

Chapter 9. Vocal Learning and Auditory-Vocal Feedback

Peter L Tyack

Sea Mammal Research Unit, Scottish Oceans Institute, School of Biology, University of St Andrews, KY16 8LB, Scotland UK

from Suthers et al. (eds.) *Vertebrate Sound Production and Acoustic Communication*
Springer Handbook of Auditory Research 53, DOI 10.1007/978-3-319-27721-9_9

Abstract

Vocal learning is usually studied in songbirds and humans, species that can form auditory templates by listening to acoustic models and then learn to vocalize to match the template. Most other species are thought to develop vocalizations without auditory feedback. However, auditory input influences the acoustic structure of vocalizations in a broad distribution of birds and mammals. Vocalizations are defined here as sounds generated by forcing air past vibrating membranes. A vocal motor program may generate vocalizations such as crying or laughter, but auditory feedback may be required for matching precise acoustic features of vocalizations. This chapter discriminates limited vocal learning, which uses auditory input to fine-tune acoustic features of an inherited auditory template, from complex vocal learning, in which novel sounds are learned by matching a learned auditory template. Two or three songbird taxa and four or five mammalian taxa are known for complex vocal learning. A broader range of mammals converge in the acoustic structure of vocalizations when in socially interacting groups, which qualifies as limited vocal learning. All birds and mammals tested use auditory-vocal feedback to adjust their vocalizations to compensate for the effects of noise, and many species modulate their signals as the costs and benefits of communicating vary. This chapter asks whether some auditory-vocal feedback may have provided neural substrates for the evolution of vocal learning. Progress will require more precise definitions of different forms of vocal learning, broad comparative review of their presence and absence, and behavioral and neurobiological investigations into the mechanisms underlying the skills.

Keywords: Auditory-vocal feedback • Compensation for noise • Lombard effect • Vocal convergence • Vocal imitation • Vocal learning • Vocal mimicry • Vocal plasticity

9.1 Introduction: Definitions of Vocal Learning

Vocal learning is usually defined as the ability of an animal to modify the acoustic structure of sounds it produces based on auditory input. Many vertebrate

species develop normal vocalizations even if they do not hear those of conspecifics, and are thought to inherit a motor program that generates the acoustic structure of each call. By contrast, some species such as humans and songbirds learn some signals from individuals with whom they interact, and these signals form a shared learned communication system. Janik and Slater (1997, p 59) define vocal learning as a modification of specific acoustic features of one's vocalizations, "as a result of experience with those of other individuals." This definition emphasizes learning to incorporate the calls of others into one's vocal repertoire. Compared to a more genetically constrained system, vocal learning can create a much more complex, open, and flexible communication system.

The classic animal model of vocal learning stems from work on oscine songbirds (Nottebohm 1970). The template model of song learning separates the process of forming an auditory memory or template through listening to a song (auditory learning) from the process of learning to produce that song by matching one's own vocal motor output to the stored template (sensorimotor learning) (Konishi 1965, 2004). For example, if a young male white-crowned sparrow (*Zonotrichia capensis*) is played recordings of the songs of an adult male conspecific at the right time, he will form memories or auditory templates of the songs he heard. As he matures and starts producing song-like sounds, he can slowly learn to match his own vocal motor output to the auditory templates, a pattern of development that Owren et al. (2011) call "reception-first" because of the need to hear the vocalization before learning to produce it. The requirement for vocal learning can be demonstrated by eliminating auditory input; if such a bird is deafened before vocal development, he will never produce normal song (Konishi 1965). A sparrow that retains normal hearing but that never has the opportunity to hear a model sound may develop song that is more normal, but that will not match wild-type songs of his population. In this case, the animal must learn to match his own vocal motor output against an auditory template that is inherited. This is often called an innate auditory template. Use of the term "innate" in this case means that it develops reliably in a species by inheritance. By contrast, many species of bird (Schleidt 1961; Konishi 1963; Nottebohm and Nottebohm 1971; Kroodsma and Konishi 1991) and mammal (Winter et al. 1973) can develop normal vocalizations without any auditory feedback. Vocalization in these species is thought to be structured by central pattern generators in the brain that control vocal motor output without requiring auditory input, a pattern of development that Owren et al. (2011) call "production-first." These species are thus classified as not having vocal learning. Note that species with vocal learning may also be able to produce some signals without auditory input. Laughter and crying in humans, for example, develop normally in hearing-impaired infants (Scheiner et al. 2006). Thus species with vocal learning may develop some vocalizations without reference to auditory input. This means that when one limits auditory input to test for its effects on vocal production, one must test all types of vocalization to fully search for vocal learning in a species.

