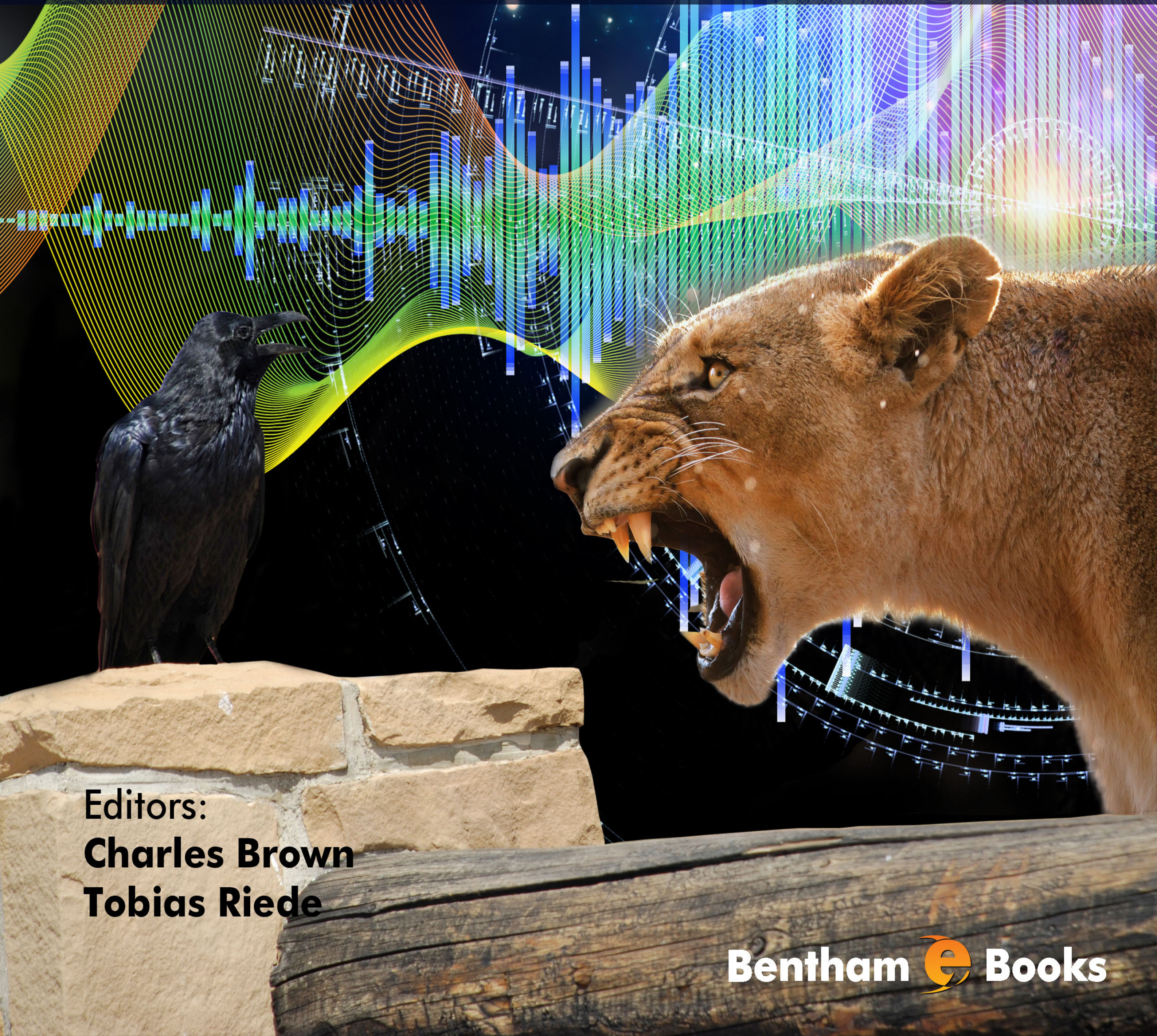
eISBN: 978-1-68108-317-9
ISBN: 978-1-68108-318-6

COMPARATIVE BIOACOUSTICS AN OVERVIEW



Editors:
Charles Brown
Tobias Riede

Bentham e Books

Comparative Bioacoustics: An Overview

Edited by

Charles Brown
Department of Psychology
University of South Alabama
Mobile, AL 36688
USA

&

Tobias Riede
Department of Physiology
Midwestern University
Glendale, AZ 85308
USA

Comparative Bioacoustics: An Overview

Editors: Charles Brown and Tobias Riede

ISBN (eBook): 978-1-68108-317-9

ISBN (Print): 978-1-68108-318-6

Published by Bentham Science Publishers – Sharjah, UAE.

First published in 2017.

Bentham Science Publishers Ltd.
Executive Suite Y - 2
PO Box 7917, Saif Zone
Sharjah, U.A.E.
Email: subscriptions@benthamscience.org

CONTENTS

PREFACE

WHY WRITE A BIOACOUSTICS METHODS EBOOK?

The intent of this volume is twofold: (a) to promote multidisciplinary and comparative analyses of sound production, propagation, and perception in biological organisms and (b) provide a source of relevant methods for the novice bioacoustician. The first goal of this initiative was approached by crossing disciplinary boundaries. The exchange between scientists in biology, psychology, neuroscience, engineering, anthropology, speech, voice, hearing, and related sciences appears critical for bioacousticians. Over the course of our careers we have collaborated with personnel trained in a wide variety of disciplines including radio broadcast engineers, speech and hearing scientists, psychologists, physiologists, anatomists, epidemiologists, ethologists, anthropologists, physicists, electrical engineers, veterinarians, zoologists, zoo keepers, vocal musicians, and sound archivists. We have collaborated with faculty members, postdoctoral fellows, as well as graduate and undergraduate students. All of these individuals brought a fresh and unique perspective to the study of bioacoustics; they all had a domain of expertise in some area of bioacoustics, and they all exhibited areas of limited familiarity to topics (and methodologies) of interest to other members of the bioacoustics community. Furthermore, the wealth of research methods available, in many instances dictates the collaborative approach.

Vocal communication is undoubtedly one of the most fascinating and complex behaviors that animals perform. It requires the analysis at all four levels (mechanism, adaptive significance, development and evolution) identified by Tinbergen (Bateson, Laland, 2013). Historically, bioacoustics research started as a small sub-discipline in ethology (*e.g*., Tembrock, 1959; Busnel, 1963; Sebeok, 1977; Nikolskii, 1984). Over the last five decades not only has an enormous amount of knowledge been accumulated (*e.g*., Kroodsma, Miller, 1996; Gerhardt, Huber, 2002; Bradbury, Vehrencamp, 2011; Hopp *et al*., 2012; Hedwig, 2014; Wiley, 2015), but new methods have been employed or developed which require more know-how than just operating a handheld microphone and tape recorder and subsequent acoustic analysis in the temporal and spectral domain. In light of the second goal *i.e*. providing a source for relevant methods, this volume can represent only a beginning. The intent of this book was also to initiate the creation of a repository for the research methodologies used by bioacousticians. The book may serve as a resource intended to help members of the bioacoustics community to communicate more skillfully with one another, to serve as a reference for training our students, to promote greater collaboration, to recruit new investigators, and to help investigators adopt new techniques to strengthen their research programs. In this respect several years ago I (CB) was contacted by Asif Ghazanfar of the Department of Psychology at Princeton University inquiring if there was a source that would essentially "coach" a graduate student or faculty member how to set up an excised larynx bench for their laboratory. We observed that there was no bioacoustical methods publication that was intended to instruct members of the research community how to learn a new technique or initiate a research project in an area in which they had no prior exposure or training, but there should be! Asif's inquiry is one of the key exchanges we had with our colleagues over the past several years that served as inspiration for developing this methods eBook.

Comparative bioacoustics is extraordinarily broad in scope. It includes the study of sound propagation, dispersion, attenuation, absorption, reverberation, and signal degradation; as well as sound detection, recognition, and classification in both marine and terrestrial organisms (including humans). This research is informed by an understanding of the mechanisms underlying sound generation and aural reception, as well as the anatomy and physiology of the organs dedicated to these functions, and their ontogeny and development. Furthermore, it includes studies of the ontogeny of vocal behavior and the relationship between the environment and the acoustic behaviors of all conspecific organisms resident in the local biome. Comparative bioacoustic researchers have developed signal processing algorithms for taxonomic classification of the acoustically conspicuous biota in the habitat, and for the calculation of biodiversity. It includes studies of the effects of industrial noise on the integrity of urban, suburban and rural terrestrial environments, and the impact of man-made noise on marine organisms. It includes the application of acoustic instrumentation for pest-control, the abatement of collisions with wildlife, and wildlife monitoring. The frequency range of the acoustic signals under consideration is equally broad. It includes studies of echolocation, and the utilization of infrasonic, sonic and ultrasonic signals.

Last but not least, the ebook format provided a unique opportunity for the development of a multimedia methods book. As bioacousticians we publish sound spectrograms, and descriptions of audio signals. We conduct measurements of the acoustic profile of a variety of habitats, and we measure and describe the mechanisms of animal sound production, but the acoustic signals that are the focus of our field are inaccessible to our readership. A key feature of this volume is our attempt to bring the field of bioacoustics fully into the 21st Century by promoting the expectation of the integration of audio and video clips into a scholarly text. In this volume readers have the ability to "hear" associated audio files by "clicking" on the speaker icon linked to a sound spectrogram or to the text itself. Where appropriate, short video clips have also been included. One goal of this volume has been to capitalize on recent developments in digital publishing in order to improve dissemination of information concerning the central methods and findings in the field of bioacoustics. One long-term goal of this effort is to encourage the creation of scholarly publications in which scientists are encouraged, and potentially expected, to archive rare audio recordings in association with their notes, measurements and observations. Audio archiving should be granted full recognition as an important component of scholarship in bioacoustics, and archived audio files give the scientific community an opportunity to study how habitats change over time in terms of their acoustic properties as the world experiences a change in climate, shifts in the density of human habitation, invasion of exotic species and so forth. For the field of bioacoustics, the benefit of such archiving is obvious and this volume strives to elevate the recognition of scholarship associated with the archiving enterprise.

AN OVERVIEW OF THIS VOLUME

This methods book would be encyclopedic if it exhaustively addressed all the domains encompassed by comparative bioacoustics. This eBook is merely a starting point for building a methodology repository. We hope that it will lead to subsequent editions that up-date and expand the methodologies presented here. Because this book is electronic where appropriate we include both video and audio files to illustrate the author's key points. This book is partitioned into four sections. Part 1 is composed of two chapters laying the ground work for an understanding of acoustics. Chapter 1 by Larsen and Wahlberg is about the physics of sound in air and how it is produced. Bioacousticians handle sound in four major contexts: sound recording, sound analysis, sound synthesis, and sound playbacks. The text is designed to achieve a "sweet spot", so that it is intended to be appropriate for novices (such as advanced undergraduate students), but to retain sufficient rigor so that the treatment is appropriate for students seeking preparation for a scholarly career in some aspect of comparative bioacoustics. The topics covered include the nature of sound waves in air, sound pressure, intensity and power. The reader is introduced to the concept of sound fields and sound propagation. The relationship between the period of the wave and wavelength, the concept of acoustic impedance, the acoustic near field and the acoustic far field are discussed. The reader is introduced to sound calibration, and the differences between a diffuse sound field, a semi-reverberant field, and a closed sound field. Frequently novices are unaware of phenomena like diffraction, the concept of an acoustic monopole, acoustic dipole, and the problem of destructive and constructive interference, and their treatment in this chapter is designed to help fledgling bioacousticians think about biological sound sources, and sound measurement with greater sophistication. The last section in this chapter considers the idea that it may be desirable for animals to beam or direct their signals towards an intended recipient, the design of a biological acoustic horn, and the strategies that investigators may employ to strive to measure the directional properties of animal acoustic displays is discussed.

