemble the rest of the gibbon's vocal repertoire; however it does closely resemble the bleating sound heard from sheep. This sound is not found in the repertoire of the wild gibbons or the Henry Vilas Zoo gibbons. Figure 4.8 shows several bleats from the male. A total of 74 samples were recorded. Figure 4.9 illustrates the variation in amplitude which exists in these calls. The bandwidth of this vocalization ranges from 500Hz to 11 kHz and the duration ranges from .4 to 1.8 seconds.

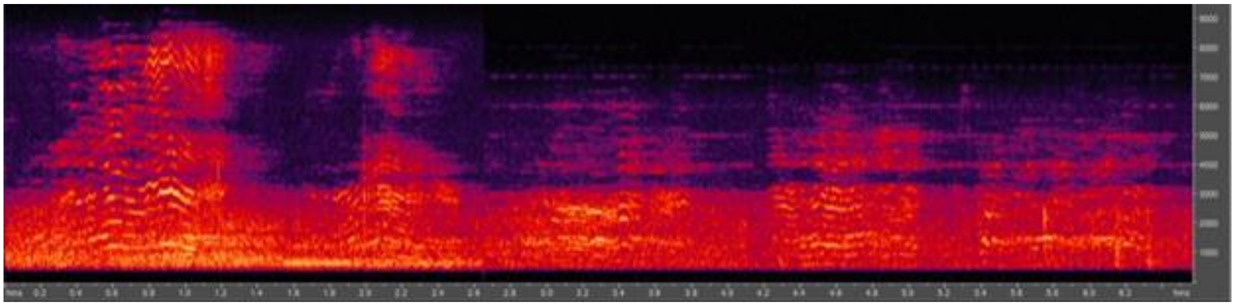


Figure 4.8: A spectral image of 5 of the bleats recorded from the Racine Zoological Society. This call was not found in any of the wild gibbon recordings or the Henry Vilas Zoo gibbons.

This call was typically given midday during a play bout that rarely lasted more than five minutes. The initial behavior moves from rougher grooming to both gibbons locking their hands and feet together. Upon grasping one another, they roll around and bare their teeth. During this time, the behavior moves from a more playful nature to a more aggressive (but still playful) nature and the more submissive individual (usually the male) utters the bleating sound. As the roughness of the play behavior escalates, the amplitude of this call increases until the

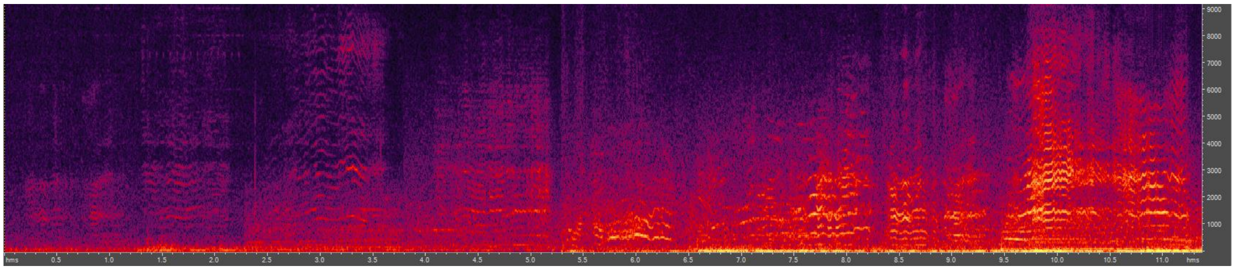


Figure 4.9: The amplitude of the bleat vocalizations is highly variable. The beginning of a bout is very quiet. As the play activity continues the strength of the bleat increases.

aggressor backs down and moves away. While the development of this call is unclear, the function of the call appears to be a stop signal to prevent any physical harm to each other.

Validation

Acoustic unit categorization of the gibbon's vocal repertoire is supported by previous research classifying gibbon calls. That these acoustic unit categories can be considered subsets of the previous classifications suggests they are part of a finer grained analysis of the gibbon's vocal repertoire. Additional support is gained by comparing the manual segmentation of the acoustic units to the computational segmentation of the acoustic units. This support, shown in the CSSM, is particularly important as it reduces the likelihood of observer bias in acoustic unit description by classifying calls based upon measureable, acoustic properties. Finally, Zipf's law is utilized as an extra statistical test.

Cepstral Self-Similarity Matrix

CSSM's are used to calculate the similarity between two acoustic units. As part of this process, Mel-Frequency Cepstral Coefficients are calculated by forming feature vectors from overlapping Hamming windows. Discrete Fourier transforms are used to calculate the log of the power spectrum for each window. The cepstral coefficients are then calculated by taking another discrete Fourier transform. The resulting image is a matrix of color coded pixels. The warm colors (red, orange and yellow) indicate strong similarity in features, while the cool colors (green and blue) indicate strong differences in features. The diagonal running from the lower left hand corner to the upper right hand corner illustrates a given acoustic unit being compared to itself. In other words, this diagonal represents the CSSM visual description of a given acoustic unit. As shown in figures 4.10 & 4.11, the CSSM approach largely agrees with the 27 acoustic units identified in the manual analysis. While the manual and computational classifications largely agreed with one another, there were a few minor differences with boundary placements. Though differences did exist in a few cases, they are likely due to human perceptual error and were not considered significant enough to invalidate acoustic unit comparisons.

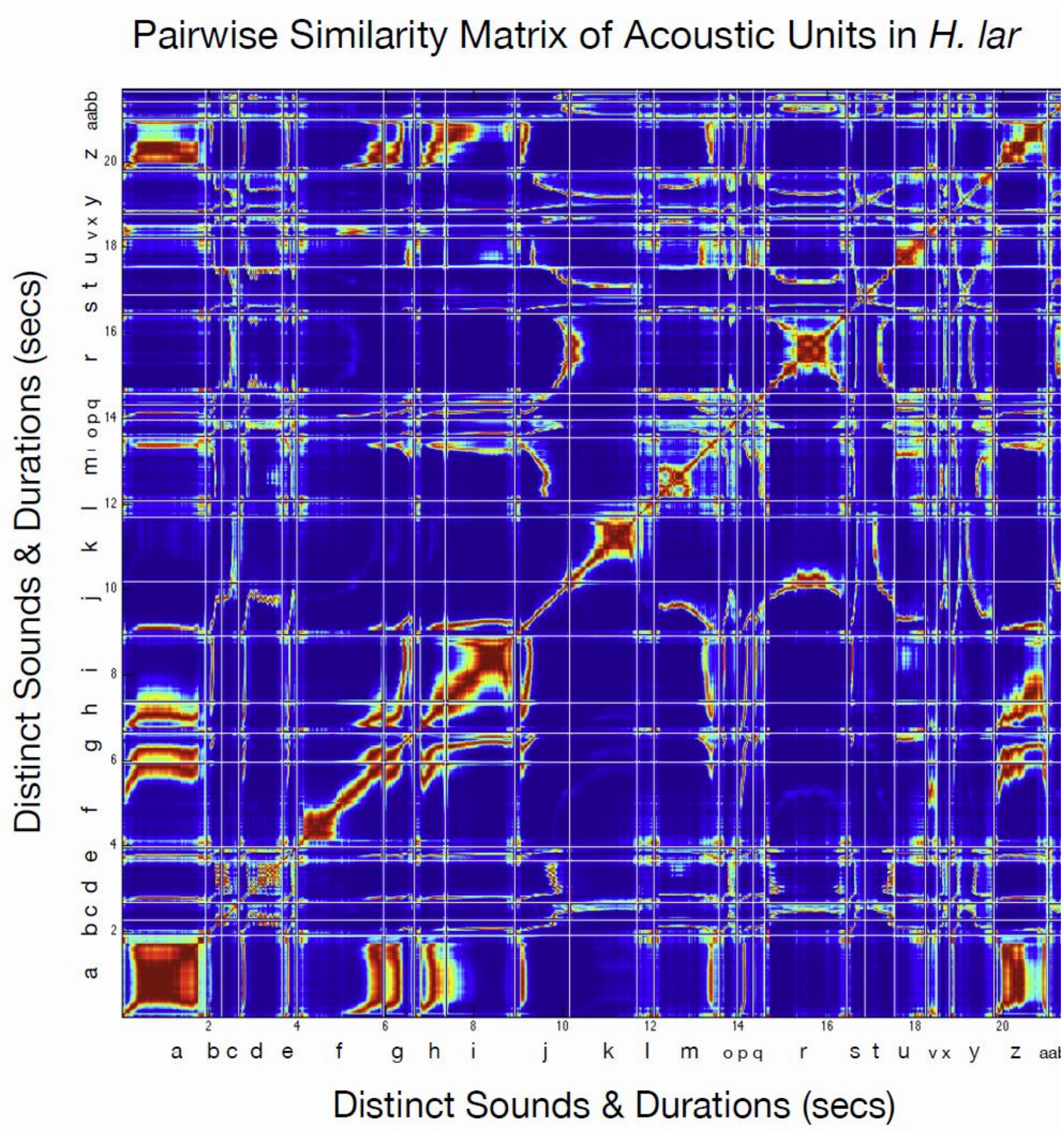


Figure 4.10: A similarity matrix of the 27 classified acoustic units displays the distinctiveness of each category as well as validates the human classification. Acoustic unit /n/ is not shown in this figure because it was determined to be an artefact of vocal production.

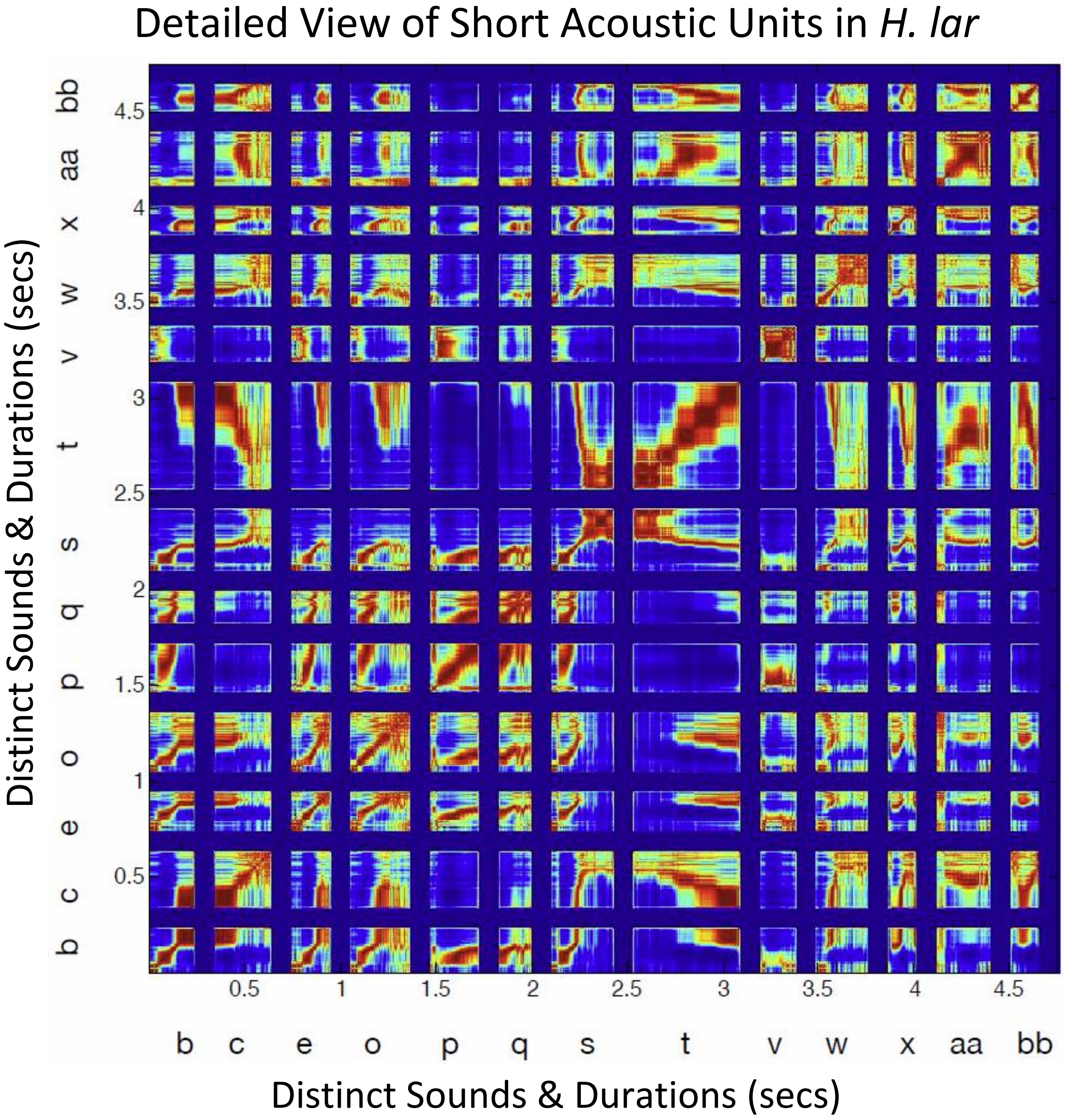


Figure 4.11 – A similarity matrix of the 13 shortest acoustic units in the white-handed gibbon's vocal repertoire. These units also appear in Figure 4.10; however this is a more detailed view of each unit.

Zipf's law

Zipf's law is a power law which predicts the occurrence of an entity be proportional to its' rank. It is used here to demonstrate there is potential communicative value in the acoustic unit analysis of gibbon vocalizations. As shown in Figure 4.12, the breakdown of acoustic units does predictably follow Zipf's law with an R^2 value of 0.9257. While the value of demonstrating this with animal vocalizations has been controversial (See Chapter 3 for a review), we believe it is still a useful statistic to support further examination of our data set.

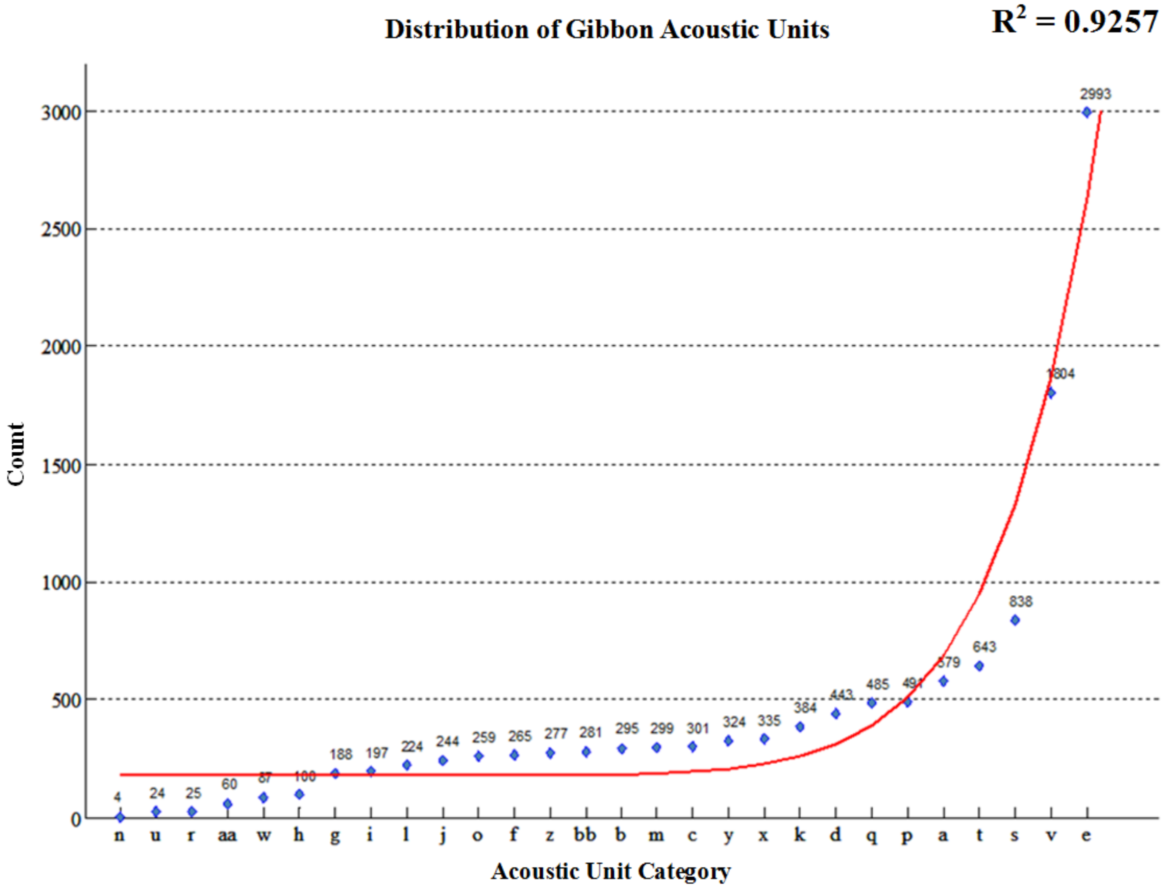


Figure 4.12: The distribution of gibbon acoustic units follows Zipf's law. While this is insufficient evidence for claims of language-like structure, it is important for demonstrating the potential in communicative value in the gibbons' vocal repertoire.

Discussion

Our findings suggest there is more complexity within white-handed gibbon vocalizations than previously thought. The manner in which this complexity has been confirmed also provides a promising future for automating the analysis of vocalizations from other species. This automated approach could be useful for many types of research and may be especially valuable to studies focused on understanding the meaning of calls. For gibbons, further work examining the transitional probabilities between acoustic units, as well as further variation in usage of acoustic units during predator presentation is necessary. Additionally, vocalizations given during other behaviors such as social interactions and feeding should be examined. Identifying how these acoustic units are used in different behavioral situations will help to address important questions about the vocal communication system of gibbons.

Examining the interior structures of the vocal repertoire has revealed many subtleties within the calls. The numerous variances found could have significant implications in how white-handed gibbons communicate. For example, behavioral specificity of particular sounds can help to explain how we expect them to be used. This could help to clarify why so many combinations can be found when eliciting a common response such as a particular predator species.

Additional complexity within the vocal repertoire of white-handed gibbons has been demonstrated here by using a novel combination of computational, linguistic, and zoological approaches. We believe the computational approach used here is the first of its kind to be applied to understanding vocal complexity in primate vocal communication. The robustness of this approach has provided independent, empirical support for splitting the previously described sound groups into 27 distinct, phoneme-like clusters. While the gibbons' mental representation of these clusters has yet to be confirmed, and thus do not pass the linguistic requirements to be considered a phoneme, their predictable occurrence in a given acoustical environment suggests they are being perceived as distinct variants of a sound. For example, in the ‘oo’ category (Raemaekers & Raemaekers 1984; Clarke et al. 2006), we identified two of the ten acoustic units as /a/ and /i/. In examining the use of these sounds, /a/ always occurred before another /a/ or /j/ and after /g/ or /i/. A /i/ always occurred before /a/ and after /h/ or /g/. Examination of these environments is discussed in Chapter 6. Furthermore, there is a growing body of literature that suggests apes are capable of remembering past events (Martin-Ordas et al. 2013) which support the theory of higher level cognitive functions such as mental representations of sounds. Determining whether this is also true for a lesser ape requires further research.

Analysis with a CSSM offers further support for splitting the general sound categories into acoustic units. Upon casual visual inspection of a spectral image, /a/ and /i/ may appear somewhat similar; however upon a more detailed inspection of a CSSM comparison, it becomes apparent that the acoustical features which define the sound are unique. When expanded to include the rest of the vocal repertoire, we demonstrate that all 27 acoustic units are unique. Furthermore, the categorization from this computational analysis was largely congruent with human observer categorization, suggesting future work may be done via an automated approach. Automating the analysis of vocalizations offers several advantages, including being more time efficient and less biased.

Behavioral novelties and vocal production

Beyond the scope of the previously described vocal repertoire, a novel sound was discovered in a pair of captive gibbons at the Racine Zoological Society. The sound produced is akin to a sheep bleat and is uttered during a specific behavioral activity. It is unlikely that the distinctness of this sound is due to any physiological damage to the gibbons’ vocal tract as the rest of their vocal repertoire is acoustically similar to wild gibbons and other captive gibbons.

While the origin of this call is unclear, we do propose several possibilities for why this has not been observed previously. All of these possibilities require further research as no known origin of this call currently exists. First, the amplitude of this call is very low and as such an observer would have to be within 10-15 feet of the gibbons to hear it. In the wild, this is highly unlikely. Second, unlike most of their other calls, this bleat was typically observed either mid-day or in the afternoon. This is generally not the best time to record in the wild due to increased ambient noise and decreased sound transmission quality from humidity and warmer air (Waser & Brown, 1984). Third, this call may be a function of familial play. In the wild, the female would have dispersed to a new territory away from her father. A causal link with a familial relationship is also consistent with why this was not observed in the other pair of captive gibbons in this study. Fourth, it is possible that this call was learned by the adult male early in his life when he was brought into captivity from the jungles of Thailand and subsequently passed onto his daughter who he has been housed with since her birth. Another possibility is that this call is used by infants, and older gibbons learn to stop using it as they mature. The male was brought into captivity at a young age and it is possible he never learned to stop using this call, which was subsequently passed onto his daughter.

Describing these bleats as novel has its drawbacks, especially when the categorization is based on a data set that is as small as one pair of gibbons. To help strengthen our findings, a large number of experts across several other zoos and sanctuaries have been consulted. These experts, which are the main caretakers of their respective gibbons, were shown a video of the Racine gibbons bleating and fighting and they were asked whether this behavior or sound was found in their pairs of gibbons. Currently no observations of the call or behavior which leads up to this call have been made with 40 other gibbons under the care of 6 other zoos and sanctuaries. One other pair of gibbons does engage in similar chasing behavior and unusual call production though how similar the sounds are to one another will require further research. Based upon our responses from others, we believe this behavior and call is a rare event in gibbon social behavior. One contributing factor that may be relevant to why the Racine gibbons consistently do this, but most of the other gibbons do not is the fact that they are related to one another whereas most of the other pairs are not. It is interesting to note that the one pair that does appear to behave somewhat similarly is a pair of half-brothers who have also been housed together for several decades. Given the rarity of housing related gibbons together, it would not be expected that many would have had the chance to witness such an irregular behavior. Functionally, it is plausible that all gibbons can produce these calls and behavior and this familial dispute could mark a point in which the offspring need to leave the group and search for their own territory. In captive settings, the ability to disperse is restricted and therefore this behavior is more persistent. Future work should focus on related pairs of gibbons to further test the hypothesis that this behavior is related to familial disputes.

Atypical production of acoustic units

In addition to the unusual bleat sounds, another interesting discovery on vocal usage was observed. The gibbons at the Henry Vilas Zoo were regularly observed violating gender specific rules when giving great calls and codas. Female great calls were regularly produced by the male and male codas were regularly produced by the female. These gender role changes either occurred for the whole call or part of the call. For instance, the male may start a great call and the female would join in and finish it or both gibbons might give the whole great call together. The gender interchange in call production suggests that there is vocal learning involved in call usage. Previous work with gibbons has suggested that gibbon song had developed from duet-splitting (Geissmann, 2002). Our results support this previous finding that gibbons are capable of producing the full repertoire of acoustic units, but they learn to use their gender specific calls. Since both of the gibbons at the Vilas Zoo were born in captivity, they may not have been exposed to stereotypical calls and thus unable to learn what call they should produce. Previous work on vocal development in agile gibbons (*Hylobates albibarbis*) suggests that song development is heavily influenced by intra- and inter-specific communication (Cheyne et al., 2007) as well as a particularly important role which mothers play in their daughters song development (Koda et al., 2013). Evidence from the pair of white-handed gibbons at the Vilas Zoo suggests their vocal development may also be influenced by adult gibbons.

Summary

Novel use of CSSM methods was successfully implemented here to analyze the vocal repertoires of white-handed gibbons on an acoustic unit level. By examining the interior structure of the calls, we were able to identify 27 acoustic units, instead of the 6 – 7 broad categories that were previously described (Raemaekers *et al.*, 1984; Clarke *et al.*, 2006). The difference in the number of sound descriptions mainly comes from examining the large category of 'other calls' in Clarke *et al.* (2006) and by separating out the different acoustic units that made up the six remaining categories from Raemaekers *et al.* (1984).

This work both supports and furthers these previous analyses (See figure 4.10) by categorizing the variation found within each of the major groups of call types. To support a finer grained analysis, novel computational methods were used. One of the major problems with forming smaller categories is the amount of time it takes to manually analyze and hand-label each unit. To properly label each category, one must also be able to recognize the difference between individual variations in a given sound versus categorical differences between sounds. This categorization becomes increasingly more difficult as more individuals are added to a data set. Determining boundaries between variation is often difficult because each individual may have a distinct accent (Spillmann et al, 2010, Mitani et al, 1996) so the same unit of sound may actually sound like it is potentially a new category. Being able to decipher between what is truly a distinct category and what is individual variation is complex even for researchers who observed the animals as they were being recorded. Adding to complexity of individual variation is our human perception of these calls. By using a more empirical approach, we can side-step the issue of observer bias.

Examining gibbon vocalizations in more detail opens the door for exploring whether any higher level linguistic structure exists as well as looking for meaning that has previously been masked. Being able to understand the communication systems of other organisms has been a long standing human interest and this research takes another step towards that goal. This work has also demonstrated that white-handed gibbon vocalizations are much more complex than previously described. In addition, this work provides the necessary foundation for determining whether previous claims of syntax in white-handed gibbon vocalizations were warranted.

Additionally, this work explores a previously undocumented call produced during a given behavioral activity. The quietness of this call compared to the rest of the gibbons' repertoire suggests it is meant to inform only the closest individual. The function of such a call is likely to mediate the play behavior to ensure it does not escalate too far. The variation in amplitude does increase as the duration of the play behavior increases. Early, quiet versions of this call may play a role as a precursor to future, louder calls when play behavior escalates. Further work examining this call is needed.

Chapter References

Baker, M. C., & Logue, D. M. (2003). Population differentiation in a complex bird sound: A comparison of three bioacoustical analysis procedures. *Ethology, 109*(3), 223-242.

Bartlett, T. Q. (2003). Intragroup and intergroup social interactions in white-handed gibbons. *International Journal of Primatology, 24*(2), 239-259.

Becker, M. L., Buder, E. H., & Ward, J. P. (2003). Spectrographic description of vocalizations in captive Otolemur garnettii. *International Journal of Primatology, 24*(2), 415-446.

Blumstein, D. T., & Munos, O. (2005). Individual, age and sex-specific information is contained in yellow-bellied marmot alarm calls. *Animal Behaviour, 69*, 353-361.