We humans appreciate the flexibility of our language, but the evolutionary origins of human vocal learning are obscure, as no nonhuman primates show strong evidence for vocal learning. At least three taxa of birds are skilled at vocal learning, and among mammals we have evidence that bats, dolphins, elephants, humans, and seals are as well (Fitch and Jarvis 2013). Research on nonhuman primates and other mammals and birds has shown some minor mod-

ification of calls based on auditory input; these changes are similar in some ways to classic vocal learning and different in other ways. There are some additional ways in which auditory input affects vocalizations, such as changes by most birds and mammals to compensate for noise or interference (Brumm and Zollinger 2011), and specialized mechanisms such as Doppler compensation in some bats (Metzner et al. 2002). The lack of clear definitions of different forms of vocal learning has led to confusion and disagreement about which species show vocal learning and which do not. This scientific uncertainty interferes with understanding the evolution of vocal learning.

This chapter discusses a broad range of phenomena that have led to connections between auditory input and vocal motor output. Nottebohm and Liu (2010, p. 3) define vocal learning as modification of one’s vocalizations “by reference to auditory feedback,” a definition that would include any of these phenomena. In this chapter, this broadest category is called “auditory-vocal feedback.” However, as discussed in the preceding text, other authors limit the term vocal learning to the development of calls that match those of other individuals with whom the learner has interacted. Following a distinction highlighted by Arriaga and Jarvis (2013) and using a terminology suggested by Fitch and Jarvis (2013), this chapter distinguishes between complex vocal learning, such as the classic cases of humans and songbirds, in which individuals can learn to produce new vocalizations by matching an auditory template formed from new sounds that they hear, from a more limited form of vocal learning, in which individuals use auditory input to modify the acoustic features of production-first vocalizations.

Janik and Slater (1997) also distinguish vocal production learning, which involves changing acoustic features of vocalizations, from learning a new use for a preexisting vocalization, or learning to comprehend the context in which a vocalization is produced. These distinctions are useful to discriminate vocal production learning from vocal usage or comprehension learning. However, this book focuses on sound production and this chapter only discusses vocal production learning, so here it is called by the more common name, vocal learning. Readers with any questions about this distinction should read Janik and Slater (1997).

9.2 Taxonomic Scope for Review of Vocal Sound Production

The word “vocal” derives from the human voice, but this chapter is a comparative review, so my definition of “vocal sounds” includes those from animals that produce sound by forcing air past membranes whose tension can be controlled. Frogs produce sound by moving air past vocal cords in the trachea into a vocal sac (Gans 1973), a mode of production that fits this definition of “vocal.” The sound production organs of birds and mammals use air under pressure from the respiratory system to vibrate soft membranes in the airways. The sound production organ in birds is called the syrinx; that in toothed whales, the phonic lips (Cranford et al. 1996; Madsen et al. 2011); and that in most other mammals, the larynx. The membrane tension in all of these sound production organs is controlled by muscles that are typically well innervated, and some avian and

mammalian species have articulatory control of acoustic filters that lie above the sound production organ itself. Control of pneumatic sound production in these taxa requires sophisticated simultaneous coordination of pressure in the lungs, tension in the vocal folds, and configuration of the acoustic filters. Neural control of the sound source and filters is often highly developed in many of these taxa (Fitch and Suthers, Chap. 1; Taylor et al., Chap. 8).

Mammals or birds with a production-first vocal development have been reported to develop species-typical vocalizations with no auditory feedback. However, in testing species whose vocal repertoires are less well known, researchers must also be careful not to miss a specific category of vocalization that may be learned. For example, if investigators just studied development of calls in oscine songbirds deprived of auditory feedback and missed testing song, they might fail to uncover evidence for vocal learning. In addition, evidence that production-first species never use auditory-vocal feedback to fine-tune their vocal production against an innate auditory template is not iron clad. Early studies that minimized the effect of deafening on vocal development are starting to be updated by studies that use more sophisticated acoustic analyses to show subtle differences. For example, Romand and Ehret (2004) demonstrated that although deafened kittens do develop species-specific meows, the meows of deafened or isolated kittens differed acoustically from those of normal kittens, suggesting the role of social experience and auditory feedback in fine-tuning meows. Thus auditory-vocal feedback may function to stabilize the acoustic structure of species-specific vocalizations in settings where this is important.

Janik and Slater (1997) differentiate vocal learning involving changes in the duration or level of a call from those that involve changes in frequency. They argue that simple control of exhalation can control duration and level, but that matching frequency requires more complex control. This chapter also argues for different levels of complexity of vocal learning, and agrees that matching of acoustic features involving coordination among respiratory, laryngeal, and articulatory muscles, like those leading to differences in frequency matching, clearly involve a more complex form of vocal learning than those that simply involve control of respiratory muscles.

However, a simple differentiation between duration/level parameters and frequency parameters may not suffice to distinguish simple versus complex forms of vocal learning. The intensity of vocal sounds is driven by increased pneumatic pressure from the lungs, and in the absence of counter-adjustments, increased pressure also leads to increased frequency in some species. The larynx in mammals and syrinx in birds can transfer more acoustic power at higher frequencies (Titze 1994). These biophysical relationships lead to a correlation between the source level and frequency of sounds in a diverse set of species (Au et al. 1995 ; Nemeth et al. 2013). Therefore, any evaluation of complexity of matching needs to account for these linkages between acoustic features of the sound production system.

