

6 Primate Vocal Production and Its Implications for Auditory Research

W. Tecumseh S. Fitch

CONTENTS

I. Introduction.....	87
II. Principles of Primate Vocal Production	89
A. The Myoelastic–Aerodynamic Theory of Phonation.....	89
B. The Source–Filter Theory of Vocal Production.....	90
C. Nonlinear Dynamics in Vocal Production.....	93
D. Summary of Vocal Production.....	94
III. Primate Vocal Adaptations.....	94
A. Laryngeal Air Sacs.....	95
B. Vocal Membranes.....	96
C. Vocal Tract Elongation.....	96
IV. Putting It to Work: Using the Production Perspective for Analysis and Synthesis.....	97
A. Why Synthesis?.....	97
B. Linear Prediction in Bioacoustics.....	98
C. Analysis and Synthesis Using Linear Prediction.....	100
V. Conclusion.....	102
References.....	103

I. INTRODUCTION

Vocal production forms an important nexus in acoustic communication, lying at the intersection of physics, physiology, neurobiology, and evolution. The foundation of vocal production is provided by physical acoustics, which specifies the basic principles that underlie any sound production system, including those of nonlinear dynamics (for understanding phonation) and linear acoustics (sound propagation through the vocal tract and into the environment). An equally important foundational field is comparative anatomy and physiology. There is considerable variety in the structure and function of primate vocal anatomy that must be grasped before the diversity of primate vocal repertoires can be fully understood. Acoustics and phys-

iology provide the necessary foundation for understanding the function of the vocal mechanism that the primate nervous system has evolved to control (in terms of motor function) and perceive (in terms of auditory function). Finally, the functioning of the vocal apparatus influences the evolution of primate vocal communication systems. This has negative and positive aspects, because the details of the vocal production system both constrain the signals that can be produced and provide adaptive avenues within which new signals evolve. Thus, an understanding of the basic mechanisms of vocal production is an important component of the ethological approach to auditory perception and the evolution of communication.¹⁻³

The intent of this chapter is to demonstrate the relevance of vocal production to scientists mainly interested in primate acoustic communication and auditory neuroscience. Two points will be emphasized, one theoretical and the other practical. First, it is now well understood that different species have different needs and have thus evolved anatomical and neural specializations that are fine-tuned to the physical and social environments in which the species normally functions.⁴ An obvious example is provided by the ultrasonic vocalizations of mouse lemurs^{5,6} which allow them to communicate without being overheard by predators. The 233 species of primates, divided into 13 different families,⁷ live in a very wide variety of habitats from mangrove forest to dry savanna and have an equally impressive diversity of vocal anatomy and vocal communication systems. An auditory neuroscience that wishes to move beyond the study of "the monkey" will need to appreciate the rich variability among the many different non-human primate species and will use these many experiments of nature to achieve a deeper understanding of the evolution of primate communication and the neural mechanisms that subserve it.

Second, and more practically, an understanding of vocal production is necessary for the application and development of new analysis and synthesis techniques for studying vocal communication and auditory perception. Major progress in understanding speech perception was made only after a deep understanding of the physics and physiology of speech production had been attained, particularly the codification of the source-filter theory of speech production.^{8,9} This understanding allowed researchers to create synthetic speech signals, systematically varying particular acoustic parameters and observing their perceptual effects. This approach was fundamental to our current understanding of speech perception,¹⁰ and it seems likely that the same will be true in the (still nascent) field of primate vocal production. Fortunately, much of our hard-won understanding of speech production appears to apply to non-human primate vocal production. This means that the tools and concepts of speech science can be adapted, with some modifications, to the study of non-human primate communication.

The following condensed overview of primate vocal production focuses on aspects of vocal production that seem likely to be important to auditory neuroscientists. To conserve space, we will assume some understanding of speech acoustics, as there are many easily accessible introductions to this topic.^{1,11,12} A survey of the considerable variability observed in the primate larynx and vocal tract follows the brief introduction to vocal acoustics. Owing to their strong reliance on acoustic communication, primates exhibit a wide variety of interesting vocal modifications, which, although poorly understood at present, provide a morphological basis for the

wide variation in vocal repertoires. This vocal diversity is probably correlated with variations in the perceptual mechanisms associated with species-specific vocal perception. Finally, a description of the implications of vocal production for analysis and synthesis of primate vocalizations focuses on the utility of linear prediction for analysis of vocal tract function and for the synthesis of natural-sounding acoustic stimuli and addresses both the power and potential pitfalls of this technique. When properly used, such techniques hold substantial promise for furthering the empirical study of primate vocal communication.

II. PRINCIPLES OF PRIMATE VOCAL PRODUCTION

The vocal tract of a terrestrial vertebrate is made up of three components: the lungs, which provide an air flow and the source of energy for phonation; the source, which converts this flow into sound; and a vocal tract, which filters this sound before releasing it into the environment. In most vertebrates and all terrestrial mammals, the source is made up of a pair of vocal folds contained within the larynx. (For a review of vertebrate vocal production in general, see Fitch and Hauser;¹³ reviews of mammalian laryngeal anatomy are available in Negus¹⁴ and Harrison.¹⁵)

A. THE MYOELASTIC-AERODYNAMIC THEORY OF PHONATION

Phonation is the process whereby the steady flow of air out of (or in some cases into) the lungs is converted into sound — that is, rapid pressure fluctuations in air. In primates, this process is accomplished by the vocal folds, which are set into self-sustained oscillations by the air flow. The vocal folds move rapidly in and out of the air stream, gating it into a series of pressure pulses or "puffs" of air and thus creating sound waves with a repetition frequency corresponding to the rate of vocal fold vibrations. This rate is termed the fundamental frequency (abbreviated F_0) and is the primary determinant of the pitch of the emitted sound (thus the term *pitch* should never be used to refer to other frequency components of a vocalization, such as the spectral peaks associated with formants). Typically, the spectrum of the sound created at the larynx includes a prominent component at the fundamental frequency, along with higher spectral components, termed harmonics, at integral multiples of F_0 .

Some early researchers believed that each pulse emitted from the larynx was accompanied by an active tensing and relaxation of the vocal fold musculature, the so-called "active" theory of phonation. This theory was proven false for humans by the demonstration of phonation in larynges excised from cadavers and thus deprived of nervous stimulation.^{16,17} In such an excised larynx preparation, the vocal folds of a "dead" larynx can be set into normal vibration by simply adjusting their anatomical position and passing warm humid air between them at the proper flow rate and pressure. Such experiments led to the explicit formulation of the myoelastic-aerodynamic theory by van den Berg,¹⁷ which has subsequently been verified with minor modifications by many researchers.¹⁸⁻²⁰ This theory holds that phonation results from the interaction between myoelastic restoring forces within the vocal folds and aerodynamic driving forces exerted by air flow on the exposed surfaces of the vocal folds. These forces together establish a self-oscillating system, meaning

that no active firing of neurons at the rate of vocal fold vibrations is necessary for phonation. Instead, neural signals to the laryngeal musculature (carried by branches of the vagus nerve) control fundamental frequency by varying the tension of the folds and their configuration.

Although the myoelastic-aerodynamic theory is widely accepted for human phonation and probably applies to most other mammals as well,^{21,22} at least one example of active phonation is known in mammals: purring by domestic cats. Purring is accomplished via active contractions of the laryngeal musculature at a rate (20–30 Hz) duplicating the purr F_0 .^{23,24} Because of the limitations on the possible rates of muscular contraction, such active phonation can only be expected to rates below about 40 Hz.²⁵ Purring is seen in most other felids,²⁶ and purr-like vocalizations are seen in some non-human primates (e.g., squirrel monkeys, *Saimiri*?). Therefore, the possibility that very low-frequency primate vocalizations might be accomplished via an active process should not be ignored. However, the vast majority of primate vocalizations have F_0 s considerably exceeding the limits of muscular contraction and are almost certainly produced using the more passive, myoelastic-aerodynamic mechanism.