Wahlberg and Larsen expand their treatment on the physics of sound and sound sources in chapter 2. This chapter is devoted to an in depth treatment of sound propagation both in air and in water. In nature the sound wave incident at the receiver is reduced in amplitude relative to the emitted signal, and more importantly the path from sender to recipient is rarely straight, and the received signal is typically distorted, degraded or blurred due to frequency dependent absorption of the sound wave by the medium within which it is propagated, and also by scattering, reflection and absorption of the signal with large and small objects encountered in the environment during its propagation. The reader is introduced to the ideas of geometric attenuation, excess attenuation, absorption by the medium and sound propagation from one medium to another. In the natural world the media is rarely homogeneous. Both meteorological and underwater conditions change the velocity of sound propagation, and the media becomes stratified causing refraction or curvature in the acoustic ray. To complicate matters further the presence of water currents and wind in the atmosphere cause turbulence in the media. The media moves in loops or eddies producing distortions of the wavefront. The efficacy of both echolocation systems and acoustic communication systems is dependent upon specializations in signal emission strategies, or perceptual processing strategies devoted to the task of assessing the likelihood that differences between two successive waveforms incident at the receiver are due to instabilities in the propagation of the signal in the habitat, or if these differences are due to changes in the waveform reflected from the target in the case of echo location systems, or are due to intended differences in the “design” of the emitted signal in the case of acoustic communication systems. These issues are of fundamental importance for many questions addressed in comparative bioacoustics.

Part 2 is composed of three chapters which focus on vocal production in terrestrial vertebrates. Chapter 3 by Brown and Riede reviews the biomechanics of sound production by the larynx. The myoelastic-aerodynamic theory of voice production accounts for most of the vocalizations produced by frogs, reptiles, and mammals. Accordingly, sound is generated by a repeating cycle in which the glottis opens and closes disrupting or modulating the transglottic air flow. The larynx with a small range in variation in physical size is remarkable in its capacity to allow for a very broad range in the fundamental frequency of oscillation both within a species, and across taxa. This phenomenon is dependent upon the viscoelastic properties of the superficial layers of the vocal folds, the layered composition of the vocal folds, and the capacity to selectively adjust the stiffness of the tissue layers within the folds. The oscillation of the vocal folds behaves almost like a string. Stretching the vocal folds causes the tissue to vibrate at higher rates as the tissue becomes stiffened much like a string. Comparative mammalian data show that there are significant species differences in the composition of the vocal folds that should impact on their viscoelastic properties. Species differ with respect to the number of layers of the lamina propria within the vocal folds, and the distribution and density of fibrous proteins, hyaluronan and fat cells. Differences in the composition of the vocal folds will alter their biomechanical properties, and the biomechanical properties of the vocal folds can be measured by two complimentary experimental procedures. In the first procedure the vocal folds are dissected from the larynx, and tissue stress is calculated using force and length measurements. This procedure requires prior macroscopic and microscopic investigation of the morphology of the vocal folds for the species in question so that the tissue is dissected appropriately. The excised larynx procedure is the second experimental manipulation commonly used to explore species differences in the biomechanics of the larynx. In this preparation the laryngeal complex is mounted on a hollow tube, and the passage of air through a "pseudo-trachea" initiates oscillations of the vocal folds much like during natural vocal behavior.

The syrinx, the avian vocal organ is a unique sound source among tetrapods. Chapter 4 by Goller reviews sound production and modification in birds, and the methodology employed to study the vocal behavior of birds. Birds are equipped with one or two sound sources within the syrinx. In songbirds, the left and right syrinx are controlled independently permitting the possibility of simultaneous or asynchronous phonation. There is substantial morphological variation in the shape of the cartilages, muscles and membranes and labia in the syrinx between different families of birds, and we begin slowly to make sense of how the morphological variability facilitates a species' typical vocal repertoire. Investigating sound production in birds is a challenge because structures are very small. Goller describes a number of techniques which have been developed. For example, airflow has been measured in spontaneously singing birds with small microbead thermistors implanted in the airway. Fine wire EMG electrodes permit recording the activity of syringeal muscles, and the nerves innervating these muscles can be severed to observe their role in the control of singing. After sound is generated by the syrinx, it travels through a vocal tract before radiated from the beak. The suprasyringeal vocal tract filters or shapes the sound generated in the syrinx. Some aspects of the vocal tract are highly dynamic, including the beak and hyoid movements which have been studied by x-ray filming.

The idea that the geometry of the airway shapes the sound produced by the syrinx or larynx is known as the source-filter theory. This theory is addressed by Hunter and Ludwigsen in chapter 5. The authors introduce essential concepts of acoustic filters and resonances beginning with two archetypes of acoustic filters: the Helmholtz resonator and the pipe resonator. Knowing the dimensions of the respective air filled cavity allows acousticians to make fairly precise predictions about its resonance characteristics. These calculations are explained in simple mathematical terms for the Helmholtz and the pipe resonator. The text also presents an approach which allows investigators to empirically determine the resonance or filter properties of an organism's airway. The organism's filter is acoustically excited either by one of three signals: a swept sine waves, white noise or by a pulse. This allows the demonstration of resonance frequencies in the frequency spectrum of the radiated sound. Finally, the authors apply these concepts to human and animal vocalizations and demonstrate how the properties of the source and filter can be extracted using spectrograms and power spectra. The sounds we deal with are most often those radiated from the lips of a speaking human or animal subject or from the beak of a bird. The sound therefore contains features that pertain to source characteristics and others that can be related to filter properties. To fully understand communication signals, bioacousticians must consider the properties of the source, and the properties of the filter, both parameters reveal relevant information about the sender. However, as noted in chapter 5, it is also important to understand the limitations and trade-offs of a source-filter approach to the acoustic analysis of vocal production.

Part 3 is focused on sound analysis methodologies. This section is comprised of five chapters. The topics discussed include sound classification using animal preference methods and computational approaches. The latter encompass sound conditioning, nonlinear dynamics, and sound classification using artificial neural networks and hidden Markov model methodologies.

Clearly bioacousticians cannot depend solely on the output of automated sound classification systems, and additional methods are needed to validate the output of software classification systems, to determine how well they line up with the perceptual judgments and preferences of living organisms. A variety of behavioral research methods focused on acoustic preference testing are described by Riebel in chapter 6. Vocal displays are often used to advertise for a mate, and many species are attracted to mating signals, and the opportunity to hear these calls can be used as a reward or reinforcement in operant conditioning tasks. Acoustic preferences can be measured through a wide range of assays including phonotaxis tests, loudspeaker approach assays using a maze, nest cavity choice tests, copulation solicitation assays, nest building assays, and vocal responsiveness assays. In addition to passive exposure tests, the sound playback can be triggered in operant preference tests using key pecking and perch hopping technologies. Behavioral preference tests can be correlated with physiological measures including heart rate, hormone profiles, maternal resource allocation and fMRI imaging studies. Riebel argues that these methods are not all equally suitable for every research species or research question, and researchers should seek to use a combination of methods to establish the internal and external validity of their data sets.

Chapter 7 by Stoddard and Owren reviews best practices for sound conditioning and preprocessing. The appropriate use of filters is essential before a recorded sound can be analyzed. Filters, either hardware or digital, include band-pass, low-pass, or high-pass versions. They all apply frequency-dependent attenuation to the recording to emphasize a portion of the audio spectrum, remove or minimize noise, and define the upper and lower frequency boundaries of the recording. Filters can be used to search for signals embedded in background noise and improve the signal-to-noise ratio, remove 60 cycle hum or low-frequency wind noise from a recording, and simulate environmental excess attenuation on broadcast signals in nature. The misuse of filters or the failure to use antialising filters during analog to digital conversion can produce strikingly unintended effects, and chapter 7 discusses these issues which are part and parcel to everyday work in a bioacoustics laboratory.

Animal vocalizations are produced by vocal fold oscillations that range from nearly periodic vibrations to noisy signals generated by chaotic aperiodic oscillations. The field of nonlinear dynamics has helped to make sense of what we perceive as voice breaks, rough and creaky voice, or as screams. Video recordings of the patterns of vibration seen in both human and animal vocal folds show several distinctly different patterns of oscillation. In chapter 8, Tokuda introduces the concept of nonlinear attractors and bifurcation in order to explain acoustic phenomena such as subharmonics, biphonation, deterministic chaos, and frequency jumps. These phenomena all occur in normal phonation as well as in disorder voicing. The data reviewed in this chapter shows how the quantification of nonlinear properties of animal vocalizations, and the investigations based on temporal analytic approaches can be extremely fruitful.

The next two chapters introduce methods devoted to software classification of speech and animal vocalizations. Bioacousticians have traditionally sorted animal calls according to subjective perceptual properties, spectrographic features, or statistically analyses of cardinal features such as duration or fundamental frequency. Johnson and Clemins (Chapter 9) provide an overview of a variety of approaches bioacousticians have employed to analyze and classify animal vocalizations. They identify some of the advantages and limitations of statistical models, template matching models, spectrogram cross-correlation methods, and several machine learning pattern recognition approaches suitable for sound classification. One obvious approach is to adopt the signal classification system employed by speech recognition technology to the task of classifying animal vocalizations. Speech recognition technology, available on most smart devices, takes an acoustic input and assigns it to a phoneme and subsequently to a letter in the alphabet. This technology, founded on the Hidden Markov Model (HMM) signal classification system, is described in detail. The HMM is a state machine, it takes acoustic input sequences, and estimates the most likely corresponding state sequences that produced the input. Though developed for the mainstream application of speech recognition, this approach is flexible and it can be adopted for the automated detection of animal vocalizations, species recognition, call-type classification, and individual identification. Throughout this chapter the authors guide the reader on how the HMM system can be applied to detect, sort, cluster, and classify animal sounds. This tool holds promise for monitoring, conducting a vocalization survey, and census on free ranging populations of animals in remote habitats.

Chapter 10 by Mercado and Sturdy describes a second computational method for sorting sound patterns based on artificial neural networks. Computation neural networks are inspired by the kinds of processes that occur within the brain. The output from one unit in a network serves as an input to another, and the outputs can be weighted to simulate variations in the strengths of synapses in brains. Neural networks are programs for recognizing patterns. One major advantage of this approach is that the network permits nonlinear classification of an acoustic dimension that superficially appears to be continuous, and the expression of nonlinear classification resembles categorical processing for some perceptual phenomena. Neural networks hold promise for identifying natural auditory categories that would not be detected by traditional statistical and spectrographic classification analyses. However, the output of the network depends on the initial architecture of the network and the learning algorithm used to "train" the network. Researchers must decide how to select the initial parameters for the network without clear guidelines on how to select the most appropriate architecture for the task at hand. By comparing the classification of sounds by animals with the classification of sounds by neural networks it is possible to determine if the network and animals are processing similar or dissimilar acoustic cues.

Bioacousticians frequently collect an unusual and interesting inventory of animal sounds over the course of their careers, and while their research methods, results and conclusions are preserved in their scholarly publications, the animal signals which are the basis of their inquiries frequently are not. Thus, it is important for members of the comparative bioacoustic community to intermittently take stock of their inventory of recorded sounds and to submit them to a sound archive. Part 4, composed of one chapter, is focused on sound recording and archiving. Webster and Budney in chapter 11 provide guidelines for collecting and identifying sound media specimens, and good practice for recording notes and observations, and the sounds themselves pertinent to their submission to a sound archiving facility. Archived sounds of focal species, and samples of biotic sounds in endangered or threatened habitats provide the bedrock for the next generation of bioacousticians to monitor how well the soundscape has been preserved or changed in relationship to naturally occurring events such as earthquakes, or other events such as climate change, invasive species, and human activities.

REFERENCES

Bateson, P., Laland, K.N. (2013). Tinbergen's four question: an appreciation and an update. *TREE 28, 12.*

Busnel, R.G. (1963). *Acoustic Behavior of Animals.*. New York: Elsevier.

Gerhardt, H.C., Huber, F. (2002). *Acoustic Communication in Insects and Anurans.* Chicago, USA: The University of Chicago Press.

Bradbury, J.W., Vehrencamp, S.L. (2011). *Principles of Animal Communication.* (2nd edition.). Sunderland MA: Sinauer Associates.

Sebeok, T.A. (1977). *How Animals Communicate.* Bloomington: Indiana University Press.

Kroodsma, D.E., Miller, E.H. (1982). *Acoustic Communication in Birds.* Vol. 1 and 2. Production, perception, and design features of sounds. New York: Academic Press.