As mentioned in Sect. 9.1, the primary distinction made in this chapter regarding complexity of vocal learning contrasts limited versus complex vocal learning. In the former, auditory feedback is used to modify existing features of a production-first call. In the latter, an auditory template of a sound is first learned from hearing a model (reception-first), and the animal then learns to match its vocal motor output to match the template. Either form may involve matching frequency and timing and amplitude to varying degrees of precision

and complexity. Voice onset time provides an example from human speech of timing cues that may require complex vocal learning for a precise match. Other species such as sperm whales (*Physeter macrocephalus*) appear to learn the timing of brief transient vocalizations to create a complex repertoire of calls (Rendell and Whitehead 2005; Rendell et al. 2012). This chapter does not prejudge which acoustic features provide the complexity of vocal learning in different species, but rather distinguishes learning to modify existing calls from the ability to learn to develop new calls, which enables a more open-ended communication system.

9.3 Vocal Mimicry: Copying Sounds of Other Species

Some of the best evidence for complex vocal learning comes from animals under human care that imitated the sounds of humans or non-conspecifics. When an animal is introduced to an environment in which it is exposed to new sounds that are not part of its normal vocal repertoire, and when it makes precise imitations of these novel sounds, there are few alternative explanations than the animal has learned to produce the sounds it has heard. This process of learning through audition to create a new vocal motor pattern is clear evidence for complex vocal learning. It has been known for centuries that some songbird (Klatt and Stefanski 1974) and parrot (Pepperberg 2010) species kept by humans can copy the sounds of humans very precisely. Among mammals, there are cases of a harbor seal (*Phoca vitulina*; Ralls et al. 1985) and an Indian elephant (*Elephas maximus indicus*; Stoeger et al. 2012) imitating speech. African elephants (*Loxodonta africana*) have also been reported to imitate the sounds of a truck (Poole et al. 2005). Bottlenose dolphins (*Tursiops truncatus*) have also proven skilled at imitating synthetic computer-generated frequency modulation patterns (Richards et al. 1984). By contrast, intensive attempts to train nonhuman primates to imitate speech have failed to provide strong evidence for vocal imitation (Kellogg and Kellogg 1933; Hayes 1951; Hayes and Hayes 1952).

The ability of some nonhuman species to imitate speech is remarkable. Not only do these animals have to form auditory templates very different from those of their own species, but they also must adapt their sound production organs to produce sounds very different from those the organ usually produces. It has been argued that the relatively minor differences between the vocal tracts of anthropoid apes vs. humans may prevent apes from producing speech (Lieberman 1984), but birds such as mynah birds are able to adapt syrinx and beak to reproduce complex features of speech such as fundamental frequency contours, formants, and consonants such as fricatives and plosive bursts (Klatt and Stefanski 1974). Similarly, the male Asian elephant that imitated speech was able to match the first two formant frequencies of his trainers quite precisely, even though the large size of his vocal tract normally generates much lower formant frequencies. He accomplished this match by inserting the tip of his trunk into his mouth just before imitating speech. This method of changing the vocal tract to produce unusual formants has not been reported for other elephants, nor does this elephant perform this action when making normal elephant vocalizations. This kind of creative manipulation of the vocal tract to change the frequency

spectrum of a call is rare. Orangutans have been reported to hold leaves to their mouth, extending their vocal tract, to lower the frequency of a call (Hardus et al. 2009). There is scant evidence for nonhuman primates producing new vocalizations to match an auditory model, but there is some evidence involving nonvocal sounds produced by airflow past the lips. Orangutans have been reported to imitate human whistling spontaneously (Wich et al. 2009). An Asian elephant showed similar vocal creativity, putting her trunk against her mouth to produce whistles, a technique that was reported to have been learned from the elephant that originated the technique (Wemmer and Mishra 1982).

9.4 Evolution of Vocal Mimicry: Learning to Copy Sounds of Other Species

As discussed in Sect. 9.3, perhaps the best evidence for complex vocal learning comes from situations in which animals mimic human speech or artificial sounds synthesized by humans to incorporate acoustic features that differ from the subject's pre-exposure repertoire. This behavior of copying a sound that is not produced by a conspecific is called vocal mimicry (Baylis 1982). In a review of vocal mimicry, Kelley and Healy (2011) point out that about 20% of songbird species have been reported to mimic sounds from non-conspecifics.

There is a certain irony that we humans, who pride ourselves on our abilities for vocal learning, require other species to match our speech for us to recognize their abilities for complex vocal learning. Vocal mimicry seems to reflect an unusual lack of constraint on vocal learning, and it seems likely that many species that have evolved skills for learning conspecific sounds might fail to imitate sounds of other species. On the other hand, social interaction appears to foster more open vocal development, for example, leading a white-crowned sparrow housed with singing males of the same and different species to be more likely to copy heterospecific song than if it just heard recorded songs (Baptista and Morton 1981). Kelley and Healy (2011) note the prevalence of mimicry among animals held in captivity, especially those from long-lived species with strong social bonds among fluid groups, and they suggest that animals that rely on copying conspecifics to maintain social relationships may broaden this pattern, copying human caregivers to maintain heterospecific social relationships in captivity. The next section examines the specific settings in which some of the best evidence for mimicry by mammals and birds of human signals was observed. These include cases involving immediate mimicry of signals and cases suggestive of a temporal separation between formation of an auditory template of a sound and learning how to produce a sound that matches the template.