Although the most common mode of vocal production in primates is phonation, other sound-generating sources are also available. The most primitive source in vertebrates is exemplified by hissing, a sound produced by members of all tetrapod classes. In hissing, air flows within a narrow constriction, generating turbulence. Turbulent noise has a very broad and aperiodic frequency spectrum. Human whispering is produced in this way, with the turbulence formed at the larynx between the non-vibrating vocal folds, and various fricative consonants (e.g., /s/, /sh/, /f/) are produced via constrictions formed elsewhere in the vocal tract (at the teeth or lips).

B. THE SOURCE-FILTER THEORY OF VOCAL PRODUCTION

Sound produced in the larynx must travel through the air-filled chambers of the throat (pharynx) and the nasal and oral cavities before being emitted into the environment (Figure 6.1). These air-filled chambers collectively comprise the vocal tract. Like any volume of air, the air contained within the vocal tract has multiple natural modes of vibration or resonance frequencies at which it prefers to vibrate. These vocal tract resonances are referred to as formants (from the Latin *formare*, "to shape"). Formants act as filters that shape the spectrum of the sound created by the source, preferentially passing energy at the formant frequencies. Human speech depends strongly on formants: vowels in all human languages are distinguished by their formant frequencies, and highly accurate speech perception is possible using synthetic signals that discard all other acoustic information besides formants.²⁸ Formants are manifested acoustically as broad peaks in the spectrum, often encompassing many harmonics. Formants are often particularly prominent in aperiodic (harsh or "noisy") sounds, where the sound source has a very broad frequency spectrum and no harmonics that could potentially be confused with the formants.

The source-filter theory of speech production^{28,29} is based on the observation that the contributions to speech of the laryngeal source and the vocal tract filter are clearly separable and additive. The central tenet of this theory is that source and



FIGURE 6.1 Schematic of the source-filter theory of vocal production. Sound production starts with the source, which turns the flow of air from the lungs into acoustic energy. In mammals, the source is typically the larynx, which contains the vibrating vocal folds, or vocal cords. These slap together to create a source signal composed of a fundamental frequency and many integer-multiple harmonics of that frequency. This sound then passes through the vocal tract, composed of the pharyngeal, oral, and nasal cavities. The air in these cavities possesses numerous vocal tract resonances, or formants, which filter the source signal to produce an output signal that is the combination of source and filter.

filter are independent. In humans, this is readily demonstrated by the fact that one can sing many different pitches (indicating different rates of vocal fold vibration) with a single vowel or produce many different vowels (indicating different formant frequencies) on the same pitch. The independence of source and filter in vocal production comes as a surprise to physicists familiar with the acoustics of wind instruments, because virtually all wind instruments rely on a strong coupling between source and "filter." Thus, the vibration frequency of the lips on the trombone (which are the source of acoustic energy) is largely determined by the frequencies of the resonances of the air column in the instrument, which are in turn determined by its length. Thus, in wind instruments (trumpets, flutes, clarinets, etc.) source-filter theory does not apply. This fundamental difference is often overlooked, providing a perennial source of confusion in the bioacoustics literature (see Fitch and Hauser¹ for more detail).

The existence of a vocal tract between the larynx and the environment is an anatomical fact for all primates, and the laws of physics dictate that the vocal tract will have resonant frequencies. Thus, it is somewhat surprising that the investigation and appreciation of formants in primate vocal production, despite work by early researchers (e.g., Lieberman²⁹ and Andrew³⁰), are fairly recent phenomena that have blossomed only in the last decade. The reasons for this neglect are obscure. Researchers may have been influenced by the notion that speech is special and, given the central importance of formants to speech, that formants are also special (uniquely human). More pragmatically, it is difficult to demonstrate conclusively that a particular spectral peak represents a formant (i.e., results from a vocal tract resonance) rather than a concentration of spectral energy originating from the source. The most conclusive demonstration of formants comes from experimental studies of vocalizations produced in light gases such as helium. Because the speed of sound is faster in a helium/oxygen atmosphere, formant frequencies are increased, and spectral analysis will show an upward shift of formants in this situation. Experiments with light gases have demonstrated the importance of vocal tract filtering in bird song,^{31,32} and cetacean clicks.³³ Similar work demonstrates that formant-like spectral peaks in

frog vocalizations are *not* the product of vocal tract filtering.³⁴ To the author's knowledge, no research on non-human primates vocalizing in helium has been published, and other methods have been used to demonstrate the existence of formants in non-human primate vocalizations.

Early workers observed spectral peaks in the vocalizations of apes, noted that the peaks were in roughly the appropriate frequency ranges given the animal's vocal tract length, and hypothesized that these features represented formants.³⁹ Later workers documented correlations between spectral peaks and facial movements, providing further evidence that such peaks represented formants in macaques and baboons,^{30,35,36} and researchers began to apply signal processing techniques appropriate for formants to non-human primates.³⁷ Perhaps the strongest evidence for formants in non-human primate calls came from the demonstration of strong correlations between vocal tract length, as measured from radiographs of anesthetized macaques, and formant frequencies, as measured from the same individuals' calls.³⁸ Because resonance frequencies are determined primarily by the length of the vocal tract, a tight correlation between multiple spectral peaks and vocal tract length can only be explained if the peaks indeed represent formants. The calls studied were aperiodic pant-threat calls, thus avoiding any potential confusion between formants and harmonics (Figure 6.2). This technique has since been applied successfully to other mammals,³⁹⁻⁴⁰ as well as to other primates (Fitch, in preparation; J. Fischer, personal communication). It is now widely accepted that the calls of many non-human primates possess formants, and given the clarity of macaques to perceive formants,⁴¹ recent research has begun to focus on the possible roles of formants in primate communication systems.⁴²⁻⁴⁵ This research endeavor has been significantly aided by an understanding of the production mechanism (see below).

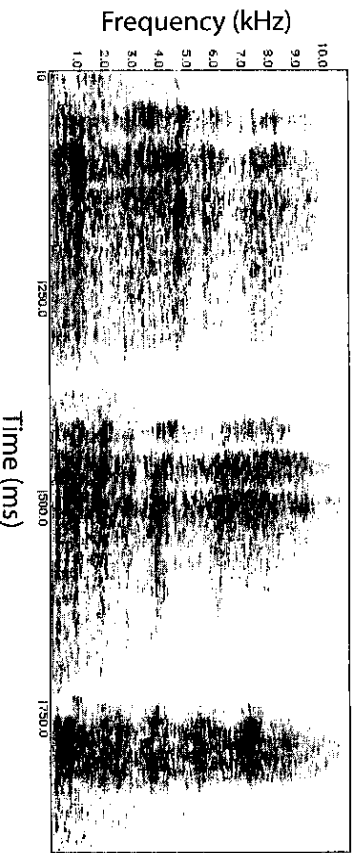


FIGURE 6.2 Spectrogram of a pant-threat vocalization from three different rhesus macaques. These calls are not clearly voiced and thus do not have a fundamental frequency and harmonics and show formant frequencies clearly. The clearly visible horizontal areas indicate the formant frequencies or vocal tract resonances. The calls come from females of increasing weight: 2.6 kg, 5.3 kg, and 9.2 kg, from left to right. The x-axis is time in msec; the y-axis is frequency in kHz.

C. NONLINEAR DYNAMICS IN VOCAL PRODUCTION

The most recent advance in our understanding of non-human vocal production has come from the application of the principles and tools of nonlinear dynamics to animal vocalizations⁴⁶⁻⁴⁷ (see Fitch et al.⁴⁸ for a more detailed introduction). Because the phonation mechanism is irreducibly nonlinear (as shown, for example, by the presence of harmonics in the source spectrum), it is surprising that the first applications of nonlinear dynamical theory to vocal production are quite recent, appearing first for human baby cries⁴⁸ and soon after for adult human pathological voices.⁴⁹ These seminal insights were quickly appreciated and augmented by leading speech researchers;⁵⁰ however, the significance of this research for normal speech is limited because healthy adults typically avoid the bifurcations, subharmonics, and chaos that are signatures of nonlinear dynamics. This limitation does not appear to apply in the case of non-human vocalizations, where such irregular phenomena are quite common (Figure 6.3).⁴⁵⁻⁴⁷ Thus, nonlinear dynamical theory is probably of greater significance in non-human primate vocalizations than in speech.