Hedwig, B. (2014). *Insect Hearing and Acoustic Communication.* Berlin: Springer.

Hopp, S.L., Owren, M.J., Evens, C.S. (2012). *Animal Acoustic Communication.* Berlin: Springer.

Nikol'skii, A.A. (1984). *Sound signals of Mammals in Evolutionary Process.* (p. 199 pp). Moscow, Nauka: (In Russian)

Tembrock, G. (1959). *Tierstimmen. Eine Einführung in die Bioakustik.* Wittenberg, Ziemsen Verlag

Wiley, R.H. (2015). *Noise Matters: the Evolution of Communication.* Cambridge MA: Harvard University Press.

Dedication

Early on in the development of this volume, our colleague, Michael J. Owren, was a co-editor and a visionary advocate of this project. Michael died (July 9, 1955 to January 15, 2014) prematurely at the age of 58 at his home in Atlanta, Georgia. We have followed Michael's career since his days as a graduate student at Indiana University. We shopped for didgeridoos with Michael in Adelaide, and wagered beers on the outcomes of football games when our institutions faced one another on the gridiron. We shared many meals with him at conferences and meetings, and miss him profoundly at both the personal and professional level.

Our colleague, Drew Rendall, a bioacoustician at the University of New Brunswick, has written a memorial of Michael for the *Emotions Researcher* coordinated with Andrea Scarantino at Georgia State University, which we reprint here:

Michael J. Owren

Michael J. Owren, a teacher and scientist who analyzed the biological foundations of animal and human communication, died on January 15, 2014 at his home in Atlanta, Georgia.

Michael was born on July 19, 1955 in Oslo, Norway, the third child of Leif and Ingrid Owren. He was raised in College, Alaska; Hanover, New Hampshire; and Bergen, Norway. He received his B.A. in Psychology from Reed College and his Ph.D. in Experimental Psychology from Indiana University.

Michael taught psychology and neuroscience for over 25 years, first while doing post-doctoral work at the University of California, Davis, and later at the University of Colorado at Denver; the University of Otago (New Zealand); Reed College; Cornell University; and Georgia State University. At the time of his death, he was an Adjunct Professor at Emory University.

Michael had a vigorous scientific career focused on understanding the nature, scope and mechanisms of non-linguistic communication. He thought closely and carefully about focal phenomena in systems of vocal production and perception and his empirical studies are widely recognized for their unparalled rigor and attention to detail.

He was also a skilled developer of novel research technologies and a sophisticated theoretician. On the methods side, he pioneered the application of spectral analysis techniques developed in speech science to the study of animal communication (see for instance his "Some analysis methods that may be useful to acoustic primatologists").

Based on the example of his own research, and on his detailed tutorials for their appropriate use and application, such techniques were widely embraced and became a standard part of the analytic toolkit of animal bioacousticians.

In his theorizing efforts, Michael was particularly invested in delineating and clarifying core constructs that undergird the theoretical foundations of the field of animal communication, and in this, as in everything else, he brought exceptional clarity of thought, expression and vision.

Michael and I jointly developed a heterodox theory of the origins and evolution of signaling systems in animals and humans (see for instance our collaborative papers “Sound on the Rebound” and “An Affect-Conditioning Model of Non-Human Primate Vocal Signaling”).

The theory, dubbed the “affect-induction model”, emphasizes that many animal vocalizations, and some forms of nonlinguistic vocal communication in humans such as laughter, “work” by influencing relatively low-level processes of attention, arousal, emotion, and motivation in the listener rather than the kind of high-level intentional and respresentational processes that support complex language in humans.

We distinguished two mechanisms of such influence, in particular. In some cases, the signal itself has acoustic properties that have a direct impact on the affective states of the recipient. Young vocalizers, for instance, can generate aversive sounds like crying, shrieking, or other kinds of loud and extravagant sounds, which directly motivate caregivers to pay attention and take action to turn off the source of the noxious stimulus.

In other cases, the signal is not high-impact by virtue of its acoustic properties alone, but it influences the affective state of the recipient by virtue of its association with social experiences that have positive or negative consequences, thereby leading to conditioned affective responses.

Dominant monkeys can, for instance, exploit social conditioning processes by pairing distinctive threat calls with subsequent physical attack on subordinate rivals, in future using the threat call alone to intimidate those individuals.

Michael applied these insights to the understanding of human laughter, working closely with Jo-Anne Bachorowski in this enterprise (see for instance two of their papers “The Acoustic Features of Human Laughter” and “Not All Laughs Are Alike”).

They proposed that laughter “works” by being associated with positive events – *e.g.* a joke, a happy meal with friends – and becoming a conditioned stimulus for those events. Since laughter breeds more laughter, laugh production creates positive and reciprocally sustaining affective states that can be used for fostering cooperation and diffusing conflicts.

The affect-induction model was creatively applied by Michael to a large domain of experimental settings, ranging from alarm calling and food calling in nonhuman primates, domestic cat meowing, infant babbling and human laughter (notable publications here include "The Acoustic Features of Vowel-Like Grunt Calls in Chacma Baboons", "Salience of caller identity in rhesus monkey (*Macaca mulatta*) coos and screams: Perceptual experiments with human listeners" and "Asymmetries in the individual distinctiveness and maternal recognition of infant contact calls and distress screams in baboons").

The model challenges the standard interpretation of non-linguistic signals as providing veridical information to recipients, suggesting that they can have a much more direct impact on recipients' responses and in ways that are not always aligned with receiver interests (see "What Do Animal Signals Mean?" and "Communication Without Meaning or Information" for an exploration of some of the tensions with the received view). But it also shows how low-level processes of influence can pave the way for more complex representational communication like that epitomized by the semantic qualities of human language.

In addition to its academic recognition, Michael's work generated interest in the popular media, as in a 2003 Chicago Tribune article that described his feline communication research as the "how of the meow," and an NPR interview on his work with Marina Davila Ross investigating the evolutionary roots of laughter:
http://www.npr.org/templates/story/story.php?storyId=104952197.

Michael loved teaching, and was a mentor to many undergraduate and graduate students. Outside the classroom, he was a life-long runner. For a while, he also sang professionally, performing during his time in Denver with an a cappella group known as Cool Shooz. To his friends and family, Michael was known for his intelligence, dry wit, and knowledge of everything. From beer to basketball to politics and world geography, Michael was the guy everyone wanted on their Trivial Pursuit team.

Michael is survived by his three siblings, Turid Owren of Portland, Oregon; Henry Owren, also of Portland; and Thomas Owren, of Bergen, Norway; as well as thirteen nieces and nephews who loved spending time with their Uncle Michael. They, along with his many students, colleagues, and friends, will miss him greatly. I will too…

Drew Rendall

Michael Owren

List of Contributors

Charles Brown	Department of Psychology, University of South Alabama, Mobile, Alabama, USA
Gregory F. Budney	Macaulay Library, Cornell Lab of Ornithology, Ithaca, New York, USA
Patrick J. Clemins	Department of Computer Science, University of Vermont, Burlington, Vermont, USA
Franz Goller	Department of Biology, University of Utah, Salt Lake City, Utah, USA
Eric Hunter	Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan, USA
Michael T. Johnson	Department of Electrical and Computer Engineering, University of Kentucky, Lexington, Kentucky, USA
Ole Naesbye Larsen	Department of Biology, University of Southern Denmark, Odense, Denmark
Dan Ludwigsen	Physics Department, Kettering University, Flint, Michigan, USA
Eduardo Mercado III	Department of Psychology, The State University of New York University at Buffalo, Buffalo, New York, USA
Michael J. Owren	OSV Acoustical Associations LLC and Emory University, Atlanta, Georgia, USA (Now deceased)
Katharina Riebel	Institute of Biology, Leiden University, Leiden, The Netherlands
Tobias Riede	Department of Physiology, Midwestern University, Glendale, Arizona, USA
Phillip K. Stoddard	Department of Biological Sciences, Florida International University, Miami, Florida, USA
Christopher B. Sturdy	Department of Psychology, University of Alberta, Edmonton, Alberta, Canada
Isao T. Tokuda	Department of Mechanical Engineering, Ritsumeikan University, Shiga, Japan
Magnus Wahlberg	Department of Biology, University of Southern Denmark, Odense, Denmark
Mike Webster	Macaulay Library, Cornell Lab of Ornithology, Ithaca, New York, USA

Part I
Sound Properties

CHAPTER 1

Sound and Sound Sources

Ole Næsbye Larsen* and **Magnus Wahlberg**

Department of Biology, University of Southern Denmark, Odense M, DK-5230, Denmark

Abstract: There is no difference in principle between the infrasonic and ultrasonic sounds which are inaudible to humans (or other animals) and the sounds that we can hear. In all cases, sound is a wave of pressure and particle oscillations propagating through an elastic medium, such as air. This chapter is about the physical laws that govern how animals produce sound signals and how physical principles determine the signals' frequency content and sound level, the nature of the sound field (sound pressure *versus* particle vibrations) as well as directional properties of the emitted signal. Many of these properties are dictated by simple physical relationships between the size of the sound emitter and the wavelength of emitted sound. The wavelengths of the signals need to be sufficiently short in relation to the size of the emitter to allow for the efficient production of propagating sound pressure waves. To produce directional sounds, even higher frequencies and shorter wavelengths are needed. In this context 'short' is measured relative to the size of the sound source. Some sound sources, such as dipoles and pistons, are inherently directional, whereas others, such as monopoles, are inherently omnidirectional.

Keywords: Bioacoustics, ka product, Sound production, Sound source.

1. INTRODUCTION

This chapter is about sound and how sound is produced. To start out, let us consider sound as everything we can hear, *i.e.* the entity that through the ears of normally functioning humans gives rise to our sense of hearing. Through our hearing we form an ever changing mental map of our environment consisting of a number of sound sources localized in space. It is common experience that if we

* **Corresponding author Ole Næsbye Larsen:** Department of Biology, University of Southern Denmark, Odense M, DK-5230, Denmark; Tel: +45 6 550 2444; Email: onl@biology.sdu.dk

Charles Brown and Tobias Riede (Eds.)

move away from a stationary and constant sound source, its sounds become progressively fainter. Eventually we cannot hear them any longer, although somebody standing closer still can. This means that our hearing has a volume threshold, below which we cannot hear the sound. On the other hand, unfortunate individuals standing too close to an explosion will experience intense pain or may lose their sense of hearing (at least temporarily) because their eardrums are ruptured by the high sound volume. So, there is both a lower and an upper limit to the volumes of sound that our ears can analyze. Similarly, there are also limits to what pitch we can hear. High-pitched sound above our upper pitch limit is traditionally referred to as *ultra-sound* while sound below our lower pitch limit is called *infra-sound* (see Fig. **1**). On top of this, our hearing abilities deteriorate with age, so that especially our abilities to hear high-pitched sounds deteriorate as we get older – *i.e.* the limit to ultra-sound varies with individual age.

Such an anthropocentric grouping of sound into ultrasonic, sonic, and infrasonic sounds is far too limiting in bioacoustics. In principle there is no difference between, on one hand, those sounds that we can hear, and on the other hand the infrasonic components of the low-pitched rumble of elephants and the ultrasonic high-pitched cries of bats, none of which we can hear. For each hearing animal, there is a certain range of sound volumes and pitches that can be detected. So, for each species we can construct a diagram like that for humans in Fig. (**1**), with the meaning of ultra-sound and infra-sound differing from species to species. Therefore, in Chapter 1 and 2 of this book we will refrain from the use of the terms ultra- and infrasound. Instead, we will focus on the physical properties that really matters for sound production and propagation. Very often the most important property boils down to the ratio between the wavelength of the emitted sound and the size of the sound source or the range from the sound source to the receiver.

So far we have referred to sound in qualitative terms of subjective perception and used terms like *volume* and *pitch* to describe attributes to an auditory event perceived by humans. However, it is much more productive to define sound in objective terms of physics, since then we are not limited by arbitrary species dependent values. Here we use objective measurements to describe attributes to physical sound events, like the units on the axes in Fig. (**1**). Objective sound events outside an organism normally cause subjective (perceived) auditory events inside the organism (Blauert, 1997). The study of this phenomenon is known as *psychoacoustics* and this subject is dealt with in Chapter 10.