Bohn, K. M., Schmidt-French, B., Schwartz, C., Smotherman, M., & Pollak, G. D. (2009). Versatility and stereotypy of free-tailed bat songs. *Plos One, 4*(8).

Buck, J. R., & Tyack, P. L. (1993). A quantitative measure of similarity for Tursiops truncatus signature whistles. *Journal of the Acoustical Society of America, 94*(5), 2497-2506.

Cheyne, S. M., Chivers, D. J., & Sugardjito, J. (2007). Covarvariation in the great calls of rehabilitant and wild gibbons (*Hylobates albibarbis*). *Raffles Bulletin of Zoology, 55*(1), 201-207.

Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *Plos One, 1*(1).

Cleveland, J., & Snowdon, C. T. (1982). The complex vocal repertoire of the adult cotton-top tamarin (*Saguinus oedipus oedipus*). *Zeitschrift Fur Tierpsychologie-Journal of Comparative Ethology, 58*(3), 231-270.

Coen, M. (2006). *Multimodal Dynamics: Self-supervised learning in perceptual and motor systems*. Unpublished Doctoral dissertation, Massachusetts Institute of Technology.

Coen, M. (2006). *Self-supervised acquisition of vowels in American English*. Paper presented at the Proceedings of the Twenty First National Conference on Artificial Intelligence (AAAI), Boston, MA.

Coen, M. (2007). *Learning to sing like a bird: Self-supervised acquisition of birdsong*. Paper presented at the Proceedings of the Twenty Second National Conference on Artificial Intelligence (AAAI), Vancouver, British Columbia.

Cowlishaw, G. (1992). Song function in gibbons. *Behaviour, 121*, 131-153.

Darwin, C. R. (1871). *The descent of man, and selection in relation to sex*: London.

Dassow, A. M. (2010). *Vocal repertoire of the white-handed gibbon (Hylobates lar).* Unpublished Masters of Zoology, University of Wisconsin-Madison.

Fee, M. S., Shraiman, B., Pesaran, B., & Mitra, P. P. (1998). The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature, 395*(6697), 67-71.

Gamba, M., & Giacoma, C. (2007). Quantitative acoustic analysis of the vocal repertoire of the crowned lemur. *Ethology Ecology & Evolution, 19*(4), 323-343.

Geissmann, T. (2002). Duet-splitting and the evolution of gibbon songs. *Biological Reviews, 77*(1), 57-76.

Geissmann, T., & Nijman, V. (2006). Calling in wild silvery gibbons (Hylobates moloch) in Java (Indonesia): Behavior, phylogeny, and conservation. *American Journal of Primatology, 68*(1), 1-19.

Gros-Louis, J. J., Perry, S. E., Fichtel, C., Wikberg, E., Gilkenson, H., Wofsy, S., et al. (2008). Vocal repertoire of Cebus capucinus: Acoustic structure, context, and usage. *International Journal of Primatology, 29*(3), 641-670.

Hammerschmidt, K., & Fischer, J. (1998). The vocal repertoire of Barbary macaques: A quantitative analysis of a graded signal system. *Ethology, 104*(3), 203-216.

Henderson, E. E., Hildebrand, J. A., & Smith, M. H. (2011). Classification of behavior using vocalizations of Pacific white-sided dolphins (Lagenorhynchus obliquidens). *Journal of the Acoustical Society of America, 130*(1), 557-567.

Hsu, M. J., Chen, L. M., & Agoramoorthy, G. (2005). The vocal repertoire of Formosan macaques, Macaca cyclopis: Acoustic structure and behavioral context. *Zoological Studies, 44*(2), 275-294.

Ji, A., Johnson, M. T., Walsh, E. J., McGee, J., & Armstrong, D. L. (2013). Discrimination of individual tigers (Panthera tigris) from long distance roars. *Journal of the Acoustical Society of America, 133*(3), 1762-1769.

Koda, H., Lemasson, A., Oyakawa, C., Rizaldi, Pamungkas, J., & Masataka, N. (2013). Possible role of mother-daughter vocal interactions on the development of species-specific song in gibbons. *Plos One, 8*(8).

Kogan, J. A., & Margoliash, D. (1998). Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. *Journal of the Acoustical Society of America, 103*(4), 2185-2196.

Koren, L., & Geffen, E. (2011). Individual identity is communicated through multiple pathways in male rock hyrax (Procavia capensis) songs. *Behavioral Ecology and Sociobiology, 65*(4), 675-684.

Lartillot, O., & Toiviainen, P. (2007). *A Matlab Toolbox for Musical Feature Extraction from Audio*. Paper presented at the Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France.

Mallat, S. (1999). *A wavelet tour of signal processing*: Academic press.

Marshall, J. T., & Marshall, E. R. (1976). Gibbons and their territorial songs. *Science, 193*(4249), 235-237.

Martin, W., & Flandrin, P. (1985). Wigner-ville spectral-analysis of nonstationary processes. *Ieee Transactions on Acoustics Speech and Signal Processing, 33*(6), 1461-1470.

Martin-Ordas, G., Berntsen, D., & Call, J. (2013). Memory for distant past events in chimpanzees and orangutans. *Current Biology, 23*(15), 1438-1441.

Masataka, N. (1982). A field study on the vocalizations of Goeldi's monkeys (*Callimico goeldii*). *Primates, 23*(2), 206-219.

Matthews, J. N., Rendell, L. E., Gordon, J. C. D., & Macdonald, D. W. (1999). A review of frequency and time parameters of cetacean tonal calls. *Bioacoustics, 10*(1), 47-71.

Mitani, J. C., GrosLouis, J., & Macedonia, J. M. (1996). Selection for acoustic individuality within the vocal repertoire of wild chimpanzees. *International Journal of Primatology, 17*(4), 569-583.

Mitani, J. C., & Marler, P. (1989). A phonological analysis of male gibbon singing behavior. *Behaviour, 109*, 20-45.

Narins, P. M., Lewis, E. R., & McClelland, B. E. (2000). Hyperextended call note repertoire of the endemic Madagascar treefrog Boophis madagascariensis (Rhacophoridae). *Journal of Zoology, 250*, 283-298.

Oppenheim, A. V., & Schafer, R. W. (2004). From frequency to quefrency: A history of the Cepstrum. *Ieee Signal Processing Magazine, 21*(5), 95-106.

Oswald, J. N., Rankin, S., Barlow, J., & Lammers, M. O. (2007). A tool for real-time acoustic species identification of delphinid whistles. *Journal of the Acoustical Society of America, 122*(1), 587-595.

Raemaekers, J. J., & Raemaekers, P. M. (1984). The ooaa duet of the gibbon (*Hylobates lar*) - A group call which triggers other groups to respond in kind. *Folia Primatologica, 42*(3-4), 209-215.

Raemaekers, J. J., & Raemaekers, P. M. (1985). Field playback of loud calls to gibbons (*Hylobates lar*) - Territorial, sex-specific and species-specific responses. *Animal Behaviour, 33*(MAY), 481-493.

Raemaekers, J. J., Raemaekers, P. M., & Haimoff, E. H. (1984). Loud calls of the gibbon (*Hylobates lar*) - Repertoire, organization and context. *Behaviour, 91*, 146-189.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review, 101*(1), 129-156.

Root-Gutteridge, H., Bencsik, M., Chebli, M., Gentle, L. K., Terrell-Nield, C., Bourit, A., et al. (2014). Identifying individual wild Eastern grey wolves (Canis lupus lycaon) using fundamental frequency and amplitude of howls. *Bioacoustics, 23*(1), 55-66.

Saar, S., & Mitra, P. P. (2008). A technique for characterizing the development of rhythms in bird song. *Plos One, 3*(1).

Spillmann, B., Dunkel, L. P., van Noordwijk, M. A., Amda, R. N. A., Lameira, A. R., Wich, S. A., et al. (2010). Acoustic properties of long calls given by flanged male orang-utans (Pongo pygmaeus wurmbii) reflect both individual identity and context. *Ethology, 116*(5), 385-395.

Taylor, A. M., Reby, D., & McComb, K. (2008). Human listeners attend to size information in domestic dog growls. *Journal of the Acoustical Society of America, 123*(5), 2903-2909.

Tchernichovski, O., Lints, T. J., Deregnaucourt, S., Cimenser, A., & Mitra, P. P. (2004). Studying the song development process rationale and methods. *Behavioral Neurobiology of Birdsong, 1016*, 348-363.

Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour, 59*, 1167-1176.

Thomson, D. J. (1982). *Spectrum estimation and harmonic analysis.* Paper presented at the *Proceedings of the IEEE.*

Van Parijs, S. M., Parra, G. J., & Corkeron, P. J. (2000). Sounds produced by Australian Irrawaddy dolphins, Orcaella brevirostris. *Journal of the Acoustical Society of America, 108*(4), 1938-1940.

Wanelik, K. M., Azis, A., & Cheyne, S. M. (2013). Note-, phrase- and song-specific acoustic variables contributing to the individuality of male duet song in the Bornean southern gibbon (Hylobates albibarbis). *Primates, 54*(2), 159-170.

Waser, P. M., & Brown, C. H. (1984). Is there a sound window for primate communication. *Behavioral Ecology and Sociobiology, 15*(1), 73-76.

“Chimpanzee pant-hoots, gibbon song vocalizations, as well as the loud calls of other apes and Old World monkeys, and human singing are typically accompanied by locomotor displays and have been suggested to be homologous features that can be traced back to a behavioural pattern already present in the common ancestor of the catarrhine primates. This behaviour is originally believed to serve the purpose of displaying and possibly reinforcing the unity of a social group towards other groups.” (Geissmann, 2009)

Chapter 5: Markov chain representation of acoustic sequences in wild vs. captive gibbon responses to predator models

Abstract

Considerable attention has been focused on whether or not animal vocalizations possess the required criteria attributed to semantic or functionally referential information. Whereas many studies have demonstrated strong evidence for the potential existence of referential signals in various species, few have explored this possibility with lesser apes. In this chapter, we review alarm calls in wild and captive white-handed gibbons (*Hylobates lar*) for motivational and referential content. We explore the benefits and drawbacks to segmenting and sequencing vocal bouts using an acoustic unit approach and a general sound category approach. In doing so, we illustrate a vast amount of information has been overlooked previously. Upon examining calls using an acoustic unit approach, we discover both motivational and referential aspects to the warning calls of white-handed gibbons. This information elucidates additional meaning in the vocal responses of gibbons and contributes to a growing body of literature which highlights the complex nature of vocal communication in non-human species.

Introduction

Deciphering meaning in non-human animal vocalizations has been the focus of considerable research (Struhsaker, 1967; Seyfarth et al., 1980a, b; Evans et al., 1993a; Slocombe & Zuberbühler, 2006; Arnold & Zuberbühler, 2008; Stephan & Zuberbühler, 2008; Schel et al., 2009; Zuberbühler, 2009; Rendall et al., 2009; Ouattara et al., 2009a, b, c; Casar & Zuberbühler, 2012; Candiotti et al., 2012). Often times, efforts to ‘*learn*’ meaning associated with vocalizations are aided by carefully planned playback experiments. Possible meanings emerge by stimulating the focal species into uttering context specific calls, such as those gathered by observing an interaction with a live predator or by presenting a predator model to the focal animal (Ploog et al., 1967; Gautier, 1974; Lieblich et al., 1980; Seyfarth et al., 1980; Snowdon et al., 1983; Evans, 1993b; Clarke et al., 2006). The association between call sequence and environmental event is referred to as being 'functionally referential' in animals or 'semantics' in humans (Townsend & Manser, 2013; Evans, 1993b). The main differences between the two stems from what cognitive processes are being employed. For many animals, the neural pathways involved in vocal production and comprehension are still unknown. There is also a great deal of variation that can occur within the vocal output and it is not well understood whether these differences convey meaningful changes to the message.

Beyond what cognitive processes are being used, an additional challenge in deciphering meaning involves how to parse the vocalizations gathered from playback experiments. First, determining what constitutes a unit of sound or acoustic unit in an animal's vocal repertoire requires insight into the animal's mind or the acceptance of certain assumptions. This is an issue which has been discussed in a previous paper of ours in which we presented a novel computational method to address this problem (Dassow et al., in press; also see Chapter 4). We utilize our previous approach in this chapter to address questions regarding the information content of white-handed gibbon (*Hylobates lar*) vocalizations. Second, determining the beginning and ending of acoustic sequences or 'word-like' equivalents is particularly challenging given the duration of each response. For instance, Figure 5.1 shows an initial 'word' in gibbon which includes very few of the acoustic unit's in their vocal repertoire, yet this sequence may extend for over two minutes in duration. The length of this sequence greatly exceeds any

---

[1] Words of this length do exist, but they are either not commonly used or are scientific descriptions used in chemistry. For example the chemical name of titin, a type of protein, is 189,819 letters long. This is not recognized as a word in the Oxford English Dictionary and not used in conversation. The longest recognized word in the OED is pneumonoultramicroscopicsilicovolcanokoniosis, at 45 letters long (OED online, 2014).

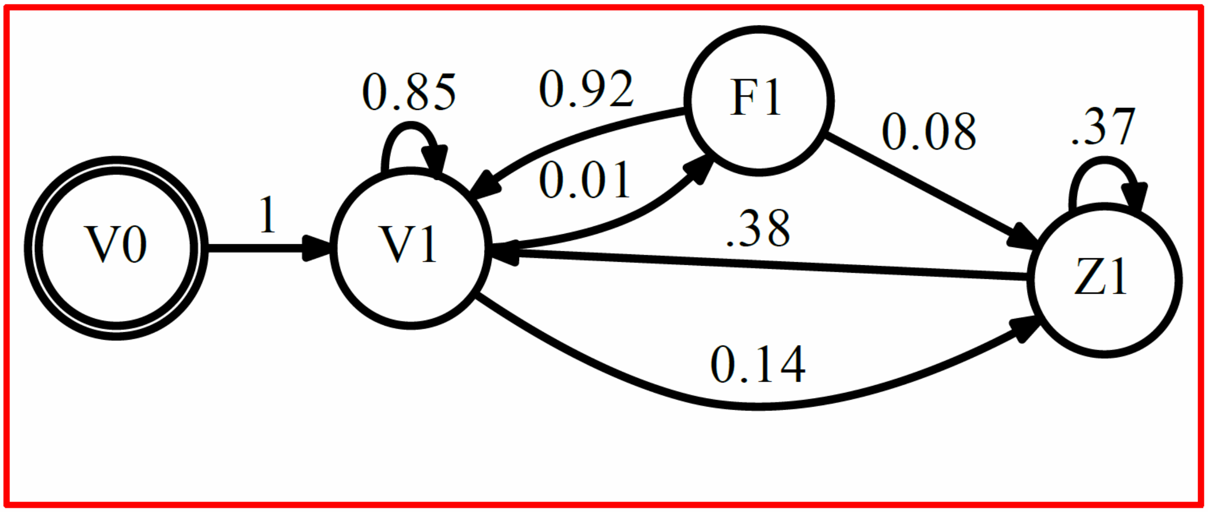


Figure 5.1: This Markov chain of a common introductory sequence to predator warning calls shows the transitional probabilities between the two commonly uttered acoustic units, /v/ and /z/, and the uncommon occurrence of the acoustic unit /f/. The duration of this sequence can exceed two minutes which suggests gibbons do not use word-like equivalents or there are temporal cues involved in their calls.

recognized, useable word in human languages[1]. An additional matter is how to treat rare occurrences of acoustic units in sequences of calls. In the case of the /f/ acoustic unit in Figure 5.1, is it representing an acceptable variation in vocal production or is it altering the meaning of the sequence? Lastly, determining whether calls are learned or innate may significantly influence how a call appears, but determining this often times require a separate line of research. Previous work has addressed questions regarding learning with some species of monkeys. For example, cross fostering studies with Japanese macaques (*Macaca fuscata*) and rhesus macaques (*Macaca mulatta*) have demonstrated a limited ability to learn calls outside of their own species' vocal repertoire (Masataka & Fujita, 1989). Other studies have shown a regression to infant calls in squirrel monkeys (*Saimiri sciureus*) (Symmes et al., 1985; Newman & MacLean, 1982) and pygmy marmosets (*Cebuella pygmaea*) (Snowdon, 1989) either from the birth of a new infant into the social group or a brain lesion which suggests at least some plasticity in vocal output. To date, these types of questions have only been addressed by speculation for lesser apes.

This chapter examines the warning calls of wild and captive white-handed gibbons (*Hylobates lar*). Previous work established a richer repertoire of acoustic unit's in gibbons. Building on this previous work, we examine the warning calls using both a general sound unit approach (Raemaekers & Raemaekers 1984; Raemaekers et al., 1984; Raemaekers & Raemaekers 1985; Clarke et al. 2006) and an acoustic unit approach (See Chapter 4). Two main questions are addressed: 1) What additional variation exists in the warning calls of wild gibbons when comparing sequences of general sound categories to acoustic unit categories? and 2) Are the warning calls of captive gibbons different than the warning calls of wild gibbons? The first question addresses the level of complexity within the utterances and the second question addresses whether learning or pathology of captivity plays a role in predator warning call production and usage. Previous research with wild agile gibbons (*Hylobates agilis*) has concluded that learning may play a role in the development of the female great call (Koda et al, 2013); however the experiment conducted by Koda and colleagues did not address the scientific goal of examining predator warning calls. The purpose for this current analysis is to explore linguistically and comparatively the types of sounds used in repeated sequences of predator warning calls.

Methods

Data collection

Wild gibbon vocalizations were gathered by Esther Clarke from April 2004 to August 2005 at the Khao Yai National Park in Thailand. Responses to predator models of clouded leopards, reticulated pythons and tigers were collected from 12 groups of gibbons. The following table shows the number of times each predator model was displayed to each group (See Clarke et al., 2006 for more details). Only responses from adult gibbons were coded and used for the current analysis.

Group	Composition	Clouded Leopard	Tiger	Reticulated Python
A	3AM, 1AF, 1JM	1	1	0
B	1AM, 1AF, 1SAM	1	1	1
C	1AM, 1AF	1	1	0
D	2AM, 1AF, 1JM, 1I?	1	1	0
H	2AM, 1AF, 1JM, 1I?	1	1	0
J	2AM, 1AF, 1JF, 1I?	1	1	0
M	1AM, 1AF	2	0	0
N	2AM, 1AF, 1JM	1	1	1
NOS	2AM, 1AF, 1J/SAM, 1I?	1	0	0
R	1AM, 1AF, 1JF	2	1	1
T	1AM, 1AF, 1JF	0	1	0
W	1AM, 1AF, 1J/SAM, 1JF	2	1	0
Total		**14**	**10**	**3**

Table 5.1: Twelve wild gibbon groups were recorded by EC after being exposed to a model of a natural predator. The vocal responses from each group were recorded. The above table indicates the number of times each group was tested with a particular predator model. Group composition consisted of one to three adult males (AM), one adult female (AF), and a variable number of offspring. Subadult animals are represented as SA, juvenile animals as J, and infant animals as I. The sex of the infant animals is unknown.

Zoo gibbon vocalizations were gathered by Angela Dassow and Michael Coen from August 2012 to September 2013 at the Racine Zoological Society and the Henry Vilas Zoo. Each zoo houses a pair of adult gibbons. The pair at the Racine Zoo is a father-daughter pair. The father is a wild born gibbon from Thailand, whereas the daughter was born in captivity. These gibbons were exposed to a live white cockatoo, and veterinarian. The gibbons at the Henry Vilas Zoo are both captive raised and they were exposed to a live white cockatoo, and veterinarian. Presentation of these stimuli occurred no more than once per week to minimize any potential stress to the gibbons. Stimuli were presented visually to the gibbons at the exterior perimeter of the gibbons' enclosure. Each encounter was under the supervision of the zoo staff.

The zoo recording equipment included two Vidpro XM-55 Condenser Shotgun Microphones connected to a four-track Zoom H4n digital recorder, sampling at 44.1kHz per channel, a Blue Yeti Pro USB Condenser Microphone, also sampling at 44.1 kHz, and a shock mount to reduce ambient noise interference. Encounters were filmed in high-definition with a

Canon VIXIA HF M500 HD image stabilized camcorder equipped with a DM-100 shotgun microphone. The microphones were distributed equidistantly around the circumferences of the gibbon enclosures. During each recording the gibbons were free to move around their enclosure and were not exposed to any toxic agents or restrained for any invasive medical procedures. This work with captive animals was approved by the University of Wisconsin-Madison's Research Animal Resource Center (RARC) and Institutional Animal Care and Use Committee (IACUC protocol number is L00452-0-08-12), Gibbon Species Survival Plan (SSP) Coordinator and Directors of the Henry Vilas Zoo and Racine Zoological Society which are both Association of Zoos and Aquariums (AZA) accredited zoos.

General vs. acoustic unit categories

Previous research has established the existence of vocal variation in predator warning calls of wild gibbons (Clarke et al., 2006). This research examined a sequence of ten general sounds produced by each gibbon after the initial sequence of 'hoo' calls were uttered from visually detecting the presence of a predator model. The recordings obtained for this previous analysis are used again, but the entire acoustic unit sequences are analyzed. Acoustic units are included from the start of vocalizing (i.e. directly after a predator model was placed on the forest floor) to the first natural break in a vocalizing bout (period of silence greater than 5 seconds in length). The acoustic unit categories for this research are described in Chapter 4 and in Dassow et al. (in prep).

Words/morphemes vs. ergodic acoustic sequences

In bioacoustics research, it is tempting to draw parallels or direct correlations to linguistic terms commonly used in human language. Terms, such as words or morphemes, have specific properties which define each. For example, morphemes are the smallest unit of grammar in a language (Payne, 1997). The word "unbelievable" contains three morphemes, “un”, “believe” and “able”. These morphemes cannot be further subdivided without losing their meaning. The rules which govern these divisions require knowledge of the linguistic context that is not present, or at best assumed, in non-human animals.

In lieu of using such poorly fit terms to describe the acoustic sequences of gibbon calls, we adopt the use of ergodic theory to locate potential boundaries in sequences of acoustic units. Ergodic theory examines one-way state transitions throughout time (Cogburn, 1984). In the case of gibbon vocalizations, ergodic theory can be used to determine the likelihood of a given acoustic sequence ending and another sequence beginning. In doing so, we identify subsequences of acoustic units which are likely to have meaning in the gibbon’s repertoire. It is important to note that these subsequences are significantly longer than a word in human language and the two should not be misinterpreted as being equivalent. As described in further detail in the discussion section, these subsequences likely contain additional subdivisions; however more data will be necessary to identify any other structure.

Markov models

Variations of Markov models exist for numerous problems. Hidden Markov Models (HMMs) have previously been used in bioacoustics research (Harris & Skowronski, 2006; Brandes, 2008; Ren et al., 2009; Brown & Smaragdis, 2009; Pace et al., 2012; Ji et al., 2013) to examine properties of vocalizations. Markov decision processes (MDPs) and partially observed Markov decision processes (POMDPs) are useful for problems involving a choice of optimal actions (Kaebling et al., 1998). Markov chains are used when describing a sequence of linked events. Vocal sequences are not necessarily optimal, so MDPs and POMDPs are not the best model to use for analysis. While useful for other types of bioacoustics research, HMMs are also not the appropriate model to use for sequence analysis since the entire string of acoustic units could be identified in each of the warning calls. As such, Markov chains offer the best approach for examining vocal sequences. Markov chains are calculated using ergodic theory to determine what natural breaks in the sequence occurs (Cogburn, 1984). A natural break is defined as a point in the sequence which has a low transitional probability in a single, forward direction. By calculating the transitional probabilities and building Markov chains, the subsequences representing meaningful differences in gibbon calls is examined.

Identifying subsequences from wild gibbons using a Markov model

The complexity of predator warning calls in gibbons is examined here using an acoustic unit approach on the entirety of the calling bout. This differs from a previous analysis (Clarke et al., 2006) by including significantly longer portions of the calling bout and also by utilizing the vast amount of variation which exists within each of the general sound categories. Markov chains are used to compare the sequences of acoustic units uttered in each call (Cogburn, 1984). The context categories for wild gibbons include 3 predators (clouded leopard, tiger & reticulated python). The calls from the adult male and adult female are used for analysis. Though the sub adult and juveniles also call, they were not utilized in this analysis due to their infrequent and partial responses to the predator models.

Acoustic sequences are identified by labeling the acoustic unit's uttered by the adult male and adult female of each group. Identification of females and males are made based upon tracking unique vocal signatures from the female great call and male coda. Similarities in patterns are collapsed into single pathways and differences are reflected by including multiple possibilities for a given transition state. Each state may have a variable number of options or transitions that can occur to take it to the next state. These differences and similarities are accumulated to track the total number of possibilities that can occur for any given context. To carry this out more efficiently, the sequence data for each adult gibbon were input into text files and run through a Markov chain model. The summary states leftover from the Markov reduction are presented as the potential vocal patterns that could carry possible meaning. The probability of moving from state to state is calculated for each transition. Subsequences were identified between acoustic units' which had low transitional probabilities moving in a single, forward direction. For example, Figure 5.3 shows three subsequences outlined with a red box. Between each box exists transitions that rarely occur such as the transition between V1 and E2 at a probability of 0.005. Within each box exists transitions that occur at high probabilities such as the transition b V1 and Z1 at a probability of 0.55.

An example of the summary states for clouded leopard responses can be found in Figure 5.2. This example demonstrates the variation which occurs within a clouded leopard warning under the general sound category approach used by Clarke et al. (2006). Alternative summary states, in Figure 5.3, are developed based on the acoustic unit analysis. These vary drastically from those shown in Figure 5.2 by incorporating the acoustic unit analysis of the calls as well as a lengthier subset of the calls. In other words, previous work was based upon a total of 10 notes (See Table 5.2) whereas the modified Markov chains include the first several hundred acoustic units'. By incorporating more units', subtle variation which occurs within these calls is established.

As observed in Figure 5.3, this additional variation also leads to a more complex view of existing transitions. To evaluate these transitions further, male responses are separated from female responses. Male responses have less variation in their basic call structure and thus provide potentially valuable insight into what core structure is needed for producing a successful warning call. Male calls are then compared to female calls to identify what variations in sequences exist. A future playback study will be needed to confirm whether each sequence variation contains a meaningful difference.

Comparing subsequences used in the wild to subsequences used in captivity

Environmental factors that influence vocal bouts vary drastically between wild and captive gibbons. We explore the warning calls of two pairs of captive gibbons to determine whether gibbons are capable of producing novel vocalizations for cues which are not part of their native home range. Markov chains are built following the methods outlined in the previous section.