The fundamental insight of nonlinear dynamics is that simple deterministic systems can produce very complex and unpredictable output. Although initially understood for weather patterns by Lorenz,⁵¹ the importance of this insight for biological systems was recognized by biologists^{52,53} and has since played an important role in many branches of biology, including physiology⁵⁴ and neuroscience.^{55,56} In vocal production, the two vocal folds provide the simple system of two coupled oscillators, and all of the classic nonlinear phenomena (bifurcations, subharmonics, biphonation, and deterministic chaos) have been observed in this system. A hallmark of nonlinear dynamical systems is deterministic chaos, a very broad-spectrum, unpredictable behavior that nonetheless bears traces of order. It is now clear that the

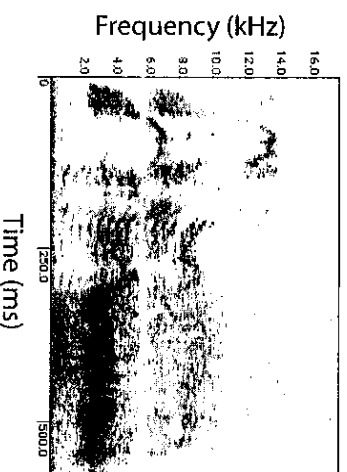


FIGURE 6.3 Spectrogram of a scream vocalization from a juvenile rhesus macaque, illustrating a series of nonlinear transitions. The call starts with a relatively low fundamental frequency, with clear harmonics, visible as horizontal stripe, but quickly transitions to a very high-frequency (around 7 kHz) tonal chirp. At about 100 msec, it returns to the lower frequency regime until around 300 msec, when it transitions to an irregular chaotic regime, with a clear concentration of energy around 3 kHz, which continues for the rest of the call. The x-axis is time in msec; the y-axis is frequency in kHz.

screen vocalizations that are produced under highly aversive circumstances by most mammals (including humans) represent deterministic chaos. A second signature of nonlinear systems is provided by bifurcations: abrupt transitions from one oscillatory regime to another (Figure 6.3). Such bifurcations can easily be observed in spectrograms of primate vocalizations, particularly in screams, but also in other vocalization types.⁴⁵ They may appear as "windows" of tonality in otherwise aperiodic vocalizations or as windows of chaos in otherwise tonal calls. Thus, nonlinear phenomena may provide a neutrally "cheap" way to generate a variety of vocal signals and thus to enrich the vocal repertoire.

Although we still have little detailed understanding of the role of nonlinear phenomena in perception, these new insights into vocal production may have interesting implications for auditory neuroscience. First, seemingly disparate phenomena (such as subharmonics or biphonation, which look tonal, and chaos, which looks noisy) often sound surprisingly similar to the human ear. From the perspective of motor theories of perception,⁵⁷ this is not surprising. Because the underlying motor-control signals for these superficially different call types might be identical, they would be perceived as similar. If this perceptual equivalence also applies for non-human primates, it would be intriguing to investigate at what level of the nervous system this perceptual equivalence is manifested. Second, nonlinear phenomena may provide distinctive vocal signatures to highlight identity or group membership.⁵⁸ Finally, nonlinear phenomena such as screams are often produced in aversive situations and are perceived as unpleasant. It has been hypothesized that their unpleasantness derives from their unpredictability in much the same way that prey species elude capture via unpredictable escape routines.⁵⁹ By this hypothesis, it is impossible to habituate to, and thus to ignore, an ever-changing "protean" sound.^{44,45}

D. SUMMARY OF VOCAL PRODUCTION

Air flow from the lungs is converted into sound by the source, typically the vibrating vocal folds housed in the larynx. The folds can vibrate in several modes. In the vibratory regime most resembling speech, the folds produce a periodic output consisting of a fundamental frequency and its integer-multiple harmonics. Such sounds have a clear pitch determined by the fundamental frequency. Other vibratory regimes of the larynx include deterministic chaos, which underlies screams and other irregular vocalizations. The source-generated sound then passes through the vocal tract, which filters out some frequencies while selectively allowing the passage of others which frequencies. The resulting spectral peaks correspond to vocal tract resonances or formants, which are present in many non-human vocalizations, and can be discriminated by macaques. Given the central importance of formant frequencies in human speech perception, it will be very interesting to investigate the degree to which non-human primates use formant frequencies in their communication systems.

III. PRIMATE VOCAL ADAPTATIONS

The principles described above apply to all primates (and indeed to most mammals). Within this unifying framework, however, members of the primate order exhibit an

impressive diversity of vocal adaptations: morphological peculiarities that play a role in vocal production. A well-known example is the huge larynx and hyoid apparatus of howler monkeys (genus *Alouatta*) which, adjusted for body size, is the largest larynx among primates. The howler hyolaryngeal source fills the entire space between the jaw and sternum and allows these relatively small monkeys to produce extremely loud, low-pitched vocalization.^{60a} Such differences in the relative size of the larynx can be observed even in closely related species.⁶² Following is a brief review of three additional types of adaptations: air sacs, vocal membranes, and vocal tract elongation. The introduction of new imaging tools⁶³ to the study of primate vocal anatomy should lead to rapid progress in understanding such adaptations.

A. LARYNGEAL AIR SACS

Many non-human primates possess out-pouchings of the epithelium lining the larynx or vocal tract into diverticula, termed *air sacs*.^{14,64} These can be small (the size of a pea in macaques) or large (6 liters in orangutans). In most of these, the opening of the sac is within millimeters of the vocal folds, which strongly suggests an acoustic function. The most common type of air sac is a subhyoid sac, a thin-walled diverticulum lining the hollowed-out body of the hyoid bone and opening into the larynx directly above the vocal folds. This type of sac is seen in almost all Old World monkeys (including vervets, baboons, and macaques) with the exception of a few colobine species.⁶⁵ It is also seen in some New World monkeys (*Aotus*, *Pithecia*, *Ateles*, *Lagothrix*, and *Alouatta*), but not in such common laboratory species as *Saimiri* or *Cebus*.⁶⁵ Although humans do not normally have air sacs, they do sometimes appear congenitally or with disease, a condition known as laryngocele, which is observed in some 2% of humans.⁶⁶ Laryngoceles duplicate precisely the location and structure of great ape air sacs.

Unfortunately, except for the work of Gautier⁶⁷ demonstrating a decrease in amplitude of calls from a DeBrazza's monkey (*Cercopithecus neglectus*) with a deflated sac, few empirical data are available to distinguish among the many hypotheses that have been offered for air sac function.¹ These hypotheses include the possibility that air sacs serve as impedance-matching systems, allowing low-frequency sounds to be more efficiently emitted into the environment (as proposed in frogs^{68,69}); that they serve as accessory lungs to increase the amount of air in an expiration and thus to increase pressure⁷⁰ and/or prolong vocalization;¹ or that they serve as coupled resonators to support phonation at particular frequencies. Less plausible hypotheses include the notion that they serve as oxygen storage mechanisms;¹⁴ — implausible because the sacs must be filled with expired air that will be poor in oxygen. Inflated air sacs could even serve as inflatable life-preservers during swimming! (Walrus air sacs appear to serve this function, along with their use in sound production.⁷¹) Because air sacs may support different functions in different species (or even in the same individual), it is currently difficult to determine which of these hypotheses, if any, applies in any particular species or call type.

An important reason that air sac function deserves scientific attention is that all great apes possess large inflatable air sacs.^{70,72,73} but humans lack them. Parsimony leads to the conclusion that the common ancestor of all these species, which would

also be the common ancestor of apes and humans, possessed an air sac (other hypotheses require its independent evolution in separate ape species). Thus, the common ancestor of chimpanzees and humans had an air sac, which was lost in humans after our divergence from chimps. This loss of a vocal adaptation is as striking as the descent of the human larynx, but has received much less attention.⁷⁴ Understanding why humans lost these air sacs, of course, requires an understanding of the role of air sacs in extant primates.