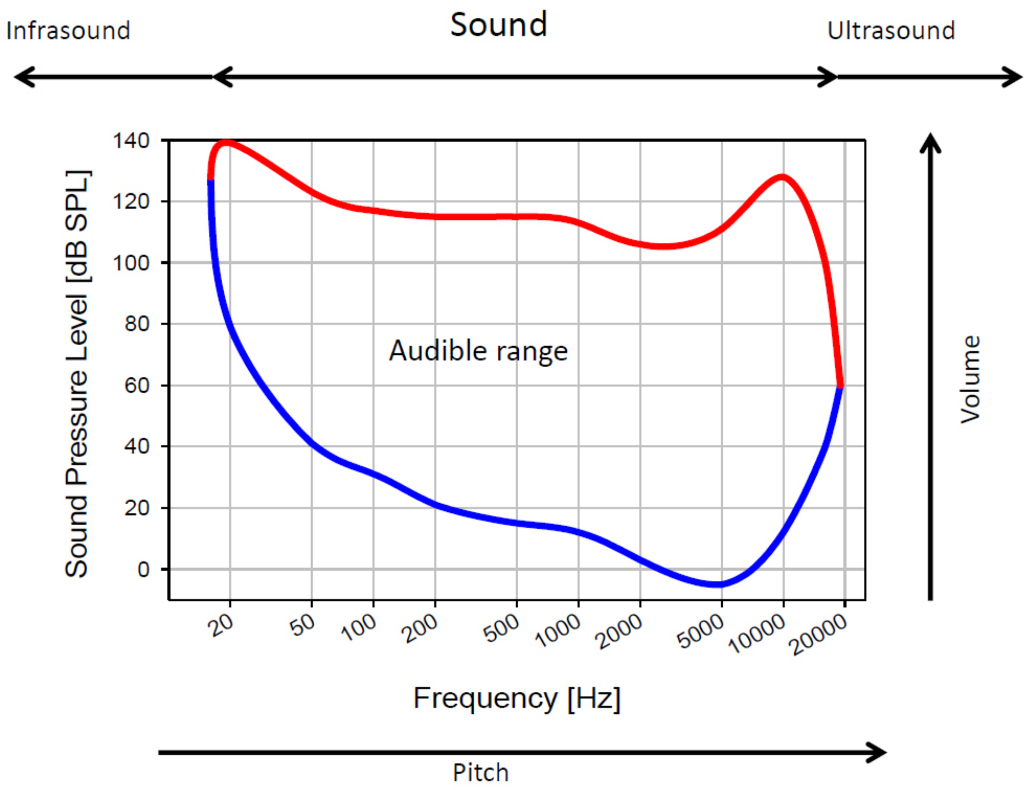


Fig. (1). An anthropocentric definition of infrasound, sound, and ultrasound is shown in relation to a map of sounds with pitch and volume perceptible to young and healthy humans. Perception by definition is subjective. The audible combinations of pitch and volume are delimited by the blue curve (faintest perceptible sounds of different frequencies) and the red curve (strongest perceptible sounds that may impair human hearing). A more accurate and objective description uses physical units of frequency measured in hertz (Hz) and sound pressure level measured in decibels (dB). Redrawn and modified from Wolfe *et al.* (2012).

In physical terms 'sound' can be defined, for instance, as '*mechanical vibrations and waves of an elastic medium*' (Blauert, 1997), or more generally as '*compressional oscillatory disturbances that propagate in a fluid*' (Jacobsen, 2007). What does it mean? Let us break up this statement in two components, the

CHAPTER 2

Propagation of Sound

Magnus Wahlberg* and **Ole Næsbye Larsen**

Department of Biology, University of Southern Denmark, Odense M, DK-5230 Denmark

Abstract: As an acoustic signal travels from the source to a receiver, it is affected by a variety of physical processes, all dictated by properties of the signal and the environment. The signal energy is weakened by geometric attenuation as well as absorption by the medium. The temporal and spectral properties can be modified by sound absorption, refraction, and interference from multi paths caused by reflections. The path from the source to the receiver may be bent due to refraction. Besides geometrical attenuation, the ground effect and turbulence are the most important mechanisms to influence communication sounds for airborne acoustics and bottom and surface effects for underwater sounds. Refraction becomes very important close to shadow zones. For echolocation signals, geometric attenuation and sound absorption have the largest effects on the signals.

Keywords: Echolocation, Excess attenuation, Ground effect, Refraction, Reverberation, Scattering, Sound absorption, Sound attenuation, Transmission loss, Turbulence.

1. INTRODUCTION

As a sound signal propagates away from a sound source its intensity is reduced due to a number of factors. The sound energy is distributed over a larger and larger area, and some of it is transformed to heat due to absorption of the medium. In addition, the sound signal experiences scattering and reflections by the bottom and surface for underwater sounds and by the ground for airborne acoustics and from interaction with small and large objects on its way from the source to the

* **Corresponding author Magnus Wahlberg:** Department of Biology, University of Southern Denmark, Odense M, DK-5230, Denmark; Tel: +45 22163950; Email: magnus@biology.sdu.dk

Charles Brown and Tobias Riede (Eds.)

receiver, creating a 'tail' of echoes blurring the temporal structure of the received signal. If there are spatial variations in the speed of sound, the sound paths are bent towards lower speeds due to refraction. Turbulence in the air causes further reduction and distortion of the signal envelope. The *path that sound signals take from source to receiver is seldom a straight one*, and the *received sound signal is usually a weaker and distorted version of the emitted one*. The total loss of sound intensity from source to receiver is known as ***transmission loss***, while the signal distortion in terrestrial environments has been termed *degradation* (Wiley and Richards, 1978) or *blurring* (Dabelsteen *et al.* 1993).

The features in the environment that influence transmission of sound in air was first described to bioacousticians in the 1970s (*e.g.* Chappuis, 1971; Marten & Marler, 1977; Piercy *et al.*, 1977; Michelsen, 1978; Wiley & Richards, 1978) and treated further in some seminal papers from the 1990's (Forrest, 1994; Embleton, 1996). For the underwater acoustic environment, bioacousticians initially relied on seminal books on the physical properties of underwater sound by *e.g.*, Urick (1967 and 1983) and Clay & Medwin (1977). In the 1990s, Medwin and Clay (1998) provided a more extended treatment of underwater sound transmission, covering many issues of direct interest for bioacousticians. In the early 2000s there has been a fast development in modelling techniques, many of which are well described in many books and papers (*e.g.*, Kastnelson *et al.*, 2012; Attenborough *et al.*, 2007). Developments in the understanding of outdoor and underwater acoustics during the past 20 years merit a new summary that we hope will help newcomers to the field.

The basic physical laws governing the attenuation of acoustic intensity are well-known and quite easy to grasp. Therefore it is often possible to give some 'rules-of-thumb' and 'ballpark estimates' of how much intensity is reduced and how much the signal envelope is distorted while propagating through the medium. However, in many situations with uneven, patchy scatters and a complicated spatially varying sound speed, the propagation of an acoustic field can become extremely complicated. Even though modeling can help us understand what the sound field looks like, in many situations measurements are necessary to obtain reliable assessments of the transmission loss and ray paths. In addition, applications of outdoor and underwater acoustics to realistic bioacoustics applications are still in development, so you should not be surprised that measured values may deviate appreciably from model ones (see *e.g.* Attenborough *et al.*, 2007 for in-air examples and DeRuiter *et al.*, 2010 for an example of porpoise signals propagating through water).

Sound propagation is most accurately described by the ***acoustic wave equation***. In its original shape, this is a four-dimensional (three spatial coordinates and time) differential equation of the second order. In many situations, a ***symmetric geometry*** can simplify the equation to facilitate the determination of the spatial variability of the acoustic field. For example, for an omnidirectional sound source in a homogenous medium the spatial variation in the sound field can be described by one coordinate: the range to the source.

Even though sound and light are described by fundamentally different physical laws, mathematical formulations of how they propagate show several resemblances. Light is a transversal wave that can propagate without a medium, whereas sound is a longitudinal wave that needs a medium to propagate. In spite of this, many of the principles derived from studies of light can be used to describe how sound propagates. A simple propagation model uses ***ray tracing***. A sound wave ray is a curve in space that describes the trajectory of a point on the wave front. One could also say that the ray visualizes sound wave propagation as a small ***acoustic particle*** travelling along a narrow beam or ray in discrete steps, and bouncing off from or being refracted through surfaces in the same way as we usually visualize light rays. In optics, it is common to switch between 'particle' and 'wave' descriptions of light propagation, depending on which phenomena are studied (cf. Chapter 1, section 4). The same holds true for acoustic waves: depending on the frequency content of the signal and the properties of the medium, we may more accurately describe sound paths as waves travelling through the medium (by solving the wave equation) or by thinking of it as acoustic particles following acoustic rays (Medwin & Clay, 1998; Kinsler *et al.*, 2000). Another helpful tool is to use ***Huygens' Principle*** to construct the wave front at a certain time by combining omnidirectional sound emissions centered at the wave front slightly earlier.

Sound is used by animals not only for communication, for warning of danger, or

Part II

Vocal Production

120 *Comparative Bioacoustics: An Overview*, 2017, 120-164

CHAPTER 3

An Introduction to Laryngeal Biomechanics

Charles Brown[1,*] and **Tobias Riede**[2]

[1] *Department of Psychology, University of South Alabama, Mobile, AL 36688, USA*

[2] *Department of Physiology, Midwestern University, Glendale, AZ 85308, USA*

Abstract: Laryngeal sounds in most frogs, in reptiles and most mammals are produced by the interaction between an airstream through the larynx and soft tissue vocal folds positioned laterally in the larynx. This produces a sound characterized by a fundamental frequency (F0), a spectrum of higher frequencies, amplitude and duration. The vibrating vocal folds disturb the airstream so that acoustic waves are generated which travel along the vocal tract from which a small portion of sound energy is radiated from mouth or nostrils. Laryngeal muscles are used for posturing of vocal folds, they adduct and abduct, or elongate and shorten them. Not only posturing and length changes of vocal folds affect the acoustic properties of a voice, but their morphology is also an important determinant of the vocal output. Vocal folds in frogs, reptiles and mammals are composed of several layers of tissue. An epithelial layer covers a lamina propria. The lamina propria itself can be composed of more than one layer, and the number of layers varies by species. The thyroarytenoid muscle, the third distinct structural part of a mammalian vocal fold, is located lateral to the lamina propria. The cellular and acellular morphology of vocal folds determines their viscoelastic properties, and therefore are critical in determining how the tissue responds to changes in airflow, posturing, and tension. The effect of multiple aspects on sound output, for example, (a) active movements facilitated by laryngeal muscles, (b) vocal fold morphology, (c) vocal fold viscoelastic properties and (d) vibration characteristics, can be studied in isolation, but the full picture of laryngeal biomechanics requires the investigation of the whole organ in action. One approach which we discuss here is the excised larynx experiment. Although this approach cannot reproduce natural vocal behavior, it helps reveal important aspects of the vocal fold functional morphology. The use of perfused *in situ* larynx preparations and the differentiated stimulation of motor efferent fibers of intrinsic laryngeal muscles, has helped characterize the

[*] **Corresponding author Charles Brown:** Department of Psychology, University of South Alabama, Mobile, AL 36688, USA; Tel: +1 251 460 6372; Email: cbrown@southalabama.edu

acoustic space available to the vocal organ. We also describe the production of ultrasonic vocalization in rodents, which do not rely on tissue oscillation but on a purely aerodynamic process. For ultrasonic vocalization, the vocal folds are used as a dynamic obstruction of the airway to produce sound by a whistling mechanism.

Keywords: Fundamental frequency, Hyaluronan, Myoelastic-aerodynamic theory, Stress-strain relationship, Stress relaxation, Vocal production.