9.5 Separation Between Acquisition of an Auditory Template and Learning Through Auditory-Vocal Feedback: How to Produce a Sound that Matches the Template

There is evidence in two cases of mammalian mimicry of human speech for a separation between forming the auditory template and the vocal practice of matching motor output to the auditory template. The harbor seal that imitated speech was found as a newborn pup in May 1971 and was raised in the home of a local Maine resident, who named him Hoover, until August 1971, when he was transferred to the New England Aquarium and placed in a pool with other harbor seals (Hiss 1983). In 1976, he was first reported showing sexual behavior, and was first noted to make sounds “as if talking” (Ralls et al. 1985). This seal was first reported to produce a word from human speech in 1978, when an observer “wrote in the files ‘he says “Hoover” in plain English. I have witnesses’ ” (Ralls et al. 1985, p. 1051). Hoover subsequently increased his repertoire of speech sounds, for which listeners could recognize a New England accent. The male Asian elephant that imitates human speech produces words in Korean that native speakers can classify accurately. He was raised for a long period in a setting with no other elephants where humans provided his only social interaction. It appears that he started to produce speech sounds at about 14 years of age, near his onset of sexual maturity (Stoeger et al. 2012). As in the classic songbird model, Hoover’s auditory template appears to have formed early in life, and for both the seal and the elephant, the vocal motor learning phase did not take place until sexual maturity. Steroid hormones influence vocal learning circuits in songbirds (Brenowitz and Kroodsma 1996); perhaps the timing of vocal learning in these other species suggests a role for similar hormonal control.

9.6 Adult Animals that Rapidly Imitate Novel Sounds

This section discusses evidence from non-oscine bird and mammal species of capabilities for imitating new sounds as adults with a short interval between hearing a new sound and imitating it. Todt (1975) described a new method to train African grey parrots (*Psittacus erithacus*) to imitate human speech sounds. Adult parrots taught with this method learned new speech sounds, but the shortest interval between auditory presentation of a new sound and the parrot’s first imitation was 8 h (overnight) and was generally 3 days. In this case, the mapping of auditory input to vocal motor output and matching of vocal output to the auditory template appears to take some time for consolidation. The lack of imitation until overnight suggests a role for sleep in the learning process, as has been observed for songbirds (Derègnacourt et al. 2005).

Richards et al. (1984) report much more rapid imitation for a subadult female bottlenose dolphin (*Tursiops truncatus*) trained in an imitation task. The first step in the study was to define the pre-exposure repertoire of frequency-modulated whistle sounds produced by the subject and by the one other dolphin

in the pool. The training procedure used operant conditioning in a series of successive training steps. First the dolphin was trained to produce any whistle sound after a vocalize signal was played. Then several model sounds (different from any in the pre-exposure repertoire) were introduced after the vocalize signal, and the dolphin was reinforced for responding immediately to the vocalize signal followed by the model signal by producing a whistle that matched first the duration and then frequency parameters of the model. Richards et al. (1984, p. 16) state that the subject “rapidly formed a generalized mimicry concept so that the presentation of any new model elicited an immediate attempt at imitation.” Reiss and McCowan (1993) also report spontaneous vocal mimicry by captive bottlenose dolphins of computer generated frequency contours produced within 0.5 s of the end of the contour.

This behavior of captive dolphins imitating computer-generated models has a timing quite similar to that described for bottlenose dolphins copying whistles in the wild. Each individual bottlenose dolphin learns to produce an individually distinctive signature whistle (Janik and Sayigh 2013). Janik (2000) found that bottlenose dolphins in the wild may match the signature whistle of a dolphin with whom it is interacting. King et al. (2013) showed that the average latency between one bottlenose dolphin producing its signature whistle, and another matching the whistle was <1 s. The demonstrations of vocal learning using synthetic whistle-like sounds in captive dolphins seem to tap into a mode of whistle communication that involves rapid copying of conspecific sounds in the wild. Whistle matching in dolphins qualifies as complex vocal learning, but here, as in adult humans learning a new melody or word, the processes of auditory learning and sensorimotor learning are more rapidly and tightly coupled than in the classic descriptions of vocal learning. If they can produce copies that match a new model on the first attempt, this suggests an ability to map acoustic features as heard in the model directly onto acoustic properties of the vocal motor output.