B. VOCAL MEMBRANES

Vocal membranes (often referred to as *vocal lips* after the German "Stimmlippe") are another common feature of the non-human primate vocal apparatus. Vocal membranes are thin, rostrally directed extensions of the vocal folds. They are present in a great diversity of non-human primates; however, it is not always clear from the anatomical literature which species have them. Many of the Old World monkeys have vocal membranes (e.g., *Papio*, *Macaca*, and *Cercopithecus*), as do chimpanzees⁷⁵ and gibbons.⁷⁶ Most New World monkeys also appear to have vocal membranes (*Cebus*, *Callithrix*, *Aotus*, *Saimiri*, and *Leontopithecus*).^{77,78}

Experimental investigations of the function of primate vocal membranes are lacking, but they are widely presumed to enable the production of high-frequency vocalization due to their extremely low mass.⁶⁴ This idea is supported by the fact that most microchiropteran bats also have vocal membranes, which are clearly implicated in the production of their ultrasonic echolocation calls.⁷⁹⁻⁸¹ Computer models of the vocal membranes suggest that, in addition to supporting high-frequency vocalization, they could improve efficiency of phonation and thus increase amplitude or decrease energy expenditure for loud calls.⁸² Vocal membranes may also result in increased coupling between the larynx and vocal tract, potentially leading to an increase in nonlinear phenomena.

C. VOCAL TRACT ELONGATION

A final interesting morphological oddity, vocal tract elongation, is seen in our own species and is also known in various other mammals. The oral vocal tract stretches from the larynx to the lips, while the nasal tract spans the larynx and nostrils. Thus, theoretically, the vocal tract can be elongated in three ways: (1) by extending the lips (which many primates do facultatively during calling); (2) by extending the nostrils (by elongating the nose, as in male proboscis monkeys⁸³); or (3) by lowering the larynx (as seen in humans, as well as in some deer⁴⁰ and large cats, such as lions and leopards⁸⁴). All of these changes have the same effect: by elongating the vocal tract, they lower formant frequencies. It has been proposed^{38,40,55,86} that this mechanism serves to exaggerate the impression of body size conveyed by vocalizations. Because vocal tract length is tightly tied to body size in most species,^{38,39,87} formant frequencies provide an accurate indication of body size in those species. However, once perceivers begin to rely upon this cue in judging the size of distant or obscured vocalizers, the stage is set for the evolution of a mechanism that fakes large body size by lowering the formants.⁸⁸ To the extent that such exaggeration is successful, it will spread through the population to become characteristic of the species as a whole.

Because all primates have a "free" hyoid that is not tightly bound to the skull base by a series of bony links, as in most mammals, their larynx is relatively mobile. Because all species thus far observed cineradiographically lower the larynx during vocalization,⁷⁴ it is likely that other primates lower the larynx during vocalization. By the size exaggeration hypothesis, this would be predicted to occur during agonistic vocalizations (such as growls or barks) or in territorial roars. Future observers should be alert to this possibility, as it is sometimes possible to observe laryngeal lowering with the naked eye, and then confirm it via acoustic measurements.⁴⁰

IV. PUTTING IT TO WORK: USING THE PRODUCTION PERSPECTIVE FOR ANALYSIS AND SYNTHESIS

A rich ethological approach to primate communication requires that researchers first study natural calls. This provides a source of hypotheses about call categories, the typical contexts in which calls are produced, and how listeners react to them (e.g., Goodall,⁸⁹ Struhsaker⁹⁰ and Hauser⁹¹). These hypotheses are then tested in perceptual experiments, which can include playing back calls in the field,^{92,93} testing calls in a more controlled laboratory setting,⁹⁴ or recording neural responses to calls.⁹⁵ The ultimate goal of this endeavor is to determine how animals process the complex flow of acoustic information to which they are continually exposed. What acoustic parameters are important? Which ones are ignored? How are these parameters used to categorize calls? The final section of this chapter shows how an understanding of primate vocal production can be used to inform this process. In particular, it focuses on the value of using digital sound synthesis to create natural-sounding primate calls. This allows the experimental manipulation of vocal stimuli to create, for example, call continua that vary along some specific parameter of interest. Though still in its infancy, this approach has great promise.

A. WHY SYNTHESIS?

There are two common approaches to studying auditory function in animals, both of which have corresponding approaches in research on humans. The first, which is favored by psychoacousticians, uses relatively simple signals such as sine waves or noise bursts to explore low-level auditory function. This approach has been used to uncover and explore such key phenomena as equal-loudness contours, critical bands, and the missing fundamental (see Moore⁹⁶ for an introduction). Because a linear system can be fully characterized by its response to isolated sine waves (or from its impulse response), this is a reasonable approach to take to initially characterize basic auditory function; however, the auditory system is not linear. It was quickly realized in the 1950s that tone/noise psychophysics would be inadequate to uncover the perceptual mechanisms underlying speech perception, and that richer signals were needed to explore higher level audition.¹⁰ It seems apparent that similar arguments apply to non-human primates as well (e.g., Rauschecker⁹⁵), and that tone/noise psychophysics, while playing an important role, will never suffice for a complete understanding of primate auditory function.

A second common approach, more typical for bioacousticians, is to use unmodified natural calls as stimuli in perceptual experiments. Relative to the previous approach, recorded calls have the advantage of being complex and realistic, thus presumably enlisting the natural perceptual mechanisms that animals use in their communication systems. However, recorded calls have distinct disadvantages as well. First, the vocalizing animal is in control of the call parameters, not the experimenter. This makes the construction of stimulus sets quite difficult, as the researcher must collect a large set of natural calls in order to pick out a few *post hoc* that fall along some continuum (or represent normal exemplars of some category). Second, and more important, it is virtually guaranteed that such a collection of natural calls will also vary in acoustic parameters other than the one of interest, and some of these uncontrolled variables may be significant to the primate listener, but less so (or even inaudible) to the human experimenter. Thus, when using prerecorded calls, we can never be certain of what variables the listener is responding to. It is worthwhile to consider what aspects of human speech perception would have gone undiscussed if only recordings of natural speech had been available. Cues such as formant transitions and voice onset time, and phenomena such as categorical perception, trading relations, prototype magnet effects, duplex perception, and many other central phenomena in speech perception would never have been discovered if researchers had lacked the ability to synthesize speech stimuli.

The solution to these problems is, of course, to synthesize primate calls.⁹⁷⁻⁹⁹ Modern digital sound synthesis techniques, originally developed for speech but easily modified to work with animal vocalization, allow the creation of synthetic calls that are indistinguishable from natural vocalizations. Such techniques can easily be implemented on standard desktop computers using widely available packages such as Matlab. Although commercial "monkey call synthesis" packages are not available (and seem unlikely to appear), an understanding of call production acoustics, along with the potential pitfalls of the algorithms, puts the creation of natural-sounding synthetic vocalizations within reach of any researcher interested in auditory perception. Following is an overview of one widely available technique, linear prediction, which points out both the pitfalls and power associated with this technique.

B. LINEAR PREDICTION IN BIOACOUSTICS

Linear prediction (LP) is a powerful spectral modeling technique based on the source-filter theory of speech described above. LP (often called LPC [linear predictive coding] when used for speech compression) became a standard tool for speech engineers during the 1970s.^{100,101} Given appropriate vocalizations, linear prediction makes it possible to separate a vocalization into its source and filter components, to modify one or the other, and then put them back together into a highly realistic and natural-seeming synthetic signal. Thus, researchers interested in the perception of vocal tract resonances can modify a specific formant or formants and then recombine the new filter with the original source signal.⁹⁹ Researchers interested in pitch perception can modify the LP-derived source using such techniques as phase-vocoding or PSOLA, and then recombine it with the LP-derived filter to obtain natural sounding calls.