1. INTRODUCTION

Acoustic signals play an important role in the majority of vertebrates, and many use the larynx to produce sound. A remarkable exception are birds, which all posses a larynx (Zweers *et al.* 1981) but have evolved a unique structure, the syrinx, to generate sound (Chapter 4 of this volume). The larynx is the source of vocal sounds in frogs (*e.g.*, Martin, Gans 1972; Dudley, Rand 1991; Ryan, Guerra 2014), reptiles (Riede *et al.* 2015) and mammals (*e.g.*, Titze 2000; Riede, Brown 2013), although there are numerous examples when new extralaryngeal structures have evolved and took over from or supplemented laryngeal vocal production (*e.g.*, Cranford *et al.* 1996; Madsen *et al.* 2012; Charlton *et al.* 2013). This chapter discusses the mechanisms that allow frogs, reptiles and mammals to produce sound with the larynx focusing on the relationship between biomechanics and vocal output. We focus on terrestrial mammals and refer to other work for laryngeal mechanism in marine mammals (*e.g.*, Dormer 1979; Au 1993; Cranford 2000; Adam *et al.* 2013).

The larynx has been dubbed a “vocal instrument” (*e.g.*, Titze 2008a) because it can produce a very wide range of frequencies and amplitudes, which in its size and functionality is not rivaled by any man-made mechanical instrument. Musical instruments are built by following a basic rule that the fundamental frequency (F0) range is governed by the size of the instrument. According to this rule large instruments support larger and slower vibrations and therefore produce low frequencies, but small instruments producing smaller and faster vibrations tend to make high frequency sounds. This size-dependent principle is defied by the laryngeal sound source in multiple ways. In this chapter, we explain the mechanisms most relevant for the ability to generate a wide F0 range and an enormous dynamic range. The main question that we address is how the critical acoustic feature, F0, is facilitated by the morphology and physiology of this vocal organ.

We begin with an introduction to the myoelastic-aerodynamic theory of voice production, a concept integrating active movements of the larynx as well as morphology, physical properties and phonatory function of the vocal apparatus. Active movements, which are discussed in section 2.1, are facilitated by laryngeal muscles. The transection or electrical stimulation of nerves, which supply laryngeal muscles, generates typical deficits or changes to the acoustic output, respectively. Acoustic changes allow conclusions about the function of the respective muscle or muscle group. The recording of the electromyographic activity of these muscles during spontaneous vocal and non-vocal behavior are important complementary studies. Conversely, since F0 features are closely associated with activity of laryngeal muscles, voice features provide insight into the neurophysiological control of laryngeal and respiratory motor patterns.

We continue with a discussion of vocal fold morphology (section 2.2). Vocal folds consist of epithelium, lamina propria and a muscle. This multi-layered organization is a feature of all vocal folds across mammals, reptiles and frogs. In section 2.3 of this chapter, we describe how a vocal fold, in a first approximation, can be compared with a string. The vocal fold string conceptualization is complicated because of the multi-layered morphology of the vocal folds, which is an important feature facilitating the wide F0 range. It introduces a nonlinear component into sound production which makes it difficult to predict the vocal range a larynx of a certain morphology can generate. An important approach to study the functional morphology of the larynx organ will be discussed in section 2.4. Much can be learned about the interaction of the airstream passing the larynx and vocal folds from excised larynx experiments. We will provide a step-by-step explanation of this latter approach.

Finally, ultrasonic calling and echolocation in rodents and bats respectively, utilizes a frequency range that cannot be perceived by most other vertebrates. Bats show numerous laryngeal specializations, which facilitate the high fundamental frequency of echolocation calls as well as high calling rates. Both rats and bats

CHAPTER 4

Sound Production and Modification in Birds – Mechanisms, Methodology and Open Questions

Franz Goller*

Department of Biology, University of Utah, 257 South, 1400 East, Salt Lake City, UT 84112, USA

Abstract: Elaborate and diverse sounds are an important aspect of many bird behaviors, and these sounds are generated by sophisticated multiple motor systems regulating respiration, vocal organ and upper vocal tract structures. The avian vocal organ, the syrinx, is a unique sound generator among vertebrates, and its morphology varies substantially between different taxa. In this review an introduction to our current knowledge of the peripheral mechanisms of sound production and modification is presented in light of the methodologies that have been used to study various aspects of phonatory mechanisms. Limitations of these methods are also identified and areas for future study and needed information are discussed for each participating motor system. Respiratory control determines the coarse temporal aspects of vocalizations. Rapid switching between expiration and inspiration enables birds to take mini-breaths during inter-syllable intervals up to syllable repetition rates of approx. 30 Hz. At higher rates, pressure modulation of a sustained expiratory pulse still indicates detailed respiratory involvement in the fine control of sound production. Even during rapid sequences of expiration and inspiration, gas exchange is maintained, allowing birds to sing very long songs. Although syringeal morphology has been studied for centuries, the functional aspects of this morphological variation have only recently become subject of investigation to complement efforts focused on neural control of acoustic features. The interplay of morphology, biomechanics and neural control remains a fertile ground for future investigation of song production mechanisms and differences between avian taxa. The neuromuscular control of sound production is best understood in doves and oscine songbirds. Syringeal muscles contribute to the regulation of airflow and tension of the vibrating tissues (membranes or labia), but complex biomechanical interactions make complete understanding of the control of acoustic parameters difficult. For example, the control of sound frequency in oscines arises from a complex interplay of

* **Corresponding author Franz Goller:** Department of Biology, University of Utah, 257 South, 1400 East, Salt Lake City, UT 84112, USA; Tel: +801-585-1929; Email: F.Goller@utah.edu

Charles Brown and Tobias Riede (Eds.)

muscle action, physical parameters (flow and pressure gradients) and morphological specializations (extracellular matrix design of the labia). The presence of two independently controlled sound generators in some bird taxa also creates the potential for an enhanced vocal range and for complex acoustic interactions. Once sound is generated in the vocal organ, it is modified as it exits the bird through trachea and oropharyngeal spaces. This modification can be highly sophisticated, as birds can dynamically adjust resonances to track the fundamental frequency of rapidly modulated song syllables to generate tonal sounds or give rise to complex harmonic content with formant-like quality. At each motor level, many details remain to be discovered, and a thorough understanding of the peripheral mechanisms will be required for decoding the central motor program of song generation. In addition, the morphological variation in syrinx structure across different bird taxa provides a rich source for studying functional aspects of sound generation, but also for investigating evolutionary aspects of this unique and elaborate sound producing organ among vertebrates.

Keywords: Birds, Functional morphology, Muscular control, Peripheral mechanisms of sound production, Respiration, Sound modification, Syrinx, Upper vocal tract.

1. BRIEF HISTORY OF EXPLORATION OF THE AVIAN SOUND SOURCE

The avian vocal organ, the syrinx, has been recognized as a unique sound source among tetrapods for a long time and has received substantial attention from researchers starting in the 18th century and continuing into the 21st century (*e.g.*, Hérissant, 1753; Cuvier, 1802; Savart 1826; Müller, 1847; Garrod 1877; Wunderlich, 1884; Fürbringer, 1888; Beddard, 1898; Häcker, 1900; Setterwall, 1901; Myers, 1917; Köditz, 1925; Ames, 1971; Warner, 1972a,b; King 1989; Prum 1992; Yildiz *et al.*, 2003; 2005; Miller *et al.*, 2008; Zimmer *et al.*, 2008; Düring *et al.*, 2013; Picasso and Carril, 2013; Erdogan *et al.*, 2014; Riede *et al.*, 2015). The early investigations described the morphology and histology of this organ, using it as a taxonomic characteristic and therefore focused on differences between different bird groups (*e.g.*, Wunderlich, 1888; Pycraft, 1900; Griffiths, 1994; Zimmer *et al.*, 2008). Because methodological tools were not available, early on little direct research was done on the functional morphology and the related sound production mechanisms. Nevertheless, possible sound sources and potential mechanisms for sound production in songbirds were discussed at the end of the 19th century (*e.g.*, Müller, 1878; Setterwall, 1901).

Early investigations of the functional aspects of the sound source involved highly creative approaches. A preparation of the excised gull syrinx (Rüppell, 1933) captured the importance of the positioning of the syrinx within the interclavicular air sac, and it is only recently that a similar approach has been applied to the oscine syrinx (Elemans, 2014; Elemans *et al.* 2015). Filming of the syrinx was pioneered by Paulsen (1967), and later videoscopic imaging of the *in situ* syrinx through the trachea was used to investigate the sound source in pigeons and oscine songbirds (Goller and Larsen, 1997a,b). Technical difficulties, specifically providing sufficiently high light levels for high frame rates, still hamper efforts to study the sound source *in situ* across a broad range of acoustic parameters.

The investigation of the physiological aspects of sound production was pioneered by studies on ducks, geese and chickens (Gross, 1964; 1976; Lockner and Murrish 1975; Phillips and Peek, 1975; Lockner and Youngren 1976; Gaunt and Gaunt, 1977; Brackenbury, 1977; 1978a; 1979; 1980a,b; 1989) and was expanded to songbirds and doves (Gaunt *et al.*, 1973; Brackenbury, 1978b,c). The availability of spectrographic analysis of acoustic signals also enhanced our understanding of the underlying production mechanisms (*e.g.*, Greenewalt, 1968; Stein, 1968).These studies included functional morphology but now also focused on neural as well as respiratory and syringeal control of vocalizations. A major renewed effort and advance then occurred in the early 1990s when Roderick Suthers began to record airflow through each sound generator of the oscine syrinx during spontaneous song generation (Suthers, 1990; Allan and Suthers; 1994; Suthers *et al.*, 1994). This started a period of intense investigation of oscine songbirds, which as vocal learners had already become the focus of neurobiological studies (*e.g.*, Konishi, 1965; Nottebohm, 1971; 1972; Nottebohm *et al.*, 1976). Although other bird taxa exhibit equally interesting vocal behavior, they have received relatively little to no attention (*e.g.*, Suthers, 2001; Suthers, 2009).

The current state of knowledge on peripheral mechanisms of sound production has been recently reviewed in detail from different viewpoints (Suthers and

CHAPTER 5

Source Filter Theory

Eric J. Hunter[*, 1] and **Daniel Ludwigsen**[2]

[1] *Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, USA*

[2] *Physics Department, Kettering University, Flint, MI, USA*

Abstract: Sources of sound exist all-around us. Sound sources are often accompanied with a filter, and changes in the shape of the filter, like changes in the slide of a trombone, has a strong impact on the emitted sound. We first describe two archetypes of filters, the Helmholtz resonator and the pipe resonator. The Helmholtz resonator consists of a larger cavity with a narrow opening, and the pipe resonator consists of a uniform tube.

We also describe an experimental approach that allows researchers to estimate the resonance characteristic of both types of filters. Three types of sound sources are used to test a resonator: a swept sine wave, a broadband noise, and an impulse. They can be played as an input to "excite" the resonator, and the output can be recorded. The ratio of the output over the input sound provides an image of the filter's resonance characteristics. A computational approach permits researchers to numerically predict the resonance properties of the filter based on the geometrical dimensions of the filter. The computational approach provides a reasonably accurate prediction of the resonance characteristics of both types of filters. Finally, we apply these concepts to biological systems focusing on human speech production.

Keywords: Amplify / attenuate, Filter, Formant, Mode of vibration, Resonance, Source.

1. INTRODUCTION

In our lives, we are exposed to many types of stimuli. They are emitted from a

* **Corresponding author Eric J. Hunter:** Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan, USA; Tel: +1 527 8846778; Email: eric.hunter.ejh@gmail.com

Charles Brown and Tobias Riede (Eds.)

source (*e.g.*, light from the sun or sound from a vibrating string). Filters can protect sensory organs from an overload of stimuli or from high intensity sources. For example, when we put on sunglasses, we are filtering the light that will meet our retina. The radiating source of the light is not changed, but, the wavelengths which arrive at the cornea are only those which are allowed to pass through the filter. Other wavelengths are attenuated to the point where we may not notice them. A filter can also be closely connected to a source, and then the filter will strongly affect the signal. It will shape it and determine how it is broadcasted. For example, consider the difference between singing in the shower and singing in your bedroom. Both rooms have a very different resonance acoustics due to their size, shape and wall characteristics. Another example is the filtering effect of a public address system - the size, shape and power of the speaker of a public address system determine its predominant resonance frequencies. The tone of a person's voice sounds different coming through the speakers compared with direct speech.