9.7 Evidence that Mammals Learn to Produce Conspecific Sounds

Some songbird species have open-ended learning, and can continue to learn new songs as adults and add them to their song repertoire as they age (Beecher and Brenowitz 2005). In a few avian and mammalian species, such as the oscine yellow-rumped Cacique (*Cacicus cela vitellinus*) and the humpback whale (*Megaptera novaeangliae*), not only do males retain the ability to learn new songs throughout their lifespan, but also the song of each region changes from month to month (Payne et al. 1983; Trainer 1989) and from year to year over decades (Payne and Payne 1985), with different members of a population tracking these changes in the song (Guinee et al. 1983). In the South Pacific, thematic material for humpback songs originates in waters off Australia, and spreads over several years to a series of breeding populations spread thousands of kilometers to the east (Garland et al. 2011). There is no way that this pattern of song change could occur except by vocal production learning, with whales continuously adjusting their songs based upon songs they hear. Delarue et al. (2009) report similar changes in the songs of bowhead whales (*Balaena*

mysticetus) recorded in the Chukchi Sea.

A less dramatic pattern in which the young copy acoustic features of the calls of their mothers provides evidence for vocal learning in bats. Jones and Ransome (1993) studied the greater horseshoe bat (*Rhinolophus ferrumequinum*), which emits an echolocation call with most of its energy concentrated in a narrow frequency band. As females age from years 1 to 3, the frequency of this call increases, and thereafter decreases. The frequency of calls of young bats matches that expected for the age of their mothers, with calls of pups of older mothers lower than those of younger mothers. This pattern suggests that although bats inherit a template for the basic structure of their call, they learn to fine-tune the frequency of their call by comparison with the age-specific features of the call of their mothers. Such a pattern would fit the definition of limited vocal learning as defined in this chapter.

The Janik and Slater (1997) definition of vocal production learning emphasizes individuals learning acoustic properties of calls from conspecifics. However, as in the case of the bats, demonstrating vocal production learning is more difficult when the subject produces a sound similar to that of a conspecific than when it mimics a more exotic sound. Here one must show that the new sound was not part of the pre-exposure repertoire of the subject, and that the change is not produced by maturation of the vocal tract, or by a new context that elicits for the first time a species-specific call type that the subject could have produced all along. The whale and bat examples discussed in the preceding paragraphs meet these criteria, but it is difficult for many species whose vocalizations do not change in such distinctive ways.

9.8 The Role of Auditory Input in Mammalian Vocal Development

One approach around this problem is to study vocal development in animals that are deprived of auditory input. If such an animal develops normal vocalizations, then it must have production-first vocal development that does not require vocal learning. Different kinds of deprivation can isolate different kinds of vocal learning. For example, if an animal is completely deafened, it can neither form auditory templates from the sounds of others, nor can it learn to fine-tune its vocal motor output against an inherited auditory template. If an animal's hearing remains intact, but it is isolated from conspecifics, then it can fine tune its own vocalizations against innate auditory templates, but it cannot form new auditory templates based on the sounds of other individuals. In practice, it can be difficult to eliminate the possibility of an animal experiencing the sounds of others. For example, the hearing of birds may function in the egg, or of mammals *in utero* (Horner et al. 1987). Deafening and/or isolation can also cause more generalized deficits, so it can be difficult to prove that it was only the lack of auditory input that interfered with normal vocalization. In spite of these problems, Konishi's (1965) study of song in deafened or isolated white-crowned sparrows formed the basis of the template model of song learning.

We humans are a species that depends heavily on vocal learning for speech and singing a tune, so it comes as a surprise that vocal learning is not well developed in nonhuman primates. Frequently cited evidence against a role of

auditory input in the vocal development of mammals comes from studies of squirrel monkeys (*Saimiri sciureus*). Winter et al. (1973) raised squirrel monkey infants with mothers that were normal or muted. They report that infants who heard no typical species-specific squirrel monkey sounds developed calls that were “virtually identical” with those of normal infants, and with no significant differences compared to the normal calls of adults. However, the sample size of normal monkeys was relatively small and the acoustic parameters, duration and fundamental frequency, were relatively simple by modern standards of acoustic analysis.

There has been recent interest in the question of whether mice (*Mus musculus*) may learn complex ultrasonic vocalizations. When adult male mice sense the presence of females, they produce complex sequences of repeated phrases of syllables that have been called songs (Holy and Guo 2005). Arriaga et al. (2012) found differences in songs of deafened versus normal mice, leading them to argue that male mice require auditory feedback to maintain normal songs, and they report that males altered their songs to match those of a cage mate, which they interpret as vocal imitation. By contrast, Hammerschmidt et al. (2012) studied the development of songs in deaf versus normal mice and found no difference, leading them to conclude that mice do not require auditory input for vocal development. Kikusui et al. (2011) demonstrated that two inbred strains of mice each had songs with different acoustic features. When they crossbred each strain, males raised with parents from the other strain developed songs similar to those of their genetic father, not the one whose sounds they heard. Both studies state they found no evidence for vocal learning of any sort. These diametrically opposing results suggest that the jury is still out on vocal learning in mice and that evidence for vocal learning in this species must be treated with caution. More research is needed in mice and many other species to test for differences in calls of animals with or without auditory input of normal calls to test whether species classed as non-learners may use auditory-vocal feedback to stabilize subtle acoustic features of vocalizations.