Of course, no algorithm can apply to all vocalization types, and linear prediction is no exception. LP is inappropriate for calls where the fundamental frequency is near or exceeds the formant frequencies (high whines or squeaks) or where vocal tract filtering is not evident (e.g., screams). However, linear prediction is appropriate for many calls with relatively low fundamentals or aperiodic sources and clear vocal tract filtering, and the technique has been successfully used both in analysis^{38,102,103} and synthesis^{97,99} of animal calls. Unfortunately, the value of linear prediction remains relatively unappreciated by researchers in bioacoustics and auditory neuroscience.

The power of linear prediction derives from a very strong limiting assumption about the nature of the signal: that the transfer function contains only resonances (termed *poles*) and not antiresonances (termed *zeros*). A general linear filter can contain either type of spectral feature, but linear prediction assumes an all-pole filter. Fortunately, this assumption is basically correct for speech (with some exceptions, such as nasal sounds such as /m/ and /n/, in which sound resonates in both the nasal and oral cavities). The all-pole model also appears to hold for those mammal vocalizations that have been studied to date (including one primate species). The author⁷⁴ studied vocal production in dogs, goats, pigs, and cotton-top tamarin monkeys using cinematography and found that loud calls in all of these species are made by lowering the larynx into the oral cavity and closing the velum, thus sealing off the nasal cavity and avoiding nasal antiresonances. For calls produced in this manner, the assumption of an all-pole filter appears valid. Other calls (such as pig grunts and dog whines) were produced as purely nasal vocalizations, with the larynx inserted into the nasal passage, so that the oral side-branch that is responsible for the antiresonance in nasalized speech sounds such as /m/ is avoided. Although the trachea, or the complex structure of the nasal passages, might still introduce antiresonances in pure nasals,^{104,105} these features appear to be relatively subtle and can be adequately modeled with several poles using linear prediction.¹⁰⁰

Of course, the production data of Fitch⁷⁴ apply only to the species studied, and research must proceed on a case-by-case basis. The primate air sacs described above may introduce antiresonances into the vocal tract filter¹⁰⁶ (although these are not prominent in macaques having a median subhyoid air sac³⁸). If so, these calls would require more general models that can include both poles and zeros. In the somewhat arcane terminology of speech engineering, these are often termed ARMA models, because they include both autoregressive (AR, or pole) and moving-average (MA, or zero) components. Unfortunately, the lifting of the all-pole assumption makes it quite difficult to derive accurate ARMA models based solely on the vocal signal. While linear prediction is guaranteed to converge to the optimal model (in a minimum least-squared error sense), no such guarantees are available with unconstrained models such as ARMA. Further, LP is robust to minor deviations from an all-pole system, because zeros can be adequately modeled by multiple poles.¹⁰⁰ Finally, while the correct analysis parameters for linear prediction can be estimated from very basic anatomical data (such as measurements of a skull, see below), the proper number of poles and zeros in an ARMA model is difficult to determine empirically but has a critical effect on the output from the algorithm. Thus, when justified by production data, linear prediction is the preferred technique. The available data, both acoustic

and cineradiographic, suggest that the all-pole assumption will be valid for a broad variety of primate vocalizations, underscoring the value of linear prediction in bioacoustics.

C. ANALYSIS AND SYNTHESIS USING LINEAR PREDICTION

Linear prediction designates a family of spectral estimation techniques; many different algorithms are available to perform LP analysis and different analysis packages implement different algorithms. A complete review of these algorithms or analysis packages is beyond the scope of this chapter; two widely available systems that implement LP and are recommended are Matlab (<http://www.mathworks.com/>) and Praat (<http://www.praat.org>). A precise mathematical introduction to LP is given by Markel and Gray,¹⁰¹ and many practical details relevant to bioacoustics are covered by Owren and Bernacki.¹⁰² This section provides a practical overview of how linear prediction can be used in the synthesis of mammal sounds, highlights the key parameter choices, and stresses some of the potential pitfalls to be avoided. This overview results from the author's years of experience using and teaching linear prediction, first as applied to human speech and later to vertebrate vocalizations, and integrates insights from the literature in speech science, computer music, and vertebrate vocal production.

The essential first step in any LP analysis is deriving an accurate model of the vocal tract transfer function, in which each real formant (in the real vocal tract of the animal) is fitted by a pole (in the mathematical model provided by the LP analysis). Once this is adequately accomplished, we can separate the source from the filter components, and then modify either one before recombining them; or we can cross-synthesize sources and filters from different individuals or even species. This critical and most challenging first step of LP analysis includes the initial modeling of the spectrum of the signal by an all-pole filter. At a minimum, this step requires a high-quality recording (if background noise, such as wind, strong reverberation, or crickets, is present, LP will model the noise along with the animal vocalization of interest; that is, "garbage in gives garbage out"). As mentioned earlier, the vocalization must be also be of a type suited for source-filter analysis in the first place (formants must be present and clearly outlined by the source). If these minimal criteria are met, the key parameters in an LP analysis are the bandwidth of the initial signal (the sampling rate should be adjusted so that the formants fill the range between 0 Hz and the Nyquist frequency) and the number of poles.

Given a segment of sampled sound and a prespecified number of poles, LP provides the frequencies and amplitudes for the poles that most accurately fit the overall spectrum of the sound. Although the technique is guaranteed to give the optimal solution with a given number of poles, it does *not* guarantee that these poles represent true vocal tract resonances of the vocalizer. This is because LP is critically sensitive to the number of poles chosen for the analysis. For instance, if a human vowel signal possessing four formants is subjected to a two-pole analysis, each pole will be forced to fit two formants and assume an intermediate frequency value and broad bandwidth. These pole frequencies will obviously not correspond to actual formant frequencies. On the other hand, if the same signal is subjected to a 50-pole analysis, each pole

might fit one harmonic of the source signal, rather than fitting the formants. Thus, a key to successful LP analysis is the choice of the number of poles. An incorrect choice will produce meaningless results. In my experience, choosing the correct LP parameters requires hours of experimentation and careful listening and a willingness to conclude that a particular signal, or call type, will simply not work.

Choosing the number of poles is best done with some knowledge of the vocal tract length. When the vocal tract length is known, the number of formants (N) can be roughly calculated by:

$$N = \frac{2L}{c} f_c \quad (6.1)$$

where L is the length of the vocal tract (in m), c is the speed of sound (350 m/sec in the moist 37°C air of a mammalian vocal tract), and f_c is the cutoff frequency (in Hz; e.g., the Nyquist frequency of the sampled signal). Note, however, that often no source energy exists at higher frequencies (e.g., above about 5 kHz in speech) and thus no energy is present in the acoustic signal at higher formant frequencies. Thus, signals should always be appropriately filtered and downsampled so that their spectrum fills the entire frequency range between 0 Hz and the Nyquist frequency (half the sampling rate). For human speech, a Nyquist of around 5 kHz (sampling rate of 10 kHz) often gives good results; for other species, this number may vary widely. Note also that LP often employs several poles as shaping poles (fitting the overall spectrum rather than particular formants), so the number of poles in the LP analysis will typically be greater than the number of formants suspected in the signal based on Eq. (6.1).

As a practical example, let us analyze the pant-threat vocalization of a rhesus macaque, sampled at 22 kHz. The vocal tract length of a rhesus macaque, with lips extended in the grunt or pant-threat position,¹⁰⁷ varies between 5 and 10 cm (for small juvenile and large adult males, respectively), giving between three and seven formants under the 11-kHz Nyquist frequency. Because each pole is specified by two coefficients, and given an extra shaping pole, this gives an appropriate LP order of 8 to 20 coefficients, depending on the size of the animal.³⁸ Note the lack of standardized terminology for the parameter determining the number of poles: the number of coefficients in the LP filter is twice the number of poles. Researchers must be careful to use the lowest number of poles that achieves a good fit to the spectrum. This is best determined by visual inspection of a frequency-domain plot of the LP impulse response (which shows the combined frequency estimate from all the poles) superimposed over the magnitude spectrum of the original signal.