An important filter system that will be discussed in detail in this chapter is the upper respiratory tract of vertebrate animals. For example, the shape of the human airway plays an important role as filter and resonator in speech production. Sound that we generate in the larynx (see chapter 3 of this volume) has to travel through pharyngeal, oral and nasal cavities before it radiates from the lips. During this passage, sound is shaped according to size, shape and material composition of the vocal tract. In this chapter, we first discuss two types of filters, the Helmholtz resonator and the tube-like filter. Secondly, we will discuss the source-filter theory and how it applies to human and animal vocal production. We also address some aspects relevant to source-filter guided sound analysis.

In sum, the goal of this discussion is not to provide an exhaustive review of filter acoustics in human and non-human vocal production. Recent reviews provide overviews about numerous details of the application of the source-filter theory in human and animal vocal production (*e.g.*, Titze 2000; Taylor, Reby 2010). It is our intention to provide a short introduction to essential concepts of acoustic filters and resonances in order to provide tools to the novice bioacoustician.

2. TWO FILTER ARCHETYPES IN ACOUSTICS

An acoustic system in resonance can act as a filter to amplify or reduce particular portions of the spectrum of a signal. Two simple models are used here to demonstrate this principle. The acoustic properties of human and nonhuman vocal tracts are most commonly described by one of these prototypical filters (*e.g.*, Story *et al.* 1996; Gridi-Papp 2008; Fletcher *et al.* 2006). The first filter model is the Helmholtz resonator. When one blows across the opening of a bottle with a narrow neck, a pure tone is produced. The tone can be easily predicted as explained below. The Helmholtz resonator is modeled as purely massive or springy acoustic components known as lumped elements. The second filter model is the pipe. It describes waves in the continuous medium of air inside a tube or pipe, and uses the conditions at the ends of the pipe to determine resonances.

2.1. Excitation of Resonant Systems

A resonance of an acoustic system is marked by a maximum frequency peak in energy transferred into the system from the source. If we excite a system with different frequencies, we tend to observe the most output for a given input around the resonance frequencies. There are at least three different ways to excite a biological system to discover its acoustic resonance properties: swept sine, white noise and an impulse. The *swept sine* excitation sweeps through a certain range of frequencies. The effect is that each frequency is tested, one at a time, before moving on to the next frequency in the sweep. You can use the WAV file 'sweep20-20k.wav' to test your speaker system and headphones, and your hearing as well. This audio example provides a single frequency tone at constant amplitude that rises linearly from 20 Hz to 20,000 Hz. Use caution playing this file; keep your audio system volume reasonably low to avoid any damage from the single frequency tone. There will be portions of the sound that you will not hear, because either your audio system cannot reproduce them, or your hearing is insensitive at those frequencies. A spectrogram of the swept sine tone is shown in Fig. (**1**).

Part III
Sound Analysis in Bioacoustics

Comparative Bioacoustics: An Overview, 2017, 253-301

CHAPTER 6

Acoustic Preference Methods: Assessing Mates

Katharina Riebel*

Institute of Biology (IBL) Leiden University, Leiden, The Netherlands

Abstract: Acoustic mating signals play an important role as mate attracting signals in many vertebrate and invertebrate species. Often, individuals within one population vary in the quantity and quality of their signaling effort. To test whether a signal indeed functions in mate attraction and whether variation in the signal influences this process, preference tests have been established as important research tools. This chapter reviews and explains contemporary methodology for acoustic preference testing. Special attention is given to general conceptual issues, experimental design, potential (but avoidable) experimental confounds and good testing practices.

Keywords: Acoustic mating preferences, Acoustic mating signals, Acoustic preference testing, Internal & external validity, Operant preference tests.

1. INTRODUCTION

1.1. How Mating Preferences Guide Mate Choice

Conspicuous acoustic signals are involved in mate attraction and choice in many vertebrate and invertebrate species (Andersson 1994; Bradbury & Vehrencamp 2011). Mate choice is not a single, punctuated event, but a process that requires locating and recognizing suitable (*i.e.* genetically and behaviorally compatible) individuals before eventually choosing a mate among them. Throughout, this process hinges on mate advertising signals that enable interested partners to find each other.

* **Corresponding author Katharina Riebel:** Institute of Biology (IBL), Leiden University, Nl-2333BE Leiden, The Netherlands; Email: K.Riebel@biology.leidenuniv.nl

Charles Brown and Tobias Riede (Eds.)

In species with random mate choice, detecting, recognizing and locating a species-specific mating signal (and thus its sender) would be sufficient to lead to a mating. At the very least, mating signals have therefore to code information for species recognition. In addition, selection should favor signal properties that promote localization of the sender of the signal. From this, we can hypothesize some important design features of mating signals: first, the emitted signal should match the sensory system of the intended receivers, and mating signals should not get lost by obstruction or attenuation when propagated through the environment from senders to receivers. As we will see further on, detecting and recognizing a mating signal will not always lead to a mating. Whenever some conspecifics are not accepted as partners (either because they are ignored, rejected or some individuals are chosen over another), there is non-random mate choice (Andersson 1994). Mating signals are crucial in mediating these choices: in most taxa they reliably convey not only information on species identity, but also on aspects of the sender's phenotypic quality as status, condition or age (Andersson, 1994; Johnstone, 1995).

1.2. What is this Chapter about?

To test whether a signal indeed functions in mate attraction and whether variation in the signal influences this process, preference tests have been established as important research tools. This chapter is designed to serve as a primer on the methodology of acoustic preference testing and key issues regarding potential (but avoidable) experimental confounds. The chapter's focus is on conceptual issues, experimental design and good testing practice. As a disclaimer up front: a single chapter cannot provide complete coverage of all taxonomic groups and all species specific-techniques. This would require a book (and many specialist co-authors) rather than a chapter. I therefore hope to be forgiven for predominantly focusing on birds, the taxonomic group I am most knowledgeable about and therefore can best discuss suitable examples to explain key issues pertaining to preference testing. This also means that (in line with the research efforts in the field to date) this chapter will deal formostly with acoustic signaling in air. I will however alert readers to issues that might require adjustments when running preference tests involving other substrates and other taxonomic groups.

1.3. Are Acoustic Signals Special?

Sound is but one way to advertise for a mate. Mating signals occur in all known sensory modalities (Bradbury & Vehrencamp 2011) and the study of communication in each of the different sensory modalities requires different specific methods and equipment. However, even within the same sensory modality there is no single all-purpose method and instrumentation. Sound waves can travel through air, water or substrate, each transmission medium requires specialized recording and playback equipment. Sound signaling is always an active production process: an active sound source (stridulation organ, drumming legs, vibrating vocal folds..) has to produce vibrations that change the density and pressure of the medium (air, water or substrate). This means that sound waves require and contain kinetic energy that in turn sets off the substrate to vibrate with the frequency/-ies of the source vibration/-s. Because sound waves do not drift passively but spread predictably and concentrically from the source, they can travel around objects, and receivers can locate a sender by moving along the amplitude gradient (a feat exploited by phonotaxis tests, see below). This process is often supported and accelerated by evolved perceptual mechanisms for ranging the distance to a sound source, and navigational mechanisms to move there directly (Naguib & Wiley 2001). The active range of acoustic mating signals can exceed that of other modalities: the range of elephant infrasound can be up to several kilometers (Larom *et al.* 1997), whale low frequency calls may be audible for hundreds of kilometers (Payne & Webb 1971). In contrast, visual signals normally cover only relatively short distances and that only if there are no obstructions between senders and receivers. In chemical signaling in air, non-volatile components are almost stationary, but volatile chemical substances can travel much longer distances. However, because this is by passive transmission, chemical signals show weather and wind direction specific dispersal patterns, and especially in strong winds, the source of such a signal can be difficult to locate (the same applies to chemical signaling in open turbulent water). High directionality and often a long distance range are thus a special design feature of acoustic signals, and this probably favored the evolution of so many different forms of conspicuous acoustic mate advertising displays in taxonomic groups as

CHAPTER 7

Filtering in Bioacoustics

Philip K. Stoddard[*, 1] and **Michael J. Owren**[2]

[1] *Florida International University, Miami, FL 33199, United States*

[2] *OSV Acoustical Associations LLC and Emory University, United States*

Abstract: Working in bioacoustics requires knowledge of filtering, which is the application of frequency-dependent energy attenuation. General filter types include low-pass, high-pass, band-pass, and band-stop versions, each of which involves selecting a target frequency range, corresponding corner frequencies, and an optimized combination of attenuation slope and pass-band ripple. Filters can be constructed in either analog (hardware) or digital (software) forms, the former being necessary when converting signals between these two kinds of representations. However, the latter are more flexible, less expensive, and the more common when working with digital signals. Readily available programs allow even novice users to easily design and use digital filters. Filtering applications include removing various kinds of noise, simulating environmental degradation effects, and searching for signals embedded in noise. While easily performed, each of these applications requires some background knowledge. There is also good reason to avoid unnecessary use of filtering, as it is easy to create unintended effects. This chapter discusses these and other issues in the context of the everyday work of bioacoustics.

Keywords: Analog filter, Artifact, Attenuation slope, Digital filter, Digital sampling, Gain function, Pass-band ripple.

1. INTRODUCTION

One afternoon in 1981, the screen went blank on the laboratory's real-time digital spectrograph, a recent technological advance at that time. The president and chief

[*] **Corresponding author Philip K. Stoddard:** Florida International University, Miami, FL 33199, United States; Tel: +1 305 348 378; E-mail: stoddard@fiu.edu

Charles Brown and Tobias Riede (Eds.)

engineer of the company selling this equipment diagnosed the problem over the phone. "Oh, that's just the filter board. What are you using the analyzer for? Bird song? For the frequency ranges of your signals you don't really need the filters. Go ahead and bypass the filter board." It was worrisome that an integral component of the machine was not working, but if its designer said the filter board was unnecessary, then that had to be correct. The offending circuit board was removed, the connection was bypassed, and the work continued.

Later, a colleague brought analog recordings he had made of Southern weaver bird (*Pioceus velatus*) song and we proceeded to scan the tape on the filterless analyzer. There on the screen appeared a most striking example of the avian *two-voice* phenomenon. Two distinct voices, not harmonically related, appeared almost to mirror each other. A loud, rising whistle was accompanied by a softer, falling whistle. Sometimes the two voices even crossed, forming a skewed "X" on the screen. However, that second voice was suddenly gone when publication-quality spectrograms of the weaver-bird songs were being made six months later on an even older, analog spectrograph. One explanation for that disturbing development was that the weaker, upper voice might have simply faded from the analog tape, but increasing the gain on the machine failed to reveal any trace of it. Only then did that missing filter board come to mind.

This incident illustrates the importance of *filtering* in the daily life of a bioacoustician, for whom the appropriate use of filters can be essential component of signal analysis. While filters have always been part of working with sound in the analog days, the advent of inexpensive digital technology brought revolutionary advances in this area. However, the power of digital technology has also created some risks. One is that bioacousticians may have become less savvy in this area, for example in knowing about the filtering that is now often being done automatically when signals are translated between analog and digital form. In addition, the digital world may engender less vigilance about the double-edged nature of filtering. While often used for the good, filtering can easily do harm as well—for example by actively creating artifacts and distortions in signals of interest. It is therefore important to understand the basic principles of filters, as well as some of the ways they are applied and potentially misapplied. The goal of this chapter is thus to provide an intuitive explanation of filter theory, applications of filters in bioacoustics, and some of the hazards surrounding filtering. The topic considered in greatest detail is *anti-aliasing*—a form of filtering that is central in both creating and reproducing digital waveforms, yet typically occurs almost invisibly in digital systems.

While the chapter is written at a conceptual level that largely avoids mathematical discussion, non-technically oriented readers may further benefit from also consulting other, similar tutorials on the topics covered (*e.g.*, Menne, 1989; Cook & Miller, 1992). Conversely, for those who are facile in linear algebra and complex arithmetic, the presentation is likely too simplistic. Mathematically rigorous treatments are also readily available, for example in a number of treatment of signal processing and filter theory (*e.g.*, Haydin, 1986; Stearns & David, 1988; Hamming, 1989; Oppenheim & Schafer, 1989).