9.9 Weakness of Geographical Dialects as Evidence for Vocal Learning

One information source that has been argued as evidence for vocal learning is the existence of geographical dialects in vocalizations. If a young animal learns its vocalizations from its neighbors, this certainly can lead to geographical variation and vocal dialects. There is a large literature on song dialects in birds, and vocal learning can lead to convergence of songs within an area or population and divergence across areas or populations (Catchpole and Slater 2008). Janik and Slater (1997) cover evidence for vocal dialects in mammals, and they discuss many mechanisms that do not involve vocal learning that could lead to geographic variation in vocalizations. Geographical differences in vocalizations can be generated by genetic differences between populations, by differences in social structure that affect call usage, by differences in sound transmission in different environments, and by settling of animals in sites where they hear sounds similar to those they produce (Catchpole and Slater 2008). Given how well this topic has been covered elsewhere, this chapter focuses on a

more fine-grained analysis of how calls converge among interacting individuals, which in my view provides better direct evidence for vocal learning.

9.10 Vocal Convergence as Animals Form a Group

Convergence of acoustic features of a vocalization among animals recorded before and after they form social bonds is a better test for vocal learning than evidence of geographical dialects. If the versions of the same call type from several animals differ from one another before exposure and become more similar after exposure to one another's vocalizations, then this provides stronger evidence for vocal learning than the simple demonstration of geographical variation. Note that vocal convergence may involve a call that developed without auditory input, but the fine-tuning of the call then does require feedback between auditory input and vocal motor output. The process of vocal convergence must allow fine-tuning of the auditory template as well as the converged vocalization. Vocal convergence thus qualifies as a form of limited vocal learning, by the definition given in Sect. 9.1. The fine-tuning of a species-specific vocalization to match that of other individuals helps to resolve the ethological question of why selection might favor allowing auditory input to modify the "correct" species-specific signal. Here the animal still maintains the species-typical call, but adds detail that indicates a social bond and membership in a social group. Vocal convergence has been reported for many species, including species for which the evidence of vocal learning is otherwise weak.

Some of the first evidence for vocal convergence in animals comes from birds. Most work on vocal learning in birds focuses on songs of oscine songbirds, but Mammen and Nowicki (1981) worked on the calls of oscine black-capped chickadees (*Parus atricapillus*). They captured five winter flocks of chickadees and recorded the chick-a-dee call from each individual of each group. When they rearranged the members of three of these groups into different aviaries, they showed that the calls of members of each new aviary flock converged over a period of about a month. Nowicki (1989) studied this process of convergence, and showed that within the first week of being housed together, members of the group converged on calls matching the central tendency of features within the group rather than copying any one individual. Farabaugh et al. (1994) provide similar data for vocal convergence in the contact calls of the Psittaciform budgerigar (*Melopsittacus undulatus*). They raised two groups of three unrelated male budgerigars, each group in a different cage within the same room. None of the six males shared contact call types at the start of the study. The first time one male imitated the contact call of another was recorded 1 week after they were housed together. By 8 weeks, all birds in the same cage shared the same dominant call type, and this dominant call differed across the two cages. The lack of convergence of birds that could hear one another but did not interact directly suggests that actual social interaction made it more likely for calls to be imitated in this species than just hearing the calls. This indicates that the information affecting choice of models is not just auditory, but also involves social interaction. Some of the best evidence for vocal learning in hummingbirds comes from three male Anna's hummingbirds (*Calypte anna*) that showed convergence of song syllables when housed in the same room (Baptista and Schuchmann 1990).

There is evidence for vocal convergence in all of the mammalian taxa with good evidence for vocal learning: bats, cetaceans, elephants, humans, and seals. When Boughman (1998) transferred greater spear-nosed bat females (*Phyllostomus hastatus*) from one captive group to another, their call structure changed to become more similar to that of their new group mates. This pattern of fine-tuning of calls suggests that the bat case be viewed as limited vocal learning. Similar vocal convergence has been observed when bottlenose dolphins (*Tursiops* sp.) form social bonds. As male bottlenose dolphins mature, most form alliances that last for many years. Smolker and Pepper (1999) studied whistles produced by three males as they formed an alliance in the wild. Over the 4-year study, the distinctiveness of the whistles decreased as all three dolphins converged on a shared whistle type that rarely was produced before alliance formation. Watwood et al. (2004) were not able to study the process of convergence, but they confirmed that males in nine alliances produced whistles that were more similar to those of their own partners than to those of any of the other males. Frupp et al. (2004) provide evidence that dolphin calves in the wild model their signature whistles on those of community members, and Miksis et al. (2002) show that captive dolphin calves incorporate features of manmade signals as they develop their signature whistles. This matching of novel models suggests that vocal convergence in bottlenose dolphins may represent complex vocal learning, in which dolphins learn a novel or individual-specific model as opposed to a species-specific model. The evidence for vocal convergence in elephants involves a 23-year-old male African elephant (*Loxodonta africana*) that had spent 18 years housed with two female Asian elephants (*Elephas maximus*). This male produced sounds that were very different from normal sounds of African elephants, but that matched the chirp sounds produced by Asian elephants (Poole et al. 2005). This copying of a non-species-typical call differs from the other cases of conspecific vocal convergence; copying such a novel call qualifies as a case of complex vocal learning. Vocal convergence is well known in humans and is one aspect of a phenomenon called vocal accommodation (Street and Giles 1982 ; Giles 1984). Sanvito et al. (2007) showed on the breeding ground of southern elephant seals (*Mirounga leonina*) that agonistic calls of subordinate males tended to converge on those of the local dominant male.