Once a signal bandwidth and number of poles have been determined that provide an accurate fit to the formants of the vocal tract filter, it is possible to isolate the source signal from the filter. The vocal tract filter function is captured by the LP coefficients (which can now be modified flexibly). To derive the source, we invert this filter and then use it to filter the original source signal. Many speech packages (e.g., Praat) include a special inverse filter function for doing so; in a general-purpose program, such as Matlab, that uses a digital filter framework, this involves simply

using the poles, which are the denominator coefficients, as zeros, which are the numerator coefficients. The inverse filtering process, by removing the effect of the vocal tract filter, produces an output that is an estimate of the original source signal. We now have a source signal (representing the output from the larynx, plus a radiation term) and a filter (representing the vocal tract transfer function).

The separated source and filter can now be used for further analysis or modified in various ways. Source pitch or spectral slope can be freely modified, as can individual formant center frequencies or bandwidths or the overall formant spacing. Sources can be combined with filters other than the ones they were originally associated with, allowing cross-synthesis of different call types or even creating hybrid calls by combining the source and filter of different species. Synthesis is completed by passing the source signal (derived by inverse filtering and optionally modified) back through the filter function (derived from LP and optionally modified). If no changes are made to either component, this process will reproduce the original signal almost exactly (to within round-off error), which provides a way of checking that the whole procedure is working correctly (and is strongly recommended). When done correctly, the signals produced by this process are indistinguishable from naturally produced signals (both for humans and those non-human species that have been tested; see, for example, Fitch and Kelley⁸⁹). In summary, linear prediction provides a framework allowing researchers to tear apart a signal into its source and filter components. Source and filter components can then be independently analyzed or modified and finally recombined into highly realistic synthetic calls.

As an example of the types of questions that can be addressed with this technique, rhesus macaques typically assume different facial postures during particular call types. For example, copulation calls or screams are always made with retracted lips, while coos or grunts are made with protruded lips.^{35,36} These different articulatory positions modify the vocal tract filter function, both by changing the overall vocal tract length and by subtly changing vocal tract shape. It is currently unknown whether monkeys are sensitive to these acoustic differences or whether the facial postures associated with calls simply represent the visual component of a multimodal signal, the incidental acoustic effects of which are ignored. This question could be answered by cross-synthesizing copulation scream sources with coo or grunt filters (or vice versa) and then using playback to determine how such hybrid calls are categorized. Other questions suited to this approach concern judgments of individual identity (do formants carry specific, high-quality cues to individuality, as some researchers^{1,4,108} have hypothesized?) or the role of different formants as cues to body size³⁸ or call category.⁹⁷ These questions are interesting in their own right but take on additional importance due to the central importance of formants in human speech. Understanding the perception and categorization of formant-rich vocalizations in non-human primates is thus a crucial component in the quest to understand the evolutionary precursors of speech perception mechanisms.

V. CONCLUSION

This brief review of primate vocal production has provided an overview of recent progress in understanding the physics and physiology of vocal production in

non-human primates and of the morphological variability in primate production mechanisms. Although a number of outstanding questions in the field have been highlighted, it should be clear that we now have a firm enough understanding of the basics of primate vocal production to open new realms of investigation to researchers interested in audition. To assist this cross-pollination, this chapter has provided an introduction to linear prediction, a signal processing technique based upon production principles and source-filter theory. Linear prediction and similar techniques provide an important tool that can be used to address previously unapproachable issues in primate vocal perception. In particular, because human speech relies so extensively upon formant perception, more detailed analysis of formant perception in non-human primates should eventually provide important insights into the evolutionary precursors (or lack thereof) of speech and language.⁸⁵

REFERENCES

1. Fitch, W.T. and Hauser, M.D., Vocal production in nonhuman primates: acoustics, physiology, and functional constraints on "honest" advertisement. *Am. J. Primatol.*, 37, 191-219, 1995.
2. Hauser, M.D., *The Evolution of Communication*. MIT Press, Cambridge, MA, 1996.
3. Bradbury, J.W. and Vehrencamp, S.L., *Principles of Animal Communication*. Sinauer Associates, Sunderland, MA, 1998.
4. Zoloth, S.R. et al., Species-specific perceptual processing of vocal sounds by monkeys. *Science*, 204, 870-872, 1979.
5. Zimmernann, E., Castration affects the emission of an ultrasonic vocalization in a nocturnal primate, the grey mouse lemur (*Microcebus murinus*). *Physiol. Behav.*, 60, 693-697, 1996.
6. Zimmernann, E., First record of ultrasound in two prosimian species, *Nannomysset-shoffen*, 68, 531, 1981.
7. Nowak, R.M., *Walker's Mammals of the World*. Johns Hopkins University Press, Baltimore, MD, 1991.
8. Chiba, T. and Kajiyama, M., *The Vowel: Its Nature and Structure*. Tokyo-Kaiseikan, Tokyo, Japan, 1941.
9. Fant, G., *Acoustic Theory of Speech Production*. Mouton & Co., The Hague, 1960.
10. Liberman, A.M., *Speech: A Special Code*. MIT Press, Cambridge, MA, 1996, p. 458.
11. Titze, I.R., *Principles of Voice Production*. Prentice Hall, Englewood Cliffs, NJ, 1994.
12. Lieberman, P. and Blunstein, S.E., *Speech Physiology, Speech Perception, and Acoustic Phonetics*. Cambridge University Press, Cambridge, U.K., 1988.
13. Fitch, W.T. and Hauser, M.D., Unpacking "honesty": Vertebrate vocal production and the evolution of acoustic signals. In *Acoustic Communication*, A. Simmons, R.R. Fay, and A.N. Popper, Eds., Springer, New York, 2002.
14. Negus, V.E., *The Comparative Anatomy and Physiology of the Larynx*. Hafner Publishing, New York, 1949.
15. Harrison, D.F.N., The anatomy and physiology of the mammalian larynx. Cambridge University Press, New York, 1995.
16. Mueller, J., *The Physiology of the Senses, Voice and Muscular Motion with the Mental Faculties*. Taylor, Walton and Maberty, London, 1848.