2. ANATOMY OF A FILTER

Filters are frequency-selective resonant devices (analog filters) or algorithms (digital filters) that are used to pass certain portions of the frequency spectrum of the wave form and exclude others. In general filters are used to prevent sampling-related contamination during the *analog-to-digital* (A/D) conversion, smooth the analog output of *digital-to-analog* (D/A) converters, remove noise, and change spectral balance or phase composition. In these applications, filters change signals by damping or excluding certain components while allowing others to pass. In addition, filters can play a critical role in automated signal detection and identification. In every application, however, filters can have non-intuitive effects that can pose problems for the unwary. Rather than simply removing or separating selected components of a signal, for instance, filtering can simultaneously be distorting the amplitude and phase characteristics of energy in spectral ranges meant to be left unaffected. Analog and digital filters operate according to the same principles, but it is important for readers to appreciate their limitations and differences (section 4).

CHAPTER 8

Nonlinear Dynamics and Temporal Analysis

Isao T. Tokuda

Department of Mechanical Engineering, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan

Abstract: Animal vocalizations range from tonal sounds produced by almost periodic vocal fold vibrations to completely aperiodic sounds generating noisy signals. Between these two extremes, a variety of nonlinear phenomena such as limit cycles, subharmonics, biphonation, chaos, and bifurcations have been found. This chapter introduces a concept of nonlinear dynamics and its methodology applicable to bioacoustic data. Since conventional spectral analysis is not sufficient to characterize nonlinear properties of the recorded sound signals, a temporal analysis based upon the method of nonlinear dynamics is developed. First, using a mathematical model of the vocal folds, basics of nonlinear dynamics and bifurcations are illustrated. The temporal analysis is then applied to acoustic data from real animal vocalizations. Our focus is on extracting low-dimensional nonlinear dynamics from several samples of vocalizations ranging from tonal sounds to irregular atonal sounds. We demonstrate that nonlinear analysis is a profitable approach for analyzing mammalian vocalizations with a harmonic composition or low-dimensional chaos.

Keywords: Animal vocalization, Chaos, Data analysis, Nonlinear dynamics.

1. INTRODUCTION

Animal vocalizations vary from noisy aperiodic calls to almost harmonic songs (Tembrock, 1996; Hauser, 1996; Bradbury & Vehrencamp, 1998). It is often the case that the irregular sounds come from complex vocal fold oscillations which are due to combined nonlinear effects of pressure, airflow, vocal fold tissue

* **Corresponding author Isao T. Tokuda:** Department of Mechanical Engineering, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan; Tel: +81 77 561 2832; Email: isao@fc.ritsumei.ac.jp

Charles Brown and Tobias Riede (Eds.)

elasticity, and collision between the vocal folds (Titze, 1994). According to the theory of nonlinear dynamics, such complex vocal fold oscillations can be generated from a system having only a few parameters. In fact, in the studies of human voice, a variety of voice instabilities have been found to be caused by nonlinear interaction between a small number of components. For instance, the vertical and horizontal vibratory modes of the vocal folds may be desynchronized (Berry *et al.*, 1994; Neubauer *et al.*, 2001), the desynchronization of the left and the right vocal folds (Ishizaka & Isshiki, 1976; Smith *et al.*, 1992; Steinecke & Herzel, 1995; Tigges *et al.*, 1997), or additional vibrating tissues such as vocal membranes (Mergell *et al.*, 1999) can lead to chaotic oscillations of the vocal folds. The concept of "nonlinear dynamics" has been introduced to the field of bioacoustics (Wilden *et al.*, 1998; Fee *et al.*, 1998; Fitch *et al.*, 2002; Mergell *et al.*, 1999) towards a systematic understanding of highly complex bioacoustic process. It should play an important role also in the investigation of animal vocalization.

The idea of nonlinear dynamics implies a possibility that the noisy animal utterances, initially considered to be attributed to turbulent noise or high-dimensional dynamics, may alternatively originate from deterministic nonlinear dynamics with only a small number of state variables. To detect such nonlinear dynamics, specialized techniques suitable for nonlinear data analysis should be applied to the bioacoustic signals. For animal voice signals, however, nonlinear characteristics have not been widely investigated. Narrowband spectrographic analysis has been proven useful in interpreting nonlinear phenomena such as bifurcations between different nonlinear states such as subharmonics, biphonation, chaotic episodes and limit cycles (Wilden *et al.*, 1998; Fitch *et al.*, 2002). Simply counting the number of calls that contain nonlinear phenomena has been so far proven to be most robust (Riede *et al.*, 1997, 2000; Blumstein *et al.*, 2008; Stoeger *et al.*, 2012; Pokrovskaya, 2013). Application of more specialized techniques based on, *e.g.* Lyapunov exponents and fractal dimensions can be, however, problematic, since these techniques are disturbed by nonstationarities that frequently arise in animal sounds.

The aim of the present chapter is to describe the basics of nonlinear dynamics in voice production and to introduce a series of procedures to analyze nonlinear dynamical characteristics in animal vocalization. Our methodology is based upon "deterministic nonlinear modeling *versus* stochastic linear (DVS)" modeling technique, which was introduced by (Casdagli, 1992) and was found suitable for bioacoustic data (Tokuda *et al.*, 2002). Here, we apply the DVS modeling technique to simulated data of vocal fold oscillations, and to a pathological voice, to monkey screams, and dog barks as examples of complex vocalizations.

2. NONLINEAR DYNAMICS AND ATTRACTORS

This section briefly describes the basics of nonlinear dynamics in voice production. As an example of a vocal system that produces a rich variety of nonlinear phenomena, a mathematical model of the vocal fold oscillation is utilized. Among a variety of vocal fold models, we exploit the asymmetric two-mass model proposed by Steinecke and Herzel (1995). This model is derived from the standard two-mass model (Ishizaka & Flanagan, 1972), which represents a vocal fold as a pair of two coupled oscillators. With some simplifications and introduction of an asymmetry between the left and the right vocal folds, the asymmetric two-mass-model has been derived. The vocal fold morphology and its corresponding model structure are schematically illustrated in Fig. (**1**). Despite its simplified formula having only four degrees of freedom, the model is capable of reproducing a variety of qualitative dynamics for pathological voices, such as normal 1:1 synchronization between the left and the right vocal folds, abnormal synchronization other than 1:1 entrainment, and chaotic dynamics.

As the main parameters to control the vocal fold vibrations, the asymmetry parameter $\tilde{Q}$, which determines the tension imbalance between the left and the right vocal folds, and the subglottal pressure P_s, which causes an air flow that passes the vocal folds and initiates and sustains their oscillations, are used (Fig. **1b**). Other parameter values are set to be the standard values described in detail in the original reference (Steinecke & Herzel, 1995). By setting the two parameters as $\{P_S, \tilde{Q}\}$ = {0.002, 0.8}, {0.01, 0.8}, {0.01, 0.754}, {0.009, 0.598}, {0.01, 0.598} (physical unit for the pressure is $g\ cm^{-1}\ ms^{-2}$), five types of attractors, which represent qualitatively different dynamical characteristics, are drawn in Fig. (**2**) and described in the followings (the terminology of “attractor” stands for a final stable state, to which the system dynamics evolves). To display the attractors, we use state space representation, time series, and power spectrum. In the state space, two dynamical variables (x_l, x_r) that represent displacements of the lower edges of

CHAPTER 9

Hidden Markov Model Signal Classification

Michael T. Johnson[*, 1] and **Patrick J. Clemins**[2]

[1] *Electrical and Computer Engineering, University of Kentucky, Lexington, KY 40506*

[2] *Cyber Specialist, Vermont EPSCoR, Asst Research Professor, Dept of Computer Science, University of Vermont, 23 Mansfield Avenue, Burlington, VT 05405-0215*

Abstract: After many decades of slow incremental growth, computer-based automatic recognition of human speech has recently gone through a much more rapid transition from the research lab to mainstream application, available on most of the 1+ trillion smartphones on the planet. (Worldwide, there are almost as many mobile phones as people, and about 1 in 5 of these are smartphones.) This growth has been largely fueled by the growth of raw computational power, rather than fundamental changes in speech recognition technology itself. The methods used in nearly every state-of-the-art automatic speech recognition system are based on the same statistical model that was first used for speech more than 30 years ago, the Hidden Markov Model. Hidden Markov Models are in many ways straightforward models, simple state machines that take input sequences and identify the most likely corresponding state sequences. The main strength of the approach is in its flexibility – flexibility to match sequences in a non-linear temporal pattern, flexibility to learn more detailed models if more training data is available, flexibility to connect multiple models together into longer continuous patterns, and flexibility to incorporate whatever data features and probabilistic models are best suited to the task. Nearly all of these benefits also carry over to the domain of bioacoustics, specifically to the classification of animal vocalizations. Although there are limits to this – human speech is better understood than animal communication – there is also much to gain, and many improvements that are possible by taking advantage of the large body of knowledge available through the long history of human speech processing and recognition technology. Agreeing with this idea, this chapter presents an overview of the use of Hidden Markov Models for classification, detection, and clustering of bioacoustics signals.

[*] **Corresponding author Michael T. Johnson:** Electrical and Computer Engineering, University of Kentucky, Lexington, KY 40506; Tel: (859) 257-0717; E-mail: mike.johnson@uky.edu

Charles Brown and Tobias Riede (Eds.)

Keywords: Feature extraction, Gaussian mixture models, Hidden Markov models, Signal classification, Signal detection.

1. INTRODUCTION

Historically, bioacoustics research projects have required the investment of hundreds of hours of manual labor by skilled scientists and technicians to identify, segment, and categorize individual vocalizations or vocal sequences from lengthy recordings made in the field. Signal detection and classification is a challenging task, made more difficult by the noisy and constantly changing ambient background typical of most bioacoustic sound domains. The ability to automatically and objectively classify bioacoustics sounds into categories using a small amount of manually labeled data for training has the potential to significantly reduce this labor requirement and accelerate research progress.

In addition to the impact on data preparation, automatic signal detection, classification, and clustering is essential to a wide variety of research applications. These include tasks such as species identification, call-type identification, individual identification, and matching of behavior and vocalization patterns, with methods and approaches often needing to be customized for each specific species of interest.

Technology for human speech recognition research has tackled these same challenges for many years, and bioacoustics researchers should clearly be able to leverage the advances in human speech technology and adapt them for application to other bioacoustics signals. There are some significant challenges to this, however. Human speech is far better understood than are the vocalizations of most species, which enables very specific labeling of large volumes of data for training statistical models. Most significantly, human researchers actually know what human vocalizations mean – we do not need to carry out correlational studies between human vocalizations and human behavior to try to make guesses about the intended signaling content of any particular human sound. Even so, the huge advances in temporal sequencing and statistical modeling that has turned automatic recognition of human speech from a daydream into reality over the past 30 years can be applied to the challenging domains within the field of bioacoustics with great effectiveness.

The most significant classification model used in human speech technology is the Hidden Markov Model (HMM) (Rabiner, 1989). An HMM is essentially a statistical state machine, in which each state represents a snapshot in time of a particular vocal characteristic, and the sequence of these states characterizes the temporal pattern of a time-varying vocalization. Within each state is a pattern recognition model that is able to quantitatively determine the match between the state and an observed window of data. Historically these models have been probability distributions such as Gaussian Mixture Models (GMMs), but more recently in the human speech community probability models are being replaced by Deep Neural Networks (DNNs) (Dahl *et al.*, 2012), so that the HMM is actually a hybrid between a statistical model and a neural network approach, such as those described in Chapter 10. Either way, the core strength of an HMM approach is its robust ability to integrate rapidly and inconsistently changing temporal dynamics with precise spectral models characterizing pieces of the sound, resulting in an overall probabilistic match between the HMM and a sample vocalization.

Many approaches have been used within the field of bioacoustics for the purpose of vocalization classification and detection. In addition to HMMs (Weisburn *et al.*, 1993, Anderson, 1999, Trawicki and Johnson, 2005, Clemins *et al.*, 2004, Clemins and Johnson, 2003, Clemins and Johnson, 2002, Ren *et al.*, 2009), existing methods include multivariate statistical analysis, spectrogram cross-correlation (Mellinger and Clark, 2000), matched-filtering (Stafford *et al.*, 1998), dynamic time warping (Brown *et al.*, 2006, Kogan and Margoliash, 1998, Buck and Tyack, 1993), and neural networks (Potter *et al.*, 1994, Deecke *et al.*, 1999).