In spite of the weak evidence for complex vocal learning among nonhuman primates, vocal convergence has been reported for several primate species: pygmy marmosets (*Cebuella pygmaea*; Snowden and Elowson 1999), cotton-top tamarins (*Saguinus oedipus*; Weiss et al. 2001; Egnor and Hauser 2004), and chimpanzees (*Pan troglodytes*; Mitani and Gros-Louis 1998; Marshall et al. 1999; Crockford et al. 2004; Watson et al. 2015). Sugiura (1998) also demonstrated rapid matching in a study in which Japanese macaques matched acoustic features of coo calls played back to them. Vocal convergence has also been reported for taxa with little other evidence for vocal learning. For example, Briefer and McElligott (2012) raised young goat kids in groups. After 5 weeks together, half sibs in the same group had contact calls that were more similar than half sibs from different groups, and the calls converged over time.

In most of the cases of vocal convergence described in the preceding paragraphs, unrelated animals form a grouping that interacts socially. These individuals start with slightly different versions of a vocalization, and they converge on a common group-distinctive version of the vocalization. Owren et al. (2011) raise concerns that emotional responses to being housed with strangers may pro-

duce vocal patterns that look like vocal convergence but do not involve vocal production learning. However, looking at the broad range of vocal convergence studies, it is hard to imagine mechanisms other than vocal learning by which unrelated animals could develop vocalizations that are so distinctive across groups. Vocal convergence as a limited form of vocal learning has a broader taxonomic scope than complex vocal learning.

From a neural perspective, it is an open question whether fine-tuning specific acoustic parameters of an existing call type involves overlap with circuits for complex vocal learning or whether it involves a separate circuit. Humans and the avian taxa with reception-first vocal development have neural circuits in the telencephalon specialized for complex vocal learning, which are not as obvious in non-learners. The wider spread of vocal convergence than complex vocal learning suggests either that it uses different circuits or that it can be achieved with less obvious neural circuits. Owren et al. (2011, p 10) argue that “convergence and divergence phenomena may have little to say about reception-first vocal development.” Arriaga and Jarvis (2013) suggest that fine-tuning the acoustic parameters of a call type whose central pattern generator (CPG) is located in the midbrain and/or brainstem would use neural circuits that modify the CPG based on cortical input and integrated auditory pathways. They propose that this involves different circuits from reception-first vocal learning, which involves forebrain circuits. On the other hand, learning to fine-tune a production-first call could involve overlap with the mechanisms used to produce a novel sound, with the primary differences involving how open the process selecting templates is to novel versus species-specific sounds, and how variable the system is for generating new vocalizations to match a template. Certainly if the signature whistles of bottlenose dolphins develop as reception-first calls that can match a novel model, then convergence of these signals may also be mediated by complex vocal learning circuits. Resolving these issues will require careful study of these neural circuits in species capable of vocal convergence and of complex vocal learning.

9.11 Neural Mechanisms that Enable Bats to Fine-Tune the Frequency of Echolocation Pulses Based on Auditory Input from Returning Echoes

Vocal convergence that involves fine-tuning the template for a production-first call based on auditory input from other conspecifics represents limited vocal learning. There is not much evidence on the neural basis for this form of limited vocal learning, but better evidence is available for a rapid and sophisticated form of auditory-vocal feedback used by echolocating bats to fine-tune their echolocation calls based on hearing their echoes. When a bat echolocates, it often needs to use auditory information about an incoming echo to adjust its next outgoing pulses. This puts a premium on auditory-vocal feedback rapid enough to function on time scales of a few tens of milliseconds. One of the most complex forms of feedback occurs for bats whose hearing is specialized to measure small changes in frequency over a narrow frequency band. The frequency band for which they can best discriminate frequencies (the “acoustic fovea”) is narrow enough that differences in velocity between bat and prey can lead to Doppler

shifts large enough to shift the echo returning from the prey to outside of the acoustic fovea. These bats shift the frequency of their outgoing pulses so that the Doppler-shifted echo remains in the acoustic fovea. Metzner (1989, 1993) reported on neural mechanisms for fine-tuning of vocal motor output based on auditory input studied in horseshoe bats with this Doppler-shift compensation. He found neurons within the midbrain that receive auditory input and that respond to vocal production and he proposed a mechanism for measuring the echo frequency and using this information to control the frequency of the outgoing pulse to compensate for Doppler shifts. Metzner (1996, p 252) studied the connectivity of this area, and proposed that it “serves as a link between the processing of auditory information and the control of vocalization and related motor patterns.”