17. van den Berg, J., Myoelastic-aerodynamic theory of voice production, *J. Speech Hearing Res.*, 1, 227-244, 1958.
18. Lieberman, P., *Intonation, Perception and Language*, MIT Press, Cambridge, MA, 1967.
19. Titze, I.R., On the mechanics of vocal fold-vibration, *J. Acoust. Soc. Am.*, 60, 1366-1380, 1976.
20. Titze, I.R., Comments on the myoelastic-aerodynamic theory of phonation, *J. Speech Hearing Res.*, 23, 495-510, 1980.
21. Paulsen, K., *Das Prinzip der Stimmbildung in der Wirbeltierreihe und beim Menschen*, Akademische Verlagsgesellschaft, Frankfurt, Germany, 1967, p. 143.
22. Slavit, D.H., Lipton, R.J., and McCaffrey, T.V., Glotographic analysis of phonation in the excised canine larynx, *Ann. Otol. Rhinol. Laryngol.*, 99, 396-402, 1990.
23. Frazer Sissom, D.E., Rice, D.A., and Peters, G., How cats purr, *J. Zool. (London)*, 223, 67-78, 1991.
24. Renniers, J.E. and Gautier, H., Neural and mechanical mechanisms of feline purring, *Respir. Physiol.*, 16, 351-361, 1972.
25. Hirose, H. et al., An experimental study of the contraction properties of the laryngeal muscles in the cat, *Ann. Otol. Rhinol. Laryngol.*, 78, 297-306, 1969.
26. Peters, G. and M. Hast, Hyoid structure, laryngeal anatomy, and vocalization in felids, *Zeitschrift für Säugetierkunde*, 59, 87-104, 1994.
27. Winter, P., Ploog, D., and Latta, J., Vocal repertoire of the squirrel monkey (*Saimiri sciureus*), its analysis and significance, *Exp. Brain Res.*, 1, 359-384, 1966.
28. Remez, R.E. et al., Speech perception without traditional speech cues, *Science*, 212, 947-950, 1981.
29. Lieberman, P., Primate vocalization and human linguistic ability, *J. Acoust. Soc. Am.*, 44(6), 1574-1584, 1968.
30. Andrew, R.J., Use of formants in the grunts of baboons and other nonhuman primates, *Ann. N.Y. Acad. Sci.*, 280, 673-693, 1976.
31. Nowicki, S., Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere, *Nature*, 325, 53-55, 1987.
32. Nowicki, S. and Marler, P., How do birds sing?, *Music Percept.*, 5, 391-426, 1988.
33. Annudsen, M., Helium effects on the click frequency spectrum of the Harbor porpoise, *Phocoena phocoena*, *J. Acoust. Soc. Am.*, 90, 53-59, 1991.
34. Rand, A.S. and Dudley, R., Frogs in helium: the anuran vocal sac is not a cavity resonator, *Physiol. Zool.*, 66, 793-806, 1993.
35. Hauser, M.D., Evans, C.S., and Marler, P., The role of articulation in the production of rhesus monkey (*Macaca mulatta*) vocalizations, *Animal Behav.*, 45, 423-433, 1993.
36. Hauser, M.D. and Schön Ybarra, M., The role of lip configuration in monkey vocalizations: experiments using xylocane as a nerve block, *Brain Lang.*, 46, 232-244, 1994.
37. Owren, M.J. and Bernacki, R., The acoustic features of vervet monkey (*Cercopithecus aethiops*) alarm calls, *J. Acoust. Soc. Am.*, 83, 1927-1935, 1988.
38. Fitch, W.T., Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques, *J. Acoust. Soc. Am.*, 102, 1213-1222, 1997.
39. Riede, T. and Fitch, W.T., Vocal tract length and acoustics of vocalization in the domestic dog *Canis familiaris*, *J. Exp. Biol.*, 202, 2859-2867, 1999.
40. Fitch, W.T. and Reby, D., The descended larynx is not uniquely human, *Proc. Roy. Soc. Biol. Sci.*, 268, 1669-1675, 2001.
41. Sommers, M.S. et al., Formant frequency discrimination by Japanese macaques (*Macaca fasciata*), *J. Acoust. Soc. Am.*, 91, 3499-3510, 1992.
42. Rendall, D., Rodman, P.S., and Emmond, R.E., Vocal recognition of individuals and kin in free-ranging rhesus monkeys, *Animal Behav.*, 51, 1007-1015, 1996.
43. Owren, M.J., Seyfarth, R.M., and Cheney, D.L., The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): implications for production processes and functions, *J. Acoust. Soc. Am.*, 101, 2951-2963, 1997.
44. Owren, M.J. and Rendall, D., An affect-conditioning model of nonhuman primate vocal signaling, in *Perspectives in Ethology*, Vol. 12, *Communication*, D.H. Owings, M.D. Beecher, and N.S. Thompson, Eds., Plenum Press, New York, 1997, pp. 299-346.
45. Fitch, W.T., Neubauer, J., and Herzel, H., Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production, *Animal Behav.*, 63, 407-418, 2002.
46. Wilden, I. et al., Subharmonics, biphonation, and deterministic chaos in mammal vocalization, *Biocacustics*, 9, 171-196, 1998.
47. Fee, M.S. et al., The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird, *Nature*, 395, 67-71, 1998.
48. Mende, W., Herzel, H., and Wernke, K., Bifurcations and chaos in newborn infant cries, *Phys. Lett. A*, 145, 418-424, 1990.
49. Herzel, H. et al., Chaos and bifurcations during voiced speech, in *Complexity, Chaos and Biological Evolution*, E. Mosekilde and L. Mosekilde, Eds., Plenum, New York, 1991, pp. 41-50.
50. Titze, I.R., Baken, R., and Herzel, H., Evidence of chaos in vocal fold vibration, in *Vocal Fold Physiology: New Frontiers in Basic Science*, I. Titze, Ed., Singular Publishing Group, San Diego, CA, 1993, pp. 143-188.
51. Lorenz, E.N., Deterministic nonperiodic flow, *J. Atmos. Sci.*, 20, 130-141, 1963.
52. May, R.M., Biological populations with nonoverlapping generations: stable points, stable cycles and chaos, *Science*, 186, 645-647, 1974.
53. May, R.M., Simple mathematical models with very complicated dynamics, *Nature*, 261, 459-467, 1976.
54. Glass, L. and Mackey, M.C., *From Clocks to Chaos: The Rhythms of Life*, Princeton University Press, Princeton, NJ, 1988.
55. Freeman, W.J. and Skarda, C.A., Spatial EEG patterns, nonlinear dynamics and perception: the neo-Sheringtonian view, *Brain Res. Rev.*, 10, 147-175, 1985.
56. Babloyantz, A. and Destexhe, A., Low-dimensional chaos in an instance of epilepsy, *Proc. Natl. Acad. Sci.*, 83, 3513-3517, 1986.
57. Lieberman, A.M. and Mattingly, I.G., The motor theory of speech perception revised, *Cognition*, 21, 1-36, 1985.
58. Hauser, M.D., Articulatory and social factors influence the acoustic structure of rhesus monkey vocalizations: a learned mode of production?, *J. Acoust. Soc. Am.*, 91, 2175-2179, 1992.
59. Driver, P.M. and Humphries, D.A., *Protean Behaviour: The Biology of Unpredictability*, Oxford University Press, London, 1988, p. 360.
60. Kelenan, G. and Sade, J., The vocal organ of the howling monkey (*Alouatta palliata*), *J. Morphol.*, 107, 123-140, 1960.
61. Schön Ybarra, M., Morphological adaptations for loud phonation in the vocal organ of howling monkeys, *Primate Rep.*, 22, 19-24, 1988.
62. Hill, W.C.O. and Booth, A.H., Voice and larynx in African and Asiatic Colobidae, *J. Bombay Nat. Hist. Soc.*, 54, 309-321, 1957.
63. Nishimura, T. et al., New methods of morphological studies with minimum invasiveness, *Primate Res.*, 15, 259-266, 1999.