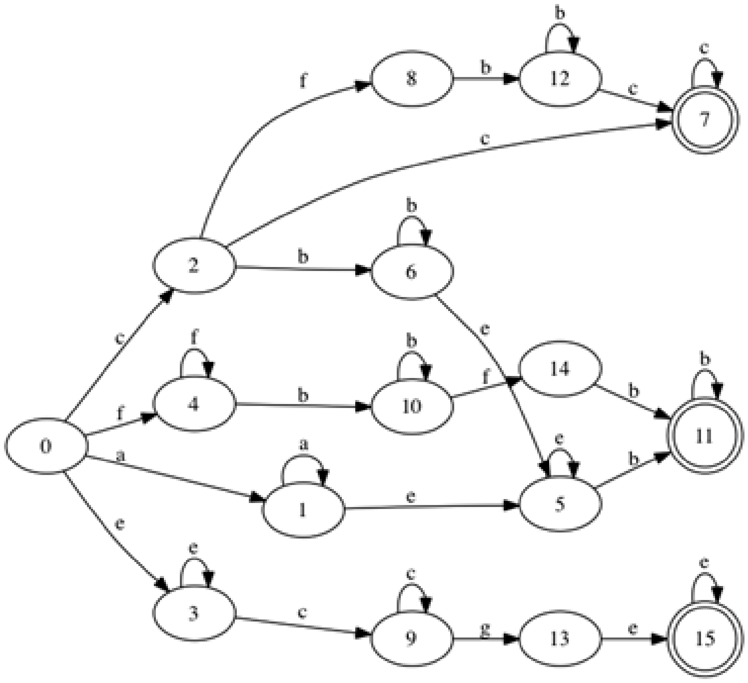


Figure 5.2: The Markov chain depicted here shows one view of the different transitions found in the clouded leopard responses of wild adult gibbons (Figure provided by Eric Raimy). The classifications of notes used for this Markov chain are described by Clarke et al., 2006 and presented in Table 5.2.

Table 5.2: A total of 10 notes were sequenced using the 6-7 gross categories of sound to build the Markov chain shown in Figure 5.2.

Group	Sex	Notes
B	F	cccccccbbe
	M	aaeeeebbbb
C	F	fffdffdccb
	M	cccccccccc
D	F	fbbbbbbbbb
	M	fffbbbfbbb
H	F	acccdbeeee
	M	eebeecefec
J	F	ccccbcccbe
	M	eeeeeccgee
N	F	ccfbbdbcff
	M	egeeeeeeec
R	M	cfbbbbbccc
W	F	ccccceecgg
	M	cbbbeeebbb

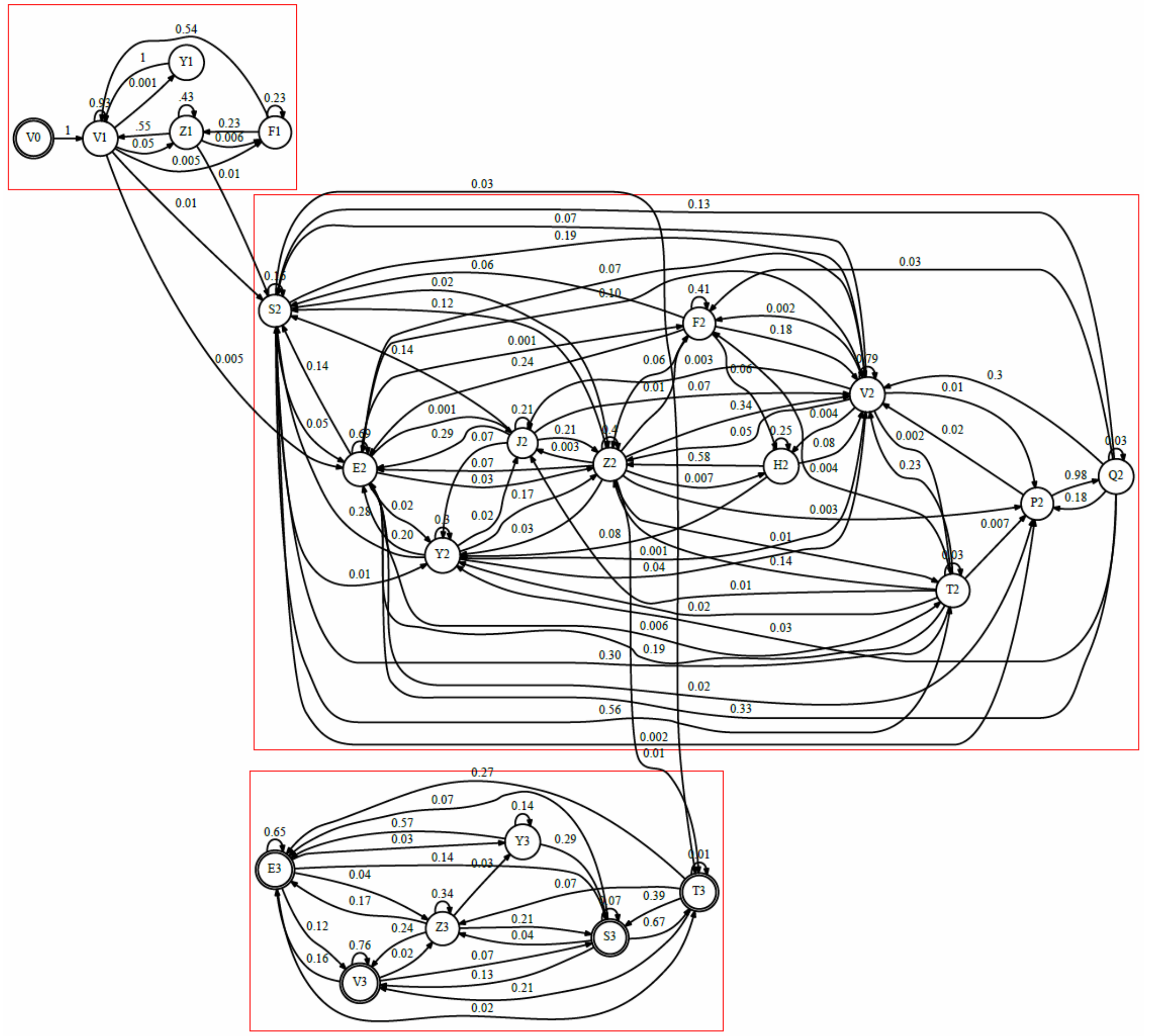


Figure 5.3: The Markov chain depicted here shows an alternative view of the clouded leopard response shown in Figure 5.2. In this figure, sequences of acoustic units are used. There is preliminary evidence to suggest that gibbon's use acoustic unit's to communicate, rather than gross sound categories.

Results

Markov chain comparison of gross categories vs. acoustic unit's

Additional variation in the predator responses are demonstrated by using an acoustic unit approach. Figure 5.3 illustrates this complexity by revealing additional groups of vocal sequences in the clouded leopard warning responses. Introductory and concluding sequences are identified, as well as an extended sequence in the main section of the call. Sequences of acoustic unit's from the snake response and tiger response also reveal further complexity in length and types of acoustic unit's used to respond to the predator model (See Figure 5.5 and 5.6-snake; 5.7 to 5.11-tiger). Beyond the variation in length and complexity of the calls, variations in male and female responses are identified.

Clouded leopard response

Acoustic unit's used in the clouded leopard response include /d/, /e/, /f/, /g/, /h/, /i/, /j/, /o/, /p/, /q/, /s/, /t/, /v/, /y/, and /z/ (Figure 5.3). A total of 6,421 units were identified across 28 adult gibbons. Acoustic unit /v/ occurred the most frequently at 3,828 times (Table 5.3; See Appendix A for a breakdown of acoustic unit's/individual). Some of the unit's rarely occurred in the male's response and were removed from the analysis of the basic call structure. Unit's /d/, /o/, /p/, /q/, and /y/ only occurred once or in one individual's response. Additional field work is necessary to determine whether these units alter the meaning of this warning response.

The basic structure of the clouded leopard response is given by the males in Figure 5.4 and includes a total of 2,551 acoustic units. Two transitional probabilities between subsequence 1 and subsequence 2 occurred. A probability of 0.007 occurs between acoustic unit V1 and S2. A probability of 0.006 occurs between acoustic units V1 and E2. The transitional probability between subsequence 2 and subsequence 3 is between acoustic units T2 and E3 at 0.14.

Acoustic unit	Total number of occurrences	Male responses	Female responses
v	3828	1735	2093
e	999	346	653
s	545	126	418
z	496	247	249
t	342	90	252
y*	63	3	60
p*	40	1	39
q*	40	1	39
f	30	0	30
j	14	0	14
h	12	0	12
i	5	0	5
o*	5	1	4
d*	1	1	0
g	1	0	1
Total	6421	2551	3869

Table 5.3: A total of 6,421 acoustic units were found in the clouded leopard response of adult male and female gibbons. Some of these units (e.g. /v/) occurred more frequently than others (e.g. /g/). Acoustic unit's /p/, /q/, /o/, and /d/ only occurred once each in the male gibbon's response and were excluded from the analysis of the basic call structure shown in Figure 5.3. Additionally, all three samples of acoustic unit /y/ were from a single male's response and were also excluded. Male N used unit /d/, male M used /p/ and /q/, and male C used /o/ and /y/.

Clouded leopard responses varied in their ending with acoustic unit's /e/, /t/, /v/ and /z/ as possibilities. Table 5.4 summarizes the total number of occurrences per subsequence of each acoustic unit in the male's response. Females produced this basic response; however they also altered the response to produce numerous variations. This varied response is illustrated in Figure 5.3 and includes far greater complexity than the basic structure of the clouded leopard warning call. Acoustic units used by the females, but not the males include /f/, /g/, /h/, /i/, and /j/. Additionally, a majority of the /o/'s, /p/'s, /q/'s, and /y/'s occurred in the female's responses.

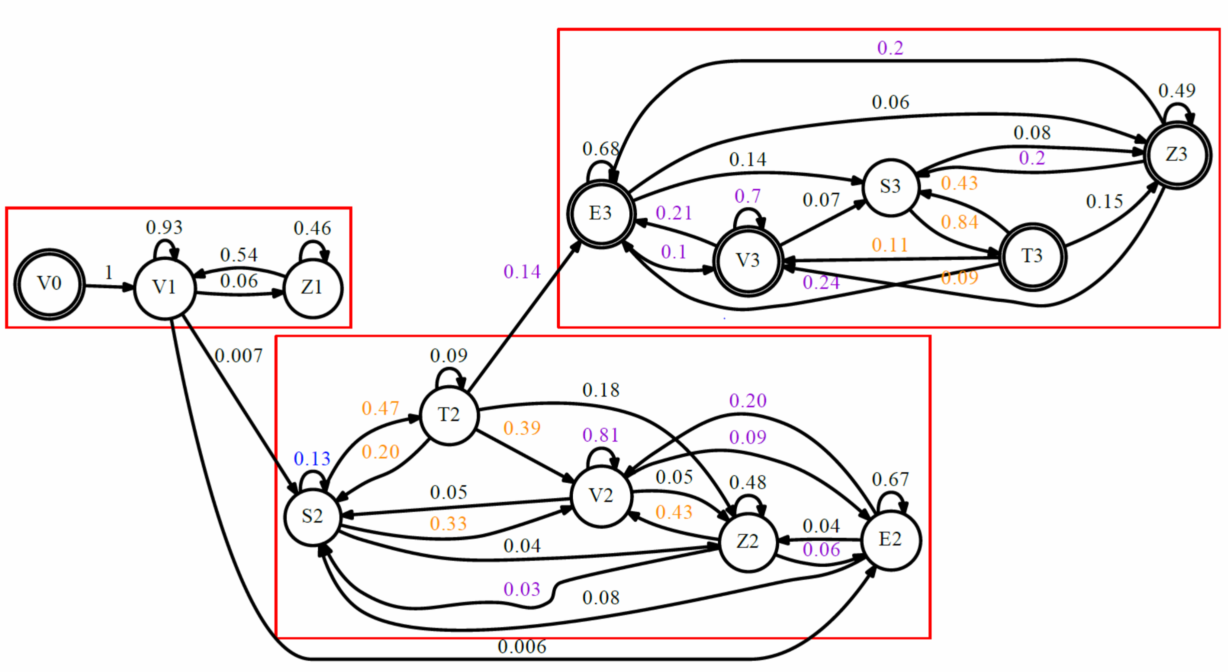


Figure 5.4: A basic clouded leopard response given by male gibbons. Females also produce this response; however they incorporate additional acoustic units into the sequences. Blue values indicate a transition which only occurs in one subsequence. Violet values indicate a 10-19% difference in transitional probabilities between subsequence 2 and subsequence 3. Orange values indicate a 20% or greater difference in transitional probabilities between subsequence 2 and subsequence 3. Black values indicate transitional probabilities between subsequence 2 and subsequence 3 differ by less than 10%.

Acoustic Unit	Subsequence 1	Subsequence 2	Subsequence 3	Total
v	1002	663	70	1735
z	116	96	35	247
e	0	226	126	352
s	0	76	50	126
t	0	44	46	90

Table 5.4: The clouded leopard response given by male gibbons consisted of five acoustic units. Subsequence 1 contained acoustic unit's /v/ and /z/. Subsequences 2 and 3 consisted of acoustic unit's /v/, /z/, /e/, /s/, and /t/.

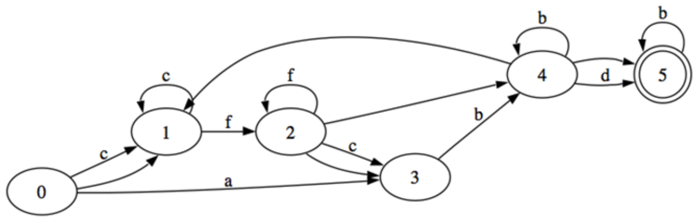
Snake model response

The snake model response uses the acoustic unit's /d/, /e/, /f/, /g/, /h/, /i/, /p/, /q/, /s/, /t/, /v/, /y/ and /z/. A total of 1,899 acoustic units were identified across six adult gibbons. Group N's response consisted of two adult males, but no adult female. Both male responses are included in this analysis. Unit /v/ occurred the most often at 1,501 times (Table 5.5; See Appendix B for a breakdown of acoustic unit's/individual). Similar to the clouded leopard response, some of the acoustic unit's only occurred once or in a single gibbon's response, which led to an overly complex view of the combined sequences (Figure 5.5). Unit's /f/, /g/, /h/, /i/, /p/, /q/, and /y/ only occurred in the females responses. Additionally, a majority of the /s/ and /t/'s occurred in the female's response. Unit's /d/, and /s/ only occurred once and unit /t/ only occurred twice in male R's response.

Acoustic unit	Total number of occurrences	Male responses	Female responses
v	1501	897	604
z	199	140	59
e	98	15	83
s*	24	1	23
t*	21	2	19
f	19	0	19
h	15	0	15
i	11	0	11
q	3	0	3
y	3	0	3
g	2	0	2
p	2	0	2
d*	1	1	0
Total	1899	1056	843

Table 5.5: A total of 1,899 acoustic units were identified in the responses of adult male and female gibbons to a snake model. In the male gibbon's responses, acoustic unit's /d/, /s/ and /t/ only occurred from male R and were excluded from the final analysis shown in Figure 5.5.

A



Group	Sex	Notes
B	F	cccfcbccfd
	M	fbbbbbbbbb
N	M	fbbbbbbbbb
R	F	abfbbbbbbb
	M	ffbdbbbbbb

B

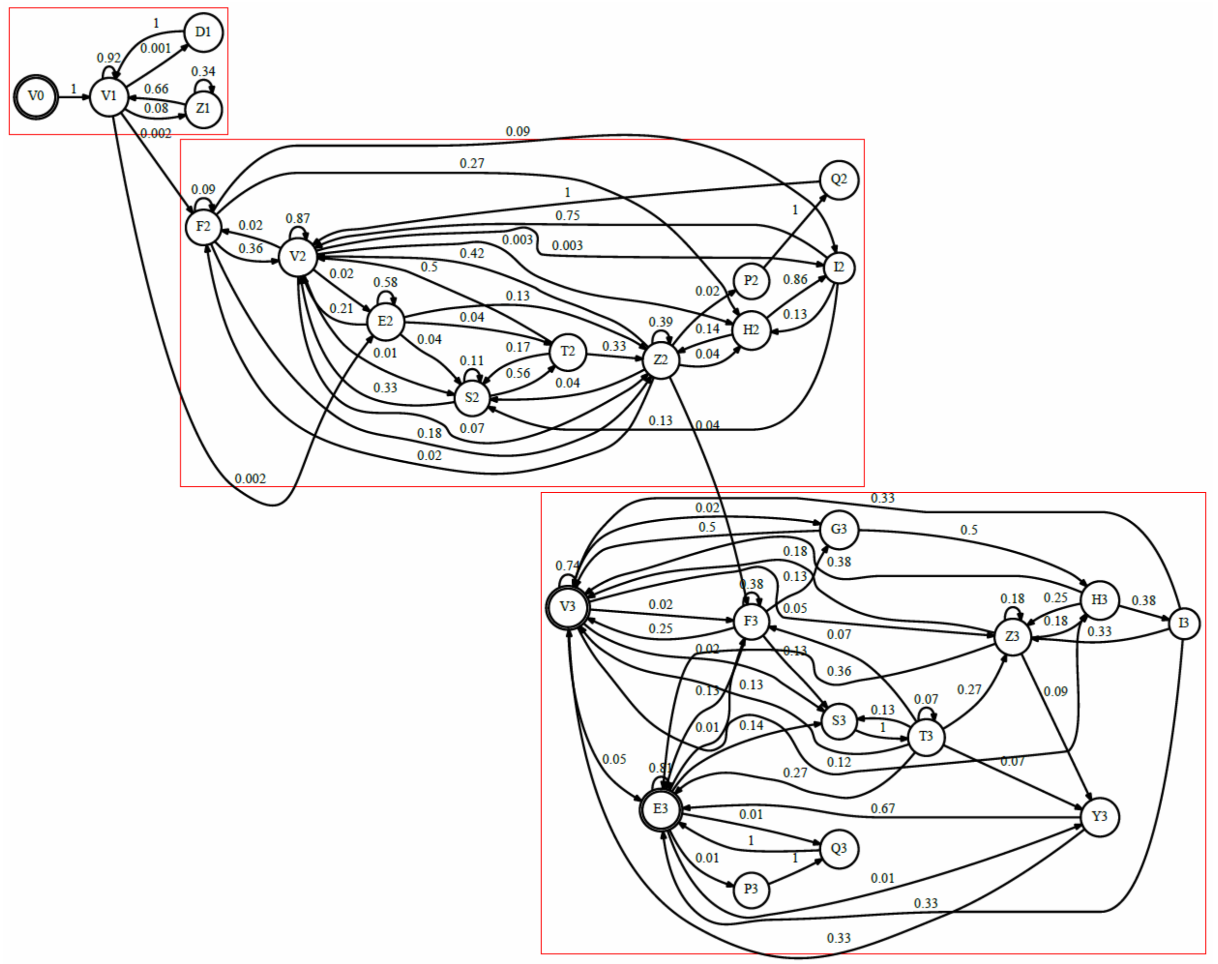


Figure 5.5 – The Markov chain depicted in panel A shows one view of the different transitions found in the snake responses of wild adult gibbons (Figure provided by Eric Raimy). The classifications of notes used for this Markov chain are described by Clarke et al., 2006 and presented in the adjacent Table. Panel B shows the same response, but uses sequences of acoustic unit's. This response appears to be significantly more complex than previously thought and includes the same general warning sequence observed in the clouded leopard response.

Unlike the clouded leopard response, the snake warning call was restricted to two subsequences for males and three subsequences for females. The male's response included a transitional probability of 0.003 between acoustic unit's V1 and E2 (Figure 5.6). This transition also occurs in the female's response; however another transition between subsequence 1 and subsequence 2 existed between V1 and F2 (Figure 5.5). This pathway has a transitional probability of 0.002. There is a transitional probability of .04 between Z2 and F3 in the female response. The basic snake call differs from the clouded leopard response by only using acoustic unit's /e/, /v/ and /z/ across two subsequences (Table 5.6). Transitional probabilities between subsequences are also different with respect to units used.

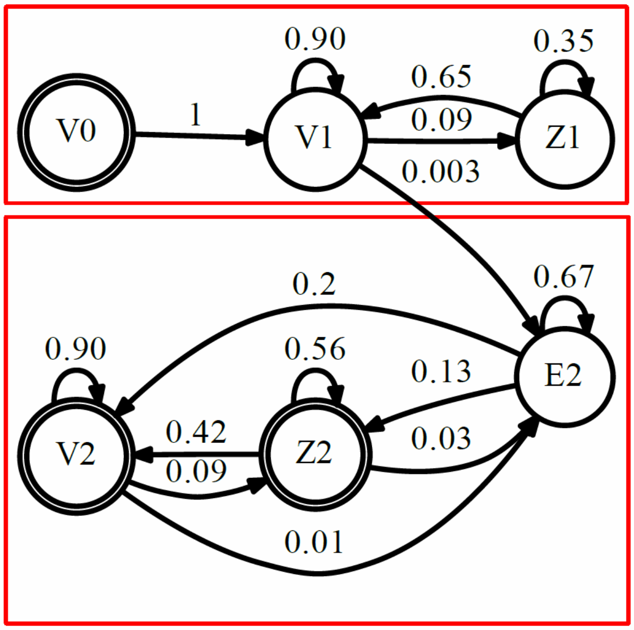


Figure 5.6: A basic snake response given by male gibbons consists of a single transition between two subsequences. Female gibbons produce a variant of this response by including a greater diversity of acoustic units in the second subsequence as well as include a third subsequence.

Acoustic Unit	Subsequence 1	Subsequence 2	Total
v	734	163	897
z	104	36	140
e	0	15	15

Table 5.6: The snake response given by male gibbons consisted of three acoustic units. Subsequence 1 contained acoustic unit's /v/ and /z/. Subsequences 2 consisted of acoustic unit's /v/, /z/, and/e/.

Tiger model response

A total of 4,940 acoustic units were identified across ten pairs of adult gibbons. Acoustic unit's used in this response include /d/, /e/, /f/, /g/, /h/, /i/, /j/, /p/, /q/, /s/, /t/, /v/, /y/, /z/, and /aa/. As shown in Table 5.7, acoustic unit /v/ appeared the most frequently and acoustic unit's /i/, /aa/, and /g/ appeared the least frequently. Analysis of these units was split amongst the groups based upon whether the groups were presented with a stationary tiger model or a running tiger model. Groups A, C, H, R and W responded to a stationary tiger model. Groups B, D, J, N and T responded to a running tiger model.

Acoustic unit	Total number of occurrences	Sitting tiger male responses	Sitting tiger female responses	Running tiger male responses	Running tiger female responses
v	4156	630	922	1707	897
z	278	91	39	101	47
s	216	22	111	19	64
e	111	22	52	15	22
t	98	21	46	10	21
p	31	1	26	0	4
q	28	1	23	0	4
f*	6	0	2	0	4
j*	4	0	2	0	2
h*	3	0	0	0	3
d*	3	0	0	0	3
y*	3	0	1	0	2
i*	1	0	0	0	1
aa*	1	0	1	0	0
g*	1	0	0	0	1
Total	4940	788	1225	1852	1075

Table 5.7: A total of 4,940 acoustic units were found in the tiger responses of adult male and female gibbons. These responses were further subdivided into *sitting* tiger responses and *running* tiger responses. *These acoustic units appeared in the original sequences, but they were removed from the final analysis due to their rare occurrences.

Stationary tiger model response

The total number of acoustic units identified from the stationary model response was 2,013 (Table 5.7; See Appendix C for a breakdown of acoustic unit's/individual). There were four acoustic unit's (/h/, /i/, /d/, and /g/), which did not occur in the stationary tiger response, but are found in the running tiger response. Unit's /aa/ and /y/ only occurred once from the females in group A and W respectively. Unit /f/ only occurred from one female, female A. Unit /j/ only occurred one time each from female A and female R. Given the rarity of these units, they were excluded from the sequence analysis.

Figure 5.7 shows the ergodic Markov chain of the acoustic unit's given by males. The transitions between subsequence 1 and 2 include V1 - S2 (.007) and V1 – P2 (.002). The transition between subsequences 2-3 includes E2 – S3 (.05). The total numbers of acoustic units per subsequence are shown in Table 5.8.

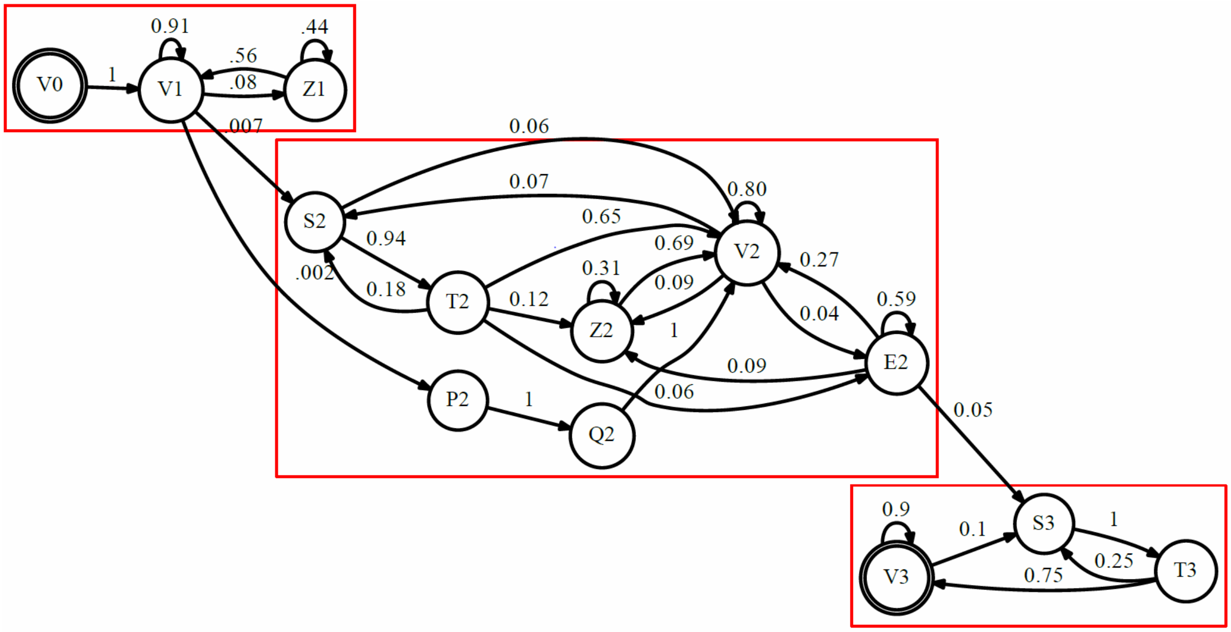


Figure 5.7: The above ergodic Markov chain represents the sequential analysis of responses given by male gibbons to a stationary tiger model. This analysis excludes the rare occurring unit's /f/, /j/, /y/ and /aa/ which only occurred in the female's response.