This chapter will provide an introduction to HMMs and specifically to the use of HMMs for bioacoustics tasks. This will include an overview of the types of tasks for which HMMs are applicable, a discussion of perceptually relevant feature extraction to use as inputs to the HMMs, an introduction to the details of training and applying HMMs, some illustrative examples, and practical notes about software implementation and tool availability.

CHAPTER 10

Classifying Animal Sounds with Neural Networks

Eduardo Mercado III[*, 1] and **Christopher B. Sturdy**[2]

[1] *Department of Psychology, University at Buffalo, SUNY, Buffalo, NY 14260, USA*

[2] *University of Alberta, Edmonton, AB T6G 2R3, Canada*

Abstract: Humans naturally classify the sounds they hear into different categories, including sounds produced by animals. Bioacousticians have supplemented this type of subjective sorting with quantitative analyses of acoustic features of animal sounds. Using neural networks to classify animal sounds extends this process one step further by not only facilitating objective descriptive analyses of animal sounds, but also by making it possible to simulate auditory classification processes. Critical aspects of developing a neural network include choosing a particular architecture, converting measurements into input representations, and training the network to recognize inputs. When the goal is to sort vocalizations into specific types, supervised learning algorithms make it possible for a neural network to do so with high accuracy and speed. When the goal is to sort vocalizations based on similarities between measured properties, unsupervised learning algorithms can be used to create neural networks that objectively sort sounds or that quantify sequential properties of sequences of sounds. Neural networks can also provide insights into how animals might themselves classify the sounds they hear, and be useful in developing specific testable hypotheses about the functions of different sounds. The current chapter illustrates each of these applications of neural networks in studies of the sounds produced by chickadees (*Poecile atricapillus*), false killer whales (*Pseudoorca crassidens*), and humpback whales (*Megaptera novaeangliae*).

Keywords: Adaptive filter, Computational modeling, Connectionism, Learning algorithm, Parallel distributed processing, Perceptron, Self-organizing.

[*] **Corresponding author Eduardo Mercado III:** Department of Psychology, University at Buffalo, SUNY, Buffalo, NY 14260, USA; Tel: +01 716 604 4424; Email: emiii@buffalo.edu

Charles Brown and Tobias Riede (Eds.)

1. INTRODUCTION

Variations in an animal's vocalizations can provide clues about how the animal uses sound, as well as qualities of the individual that is vocalizing. Bioacoustic research depends heavily on the ability to characterize these variations. Often, variations in animal sounds are salient to human listeners. Some acoustic features are imperceptible, however, such as ultrasonic frequency modulation produced by echolocating bats, and phase shifts in audible sounds. Characterizing sounds typically involves not only describing their properties, but also relating those properties to those of other sounds. This process can involve sorting sounds based on how and when they are produced, the functional contexts within which they occur, and perceptual or acoustic similarities and differences across sounds. The most common approaches to sorting sounds include subjective sorting based on auditory quality or spectrographic features, statistical comparisons of quantified features, and computational methods for assessing how sounds are distributed within a multidimensional "feature space." The current chapter focuses on the last of these approaches, emphasizing a computational method for sorting patterns known as artificial neural networks, hereafter referred to as *neural networks*.

1.1. Classifying Sounds Subjectively

Researchers typically sort animal vocalizations based on how they sound, often by comparing them to other familiar sounds. For example, sounds in the songs of humpback whales (*Megaptera novaengliae*) have been described as *snores*, *moans*, or *cries* (Winn & Winn, 1978). Some animals even get their names from the aural impressions their sounds produce in humans, including chickadees, cuckoos, and whooping cranes, to name a few. A bioacoustician may attempt to objectively classify sounds by referring to sounds with alphanumeric labels, but even then the basis for differentiating sounds is often subjective.

Subjective classification strategies invariably involve sorting an animal's sounds by mapping them onto pre-existing perceptual categories acquired through experience. In the case of auditory categorization, it is known that repeated experiences with particular sounds can change how individual neurons respond to those sounds (Weinberger, 2004), and subsequent changes in an individual's neural responses are correlated with changes in the ability to perceive differences between sounds (Recanzone, Schreiner, & Merzenich, 1993). Thus, how a researcher aurally classifies animal sounds will depend on how those sounds are processed by large networks of interconnected neurons, which will in turn depend in complex ways on past auditory experiences, which may be culturally determined. How it is that activity within complex networks of neurons give rise to variations in subjective impressions is not yet known.

Subjective visual classification of time-frequency representations is even more problematic. Identical sounds can be decomposed using a wide range of algorithms with various levels of temporal and spectral resolution, each of which can give rise to a variety of distinctive images. Typically, the adequacy of a particular set of visual representations is judged on how well it corresponds to audible differences between sounds. Thus, visual classifications inherit the subjective biases of listeners while adding further, visually based perceptual subjectivity.

1.2. Naturalistic Classification of Sounds

When animals other than humans hear vocalizations, they almost certainly do not classify them by assigning linguistic labels to the sounds. Nevertheless, non-human animals can recognize similarities between novel and familiar sounds, and can respond to novel sounds based on these similarities (Orduña, Mercado, Gluck, & Merzenich, 2005). Interestingly, at least some animals seem to be able to sort human speech and other animal sounds in ways that emulate sorting by humans (Heimbauer, Beran, & Owren, 2011; Kuhl & Miller, 1975; Sturdy, Phillmore, Price, & Weisman, 1999; Sturdy, Phillmore, & Weisman, 2000). Such overlap in sorting strategies likely reflects similarities in the neural circuits that vertebrates use to process sound, as well as common features of natural acoustic environments. Although researchers have no ability to discern what an animal's subjective awareness of vocalizations is like, psychophysical methods can be used to evaluate which vocalizations animals can distinguish, as well as how animals categorize sounds.

By combining such behavioral techniques with analyses of the acoustic features of

Part IV
Sound Recording and Archiving

CHAPTER 11

Sound Archives and Media Specimens in the 21st Century

Michael S. Webster* and **Gregory F. Budney**

Macaulay Library, Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA

Abstract: Audio recordings of birds and other animals, and also other forms of 'biodiversity media' (*e.g.*, video recordings), capture the behavioral phenotype in ways that traditional museum specimens cannot, and natural history audio/media archives hold collections of recordings that span geography, time, and taxonomy. As such, these recordings can be used for a broad range of studies in ecology, evolution, and animal behavior, and newly developed tools for collecting and analyzing these recordings promise to further increase that research potential. Moreover, the digital revolution has made it easier than ever for high quality recordings to be collected and deposited in an archive, opening the door for large-scale citizen science efforts. But this potential also brings new challenges that must be met by the research community with regard to digital standards and accessibility. We recommend that researchers and other recordists deposit their materials in a suitable archive, that sound/media archives build strong partnerships with other types of natural history collections, that these archives also embrace technological advances to make their assets more accessible, and that archives and acoustic researchers harness "the power of the crowd" through crowd-sourcing and similar approaches. In doing so, sound archives and bioacoustic research will play an ever-increasing role in understanding our natural world, including responses of natural systems to human activities, in the 21st century.

Keywords: Archival standards, Audio specimen, Conservation efforts, Databasing, Environmental recording, Field recording equipment.

* **Corresponding author Michael S. Webster:** Macaulay Library, Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA; Tel: + 1 800 843-2473; Email: MSW244@Cornell.edu

Charles Brown and Tobias Riede (Eds.)

"*I urge conservation organizations that fund avifaunal surveys in tropical forests around the world to require their participants to use tape recorders systematically. Copies of all recordings should be placed in a professionally maintained sound collection that provides easy access to researchers*." (Parker 1991)

INTRODUCTION: WHAT IS A 'MEDIA SPECIMEN'?

Biological research collections have long been the backbone of research in disparate fields including (but not limited to) ecology, evolution, systematics, functional anatomy, and conservation biology. This is because the specimens housed within these collections capture a key factor that is central to these various fields: phenotypic variation across space and time. Accordingly, biological specimens can be used to address a wide diversity of research questions (Bradley *et al.* 2014, Roche *et al.* 2014). The biological specimens in most research collections consist of physical specimens, such as study skins, pickled specimens, skeletons, and other preparations derived directly from the collected organism. These physical specimens are valuable to research because they capture key aspects of the individual's phenotype, and with modern technologies have become even more valuable, for example as the source of material for molecular genetic and chemical isotopic studies.

Yet, valuable as they are, physical specimens do a poor job of capturing one key dimension of the individual phenotype that is critically important to a wide array of research questions: the behavioral phenotype. Behavior is the mechanism by which the individual interacts with its environment, including the abiotic (weather and other conditions), biotic (predators, prey, competitors), and social (conspecifics). Although physical specimens can reveal the anatomical mechanisms that lead to a behavior, they rarely are able to reveal the behavior itself. This limitation has long been recognized. John James Audubon, for example, was an avid collector of bird specimens, but was frustrated by the unnatural repose of the traditional study skin. Accordingly, Audubon used wire to mold his collected specimens into life-like poses, and of course also created paintings depicting birds behaving under natural or near-natural conditions (Rhodes 2006). These paintings captured the behavior of birds to the best of

Audubon's abilities, but were severely limited by the technologies available in his day (pen and ink, canvas and paint).

Technology has evolved considerably in the nearly 190 years since the first publication of Audubon's *Birds of North America*. Starting with the first analog recordings of birds singing – sound-synched films recorded at a city park in 1929 (for one example, see http://macaulaylibrary.org/audio/16968) – and advancing through various forms of analog media (*e.g.*, magnetic tape) to today's digital media technologies. Thanks to these technological advances, it is now possible to capture and preserve the behavioral phenotypes of individual animals at a particular place and time. As such, these recordings are true "media specimens", analogous to physical specimens, and suitable for a wide diversity of behavioral and other biological studies (see below). Other useful forms of biodiversity media include ambient or environmental recordings that capture the sounds of an entire biological community at a particular place and time (see Pijanowski *et al.* 2011), and recordings from autonomous recorders ("acoustic sensors", see Kasten *et al.* 2012) and camera traps (Kelly 2008, Rowcliffe & Carbone 2008) that can be used for biological monitoring. All of these are scientifically valuable acoustic data, and indeed are most valuable if preserved and made broadly accessible.

Natural history media archives, including sound archives that focus primarily on bioacoustic recordings, house these biodiversity media and make them broadly accessible to the scientific community and public alike. Over a decade ago, Alström & Ranft (2003) reviewed the status of the world's natural history sound archives (see also Gaunt *et al.* 2005). They argued that such archives are critically important to biological research, particularly modern day systematics and taxonomy. A lot has happened in the intervening decade, however, including dramatic improvements in digital media technology, revolutionary new approaches to collecting and sharing media *via* the internet, and powerful new approaches to analyzing bioacoustic and other media-borne data. Accordingly, biodiversity media have become all the more relevant to 21st century research, as evidenced by dramatic increases in peer-reviewed publications that use recordings from media collections (Fig. **1**).

Comparative Bioacoustics: An Overview, 2017, 486-490

SUBJECT INDEX

Charles Brown and Tobias Riede (Eds.)

M

N

O

P

R

S

T

U

V

W

X

Charles H. Brown

Charles H. Brown currently serves as a Professor of Psychology at the University of South Alabama. Previously he held appointments as Guest Scientist, Naval Aerospace Medical Research Laboratory, Naval Air Station, Pensacola, Florida, and Assistant Professor of Psychology at the University of Missouri-Columbia. He earned his M.A. and Ph.D. with an emphasis on psychobiology at Michigan State University and was appointed CIC Traveling Scholar at Kresge Hearing Research Institute, University of Michigan. He received postdoctoral training in Ophthalmology and Otorhinolaryngology at the University of Michigan. He has a record of extramural support from the National Institutes of Health, National Science Foundation, Air Force Office of Scientific Research, and The State of Alabama Supercomputer Network. He was elected as Fellow by the American Psychological Association (Division 6), and Quality Enhancement Plan Professional Development Fellow at the University of South Alabama. His research interests are in comparative bioacoustics, normal and disorder hearing and communication, and the scholarship of teaching and learning.