64. Schon Ybarra, M., A comparative approach to the nonhuman primate vocal tract: implications for sound production, in *Current Topics in Primate Vocal Communication*, E. Zimmerman and J.D. Newman, Eds., Plenum, New York, 1995, pp. 185-198.
65. Lampert, H., Zur Kenntnis des Platyrhinenkehltopfes, *Gegenbaurs Morphologisches Jahrbuch*, 1926, 607-654, 1926.
66. Buysse, M.L., *Birth Defects Encyclopedia*, Blackwell Scientific Publications, St. Louis, MO, 1990.
67. Gautier, J.P., Etude morphologique et fonctionnelle des annexes extra-laryngées des cercopitheciinae; liaison avec les cris d'espacement, *Biol. Gabonica*, 7(2), 230-267, 1971.
68. Ryan, M.J., *The Tanager Frog: A Study in Sexual Selection and Communication*, University of Chicago Press, 1985.
69. Dudley, R. and Rand, A.S., Sound production and vocal sac inflation in the Tanager frog, *Physalidemus pustulosus* (Leptodactylidae), *Copeia*, 1991(2), 460-470, 1991.
70. Kalemien, G., Anatomy of the larynx and the anatomical basis of vocal performance, in *The Chimpanzee*, Vol. 1, G. Bourne, Ed., S. Karger, Basel, 1969, pp. 165-187.
71. Fay, F.H., Structure and function of the pharyngeal pouches of the walrus (*Odobenus rosmarus* L.), *Mammalia*, 24, 361-371, 1960.
72. Gregory, W.K., *The Anatomy of the Gorilla: The Studies of Henry Cushman Raven and Contributions by William B. Atkinson [and Others]*, Columbia University Press, New York, 1950.
73. Brandes, R., Über den Kehlkopf des Orang-utan in verschiedenen Altersstadien mit besonderer Berücksichtigung der Kehlsackfrage, *Morphologisches Jahrbuch*, 69, 1-61, 1931.
74. Fitch, W.T., The phonetic potential of nonhuman vocal tracts: comparative cinematographic observations of vocalizing animals, *Phonetica*, 57, 205-218, 2000.
75. Kalemien, G., Comparative anatomy and performance of the vocal organ in vertebrates, in *Acoustic Behavior of Animals*, R. Busnel, Ed., Elsevier, Amsterdam, 1963, pp. 489-521.
76. Neman, J. and Kalemien, G., Beiträge zur Kenntnis des Gibbonkehltopfes, *Zeitschrift für Anatomie und Entwicklungsgeschichte*, 59, 259-292, 1933.
77. Starck, D. and Schneider, R., Respirationsorgane, in *Primatologia*, Vol. III, Part 2, H. Hofer, A.H. Schultz, and D. Starck, Eds., S. Karger, Basel, 1960.
78. Neman, J., Das Stimmorgan der Primaten, *Zeitschrift für Anatomie und Entwicklungsgeschichte*, 81, 657-672, 1926.
79. Griffin, D.R., *Listening in the Dark*, Yale University Press, New Haven, CT, 1958.
80. Griffiths, T.A., Modification of *M. cricothyroideus* and the larynx in the Mormoopidae, with reference to amplification of high-frequency pulses, *J. Mammol.*, 59, 724-730, 1978.
81. Suthers, R.A., The production of echolocation signals by bats and birds, in *Animal Sonar: Processes and Performance*, P.E. Nachtigall and P.W.B. Moore, Eds., Plenum, New York, 1988, pp. 23-45.
82. Mergell, P., Fitch, W.T., and Herzel, H., Modeling the role of non-human vocal membranes in phonation, *J. Acoust. Soc. Am.*, 105, 2020-2028, 1999.
83. Napier, J.R. and Napier, P.H., *The Natural History of the Primates*, MIT Press, Cambridge, MA, 1985.
84. Hast, M., The larynx of roaring and non-roaring cats, *J. Anat.*, 163, 117-121, 1989.
85. Fitch, W.T., The evolution of speech: a comparative review, *Trends Cognitive Sci.*, 4, 258-267, 2000.

86. Fitch, W.T., Acoustic exaggeration of size in birds by tracheal elongation: comparative and theoretical analyses, *J. Zool. (London)*, 248, 31-49, 1999.
87. Fitch, W.T., Skull dimensions in relation to body size in nonhuman mammals: the causal bases for acoustic allometry, *Zoology*, 103, 40-58, 2000.
88. Krebs, J.R. and Dawkins, R., Animal signals: mind reading and manipulation, in *Behavioural Ecology*, J.R. Krebs and N.B. Davies, Eds., Sinauer Associates, Sunderland, MA, 1984, pp. 380-402.
89. Goodall, J., *The Chimpanzees of Gombe: Patterns of Behavior*, Harvard University Press, Cambridge, MA, 1986.
90. Struhsaker, T.T., Auditory communication among vervet monkeys (*Cercopithecus aethiops*), in *Social Communication among Primates*, S.A. Altmann, Ed., Chicago University Press, 1967, pp. 281-324.
91. Hauser, M.D., Rhesus monkey (*Macaca mulatta*) copulation calls: honest signals for female choice?, *Proc. Roy. Soc. London*, 254, 93-96, 1993.
92. Seyfarth, R.M., Cheney, D.L., and Marler, P., Monkey responses to three different alarm calls: evidence of predator classification and semantic communication, *Science*, 210, 801-803, 1980.
93. Hauser, M.D., Functional referents and acoustic similarity: field playback experiments with rhesus monkeys, *Animal Behav.*, 55, 1647-1658, 1998.
94. Owren, M.J., Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans. I. Natural calls, *J. Comp. Psychol.*, 104, 20-28, 1990.
95. Rauschecker, J.P., Tian, B., and Hauser, M., Processing of complex sounds in the macaque nonprimary auditory cortex, *Science*, 268, 111-114, 1995.
96. Moore, B.C.J., *An Introduction to the Psychology of Hearing*, Academic Press, New York, 1988.
97. Owren, M.J., Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans. II. Synthetic calls, *J. Comp. Psychol.*, 104, 29-40, 1990.
98. Norcross, J.L., Newman, J.D., and Fitch, W., Responses to natural and synthetic phee calls by common marmosets, *Am. J. Primatol.*, 33, 15-29, 1994.
99. Fitch, W.T. and Kelley, J.P., Perception of vocal tract resonances by whooping cranes, *Grus americana*, *Ethology*, 106(6), 559-574, 2000.
100. Aial, B.A. and Hanauer, S.L., Speech analysis and synthesis by linear prediction of the speech wave, *J. Acoust. Soc. Am.*, 50(2), 637-655, 1971.
101. Markel, J.D. and Gray, A.H., *Linear Prediction of Speech*, Springer-Verlag, New York, 1976.
102. Carrette, E., Shipley, C., and Buchwald, J., Linear prediction theory of vocalization in cat and kitten, in *Frontiers in Speech Communication Research*, B. Lindblom and S. Ohman, Eds., Academic Press, New York, pp. 245-257, 1979.
103. Owren, M.J. and Bernacki, R.H., Applying linear predictive coding (LPC) to frequency-spectrum analysis of animal acoustic signals, in *Animal Acoustic Communication: Sound Analysis and Research Methods*, S.L. Hopp, M.J. Owren, and C.S. Evans, Eds., Springer, New York, pp. 130-162, 1998.
104. Dang, J. and Honda, K., Acoustic characteristics of the human paramasal sinuses derived from transmission characteristic measurement and morphological observation, *J. Acoust. Soc. Am.*, 100, 3374-3383, 1996.
105. Dang, J., Honda, K., and Suzuki, H., Morphological and acoustical analysis of the nasal and the paramasal cavities, *J. Acoust. Soc. Am.*, 96(4), 2088-2100, 1994.
106. Hainhoff, E.H., Occurrence of anti-resonance in the song of the stamang (*Hylobates syndactylus*), *Am. J. Primatol.*, 5, 249-256, 1983.

107. Hauser, M.D., Evans, C.S., and Marler, P., The role of articulation in the production of rhesus monkey (*Macaca mulatta*) vocalizations, *Animal Behav.* 45, 423-433, 1993.
108. Rendall, D.A., Social Communication and Vocal Recognition in Free-Ranging Rhesus Monkeys (*Macaca mulatta*), University of California, Davis, 1996.

7 Developmental Modifications in the Vocal Behavior of Non-Human Primates

Julia Fischer

CONTENTS

I. Introduction.....	109
II. Impact of Methodological Advances.....	110
III. Plasticity in the Usage and Comprehension of Calls.....	113
A. The Case of Pygmy Marmoset Babbling.....	113
B. Dialects and the Question of Vocal Convergence.....	114
C. Infants' Responses to Acoustic Gradation.....	117
IV. Future Research.....	120
Acknowledgments.....	121
References.....	121

I. INTRODUCTION

Understanding the roots of human language is a fascinating topic that has generated as much interest as controversy (see References 1 to 3 for some recent contributions). A key feature of human language is that it is learned, in terms of both its production and comprehension. From a comparative perspective, this raises the question of whether and to what extent other primate species exhibit vocal learning. One of the earliest studies in this field involved an attempt to teach a chimpanzee to speak.^{4,5} Vicki, the chimpanzee raised with the Hayeses, for instance, in fact learned to utter a few "words," but the difficulties she had in mastering this task were apparently more striking than her successes.⁴

The finding that apes or monkeys have difficulties acquiring human speech, however, does not refute the possibility that learning plays an important role in the development of their own species-typical vocalizations. Most of the evidence accumulated about vocal development comes from studies on monkeys,^{6,7} while hardly anything is known about the vocal development of apes.⁸ In contrast, numerous