Acoustic Unit	Subsequence 1	Subsequence 2	Subsequence 3	Total
v	429	181	20	630
z	62	29	0	91
e	0	22	0	22
s	0	18	4	22
t	0	17	4	21
p	0	1	0	1
q	0	1	0	1

Table 5.8: The stationary tiger model response given by male gibbons consisted of seven acoustic units. Subsequence 1 contained acoustic unit's /v/ and /z/. Subsequence 2 consisted of acoustic unit's /v/, /z/, /e/, /s/, /t/, /p/, and /q/. Subsequence 3 consisted of acoustic unit's /v/, /s/, and /t/. Though acoustic unit's /p/ and /q/ only occurred once, they were included in this analysis because unit /p/ formed part of a transition between subsequences 1 and 2. This transition also occurred in the female's response.

Transitions which occurred in the sitting tiger response, but not the running tiger response include the subsequence 2-subsequence 3 transition of E2-S3, and all of the subsequence 3 transitions. Transitions in the sitting tiger response which occurred in males, but not females include: T2-S2, T2-Z2, and T2-E2. The transition between S2-T2 occurred more frequently in males than females with a probability of 0.94 for males and 0.14 for females. The transition between S2-V2 occurred less frequently for males than females with a probability of 0.06 for males and 0.71 for females.

Running tiger model response

A total of 2,933 acoustic units were identified in the running tiger responses (Table 5.7; See Appendix D for a breakdown of acoustic unit's/individual). Similar to the stationary tiger response, some acoustic unit's in the running tiger response occurred rarely or only from a single gibbon. Acoustic unit's /f/, /h/, /d/, /j/, /i/, and /g/ only occurred in the response from the female of group B. Acoustic unit /y/ only occurred in the response given by the female in group D. Unit's /p/ and /q/ only occurred four times; however they were included in the analysis because they appeared in 3 of the 5 female responses.

Figure 5.8 shows the ergodic Markov chain of the male's response to the running tiger model. Unlike the stationary tiger response, this sequence is restricted to 2 subsequences, rather than 3. The transitional probabilities between subsequences 1-2 include V1-S2 (.002) and Z1-S2 (.01). Table 5.9 shows a breakdown of the number of acoustic units per subsequence in the male gibbon's response. Transitions which occurred in the running tiger response, but not the sitting tiger response include S2-S2, V2-T2 and S2-Z2. Transitions which occurred in the male's response, but not the female's response to the running tiger include Z1-Z1 and V2-T2.

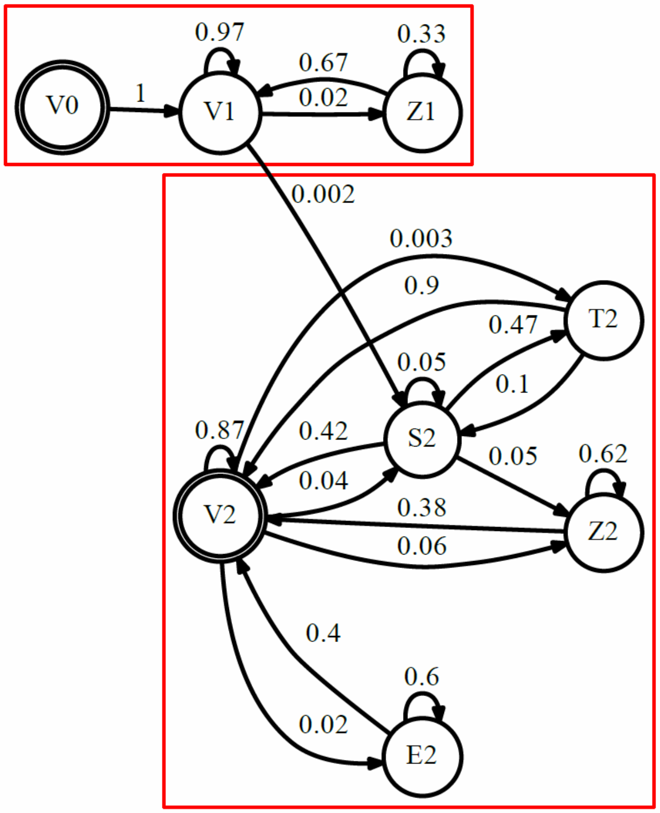


Figure 5.8: The above ergodic Markov chain represents the sequential analysis of responses given by male gibbons to a running tiger model. This analysis excludes the rare occurring unit's /f/, /h/, /d/, /j/, /y/, /i/, and /g/.

Acoustic Unit	Subsequence 1	Subsequence 2	Total
v	1395	312	1707
z	49	52	101
e	0	15	15
s	0	19	19
t	0	10	10

Table 5.9: The running tiger response given by male gibbons consisted of five acoustic units'. Subsequence 1 contained acoustic unit's /v/ and /z/. Subsequence 2 consisted of acoustic unit's /v/, /z/, /e/, /s/, and /t/.

Comparison of male responses to running tiger model vs. sitting tiger model

Figure 5.9 shows a comparison between the male gibbons' response to the running and sitting tiger (Shown separately in Figures 5.7 and 5.8). Each transition has two probabilities shown. The first probability is from the running tiger response and the second probability is from the sitting tiger response. Blue values indicate the transition only occurred in one of the groups. For example, the S2-S2 transitions is labeled "0.05/NA", indicating this transition occurred in the running tiger responses with a probability of 0.05, but the transition did not occur in the sitting tiger response. Violet values indicate a 10-19% difference in transitional probabilities between groups. For example, the Z1-Z1 transition has a 0.33 transitional probability for the running tiger response, but a 0.44 transitional probability for the sitting tiger response. Orange values indicate a 20% or greater difference between transitional probabilities. Black values indicate a less than 10% difference in transitional probabilities

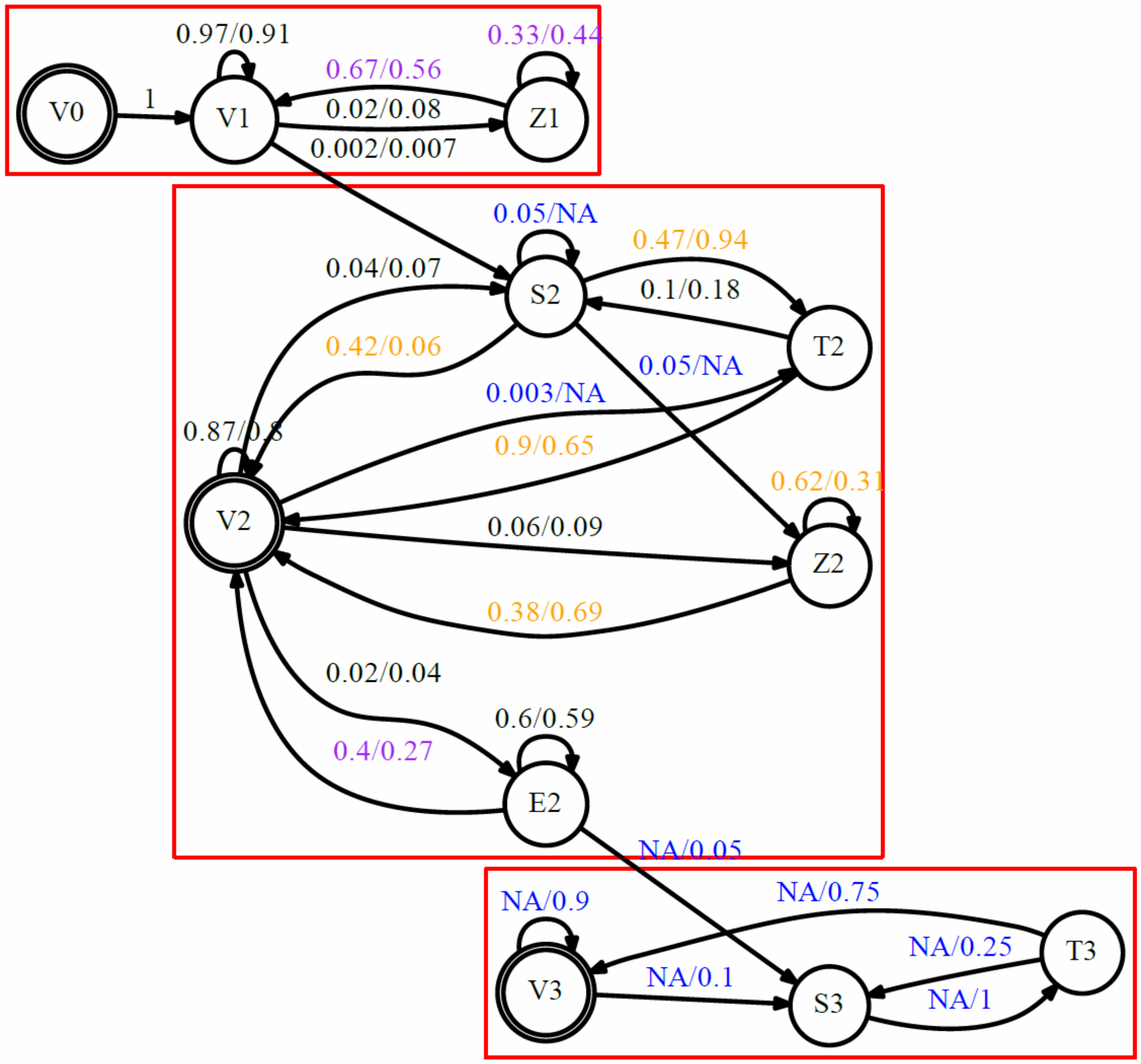


Figure 5.9: The above Ergodic Markov chain represents the comparison of the sequential analysis of sitting vs. running tiger responses from male gibbons. Blue values indicate a transition which only occurs in one group's response. Violet values indicate a 10-19% difference in transitional probabilities between sitting and running responses. Orange values indicate a 20% or greater difference in transitional probabilities between sitting and running. Black values indicate transitional probabilities between sitting and running responses differ by less than 10%.

Female responses to a running tiger model

Two additional analyses of the running tiger model response are shown in Figure 5.10 and Figure 5.11. The response in Figure 5.10 shows only the data from female B which consisted of 389 acoustic units'. Several acoustic units appeared in female B's response which was not found in any other gibbon's response to a running tiger model. These include acoustic unit's /f/, /h/, /d/, /j/, /i/, and /g/. A breakdown of acoustic unit occurrences per subsequence is given in Table 5.10. Most of the acoustic unit's excluded from the running tiger response in Figure 5.8 occurred from female B and is included in this analysis. Figure 5.11 shows the response of the four other females to the running tiger model and includes 685 acoustic units. A breakdown of acoustic unit occurrences per subsequence is given in Table 5.11.

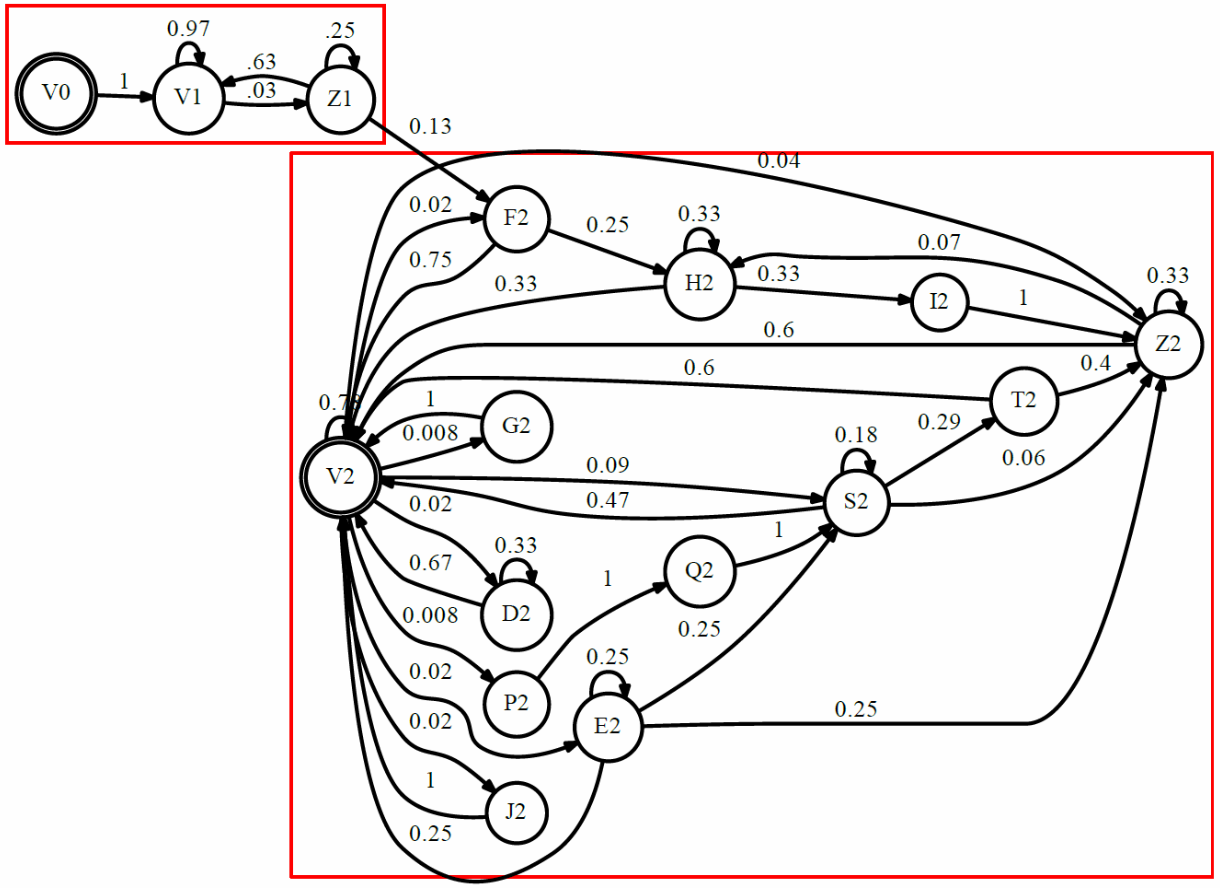


Figure 5.10: The above ergodic Markov chain represents the sequential analysis of female B's response to a running tiger model. This analysis includes the rare occurring unit's /f/, /h/, /d/, /j/, /i/, and /g/. This response was isolated due to its' distinction from other females responses in acoustic units present as well as the order of acoustic units from the rest of the females.

Acoustic Unit	Subsequence 1	Subsequence 2	Total
v	191	133	324
z	8	15	23
s	0	17	17
e	0	4	4
t	0	5	5
p	0	1	1
q	0	1	1
f*	0	4	4
h*	0	3	3
d*	0	3	3
j*	0	2	2
i*	0	1	1
g*	0	1	1

Table 5.10 – The running tiger response given by female B consisted of thirteen acoustic units'. Subsequence 1 contained acoustic unit's /v/ and /z/. Subsequence 2 consisted of all thirteen acoustic units'. *Acoustic unit's /f/, /h/, /d/, /j/, /i/, and /g/ were unique to this gibbon's response.

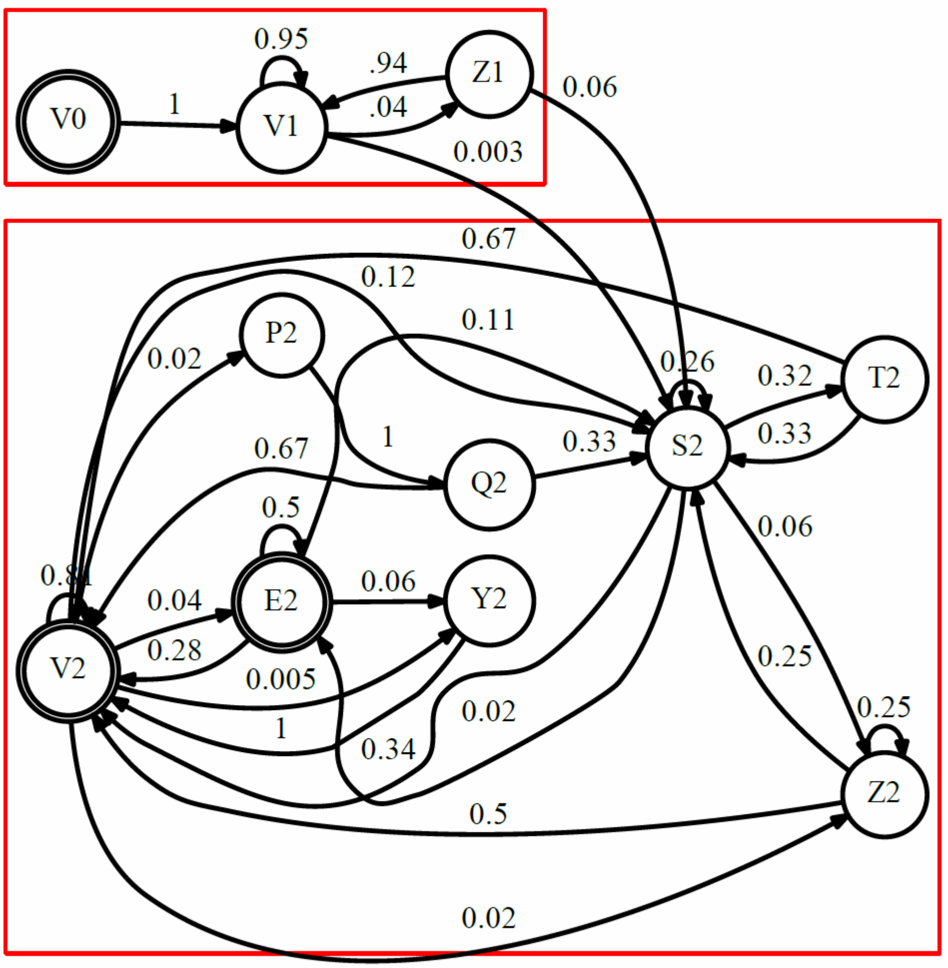


Figure 5.11: The above ergodic Markov chain represents the sequential analysis of females D, J, N and T's response to a running tiger model. This analysis does not include the rare occurring unit's /f/, /h/, /d/, /j/, /i/, and /g/ which are part of female B's response to the running tiger.

Acoustic Unit	Subsequence 1	Subsequence 2	Total
v	373	200	573
s	0	47	47
t	0	15	15
p	0	3	3
q	0	3	3
e	0	18	18
y*	0	2	2

Table 5.11 – The running tiger response given by females D, J, N, T consisted of seven acoustic units'. Subsequence 1 contained acoustic unit's /v/ and /z/. Subsequence 2 consisted of all seven acoustic units'. *Acoustic unit /y/ was unique to female D's response.

Wild vs. captive warning calls using acoustic unit sequences

What is perceived as a threat to a gibbon can vary between habitats. Predators, such as clouded leopards, are a constant threat to wild gibbons, but they pose little to no threat to captive gibbons. Similarly, perceived predators (e.g. veterinarians) pose a greater threat to captive gibbons than wild gibbons. This variation in social encounters provides a useful comparison for examining how warning calls are used and developed. Acoustic unit sequences of captive gibbon warning calls are examined below for their similarity to wild gibbon warning calls.

Henry Vilas Zoo

Veterinarian response

Markov chain representation of the Vilas gibbons' response to the presence of their veterinarian is shown in Figure 5.12. A total of 170 acoustic units were identified in this response. This response is limited to the acoustic unit's /v/ (n= 145) and /z/ (n=25), which is a common start to all predator responses. Potential reasons for the lack of a more diverse or prolonged response is provided in the discussion section. There are no transitional probabilities out of this subsequence.

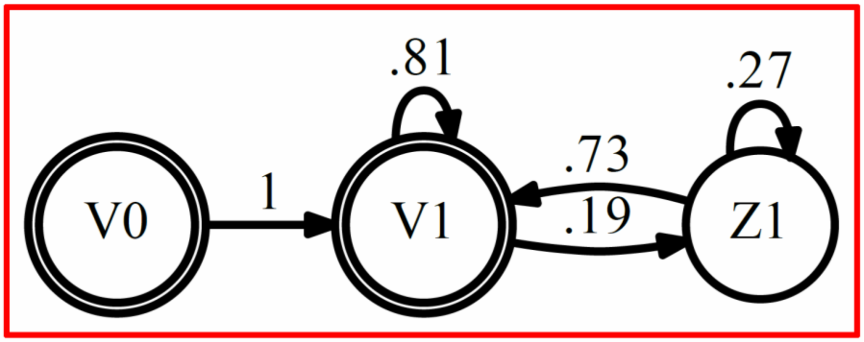


Figure 5.12: The Vilas gibbons' response to the presence of their veterinarian was limited to a general alarm response which contains the /v/ and /z/ acoustic unit's. The lack of a prolonged and more diverse response is likely due to several years of passive encounters with their veterinarian. This response suggests they view the presence of their veterinarian as someone to be wary of, but not necessarily as a direct threat.

Cockatoo response

The gibbons response to the white cockatoo (*Cacatua alba)* is the most complex of the stimuli tested. A total of 434 acoustic units were identified in this response. As shown in Figure 5.13, this response contains an acoustic unit, /bb/, not found in the predator calls used by wild gibbons. Other acoustic unit's used in this response include /e/, /s/, /t/, /v/, and /z/ (Table 5.12). Responses begin with a /z/ acoustic unit, rather than a /v/, which is unique to this vocal bout.

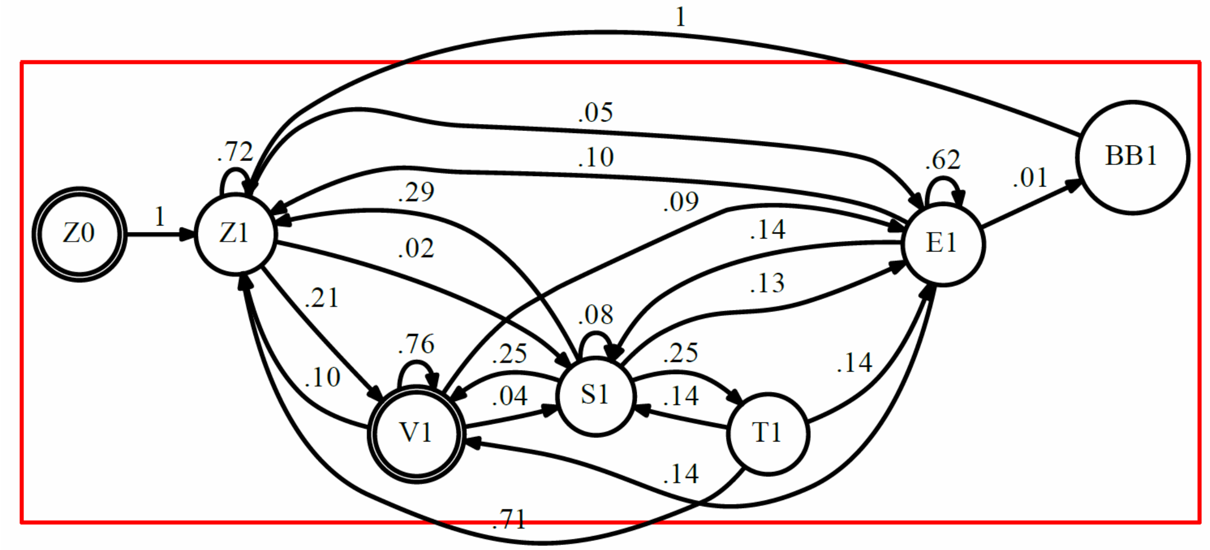


Figure 5.13: Unlike the rest of the warning calls examined, the gibbon's response to the cockatoo begins with a series of /z/ acoustic unit's. This call is given whenever the cockatoo was within visual or auditory range of the gibbons'.

Acoustic Unit	Total number of occurrences	Male response	Female response
v	184	80	104
z	145	75	70
e	73	24	49
s	24	8	16
t	7	3	4
bb	1	1	0
Total	434	191	243

Table 5.12: The Vilas gibbon's response to the cockatoo consisted of six acoustic units'. There were no transitions out of this response.

Racine Zoological Society

Veterinarian response

A response to the presence of the veterinarian is shown in Figure 5.14. This response is a combination of the general warning call given during all predator responses coupled with a female great call, but no male coda. A total of 252 acoustic units were identified in this response. Acoustic units used in this response include /a/, /e/, /f/, /g/, /h/, /i/, /j/, /k/, /l/, /m/, /v/, and /z/ (Table 5.13). This response contained two subsequences with a transitional probability of .005 between V1 and F2.

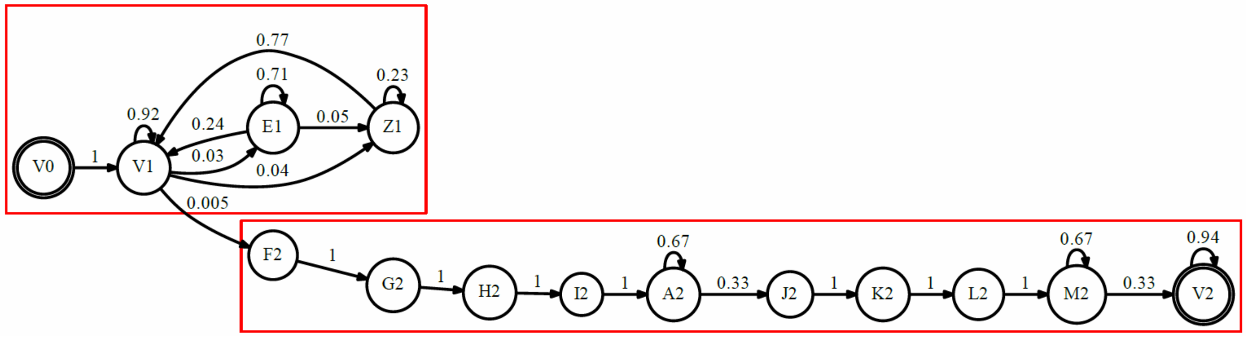


Figure 5.14: The Racine gibbons response to their veterinarian begins very similar to the Vilas gibbon's response. An additional and more varied response follows which was not observed in the Vilas pair. This response was gathered as the veterinarian was restraining the female gibbon for a routine check-up.

Acoustic Unit	Total number of occurrences	Male response	Female response
v	206	115	91
e	21	1	20
z	13	5	8
a	3	0	3
m	2	0	2
f	1	0	1
g	1	0	1
h	1	0	1
i	1	0	1
j	1	0	1
k	1	0	1
l	1	0	1
Total	252	121	131

Table 5.13: The Racine gibbon's response to their veterinarian consisted of twelve acoustic units'. There was a single transition between V1 and F2 given by the female.

Cockatoo response

Both gibbons responded to the cockatoo with a visual examination, but gave no vocal utterances. The gibbons moved across their enclosure towards the cockatoo and visually inspected it before moving back to their previous location. The male spent 14 seconds visually examining the cockatoo and the female spent 52 seconds.

Discussion

Previous work has examined predator responses of wild white-handed gibbons using the ten notes following the introductory sequence of calls. The current approach has revealed previously undetected differences in these calls by expanding the analyses to include all acoustic units uttered after the presentation of a predator model. The variation revealed by the current analysis suggests the communication system of white-handed gibbons is more complex than previously thought. A motivational, general warning is present at the beginning of all predator calls as a /v/-/z/ complex. This may function as a general alertness to all nearby gibbons. **A functionally referential, predator specific variation is present in the middle of each call.** Similarly, a referential sequence is found at the end of calls. This may function as a calming signal to reduce alertness and vigilance. The presence of both a functionally referential and motivational portion of a response has been observed in other primates (Marler et al., 1992; Blumstein, 1999), but this is the first evidence for it in gibbons.

The transitional probabilities between each group of acoustic unit sequences are very low. **It is currently unclear why the length of these sequences and the amount of repetition within them is present.** It is possible that these warning calls extend over several minutes because the predator models were present for the same duration. Additionally, a great deal of variation occurs within the first subsequence with respect to the number of /z/ acoustic unit's present. Variation in the occurrence of the /z/ acoustic unit at the beginning of the alarm calls may provide both referential and motivational information. The existence of subtle variation in acoustic unit's in a vocal bout has previously been recognized in meerkats (Manser et al., 2001, 2009). Further work exploring the variation in alarm calls is needed and should focus on whether or not variation in the number of /z/ acoustic unit's provides valuable referential information.

Wild gibbons

The responses to the tiger model varied greatly between a moving tiger model and a stationary tiger model. The units in each response were the same; however the length of the response and the transitions between subsequences varied. The general motivational warning (i.e. v-z complex) observed in the first subsequence was similar to other predator warning calls. The second subsequence consisted of the same acoustic units in both the running and stationary model; however the transitions between these unit's varied. Additionally, subsequence 2 represented the end of the running model response, whereas the stationary response had an additional subsequence. The similarity in this second subsequence is likely a referential sequence identifying either the type of animal posing the threat or the type of response necessary for deterring the predator. Since tigers are wait and ambush hunters, a running tiger would likely pose little threat to a gibbon, whereas a sitting tiger is more likely to be perceived as an animal ready to hunt. The additional subsequence in the stationary model response is possibly due to the gibbon's continued response to deter the predator from hunting. This persistent call could be a type of mobbing response, which has been observed as a method of deterring predators in many other species (Shedd, 1982; Klump & Shalter, 1984; Zuberbühler et al., 1999; Templeton & Greene, 2007) and has also been observed in white-handed gibbons during and encounter with a live tiger (Uhde & Sommer, 2002).

A noticeable deviation from the tiger responses was present in the adult female of group B. This female used more types of acoustic units than any other gibbon in her response to a tiger model. Acoustic unit's /d/, /f/, /g/, /h/, /i/, and /j/ were exclusively used by female B.

Whether these units are meaningful or not is debatable. Given the rarity of their occurrences, it is plausible that they are either not necessary or they represent flexibility within a response. Small variation in responses may be due to individuals assessing the level of threat differently. It is possible that she has had fewer encounters with tigers than the other gibbons and perceives tigers to be more of a threat. Though tigers can pose a risk to gibbons, predatory encounters are rare and therefore do not pose a great risk to a healthy, experienced gibbon (Uhde & Sommer, 2002). The variability observed in female B's response may also be due to additional factors. The distance between female B and the rest of her group could have been greater than the distance between adults of other groups. This potential isolation would have made her more vulnerable to predation. Another possibility is that she was responding to another threat she perceived in the nearby area, such as another predator or seeing the human observers. She also may have encountered a live predator earlier in the day which would have left her in a more vigilant state than the other gibbons. Further work is necessary to determine why her response was different than the others.

The gibbon's response to a clouded leopard model consisted of similar variability in acoustic units as the tiger model responses. Male responses appeared to consist of the basic call structure necessary to convey a warning call, which included five acoustic units. The other ten acoustic units in the full responses either occurred rarely or in female responses only. The variability in this response may reflect the variability in perceived level of threat the clouded leopard model posed. Similar to the tiger response, gibbons that are more familiar with encounters of live clouded leopards may not be as quick to respond to a model. The distance between the gibbons and the model may also be reflected in the response sequences. Future research examining predator responses should estimate distances between each gibbon to the predator model. Another plausible explanation for responding differently is the lack of olfactory cues associated with the model. Experienced gibbons which are familiar with clouded leopard encounters and odors may be more reluctant to produce a full warning sequence to an object which is lacking a key identifying aspect. Whether gibbons are capable of utilizing olfactory cues from predators needs to be examined experimentally.

The basic snake model response was the simplest of the wild predator warning calls with three types of acoustic units present across two subsequences. Similar to other predator model responses, both sexes gave the basic response to the presence of a snake model. Females also gave a more elaborate variation of the response which included an additional ten types of acoustic units and a third subsequence. Three of the acoustic units, /p/, /q/, and /y/, were exclusively used by a Female R. Female's B and R both used acoustic unit /g/ on a single occasion. It is unclear why there is additional variation present in the female's response. One possibility is greater vigilance to protect the young. Unlike female B, female R has a juvenile gibbon to care for. Juvenile gibbons require more parental care than older gibbons and the females exaggerated response to the snake model could reflect this heightened state of care. This response could also be an opportunity for the juvenile gibbons to learn how to appropriately respond to the threat of a large snake. Further field work is necessary to explore these potential relationships.

The average response of 34 minutes of vocalizing is lengthier than one would expect for a predator warning call. This is likely due to the experimental design, rather than representing novel information. Each group was presented with the model for approximately 20 minutes. The repetitive sequences are potentially reflecting the duration of the model presentation, rather than distinct information. Additional experiments testing whether duration of predator model presentation is correlated to length of response is still needed.

Captive gibbons

Responses from captive gibbons to the presence of a veterinarian varied greatly between each zoo. The veterinarian response from the Racine gibbons was stronger and more diverse in acoustic units used than the response from the Vilas gibbons. Veterinarian visits with the Vilas gibbons have been limited to non-invasive procedures for multiple years now, which have potentially reduced the perceived threat of the veterinarian's presence. Anecdotal evidence from the keepers at the Vilas Zoo and the veterinarian suggests that these gibbons do have a more prolonged response when they encounter the veterinarian for a procedure. Given the potential stress this could cause, we opted for passive encounters only. The female gibbon at the Racine zoo did need to be restrained for a routine check-up and this did elicit a response from both gibbons. This response included both motivational (i.e. v-z complex) and functionally referential (i.e. unique sequence which includes other acoustic units) aspects as shown in Figure 5.14. The motivational response was similar to the Vilas gibbons' response to their veterinarian. Variation in veterinarian response between the Vilas and Racine gibbons suggest gibbons may be capable of assessing the level of threat a perceived predator poses. This is supported by reports of the Vilas gibbons' previous vocal bouts given during physical encounters with their veterinarian. Follow up work should include recording the next invasive encounter the Vilas gibbons have with their veterinarian to determine whether they increase their warning call to include referential content. Additional work with the Racine gibbons is necessary to determine whether years of passive encounters with their veterinarian will lead to a reduction in warning calls in his presence.

The Vilas gibbons responded vocally to the presence of a cockatoo, despite cockatoos not being predatory in nature. Both gibbons would begin a vocal bout when the cockatoo was within visual range or auditory range. It is unknown what provokes this response, but it is possible that the gibbons cue into a warning aspect of the cockatoos call. Previous work has demonstrated the ability of interspecies utilization of warning calls (Rainey et al., 2004; Templeton & Greene, 2007; Goodale et al., 2010). The uniqueness of the response to the cockatoo should be examined further to determine whether gibbons utilize the calls of other species to indicate potential threats nearby. The call given by the cockatoo at the Vilas Zoo contained a series of abrupt screeches, whereas the cockatoo at the Racine Zoo contain a series of soft coos. The Racine gibbons did not call in the presence of their cockatoo which supports the theory that gibbons may be able to gain at least a limited amount of acoustic feature information from other species vocalizations. A wide array of other taxa are known to benefit from the use of other species warning calls (Klump & Shalter, 1984; Bradbury & Vehrencamp, 1998; Manser et al. 2001; Gill & Bierema, 2013) and gibbons may be another species to add to a continually growing list.

A few similarities exist when comparing wild calls to captive calls. For instance, the general warning call which initiated all wild predator calls is present in the captive gibbon warning calls. The length of the captive gibbon calls is also of a greater duration than expected given the urgency of warning calls. The warnings given by the captive gibbons do differ from the wild gibbons suggesting plasticity in sequence production exists with captive gibbons. Future work should focus on whether these sequences are representative of natural variation or due to learning in gibbons.

There are still many sequences with unknown meanings associated with them. Future work should focus on carefully designed playback experiments in the wild to test for potential meaning in these sequences. While these subsequences exist within the gibbons vocal repertoire, little if any overlap with language equivalents are present. Future work directed at understanding the function of these long sequences is also needed.

Appendices

Sex	Group	v	e	s	z	t	y	p	q	f	j	h	i	o	g	d	Total
F	A	250	1	57	1	7	0	15	14	0	0	0	0	0	0	0	**345**
M	A	307	0	13	5	2	0	0	0	0	0	0	0	0	0	0	**327**
F	B	93	11	21	35	19	1	1	1	1	7	11	4	0	1	0	**206**
M	B	77	65	2	1	4	0	0	0	0	0	0	0	0	0	0	**149**
F	C	18	57	62	36	40	42	2	2	0	1	0	0	4	0	0	**264**
M	C	118	42	17	91	14	3	0	0	0	0	0	0	1	0	0	**286**
F	D	217	24	48	38	31	6	5	5	3	0	0	0	0	0	0	**377**
M	D	120	0	1	25	0	0	0	0	0	0	0	0	0	0	0	**146**
F	H	129	24	5	8	4	0	0	0	4	0	0	0	0	0	0	**174**
M	H	37	4	0	0	0	0	0	0	0	0	0	0	0	0	0	**41**
F	J	104	5	13	4	9	0	0	0	0	0	0	0	0	0	0	**135**
F	M	92	72	36	7	20	0	1	1	0	1	0	0	0	0	0	**230**
M	M	124	40	6	11	9	0	1	1	0	0	0	0	0	0	0	**192**
F	M2	66	30	73	54	62	2	0	0	0	5	0	0	0	0	0	**292**
M	M2	65	35	38	27	34	0	0	0	0	0	0	0	0	0	0	**199**
F	N	194	113	17	14	15	1	3	3	12	0	1	1	0	0	0	**374**
M	N	53	41	3	15	2	0	0	0	0	0	0	0	0	0	1	**115**
F	NOS	142	35	23	0	4	0	5	6	0	0	0	0	0	0	0	**215**
M	NOS	432	0	3	31	0	0	0	0	0	0	0	0	0	0	0	**466**
F	R	146	18	5	8	4	2	2	2	4	0	0	0	0	0	0	**191**
M	R	77	9	14	0	9	0	0	0	0	0	0	0	0	0	0	**109**
F	R2	152	31	7	39	6	0	5	5	6	0	0	0	0	0	0	**251**
M	R2	82	50	2	7	3	0	0	0	0	0	0	0	0	0	0	**144**
F	W	333	98	30	1	18	0	0	0	0	0	0	0	0	0	0	**480**
M	W	65	3	18	34	5	0	0	0	0	0	0	0	0	0	0	**125**
F	W2	157	134	21	4	13	6	0	0	0	0	0	0	0	0	0	**335**
M	W2	178	57	9	0	8	0	0	0	0	0	0	0	0	0	0	**252**
		3828	**999**	**544**	**496**	**342**	**63**	**40**	**40**	**30**	**14**	**12**	**5**	**5**	**1**	**1**	

Appendix A – The above Table contains a breakdown of individual adult gibbon responses to the presentation of a clouded leopard predator model. Groups labeled with a '2' (e.g. M2) were tested twice with the same predator model on separate occasions.

Sex	Group	v	z	e	s	t	f	h	i	q	y	g	p	d	Total
F	B	478	30	1	0	0	9	11	8	0	0	1	0	0	**538**
M	B	199	23	2	0	0	0	0	0	0	0	0	0	0	**224**
M1	N	310	49	0	0	0	0	0	0	0	0	0	0	0	**359**
M2	N	191	17	0	0	0	0	0	0	0	0	0	0	0	**208**
F	R	126	29	82	23	19	10	4	3	3	3	1	2	0	**305**
M	R	197	51	13	1	2	0	0	0	0	0	0	0	1	**265**
		1501	**199**	**98**	**24**	**21**	**19**	**15**	**11**	**3**	**3**	**2**	**2**	**1**	

Appendix B – The above Table contains a breakdown of individual adult gibbon responses to the presentation of a snake predator model.

Sex	Group	v	s	z	e	t	p	q	f	j	y	aa	Total
F	A	213	43	6	4	12	8	8	2	1	0	1	**298**
M	A	190	18	18	19	18	1	1	0	0	0	0	**265**
F	C	150	10	18	8	2	5	5	0	0	0	0	**198**
M	C	136	2	33	3	1	0	0	0	0	0	0	**175**
F	H	78	19	2	22	16	5	4	0	0	0	0	**146**
M	H	40	1	3	0	1	0	0	0	0	0	0	**45**
F	R	169	25	10	1	10	8	6	0	1	0	0	**230**
M	R	161	0	13	0	0	0	0	0	0	0	0	**174**
F	W	312	14	3	17	6	0	0	0	0	1	0	**353**
M	W	103	1	24	0	1	0	0	0	0	0	0	**129**
		1552	**133**	**130**	**74**	**67**	**27**	**24**	**2**	**2**	**1**	**1**	

Appendix C – The above Table contains a breakdown of individual adult gibbon responses to the presentation of a stationary tiger model.

Sex	**Group**	**v**	**z**	**s**	**e**	**t**	**p**	**q**	**f**	**d**	**h**	**j**	**y**	**g**	**i**	**Total**
F	B	324	23	17	4	5	1	1	4	3	3	2	0	1	1	**389**
M	B	142	25	1	15	1	0	0	0	0	0	0	0	0	0	**184**
F	D	108	7	26	18	7	0	0	0	0	0	0	2	0	0	**168**
M	D	80	20	1	0	0	0	0	0	0	0	0	0	0	0	**101**
F	J	52	0	0	0	0	0	0	0	0	0	0	0	0	0	**52**
M	J	300	8	2	0	1	0	0	0	0	0	0	0	0	0	**311**
F	N	97	3	16	0	4	2	2	0	0	0	0	0	0	0	**124**
M	N	1110	34	15	0	8	0	0	0	0	0	0	0	0	0	**1167**
F	T	316	14	5	0	4	1	1	0	0	0	0	0	0	0	**341**
M	T	75	14	0	0	0	0	0	0	0	0	0	0	0	0	**89**
		2604	**148**	**83**	**37**	**30**	**4**	**4**	**4**	**3**	**3**	**2**	**2**	**1**	**1**	

Appendix D – The above Table contains a breakdown of individual adult gibbon responses to the presentation of a moving tiger model.

Chapter References

Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology, 18*(5), R202-R203.

Blumstein, D. T. (1999). The evolution of functionally referential alarm communication: multiple adaptations; multiple constraints. *Evolution of Communication, 3*(2), 135-147.

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication* (Second ed.). Sunderland, Massachusetts: Sinauer Associates, Inc.

Brandes, T. S. (2008). Feature vector selection and use with hidden Markov models to identify frequency-modulated bioacoustic signals amidst noise. *Ieee Transactions on Audio Speech and Language Processing, 16*(6), 1173-1180.

Brown, J. C., & Smaragdis, P. (2009). Hidden Markov and Gaussian mixture models for automatic call classification. *Journal of the Acoustical Society of America, 125*(6), EL221-EL224.

Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012). Context-related call combinations in female Diana monkeys. *Animal Cognition, 15*(3), 327-339.

Casar, C., & Zuberbühler, K. (2012). Referential alarm calling behaviour in New World primates. *Current Zoology, 58*(5), 680-697.

Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *Plos One, 1*(1).

Cogburn, R. (1984). The Ergodic-theory of Markov-chains in Random-environments. *Zeitschrift Fur Wahrscheinlichkeitstheorie Und Verwandte Gebiete, 66*(1), 109-128.

Evans, C. S., Evans, L., & Marler, P. (1993). On the meaning of alarm calls - functional reference in an avian vocal system. *Animal Behaviour, 46*(1), 23-38.

Evans, C. S., Macedonia, J. M., & Marler, P. (1993). Effects of apparent size and speed on the response of chickens, Gallus gallus, to computer-generated simulations of aerial predators. *Animal Behaviour, 46*(1), 1-11.

Gautier, J. P. (1974). Field and laboratory studies of vocalizations of Talapoin monkeys (*Miopithecus talapoin). Behaviour, 51*(3-4), 209-273.

Geissmann, T. (2009). Door slamming: Tool-use by a captive white-handed gibbon (*Hylobates lar*). *Gibbon Journal, 5*, 53-60.

Gill, S. A., & Bierema, A. M. K. (2013). On the meaning of alarm calls: A review of functional reference in avian alarm calling. *Ethology, 119*(6), 449-461.

Goodale, E., Beauchamp, G., Magrath, R. D., Nieh, J. C., & Ruxton, G. D. (2010). Interspecific information transfer influences animal community structure. *Trends in Ecology & Evolution, 25*(6), 354-361.

Harris, J. G., & Skowronski, M. D. (2006). *Automatic speech processing methods for bioacoustic signal analysis: a case study of cross-disciplinary acoustic research*. Paper presented at the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP.

Ji, A., Johnson, M. T., Walsh, E. J., McGee, J., & Armstrong, D. L. (2013). Discrimination of individual tigers (Panthera tigris) from long distance roars. *Journal of the Acoustical Society of America, 133*(3), 1762-1769.

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence, 101*(1-2), 99-134.

Klump, G. M., & Shalter, M. D. (1984). Acoustic behavior of birds and mammals in the predator context .1. Factors affecting the structure of alarm signals .2. The functional-significance and evolution of alarm signals. *Zeitschrift Fur Tierpsychologie-Journal of Comparative Ethology, 66*(3), 189-226.

Koda, H., Lemasson, A., Oyakawa, C., Rizaldi, Pamungkas, J., & Masataka, N. (2013). Possible role of mother-daughter vocal interactions on the development of species-specific song in gibbons. *Plos One, 8*(8).

Lieblich, A. K., Symmes, D., Newman, J. D., & Shapiro, M. (1980). Development of the isolation peep in laboratory-bred squirrel monkeys. *Animal Behaviour, 28*(FEB), 1-9.

Manser, M. (2009). What do functionally referential alarm calls refer to? In R. Dukas & J. M. Ratcliffe (Eds.), *Cognitive Ecology II*. Chicago: University of Chicago Press.

Manser, M. B., Bell, M. B., & Fletcher, L. B. (2001). The information that receivers extract from alarm calls in suricates. *Proceedings of the Royal Society B-Biological Sciences, 268*(1484), 2485-2491.

Marler, P., Evans, C. S., & Hauser, M. D. (1992). Animal signals: motivational, referential, or both *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 66-86). Cambridge: Cambridge University Press.

Masataka, N., & Fujita, K. (1989). Vocal learning of Japanese and Rhesus monkeys. *Behaviour, 109*(No. 3/4), 191-199.

Newman, J. D., & Maclean, P. D. (1982). Effects of tegmental lesions on the isolation call of squirrel monkeys. *Brain Research, 232*(2), 317-329.

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009a). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences of the United States of America, 106*(51), 22026-22031.

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009b). Campbell's monkeys use affixation to alter call meaning. *Plos One, 4*(11).

Ouattara, K., Zuberbühler, K., N'Goran, E. K., Gombert, J. E., & Lemasson, A. (2009). The alarm call system of female Campbell's monkeys. *Animal Behaviour, 78*(1), 35-44.

Pace, F., White, P., & Adam, O. (2012). *Hidden Markov Modeling for humpback whale (Megaptera Novaeanglie) call classification*. Paper presented at the Proceedings of Meetings on Acoustics.

Payne, T. E. (1997). *Describing morphosyntax: A guide for field linguists*: Cambridge University Press.

Ploog, D., Hopf, S., & Winter, P. (1967). Ontogenesis of behavior of squirrel monkeys (*Saimiri sciureus*). *Psychologische Forschung, 31*(1), 1-1.

Raemaekers, J. J., & Raemaekers, P. M. (1984). The ooaa duet of the gibbon (*Hylobates lar*) - A group call which triggers other groups to respond in kind. *Folia Primatologica, 42*(3-4), 209-215.

Raemaekers, J. J., Raemaekers, P. M., & Haimoff, E. H. (1984). Loud calls of the gibbon (*Hylobates lar*) - Repertoire, organization and context. *Behaviour, 91*, 146-189.

Rainey, H. J., Zuberbühler, K., & Slater, P. J. B. (2004). Hornbills can distinguish between primate alarm calls. *Proceedings of the Royal Society B-Biological Sciences, 271*(1540), 755-759.

Ren, Y., Johnson, M. T., Clemins, P. J., Darre, M., Glaeser, S. S., Osiejuk, T. S., et al. (2009). A framework for bioacoustic vocalization analysis using hidden markov models. *Algorithms, 2*(4), 1410-1428.

Rendall, D., Owren, M. J., & Ryan, M. J. (2009). What do animal signals mean? *Animal Behaviour, 78*(2), 233-240.

Schel, A. M., Tranquilli, S., & Zuberbühler, K. (2009). The alarm call system of two species of black-and-white Colobus monkeys (Colobus polykomos and Colobus guereza). *Journal of Comparative Psychology, 123*(2), 136-150.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980a). Monkey responses to 3 different alarm calls - Evidence of predator classification and semantic communication. *Science, 210*(4471), 801-803.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980b). Vervet monkey alarm calls - semantic communication in a free-ranging primate. *Animal Behaviour, 28*(4), 1070-1094.

Shedd, D. H. (1982). Seasonal variation and function of mobbing and related antipredator behaviors of the American robin (Turdus migratorius). *The Auk*, 342-346.

Slocombe, K. E., & Zuberbühler, K. (2006). Food-associated calls in chimpanzees: responses to food types or food preferences? *Animal Behaviour, 72*, 989-999.

Snowdon, C. T. (1989). Vocal communication in new-world monkeys. *Journal of Human Evolution, 18*(7), 611-633.

Snowdon, C. T., Cleveland, J., & French, J. A. (1983). Responses to context-specific and individual-specific cues in cotton-top tamarin long calls. *Animal Behaviour, 31*(FEB), 92-&.

Stephan, C., & Zuberbühler, K. (2008). Predation increases acoustic complexity in primate alarm calls. *Biology Letters, 4*(6), 641-644.

Stevenson, A., & Lindberg, C. A. (Eds.). Oxford University Press.

Struhsaker, T. T. (1967). Auditory communication among vervet monkeys (Cercopithecus aethiops) *Social Communication Among Primates* (pp. 281-324). Chicago, IL: University of Chicago Press.

Symmes, D., & Biben, M. (1985). Maternal recognition of individual infant squirrel-monkeys from isolation call playbacks. *American Journal of Primatology, 9*(1), 39-46.

Templeton, C. N., & Greene, E. (2007). Nuthatches eavesdrop on variations in heterospecific chickadee mobbing alarm calls. *Proceedings of the National Academy of Sciences of the United States of America, 104*(13), 5479-5482.

Townsend, S. W., & Manser, M. B. (2013). Functionally referential communication in mammals: The past, present and the future. [Review]. *Ethology, 119*(1), 1-11.

Uhde, N. L., & Sommer, V. (2002). Antipredatory behavior in gibbons (Hylobates lar, Khao Yai/Thailand) In L. E. Miller (Ed.), *Eat or be eaten: predator sensitive foraging among primates* (pp. 268): Cambridge University Press.

Zuberbühler, K. (2009). Survivor signals: The biology and psychology of animal alarm calling. *Advances in the Study of Behavior, Vol 40, 40*, 277-322.

Zuberbühler, K., Jenny, D., & Bshary, R. (1999). The predator deterrence function of primate alarm calls. *Ethology, 105*(6), 477-490.

“The patient is still able to accomplish most of the motor acts, in spite of the trembling but goes about performing them with extreme slowness. We noticed this fact in its relation to the faculty of speech; there is a comparatively considerable lapse of time between the thought and the act. One might suppose that the nervous influx cannot be set to work until after extraordinary efforts, and, in reality, the slightest movements occasion extreme fatigue.” (Charcot, 1879)

Chapter 6: A translational medical application of the CSSM framework

Abstract

Understanding speech deficits associated with parkinsonism has been a focal area of research. Methods for automated detection of speech deficits are explored in this chapter using vocalizations from a rat model. Acoustic units in the vocal repertoire of healthy rats are identified and compared with vocalizations from rats after receiving a neurotoxic brain lesion to induce nigrostriatal dopamine loss. By examining acoustic units using an automated computational analysis, we discovered rat acoustic units degrade differentially in the diseased state. Furthermore, by examining vocalizations in a Cepstral Self-Similarity Matrix view, subtle degradation which was not obvious in the spectral view becomes apparent. The current findings suggest our approach can be used as a robust detection method of vocal loss in rats.

Introduction

Parkinson’s disease (PD) is a neurological disorder that is primarily characterized by a loss of sensorimotor function, such as hypokinesia, bradykinesia, rigidity, tremor, and postural instability (Bergman & Deuschl, 2002; Braak et al., 2004; Calne 2005). The primary disease pathology that is related to these sensorimotor deficits is the destruction of dopamine neurons in the nigrostriatal pathways (Braak et al., 2004), although, it is widely believed that extra-striatal and non-dopaminergic mechanisms contribute to these deficits as well (Pavese, 2012). Importantly, PD also leads to cranial sensorimotor deficits, such as swallowing and voice problems (Darley et al., 1969 a,b; Fox et al., 2002; Ho et al., 1998; Logemann et al., 1978; Weismer, 1984; Plowmann-Prine et al., 2009; Sapir, 2014). While there is currently no cure for PD, there has been a great deal of effort to find treatments such as medications, deep brain stimulation and exercise (Sapir et al., 2008; Fox et al., 2006; D'Alatri et al., 2008; Ramig et al., 2008).

Rats have been presented as a good model organism for studying deficits associated with parkinsonism since the degeneration of dopamine cells mimic deficits observed in humans at early stages of PD (Cenci et al., 2002; Fleming et al., 2005; Meredith & Kang, 2006; Schallert et al., 2000; Tillerson et al., 2001). Previous research has successfully induced parkinsonism in rats by depleting dopamine levels unilaterally in the medial forebrain bundle of the brain through the use of dopamine neurotoxin 6-OHDA, including inducing vocalization deficits akin to those seen in humans with PD (Fulceri et al., 2006; Marshall, 1979; Ungerstedt, 1971). Some of the deficits that can be observed in rats include loss of motor limb function, gait abnormalities and asymmetries in body posture (Cenci, 2002; Tillerson et al., 2001). Detection methods and treatments for many of these signs have been examined with varying levels of success in rats and/or humans (Carella et al., 2001; Brown et al., 1999; Kane et al., 2011), while other signs are still being explored. By inducing parkinsonism in rats, variables such as home environment, age post-onset, severity of lesion and medications can be controlled and neural modulation can be studied. This also sets up an ideal environment for determining what types of vocalizations degrade in rats in order to examine the vocal deficits associated with parkinsonism.

One of the signs being explored is the sensorimotor vocalization deficits that accompany early stages of PD (Mollaei et al., 2013). It has already been established that the quality of ultrasonic vocalizations degrades in rats after receiving neurotoxin brain lesions (Ciucci et al., 2009; Ciucci et al., 2007). With respect to vocal degradation, there is a noticeable difference in the intensity, bandwidth and complexity of ultrasonic vocalizations. In rats, ultrasonic vocalizations have been categorized as 22 kHz, 40 kHz or 50 kHz calls and each refers to a different context. The 22 kHz calls are typically associated with stress such as when an individual sees a predator (Blanchard et al., 1991; Portfors, 2007), experiences or anticipates pain (Cuomo et al., 1988; Tonoue et al., 1986; Antoniadis & McDonald, 1999), perceives a potential threat (Brudzynski & Ociepa, 1992) or is defeated socially (Thomas et al., 1983). The 40 kHz calls are typically uttered by infants in response to poorly regulated body temperatures (Allin & Banks, 1971; Allin & Banks, 1972; Carden & Hofer, 1992) and the 50 kHz calls are typically uttered by individuals during positive situations such as play (Knutson et al., 1998; Knutson et al., 1999; Knutson et al., 2002), courtship (Barfield et al., 1979; Portfors, 2007) and in anticipation of feeding (Burgdorf et al., 2000). What has yet to be studied is whether those general categories contain further subdivisions which carry additional meaning. As shown in the previous two chapters, subdivision of categories reveals remarkable complexity in the vocalizations of white-handed gibbons. Evidence exists for additional complexity in rat vocalizations, as demonstrated in this chapter, and how this complexity degrades is explored. We hypothesize that acoustic units in the rats' vocal repertoire will degrade differentially (i.e. highly complex acoustic units will degrade more than simple acoustic units). We also hypothesize our Cepstral Self-Similarity Matrices (CSSM) analysis will capture the degradation of acoustic units better than a traditional spectral analysis. By studying vocal loss in a rat model, we can learn more about what contexts are affected the most as well as potentially gain insight into what analytical approaches may be helpful for humans.

Methods

Recording set-up

The rats used in this research are all 4-5 month old, sexually experienced, captive bred, male laboratory rats of the *Long-Evans* strain (Charles River Laboratories International, Inc.). Each rat was recorded separately in their home cage after being stimulated by a sexually receptive female. Vocalizations were recorded with Avisoft software at 214.285 kHz to ensure that all of the ultrasonic vocalizations were obtained above the Nyquist limit. A microphone was placed 16 cm above the male rats' cage and vocalizations were recorded for several seconds to minutes after being stimulated by the female. This process was repeated for each rat (n=86) in the study as well as multiple times per rat.

Baseline or pre-lesion vocalizations were obtained from each rat to establish a natural vocal repertoire. Whereas previous research has identified three main categories of calls, 22 kHz, 40 kHz, and 50 kHz, we are interested in examining the interior structure of calls in these groups to determine whether any subcategories exist. After baseline recordings were obtained for each rat, they underwent surgery to receive either an injection of 7 μg of 6-OHDA in the left medial forebrain bundle or a sham lesion (Ciucci et al. 2010). After a short recovery period, each rat was recorded again to obtain lesion-influenced recordings.

Baseline Acoustic Unit Classifying

To examine audio files for acoustic differences, each file was frequency lowered by a factor of 10 (resampled to 22,050 Hz) to fall within the range of human hearing. After lowering the frequency, each file was listened to twice before proceeding to define boundaries of acoustic units. This was done to ensure familiarization with the general sounds. Some acoustic unit boundaries were determined based upon the silence that preceded and followed a particular call (similar to previous work (Brudzynski et al., 1993; Brudzynski et al., 1995; Brudzynski et al., 2005)); however other acoustic units were determined to be unique if a change in frequency was observed and if this change could also be found in an acoustic unit that was isolated. For example, acoustic units ‘k’ and ‘l’ were often found together (Figure 6.1, left image); however they were also found independent of one another (Figure 6.1, central and right images) and had a noticeable difference in frequency range. Similar to segmentation of gibbon calls in chapter 4, these categories were first identified by manually processing each file and later confirmed using the Cepstral Self-Similarity Matrix (CSSM) framework.

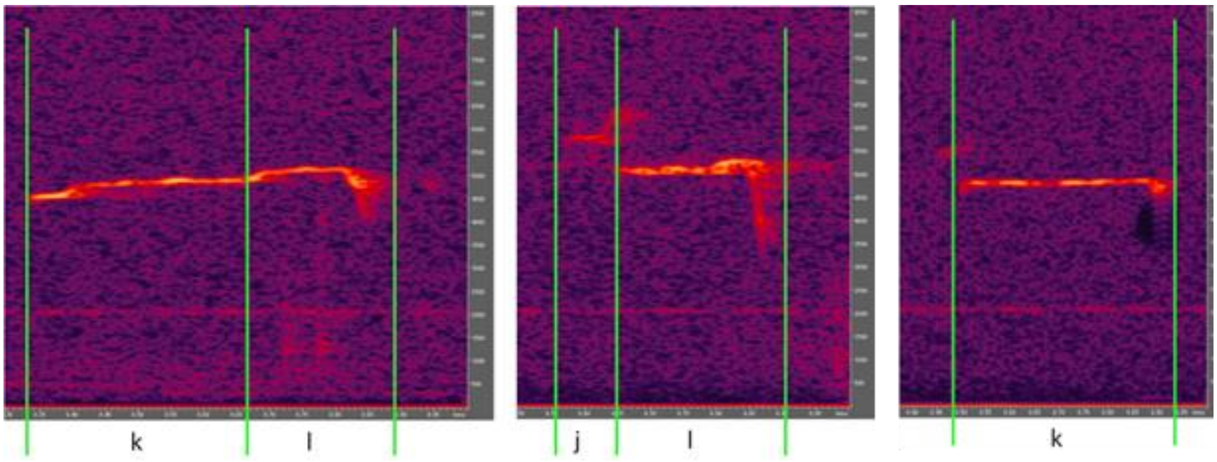


Figure 6.1: The spectrogram on the left shows acoustic units ‘k’ and ‘l’ as seemingly continuous sounds. The image in the middle shows ‘l’ following ‘j’, but separate from any other acoustic units. The image on the right shows ’k’ separate from any other acoustic unit.

Lesion Acoustic Unit Classifying

Acoustic units from lesioned rats were identified using a similar approach as the non-lesion files. Due to the lesions reducing the quality of some of the acoustic units, additional steps were also necessary to classify certain calls. For partially degraded calls, frequency measurements (start frequency (Hz), peak frequency (Hz), minimum frequency (Hz), and end frequency (Hz)) and duration were used as a comparison to baseline samples of a call. Since the acoustic features of the degraded vocalizations are altered, the acoustic unit category that has the most overlap in features was selected. Severely degraded calls which could not be compared to baseline vocalizations were classified as undeterminable.

Category comparisons

The acoustic unit categories established from the baseline or pre-lesion data are compared to the lesion files to determine which types of vocalizations are the most susceptible to degradation. Figure 6.2 shows a comparison between acoustic unit’s /j/ and /l/ before and after lesioning.

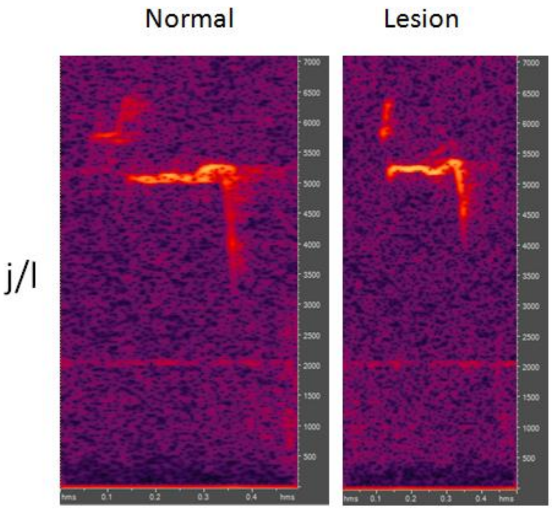


Figure 6.2: The left spectral image is samples of calls given by a male rat prior to receiving a brain lesion. The first acoustic unit is /j/ and the second is /l/. The right spectral image is acoustic unit's j & l after surgery. Note how there is some degradation of acoustic unit /j/, but little to no degradation in acoustic unit /l/.

Validation

Acoustic units were confirmed using the CSSM approach and Zipf's law. Detailed methods for each can be found in Chapter 4. Short ultrasonic vocalizations were separated from the long ultrasonic vocalizations for analysis. This was necessary to obtain a clear comparison of acoustic features.

Results

Pre-lesion categories

Baseline or pre-lesion vocalizations were obtained from each rat for a total of 773 files. A total of 29 acoustic unit categories were identified in the vocal repertoire of the male rat (Figure 6.3). While this data set only consists of young, male rats, we believe that it represents the majority, if not all, of the calls types produced by males since the plotted data follows Zipf's law ($R^2 = 0.9948$) (Figure 6.4). A majority of these calls are in the 40 kHz and 50 kHz range of ultrasonic vocalizations; however there are 3 ultrasonic vocalizations around the 22 kHz range. After these calls are frequency lowered by a factor of ten (to make them all audible to humans), the length of the calls range from 0.063 - 5.264 seconds.

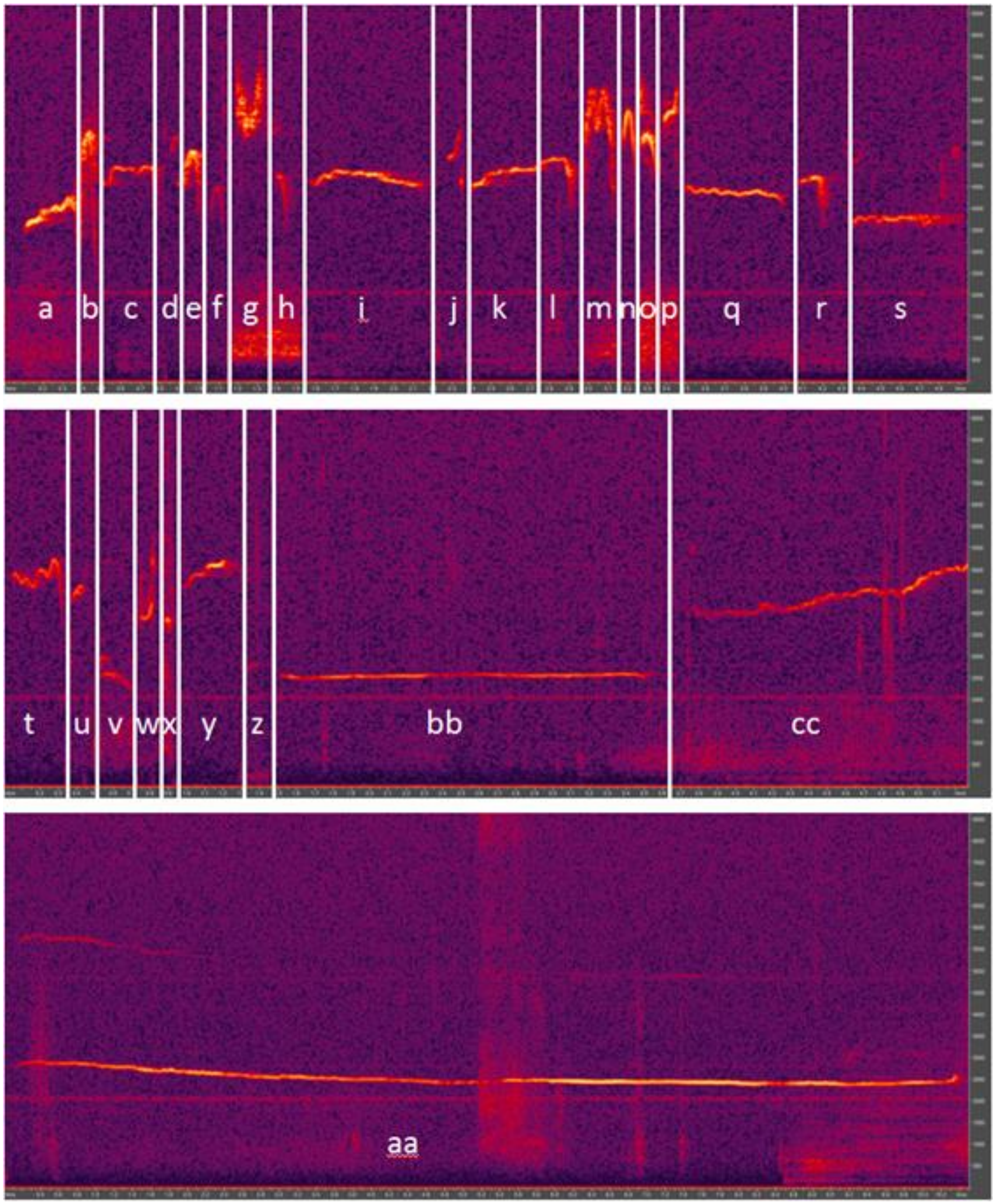


Figure 6.3: This figure depicts the variation in the vocal repertoire of the rat that has currently been identified and described. After these calls are frequency lowered by a factor of ten (to make them all audible), the length of the calls range from 0.063 - 5.264 seconds.

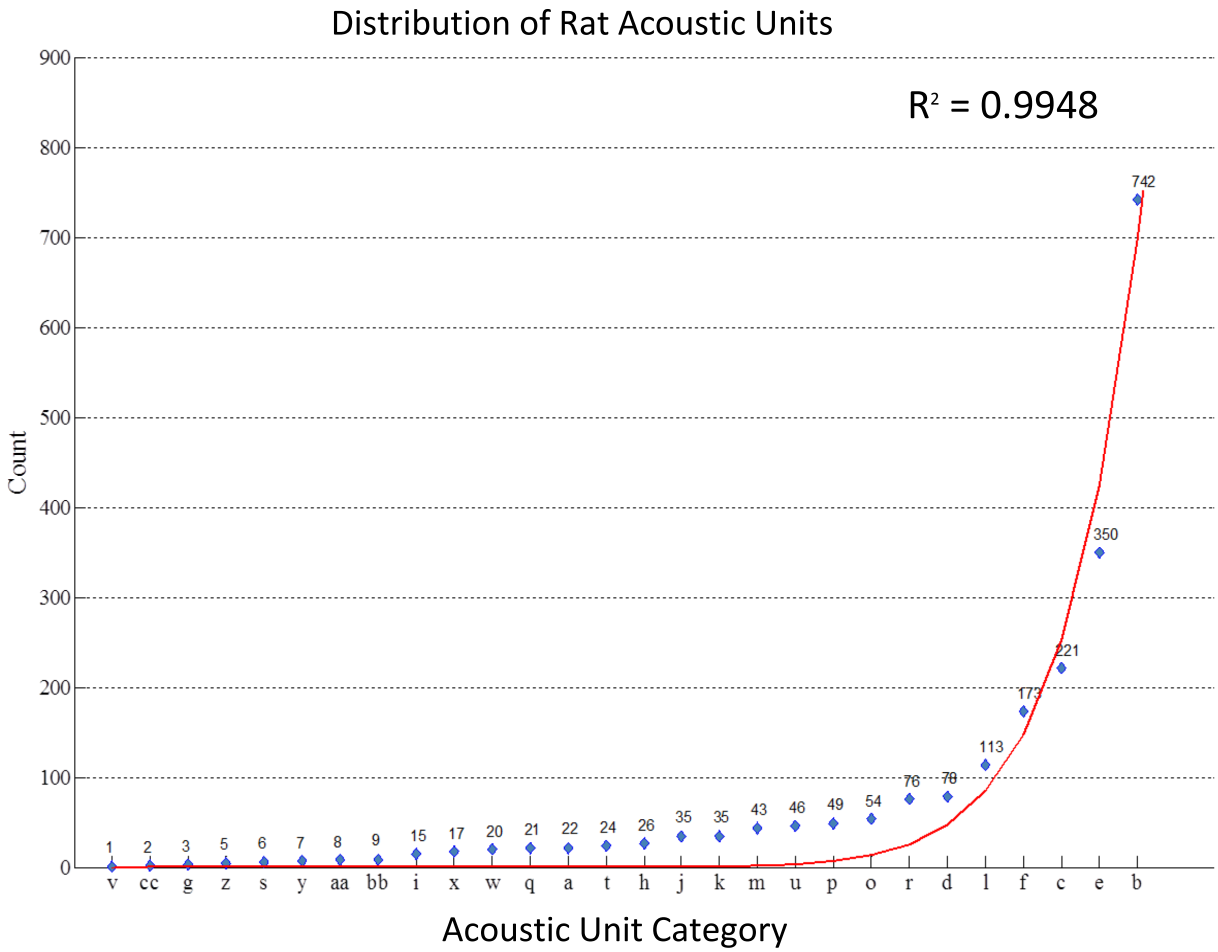


Figure 6.4: This graph shows a distribution of acoustic units which occur in the repertoire of healthy laboratory rats.

Confirmation of these categories was obtained by examining the CSSM of each file (Figures 6.5 & 6.6). The acoustic signature of each of these categories appears to be unique. Some acoustic units, such as /q/, had a great deal of variability in their CSSM signatures.

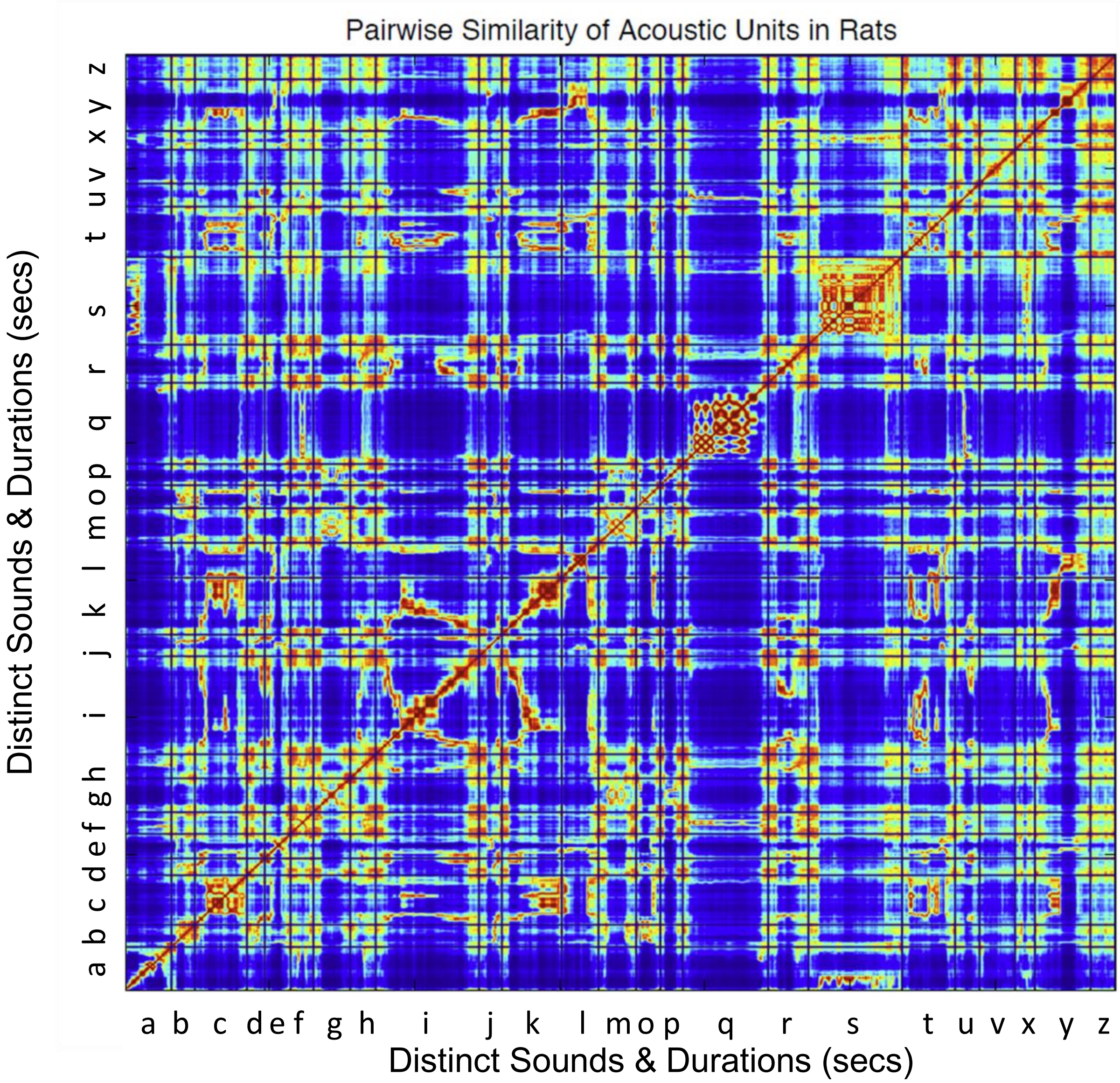


Figure 6.5: CSSM view of the vocal repertoire of healthy rats. The 26 shortest acoustic units shown here have unique signitures which are computationally distinct from one another. The remaining acoustic units are shown in Figure 6.6 and consist of the long calls.

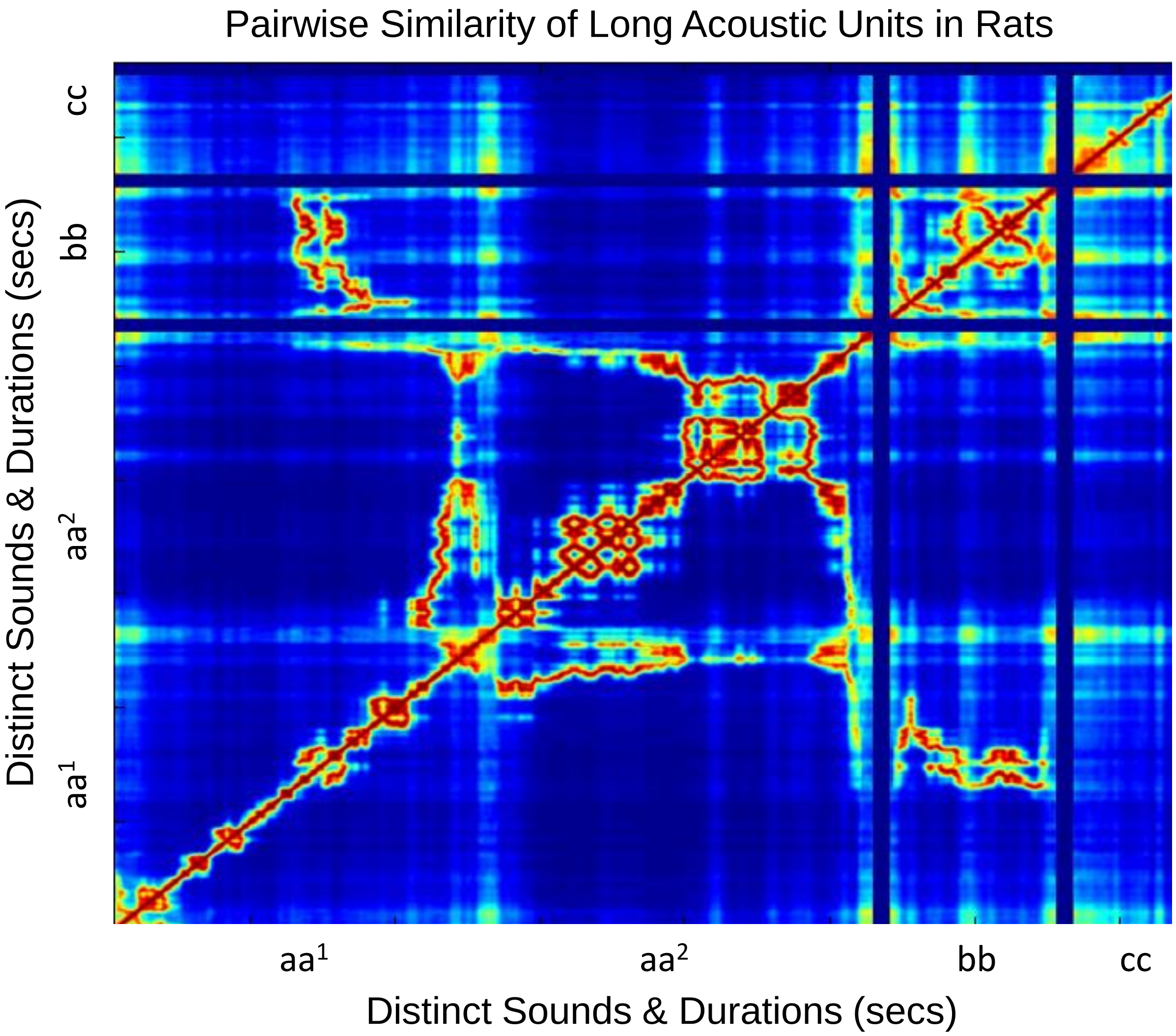


Figure 6.6: CSSM view of the longest acoustic units in the vocal repertoire of healthy rats. The aa category which was originally labeled as a single acoustic unit has two distinct portions in the CSSM view. These have been labeled aa^1 and aa^2.

Lesion categories

A total of 162 files were gathered after rats received a lesion. Of the 29 acoustic unit categories described from the pre-lesion data, 11 acoustic units (/c/, /d/, /f/, /m/, /n/, /o/, /p/, /q/, /r/, /y/, and /z/) showed no significant deterioration in the spectral view of the lesion data. Moderate degradation was recorded in 8 acoustic units (/e/, /g/, /h/, /i/, /j/, /l/, /t/, and /u/). High degradation was recorded in 7 acoustic units (/a/, /b/, /k/, /s/, /aa/, /bb/, and /cc/).

Moderate degradation was determined to be either shortening of an acoustic unit or change in frequency bandwidth. High levels of degradation are determined to be variation in 3 or more measured acoustic properties (Figures 6.7 and 6.8). The remaining 3 acoustic units (v, w, and x) were not found in any of the data sets and are likely fully degraded.

Of the 11 acoustic units which showed little to no degradation, 6 fell around the 40 kHz range, 4 around the 50 kHz range and 1 around the 22 kHz range. Of the 8 acoustic units which showed moderate degradation, 7 were around the 40 kHz range and 1 was around the 50 kHz range. The highly degraded acoustic units consisted of 5 around the 40 kHz range and 2 around the 22 kHz range. Of the three which could not be found, 2 fell around the 40 kHz range and 1 fell around the 22 kHz range.

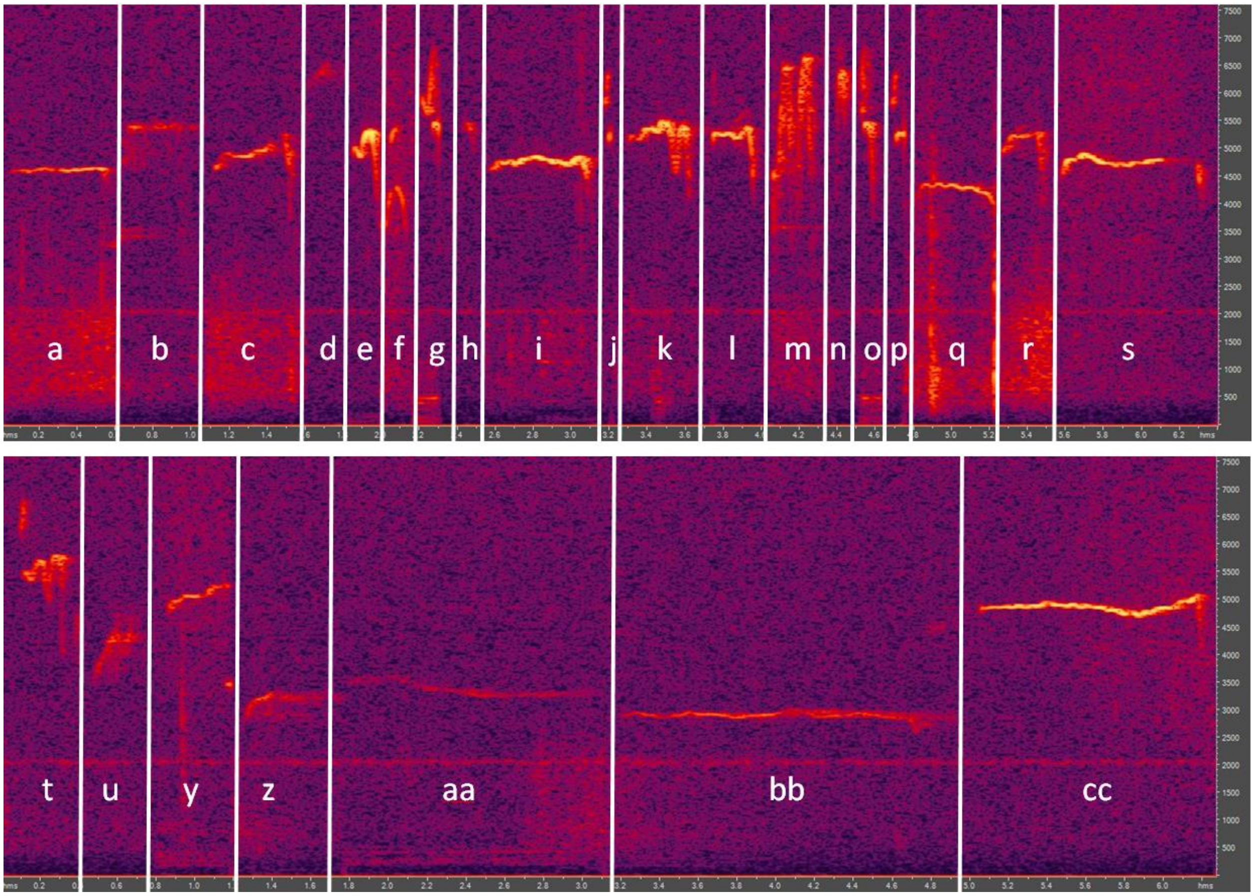


Figure 6.7: The spectral view of vocalizations from lesioned rats illustrates some degradation in acoustic units. Missing from the vocal repertoire of lesioned rats are acoustic units /v/, /w/ and /x/.

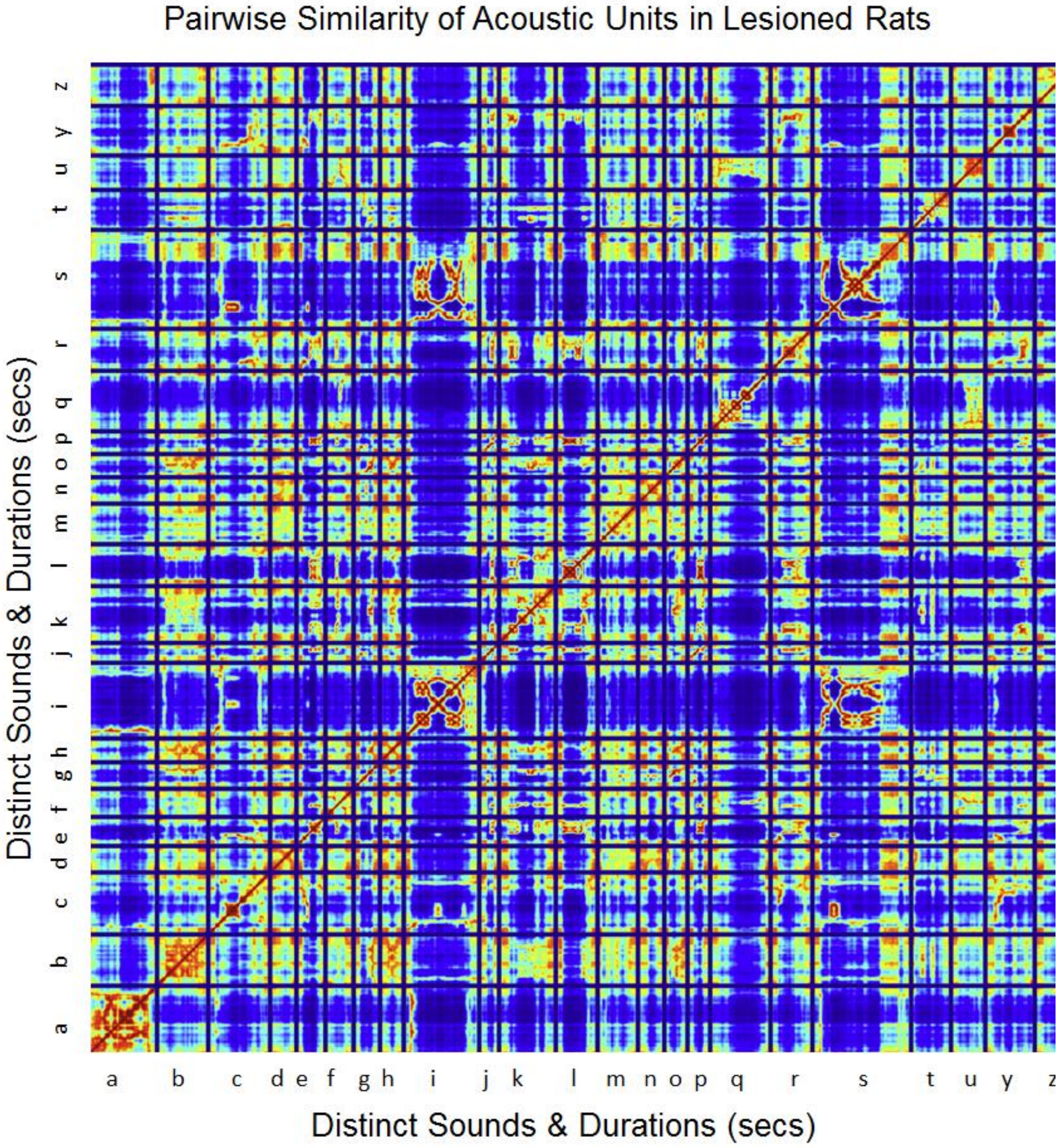


Figure 6.8: The CSSM view of vocalizations from lesioned rats characterizes the loss of acoustic complexity. Acoustic unit's /v/, /w/, and /x/ could not be found in the lesioned data

The manual analysis of acoustic units did not reveal significant change in several categories; however the CSSM analysis did identify signal degradation. The top half of figure 6.9 shows the spectral images of what a normal /q/ acoustic unit looks like (left) compared to a /q/ acoustic unit from a rat after receiving a brain lesion (right). Little variation in acoustic features is apparent in these two images; however the CSSM view of each (shown in the lower half of Figure 6.9) shows clear degradation of acoustic structure. Though some of the structure is still present, some has disappeared. Comparisons of pre and post lesion CSSM's for the other acoustic unit categories are found in Figures 6.10 and 6.11.

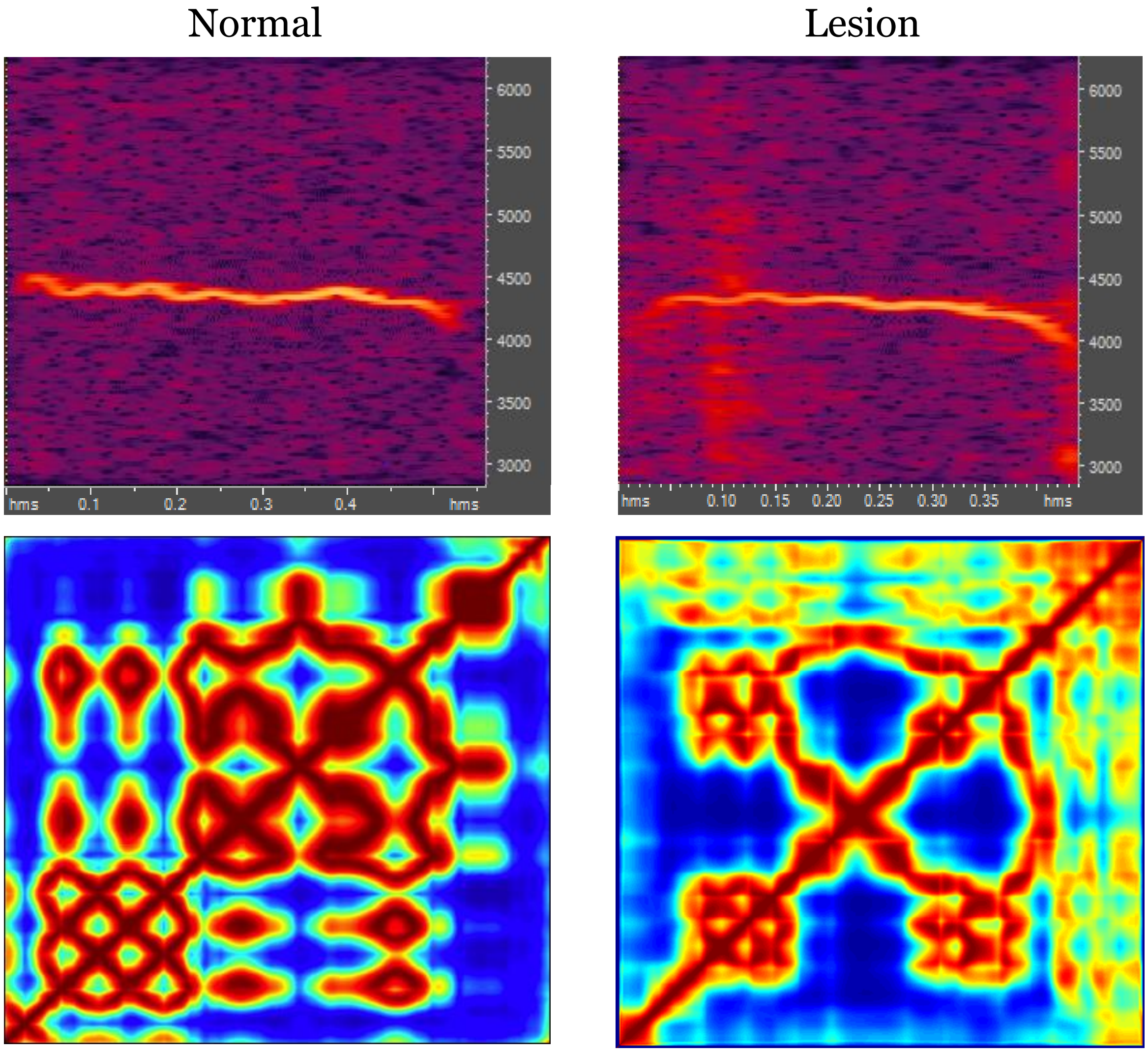


Figure 6.9: The top panels show a spectral view of acoustic unit /q/ in a normal versus a lesioned rat. These units appeared similar acoustically and visually. The lower panels show the CSSM view of the same /q/ acoustic units. The CSSM view reveals changes in complexity which is otherwise obscured by visual inspection of spectrograms. Though some of the structure still exists in the lesioned sample, significant portions are missing.

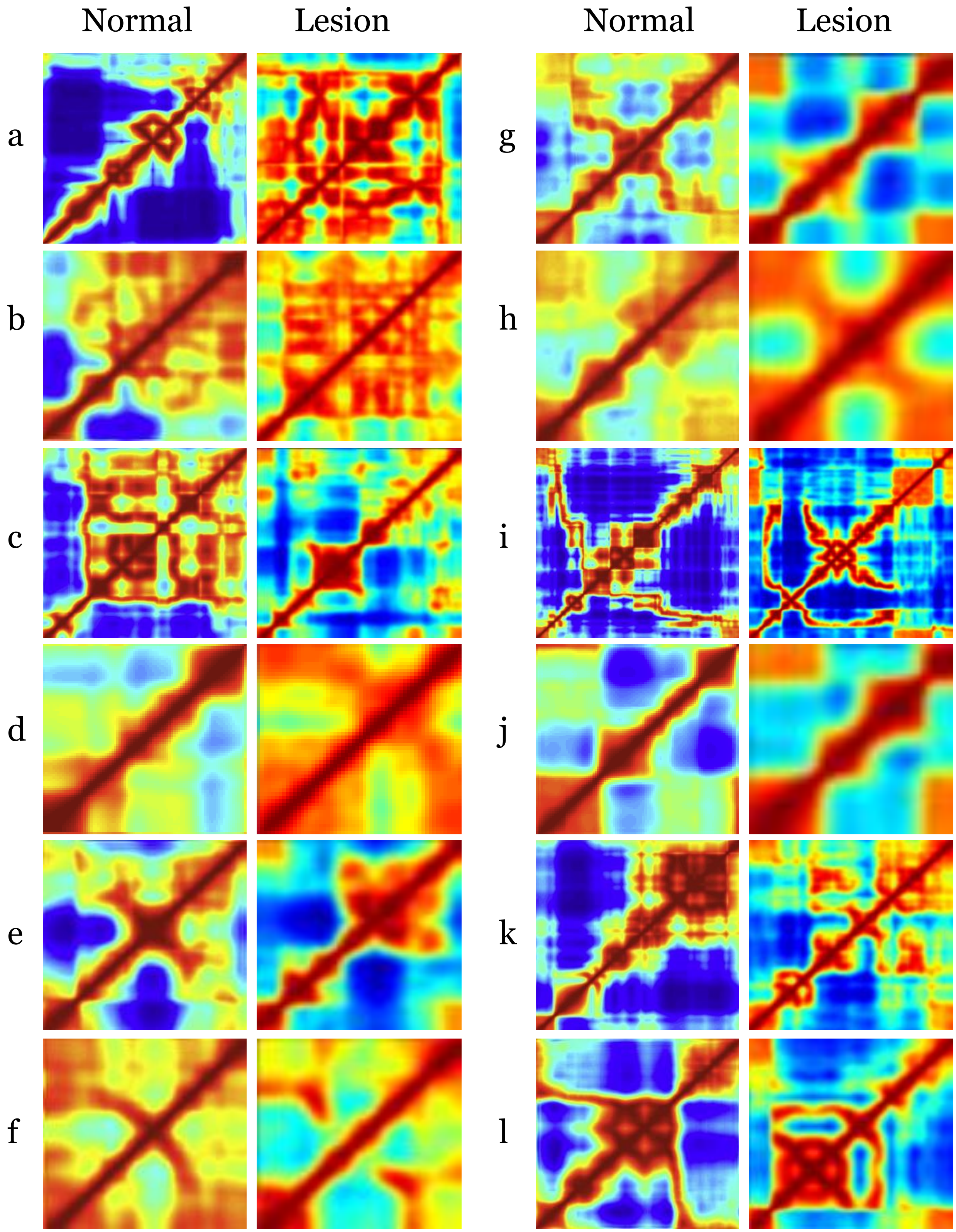


Figure 6.10: The degradation of acoustic units in lesioned rats is present in each of the above comparative samples; however certain acoustic units, such as /a/ are more variable than others, such as /j/.

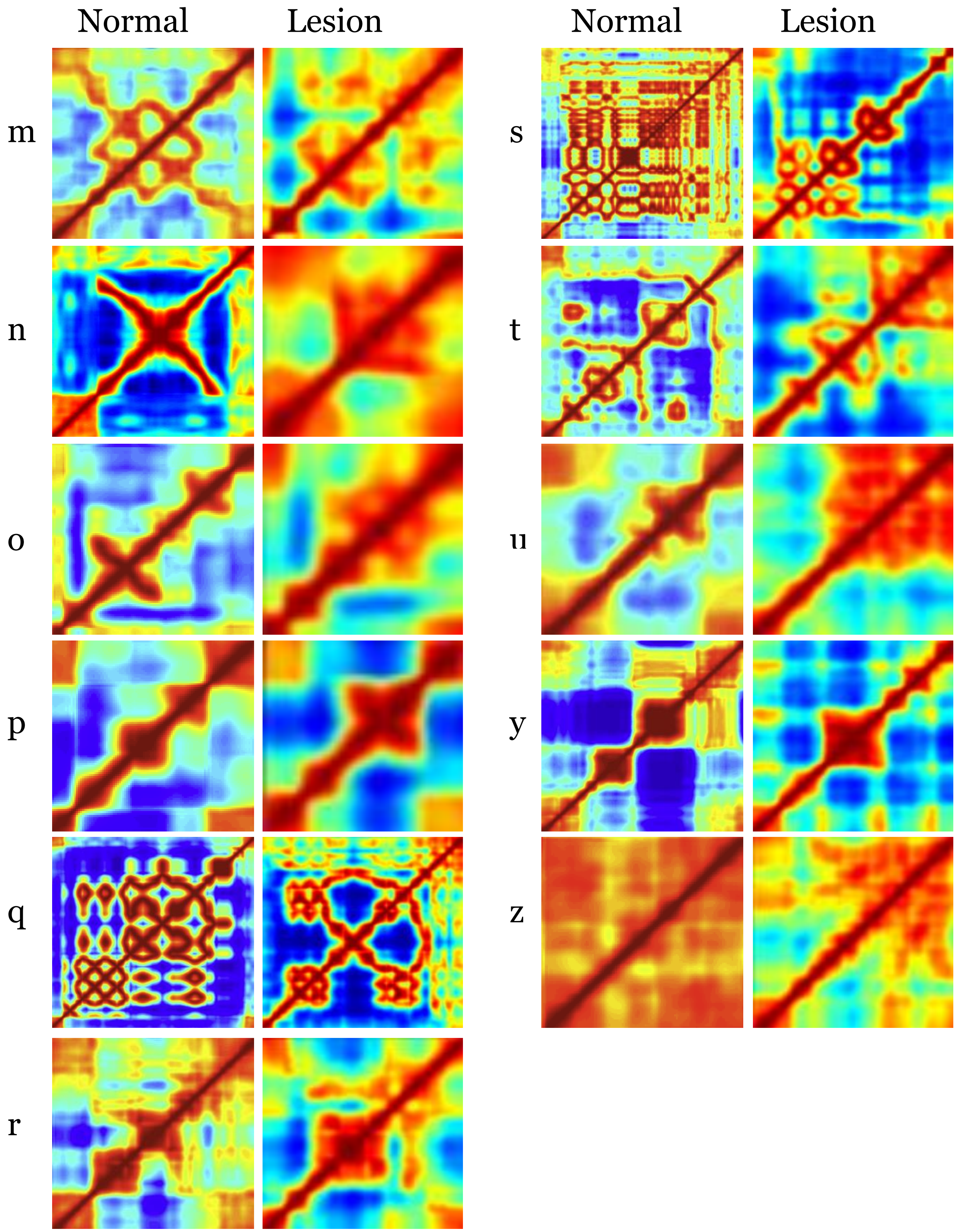


Figure 6.11: The remaining acoustic units show similar differential degradation as shown in Figure 6.10.

Discussion

Describing the baseline vocalizations of rats in greater detail has provided a foundation for examining vocal loss in rats at an acoustic unit level. The results discovered in the current study support previous research which has demonstrated a loss of vocal output in lesioned rats (Ciucci et al., 2009; Ciucci et al., 2007). Additionally, the current study has identified differential vocal loss within the acoustic units identified in the rat's natural vocal repertoire. The examination of similar acoustic units (e.g. /b/, /e/, /f/, and /n/) which vary primarily by their frequency range provides insight into what differential loss exists. Acoustic unit's /f/ and /n/ showed very little change from normal to lesion recordings, while acoustic unit /e/ showed noticeable distortion to the overall acoustic envelope. Furthermore, acoustic unit /b/ demonstrated significant degradation in its acoustic structure. This variable or differential loss in acoustic complexity exists across the entire vocal repertoire of the rat, which leads to questions regarding which treatment methods may be the most effective for regaining these lost or degraded sounds.

Three acoustic units were too degraded to identify in the diseased rats. Acoustic unit's /v/, /w/, and /x/ were only found in healthy rats. These acoustic units may be more susceptible to degradation. If so, future studies should focus on disease progression in rats and determine at what point these acoustic units are lost. An alternative explanation for why these acoustic units were not found is that they may not have been uttered during the recording time. Each of the missing acoustic units is relatively rare in the healthy rat recordings. A plausible explanation for their absence is the recording time was too short to capture each unit. To examine vocalizations at an acoustic unit level, longer recordings should be collected in the future to address questions regarding the lack of certain acoustic units. Gathering longer recordings has been problematic in the past from an analytical standpoint; however automated approaches such as the one described in this dissertation can reduce the amount of time spent examining each file for abnormalities.

The CSSM analysis provided a clearer picture of acoustic unit degradation than the spectral analysis. Subtle changes in acoustic unit composition that could not be detected via spectral images could be detected using a computational approach. For example, acoustic unit /q/ did not appear to change much in the spectral view; however the CSSM view revealed significant degradation. Detection of previously unidentifiable vocal loss may lead to earlier treatment options (Ciucci et al., 2013). This potential application has yet to be explored, but offers a promising approach for future research which complements current techniques for early detection of biomarkers such as measuring elevated levels of alpha-synuclein oligomers or microRNA brain profiling (Minones-Moyano et al., 2011; Tokuda et al., 2010).

By providing a formal, mathematical measure of disease progression, further studies examining the different physiological pathways may be examined. For example, future work could examine which pathways are involved in the production of the relatively robust group of acoustic units and what neural pathways are involved in the acoustic units most prone to vocal degradation. These studies could provide valuable insight into why differential vocal loss exists. Additionally, discovering these differences can lead to determining whether certain pathways respond better to drug vs. exercise treatment.

The acoustic unit categories described here is based upon human perception and computational description of each sound. Whether the rats perceive these units as distinct requires further research. Determining how they perceive these sounds could also assist with future work determining whether any linguistic rules exist in their vocal communication system. This will also be useful to examine whether rats use specific communication strategies to avoid using the acoustic units which have degraded. Modular variation of a subset of ultrasonic vocalizations in rats has been identified and described using subglottal pressure recordings and electromyographic (EMG) laryngeal muscle activity (Riede, 2011, 2014). Future work expanding the use of subglottal pressure and EMG activity as measurements to describe all acoustic units identified in this chapter is important for understanding why certain acoustic units degraded more than others during the current study.

Chapter References

Allin, J. T., & Banks, E. M. (1971). Effects of temperature on ultrasound production by infant albino rats. *Developmental Psychobiology, 4*(2), 149-156.

Allin, J. T., & Banks, E. M. (1972). Functional aspects of ultrasound production by infant albino rats (*Rattus norvegicus*). *Animal Behaviour, 20*(1), 175-185.

Antoniadis, E. A., & McDonald, R. J. (1999). Discriminative fear conditioning to context expressed by multiple measures of fear in the rat. *Behavioural Brain Research, 101*(1), 1-13.

Barfield, R. J., Auerbach, P., Geyer, L. A., & McIntosh, T. K. (1979). Ultrasonic vocalizations in rat sexual behavior. *American Zoologist, 19*(2), 469-480.

Bergman, H., & Deuschl, G. (2002). Pathophysiology of Parkinson's disease: From clinical neurology to basic neuroscience and back. *Movement Disorders, 17,* S28-S40.

Blanchard, R. J., Blanchard, D. C., Agullana, R., & Weiss, S. M. (1991). 22 kHz alarm cries to presentation of a predator by laboratory rats living in visible burrow systems. *Physiology & Behavior, 50*(5), 967-972.

Braak, H., Ghebremedhin, E., Rub, U., Bratzke, H., & Del Tredici, K. (2004). Stages in the development of Parkinson's disease-related pathology. *Cell and Tissue Research, 318*(1), 121-134.

Brown, R. G., Dowsey, P. L., Brown, P., Jahanshahi, M., Pollak, P., Benabid, A. L., et al. (1999). Impact of deep brain stimulation on upper limb akinesia in Parkinson's disease. *Annals of Neurology, 45*(4), 473-488.

Brudzynski, S. M. (2005). Principles of rat communication: Quantitative parameters of ultrasonic calls in rats. *Behavior genetics, 35*(1), 85-92.

Brudzynski, S. M., Bihari, F., Ociepa, D., & Fu, X. W. (1993). Analysis of 22 kHz ultrasonic vocalization in laboratory rats - long and short calls. *Physiology & Behavior, 54*(2), 215-221.

Brudzynski, S. M., & Chiu, E. M. C. (1995). Behavioral responses of laboratory rats to playback of 22 kHz ultrasonic calls. *Physiology & Behavior, 57*(6), 1039-1044.

Brudzynski, S. M., & Ociepa, D. (1992). Ultrasonic vocalizations of laboratory rats in response to handling and touch. *Physiology & Behavior, 52*(4), 655-660.

Burgdorf, J., Knutson, B., & Panksepp, J. (2000). Anticipation of rewarding electrical brain stimulation evokes ultrasonic vocalization in rats. *Behavioral Neuroscience, 114*(2), 320-327.

Calne, D. (2005). A definition of Parkinson's disease. *Parkinsonism & Related Disorders, 11*, S39-S40.

Carden, S. E., & Hofer, M. A. (1992). Effect of a social companion on the ultrasonic vocalizations and contact responses of 3-day-old rat pups. *Behavioral Neuroscience, 106*(2), 421-426.

Carella, F., Genitrini, S., Bressanelli, M., Soliveri, P., Servello, D., Broggi, G., et al. (2001). Acute effects of bilateral subthalamic nucleus stimulation on clinical and kinematic parameters in Parkinson's disease. *Movement Disorders, 16*(4), 651-655.

Cenci, M. A., Whishaw, I. Q., & Schallert, T. (2002). Animal models of neurological deficits: how relevant is the rat? *Nature Reviews Neuroscience, 3*(7), 574-579.

Charcot, J. M. (1879). *Lectures on the diseases of the nervous system: Delivered at la Salpêtrière*. Philadelphia: Henry C Lea.

Ciucci, M. R., Ahrens, A. M., Ma, S. T., Kane, J. R., Windham, E. B., Woodlee, M. T., et al. (2009). Reduction of dopamine synaptic activity: degradation of 50-kHz ultrasonic vocalization in rats. *Behavioral neuroscience, 123*(2).

Ciucci, M. R., Grant, L. M., Rajamanickam, E. S. P., Hilby, B. L., Blue, K. V., Jones, C. A., et al. (2013). Early Identification and Treatment of Communication and Swallowing Deficits in Parkinson Disease. *Seminars in Speech and Language, 34*(3), 185-202.

Ciucci, M. R., Ma, S. T., Fox, C., Kane, J. R., Ramig, L. O., & Schallert, T. (2007). Qualitative changes in ultrasonic vocalization in rats after unilateral dopamine depletion or haloperidol: A preliminary study. *Behavioural Brain Research, 182*(2), 284-289.

Ciucci, M. R., Vinney, L., Wahoske, E. J., & Connor, N. P. (2010). A translational approach to vocalization deficits and neural recovery after behavioral treatment in Parkinson disease. *Journal of Communication Disorders, 43*(4), 319-326.

Cuomo, V., Cagiano, R., Desalvia, M. A., Maselli, M. A., Renna, G., & Racagni, G. (1988). Ultrasonic vocalization in response to unavoidable aversive stimuli in rats - effects of benzodizepines. *Life Sciences, 43*(6), 485-491.

D'Alatri, L., Paludetti, G., Contarino, M. F., Galla, S., Marchese, M. R., & Bentivoglio, A. R. (2008). Effects of bilateral subthalamic nucleus stimulation and medication on parkinsonian speech impairment. *Journal of Voice, 22*(3), 365-372.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Clusters of Deviant Speech Dimensions in Dysarthrias. *Journal of Speech and Hearing Research, 12*(3), 462-&.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12*(2), 246-270.

Duchin, L. E. (1990). The evolution of articulate speech - Comparative anatomy of the oral cavity in Pan and Homo. *Journal of Human Evolution, 19*(6-7), 687-697.

Fleming, S. M., Delville, Y., & Schallert, T. (2005). An intermittent, controlled-rate, slow progressive degeneration model of Parkinson's disease: antiparkinson effects of Sinemet and protective effects of methylphenidate. *Behavioural Brain Research, 156*(2), 201-213.

Fox, C. M., Morrison, C. E., Ramig, L. O., & Sapir, S. (2002). Current perspectives on the Lee Silverman Voice Treatment (LSVT) for individuals with idiopathic Parkinson disease. *American Journal of Speech-Language Pathology, 11*(2), 111-123.

Fulceri, F., Biagioni, F., Lenzi, P., Falleni, A., Gesi, M., Ruggieri, S., et al. (2006). Nigrostriatal damage with 6-OHDA - Validation of routinely applied procedures. *Cellular and Molecular Mechanisms of Drugs of Abuse and Neurotoxicity: Cocaine, Ghb, and Substituted Amphetamines, 1074*, 344-348.

Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J. L., & Gates, S. (1998). Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurology, 11*(3), 131-137.

Inagi, K., Schultz, E., & Ford, C. N. (1998). An anatomic study of the rat larynx: Establishing the rat model for neuromuscular function. *Otolaryngology-Head and Neck Surgery, 118*(1), 74-81.

Johnson, A. M., Doll, E. J., Grant, L. M., Ringel, L., Shier, J. N., & Ciucci, M. R. (2011). Targeted training of ultrasonic vocalizations in aged and Parkinsonian rats. *Journal of visualized experiments: JoVE, 54*.

Jurgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews, 26*(2), 235-258.

Jurgens, U. (2009). The neural control of vocalization in mammals: A review. *Journal of Voice, 23*(1), 1-10.

Kane, J. R., Ciucci, M. R., Jacobs, A. N., Tews, N., Russell, J. A., Ahrens, A. M., et al. (2011). Assessing the role of dopamine in limb and cranial-oromotor control in a rat model of Parkinson's disease. *Journal of Communication Disorders, 44*(5), 529-537.

Knutson, B., Burgdorf, J., & Panksepp, J. (1998). Anticipation of play elicits high-frequency ultrasonic vocalizations in young rats. *Journal of Comparative Psychology, 112*(1), 65-73.

Knutson, B., Burgdorf, J., & Panksepp, J. (1999). High-frequency ultrasonic vocalizations index conditioned pharmacological reward in rats. *Physiology & Behavior, 66*(4), 639-643.

Knutson, B., Burgdorf, J., & Panksepp, J. (2002). Ultrasonic vocalizations as indices of affective states in rats. *Psychological Bulletin, 128*(6), 961-977.

Logemann, J. A., Fisher, H. B., Boshes, B., & Blonsky, E. R. (1978). Frequency and Co-occurrence of Vocal-tract Dysfunctions in Speech of a Large Sample of Parkinson Patients. *Journal of Speech and Hearing Disorders, 43*(1), 47-57.

Marshall, J. F., & Gotthelf, T. (1979). Sensory inattention in rats with 6-hydroxydopamine-induced degeneration of ascending dopaminergic neurons - Apomorphine-induced reversal of deficits. *Experimental Neurology, 65*(2), 398-411.

Meredith, G. E., & Kang, U. J. (2006). Behavioral models of Parkinson's disease in rodents: A new look at an old problem. *Movement Disorders, 21*(10), 1595-1606.

Minones-Moyano, E., Porta, S., Escaramis, G., Rabionet, R., Iraola, S., Kagerbauer, B., et al. (2011). MicroRNA profiling of Parkinson's disease brains identifies early downregulation of miR-34b/c which modulate mitochondrial function. *Human Molecular Genetics, 20*(15), 3067-3078.

Mollaei, F., Shiller, D. M., & Gracco, V. L. (2013). Sensorimotor adaptation of speech in Parkinson's Disease. *Movement Disorders, 28*(12), 1668-1674.

Pavese, N. (2012). PET studies in Parkinson's disease motor and cognitive dysfunction. *Parkinsonism & Related Disorders, 18*, S96-S99.

Plowman-Prine, E. K., Sapienza, C. M., Okun, M. S., Pollock, S. L., Jacobson, C., Wu, S. S., et al. (2009). The Relationship Between Quality of Life and Swallowing in Parkinson's Disease. *Movement Disorders, 24*(9), 1352-1358.

Portfors, C. V. (2007). Types and functions of ultrasonic vocalizations in laboratory rats and mice. *Journal of the American Association for Laboratory Animal Science, 46*(1), 28-34.

Ramig, L. O., Fox, C., & Sapir, S. (2008). Speech treatment for Parkinson's disease. *Expert Review of Neurotherapeutics, 8*(2), 297-309.

Riede, T. (2011). Subglottal pressure, tracheal airflow, and intrinsic laryngeal muscle activity during rat ultrasound vocalization. *Journal of Neurophysiology, 106*(5), 2580-2592.

Riede, T. (2014). Rat Ultrasonic Vocalization Shows Features of a Modular Behavior. *Journal of Neuroscience, 34*(20), 6874-6878.

Sapir, S. (2014). Multiple factors are involved in the dysarthria associated with Parkinson's disease: A review with implications for clinical practice and research. *Journal of Speech, Language, and Hearing Research*.

Sapir, S., Ramig, L., & Fox, C. (2008). Speech and swallowing disorders in Parkinson disease. *Current Opinion in Otolaryngology & Head and Neck Surgery, 16*(3), 205-210.

Schallert, T., Fleming, S. M., Leasure, J. L., Tillerson, J. L., & Bland, S. T. (2000). CNS plasticity and assessment of forelimb sensorimotor outcome in unilateral rat models of stroke, cortical ablation, parkinsonism and spinal cord injury. *Neuropharmacology, 39*(5), 777-787.

Thomas, D. A., Takahashi, L. K., & Barfield, R. J. (1983). Analysis of ultrasonic vocalizations emitted by intruders during aggressive encounters among rats (*Rattus norvegicus*). *Journal of Comparative Psychology, 97*(3), 201-206.

Tillerson, J. L., Cohen, A. D., Philhower, J., Miller, G. W., Zigmond, M. J., & Schallert, T. (2001). Forced limb-use effects on the behavioral and neurochemical effects of 6-hydroxydopamine. *Journal of Neuroscience, 21*(12), 4427-4435.

Tokuda, T., Qureshi, M. M., Ardah, M. T., Varghese, S., Shehab, S. A. S., Kasai, T., et al. (2010). Detection of elevated levels of alpha-synuclein oligomers in CSF from patients with Parkinson disease. *Neurology, 75*(20), 1766-1772.

Tonoue, T., Ashida, Y., Makino, H., & Hata, H. (1986). Inhibition of shock-elicited ultrasonic vocalizations by opioid-peptides in the rat - A psychotropic effect. *Psychoneuroendocrinology, 11*(2), 177-184.

Ungerstedt, U. (1970). Striatal dopamine release after amphetamine or nerve degeneration revealed by rotational behaviour. *Acta Physiologica Scandinavica, Supplementum 367*, 49-68.

Weismer, G. (1984). Articulatory characteristics of Parkinsonian dysarthria: Segmental and phrase-level timing, spirantization, and glottal-supraglottal coordination. In M. McNeil, J. Rosenbek & A. Aronson (Eds.), *The dysarthrias: Physiology, acoustics, perception, management* (pp. 101-130). San Diego: College-Hill.

"We are not students of some subject matter, but students of problems. And problems may cut right across the borders of any subject matter or discipline." (Popper, 1963)

Chapter 7: Conclusions & general discussion

Understanding the vocalizations of non-human species requires the collaborative input of several areas of research (Mitani and Marler, 1989; Potter et al., 1994; McCowan et al., 1999; Mellinger & Clark, 2000; Hurford, 2004; Yip, 2006; Harris & Skowronski, 2006). The studies described in this dissertation offers support for including computational and linguistic perspectives in examining bioacoustics questions. The hypothesis that white-handed gibbon vocalizations consist of acoustic units akin to human phonemes is supported by the development of an automated method of segmenting and classifying calls across a wide variety of social contexts. It is further supported by examining the sequential variation in acoustic unit use for different predator warning contexts. The development of a fully automated process was also used in this dissertation to examine differential vocal loss in a rat model for Parkinson's disease. Evidence from this dissertation advances the bioacoustics field by offering a formal and distinct way of characterizing acoustic units in non-human animals. This approach is especially helpful for addressing questions focused on understanding call meaning.

Dissertation Summary

Computational approaches to bioacoustics research

Chapter 4 examined white-handed gibbon vocalizations for acoustic units using a novel computational approach. Acoustic units were previously proposed as a fundamental unit of sound in gibbon vocal communication by using a standard bioacoustics analytical technique of hand-labeling each sound (Dassow, 2010). This chapter offers empirical support for these divisions by examining the Cepstral Self-Similarity Matrix (CSSM) of each acoustic unit. Further support for the use of a CSSM approach to segmenting and classifying vocalizations is provided by examining samples of words from English. The accurate classification of English and the agreement with the intuitive segmentation of gibbon calls suggests acoustic units do exist in the vocal repertoire of white-handed gibbons. Additionally, a novel acoustic unit is described from a pair of captive gibbons. This acoustic unit is given during a specific type of play behavior and is acoustically distinct from the rest of the gibbon's vocal repertoire. Though there is currently not enough evidence to claim vocal learning, there is a vocal novelty to this call which warrants further investigation into its origin and function. This study provides the foundation for the next two studies which offer more support for the presence of acoustic units in gibbons by exploring meaning associated with distinct sequences of calls and for the presence of acoustic units in rat vocalizations.

Linguistically inspired analysis of white-handed gibbon vocal communication

Chapter 5 re-examines meaning in white-handed gibbon vocalizations, by exploring the variation of acoustic units in response to various predator models. Previous work examined meaning by examining gross categories of sound during a short sequence of a group response (Clarke et al., 2006; Clarke et al., 2012). By identifying every acoustic unit in the first bout of calling for each predator model (clouded leopard, tiger – running and stationary, snake, cockatoo, and veterinarian), further complexity within the meaning of these calls has been discovered. Specifically, wild gibbons use an initial, motivational sequence to alarm others of the presence of a predator and then follow this motivational sequence with one or two distinct referential sequences. Within the tiger response, a distinction between moving and stationary models could be detected. This discovery adds white-handed gibbons to a list of species which use both motivational and referential context in their vocalizations (Macedonia & Evans, 1993; Evans et al., 1993; Manser et al., 2001; Seyfarth & Cheney, 2003; Hollen & Manser, 2007).

Captive gibbons differed in their response to an encounter with their veterinarian suggesting they are capable of modifying their use of warning calls to the current, perceived level of threat. The gibbons which had no physical encounter with their veterinarian over the past several years limited their response to the general motivational warning; whereas the gibbons which had a recent physical encounter with their veterinarian produced a motivational and referential response. Furthermore, the gibbons at the Vilas Zoo responded with a unique sequence to the warning call of a cockatoo suggesting gibbons are capable of utilizing the warning calls of other species. Interspecies utilization of warning calls has already been shown in a wide variety of taxa (Seyfarth & Cheney, 1990; Oda & Masataka, 1996; Zuberbühler, 2000; Fichtel, 2004; Magrath et al., 2007; Templeton & Greene, 2007; Vitousek et al., 2007; Fallow & Magrath, 2010; Seiler et al., 2013). The captive gibbon's response to the cockatoo warning call differed significantly from the other predator warning calls which suggest gibbons respond to more environmental cues than previously thought.

Differential vocal loss and recovery in a rat model of Parkinson's disease

Parkinsonism leads to a variety of sensorimotor deficits including diminished quality and quantity of vocalizations (Ciucci et al., 2007; Ciucci et al., 2009). Chapter 6 explores vocal loss in a rat model of Parkinson's disease by using the CSSM framework to identify acoustic units in healthy and lesioned rats. CSSM views of healthy rat vocalizations reveal remarkable complexity within the generally accepted categories of 22 kHz, 40 kHz, and 50 kHz calls. Tracing each of these acoustic units as they are impaired from lesioning the brain of healthy rats revealed an overall loss of acoustic complexity with some acoustic units appearing more degraded than others. Furthermore, degradation in acoustic units was detected using the CSSM approach in cases which were not identified using a spectral approach. The results of this comparison of methods suggest earlier detection of vocal loss is possible when using a CSSM analysis. Previous research has explored other factors such as alpha-synuclein oligomer levels or microRNA brain profiling as methods for early detection of PD (Minones-Moyano et al., 2011; Tokuda et al., 2010). Examining vocalizations for subtle degradation using the CSSM approach is another tool which would be useful for detecting early symptoms of PD. Earlier detection, coupled with differential loss suggests a targeted form of treatment would be useful.

Conclusions

The research described in this dissertation advances our collective understanding of vocal communication in vertebrates. By developing an automated approach to segmenting acoustic units in gibbon vocalizations, we were able to explore the complexity of their responses to predators. This led to the discovery of both motivational and referential content in their responses to predator models which adds gibbons to a growing list of species whose vocal responses contain complex signaling. Additionally, an automated approach to segmenting vocalizations led to the discovery of differential vocal loss in a rat model of Parkinson's disease.

Future directions

Examining the vocalizations of other species for acoustic units using the CSSM approach could provide insight into common, evolutionary bases for vocal complexity. Since the CSSM framework does not use acoustic features unique to gibbons, there is reason to believe it would work well for analyzing other species. Future work should focus on the following:

1) What evolutionary commonalities can be found? It has long been understood that certain aspects of vocalizations are inherently tied to anatomical constraints of organisms. For example, with rare exception, the size of an animal restricts whether it is capable of producing ultrasonic, sonic or subsonic frequencies. Examining other species using the CSSM approach may reveal additional commonalties across various groups. This examination could also extend to the lesser studied species including reptiles and fish.
2) Do other linguistic commonalities exist in vocal communication systems? Previous debates have discussed certain core linguistic properties, such as natural classes, identity computation, phonemes, or stress systems, as criteria for accepting a given species has an

advanced communication system. Using the CSSM approach, other features, such as co-articulation, may also be searched for. Higher level structure, such as morphological complexity, can also be searched for with greater accuracy.

3) What else can we learn about the communication system of white-handed gibbons?
    a) Given the variation present in the predator warning calls, follow up studies searching for additional behavioral variation which could correlate to meaning would be warranted. Specifically, looking for differences associated with group size, age of offspring present, distance between group members or other environmental factors which may influence variation in vocal responses to predator models is needed. While the presentation of predator models was controlled for to ensure minimal stress, whether or not wild predators encounters occurred in the previous 24 hours was not recorded. Setting up trail cameras to capture the movement of other organisms could offer valuable insight into the motivational states of the gibbons. Additionally, in the hour leading up to the predator presentation, it would be useful to obtain a recording of the ambient noise and conditions, including what birds are present and whether or not they are giving warning calls. Evidence from the captive gibbons suggests gibbons do respond to the warning calls of another species. This may also affect the motivational states of the wild gibbons prior to model presentation. All of these factors may help discern why there is so much variation within the predator warning calls of white-handed gibbons. This may also be used to explain why the warning calls appear to be so long.
    b) Transitional probability variation existed within the first word-like sequence of each warning call. Whether this is solely a motivational sequence or possibly also a referential sequence requires further evaluation. To determine if referential content exists, additional behavioral data will be necessary. This could be added to the suggested work outline above.

    c) Word-like sequences have been identified for predator warning calls, but what other "words" can be found in non-predatory contexts? Is it possible to find sequences specific to highly sought after food items or social encounters? Do the groups which encounter visits from pileated gibbons (*Hylobates pileatus*) produce word-like sequences to identify these similar, yet distinct primates? Some of these questions require gathering more recordings, but some could be examined using the current data set.
    d) The zoo gibbons provided a valuable insight into novel or rarely observed vocal behavior. Question which have yet to be answered are what do the bleats mean and how did they develop? Do all gibbons produce these calls as infants? Given the quiet nature of these calls, it would be unlikely to encounter them in the wild, but perhaps they are not as rare as they seem. Beyond the unusual calls, testing the captive gibbons for their ability to learn novel sequences, such as food calls, would be useful. Particularly, the ability to learn sequences could be beneficial for captive release programs. Ensuring released gibbons are fully capable of foraging independently could be assisted by monitoring the number of food associated calls given in a day. Learning novel sequences could also be used to teach predator warning calls given appropriate contextual information. These skills would be essential for the success of any captive release program.

4) What other translational medical benefits could be gained from studying rat vocalizations? Further research examining the effects various drug treatments have on regaining degraded vocalizations is necessary. Specifically, whether certain types of degraded acoustic units respond better to certain types of drugs. Also testing rats to determine whether or not they alter their communicative strategies to avoid using the degraded or diminished acoustic units is necessary. Additionally, testing whether targeted therapy is a useful treatment for regaining degraded acoustic units could provide valuable insight into whether or not this should be tested with human patients. While testing vocalizations for degradation is minimally invasive, it does require additional time from patients and researchers. If

differential degradation can be detected in humans using the CSSM approach, further research exploring its value as a potential tool for earlier detection of PD symptoms would be merited.

Chapter References

Ciucci, M. R., Ahrens, A. M., Ma, S. T., Kane, J. R., Windham, E. B., Woodlee, M. T., et al. (2009). Reduction of dopamine synaptic activity: degradation of 50-kHz ultrasonic vocalization in rats. *Behavioral neuroscience, 123*(2).

Ciucci, M. R., Ma, S. T., Fox, C., Kane, J. R., Ramig, L. O., & Schallert, T. (2007). Qualitative changes in ultrasonic vocalization in rats after unilateral dopamine depletion or haloperidol: A preliminary study. *Behavioural Brain Research, 182*(2), 284-289.

Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *Plos One, 1*(1).

Clarke, E., Reichard, U. H., & Zuberbühler, K. (2012). The anti-predator behaviour of wild white-handed gibbons (Hylobates lar). *Behavioral Ecology and Sociobiology, 66*(1), 85-96.

Dassow, A. M. (2010). *Vocal repertoire of the white-handed gibbon (Hylobates lar).* Unpublished Masters of Zoology, University of Wisconsin-Madison.

Evans, C. S., Evans, L., & Marler, P. (1993). On the meaning of alarm calls - functional reference in an avian vocal system. *Animal Behaviour, 46*(1), 23-38.

Fallow, P. M., & Magrath, R. D. (2010). Eavesdropping on other species: mutual interspecific understanding of urgency information in avian alarm calls. *Animal Behaviour, 79*(2), 411-417.

Fichtel, C. (2004). Reciprocal recognition of sifaka (Propithecus verreauxi verreauxi) and redfronted lemur (Eulemur fulvus rufus) alarm calls. *Animal Cognition, 7*(1), 45-52.

Harris, J. G., & Skowronski, M. D. (2006). *Automatic speech processing methods for bioacoustic signal analysis: a case study of cross-disciplinary acoustic research.* Paper presented at the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP.

Hollen, L. I., & Manser, M. B. (2007). Motivation before meaning: Motivational information encoded in meerkat alarm calls develops earlier than referential information. *American Naturalist, 169*(6), 758-767.

Hurford, J. R. (2004). Human uniqueness, learned symbols and recursive thought. *European Review, 12*(4), 551-565.

Macedonia, J. M., & Evans, C. S. (1993). Variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology, 93*(3), 177-197.

Magrath, R. D., Pitcher, B. J., & Gardner, J. L. (2007). A mutual understanding? Interspecific responses by birds to each other's aerial alarm calls. *Behavioral Ecology, 18*(5), 944-951.

Manser, M. B., Bell, M. B., & Fletcher, L. B. (2001). The information that receivers extract from alarm calls in suricates. *Proceedings of the Royal Society B-Biological Sciences, 268*(1484), 2485-2491.

McCowan, B., Hanser, S. F., & Doyle, L. R. (1999). Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Animal Behaviour, 57*, 409-419.

Mellinger, D. K., & Clark, C. W. (2000). Recognizing transient low-frequency whale sounds by spectrogram correlation. *Journal of the Acoustical Society of America, 107*(6), 3518-3529.

Minones-Moyano, E., Porta, S., Escaramis, G., Rabionet, R., Iraola, S., Kagerbauer, B., et al. (2011). MicroRNA profiling of Parkinson's disease brains identifies early downregulation of miR-34b/c which modulate mitochondrial function. *Human Molecular Genetics, 20*(15), 3067-3078.

Mitani, J. C., & Marler, P. (1989). A phonological analysis of male gibbon singing behavior. *Behaviour, 109*, 20-45.

Oda, R., & Masataka, N. (1996). Interspecific responses of ringtailed lemurs to playback of antipredator alarm calls given by Verreaux's sifakas. *Ethology, 102*(6), 441-453.

Popper, K. R. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Routledge and Kegan Paul.

Potter, J. R., Mellinger, D. K., & Clark, C. W. (1994). Marine mammal call discrimination using artificial neural networks. *Journal of the Acoustical Society of America, 96*(3), 1255-1262.

Seiler, M., Schwitzer, C., Gamba, M., & Holderied, M. W. (2013). Interspecific semantic alarm call recognition in the solitary sahamalaza sportive lemur, Lepilemur sahamalazensis. *Plos One, 8*(6).

Seyfarth, R., & Cheney, D. (1990). The assessment by vervet monkeys of their own and another species alarm calls. *Animal Behaviour, 40*, 754-764.

Seyfarth, R. M., & Cheney, D. L. (2003). Meaning and emotion in animal vocalizations. In P. Ekman, J. J. Campos, R. J. Davidson & F. B. M. DeWaal (Eds.), *Emotions inside out - 130 Years after Darwin's the Expression of the Emotions in Man and Animals* (Vol. 1000, pp. 32-55).

Templeton, C. N., & Greene, E. (2007). Nuthatches eavesdrop on variations in heterospecific chickadee mobbing alarm calls. *Proceedings of the National Academy of Sciences of the United States of America, 104*(13), 5479-5482.

Tokuda, T., Qureshi, M. M., Ardah, M. T., Varghese, S., Shehab, S. A. S., Kasai, T., et al. (2010). Detection of elevated levels of alpha-synuclein oligomers in CSF from patients with Parkinson disease. *Neurology, 75*(20), 1766-1772.

Vitousek, M. N., Adelman, J. S., Gregory, N. C., & St Clair, J. J. H. (2007). Heterospecific alarm call recognition in a non-vocal reptile. *Biology Letters, 3*(6), 632-634.

Yip, M. J. (2006). The search for phonology in other species. *Trends in Cognitive Sciences, 10*(10), 442-446.

Zuberbühler, K. (2000). Interspecies semantic communication in two forest primates. *Proceedings of the Royal Society B-Biological Sciences, 267*(1444), 713-718.