Studies of auditory-vocal feedback in bats show that midbrain and brainstem structures can rapidly fine-tune precise acoustic features of production-first sounds. Reliance on low levels of the brain, a few synapses from the relevant motoneurons, enables rapid processing. Vocal convergence of production-first sounds occurs over longer time scales, reducing the constraints on timing. However, vocal convergence is more complex than Doppler compensation in that it requires the formation of an auditory template, coupled with a mechanism to fine-tune a production-first vocal motor program so that the animal can reliably produce the fine-tuned call. We know that birds and humans use specialized auditory-vocal feedback mechanisms in the telencephalon to develop new reception-first calls. The bat work shows how auditory input can modulate pattern generators in the midbrain, but further research will be required to determine whether fine-tuning of production-first calls leading to vocal convergence can be achieved by similar feedback in the midbrain and brainstem, or whether this requires telencephalic pathways to support the learning and memory requirements of vocal learning. If vocal convergence does require telencephalic pathways, it will also be important to test whether this form of limited vocal learning involves the same circuits as complex vocal learning or not.

9.12 The Lombard Effect: Modifying Vocal Output in Response to Noise

The study of vocal convergence suggests that there may be a broader taxonomic range for this limited form of vocal learning than is typically credited for complex vocal learning, especially among mammals. But there may be an even more widely distributed connection between auditory input and vocal motor output. There is a problem faced by all animals that use sound to communicate, a problem that would select for a connection between auditory input and vocal motor output. This problem is communicating in varying noise conditions. Both in the ocean and in air, ambient noise levels vary over tens of decibels, leading to significant variation in the effective range of communication for a fixed signal level.

The best known compensation mechanism for noise is called the Lombard effect, named after the French otolaryngologist Etienne Lombard who discovered that humans speak more loudly when in the presence of loud noise (Brumm and Zollinger 2011). Brumm and Zollinger (2011) discuss the evidence of how

broadly distributed the Lombard effect is among birds and mammals. The Lombard effect has been demonstrated among oscine songbirds, hummingbirds, and parrots, all of which are skilled at complex vocal learning; in the domestic fowl (Brumm et al. 2009), thought not to be capable of vocal learning; and crested tinamous, which belong to the most basal group of living birds. Hotchkin and Parks (2013) report that evidence for the Lombard effect has been found in every species of mammal in which it has been studied. The presence of the Lombard effect has not been as systematically explored among basal mammals as among birds, but the broad distribution and lack of evidence for absence of the Lombard effect in mammals suggest that it was likely present in the common ancestor of mammals. Brumm and Zollinger (2011) suggest that the Lombard effect has a very old history in birds and mammals, and they argue that either it independently evolved in both taxa or originated in a common ancestor and therefore could be 300 MY old and shared among amniotes.

The neural basis for the Lombard effect has not been studied in as much detail as song learning in oscine songbirds. Neurobiological studies in cats and squirrel monkeys locate the neuronal circuits for the Lombard effect in the brainstem. Nonaka et al. (1997) demonstrated the Lombard effect in decerebrate cats, showing that sufficient neuronal circuits exist within the brainstem to support the Lombard effect. Working with squirrel monkeys, Hage et al. (2006) report that neurons in the brainstem respond both to auditory input and vocal motor output, and Hage et al. (2006) suggest that these neurons mediate the Lombard effect by integrating audiovocal information. This led Owren et al. (2011, p 7) to argue “the Lombard effect is mediated at the brainstem level, meaning its occurrence is likely uninformative with respect to the operation of higher level vocal control systems.” By contrast, Eliades and Wang (2012) showed that the Lombard effect in common marmosets (*Callithrix jacchus*) is mediated by neurons in the auditory cortex that respond to the monkey’s own vocalizations. When a monkey hears its own vocalization masked by noise, the neural response weakens, producing an error signal that predicts vocal intensity compensation in later vocalizations. This suggests that at least in primates, a parallel cortical circuit also affects modulation of call intensity based on auditory monitoring of one’s own calls.

9.13 A Broader Look at Compensation for Noise in Animal Communication

Communicating in noise is a ubiquitous problem that may require modifying the outgoing signal depending on the noise present at the time of calling. The Lombard effect was the first such compensation mechanism studied in humans and animals, but the last few decades have seen a great expansion in evidence for compensation for noise in animal communication. Communication engineers recognize a suite of mechanisms that can be used to compensate for noise, including increasing the level of the signal, length of the signal, or redundancy of the signal. If the noise is limited to within a frequency band, then the frequency of the call can also be shifted outside of the noise band, just like switching a walkie-talkie to a quieter channel. If transient sounds, including calls of conspecifics, are interfering with communication, then the caller can

wait for a quiet period to produce a call, or can shift the frequency of its own call away from the interfering frequency band.

All of the mechanisms identified by engineers to compensate for noise or interference listed in the preceding paragraph have been documented in animals that communicate with sound. One strategy involves waiting to produce a signal until the noise level reduces, or timing vocalizations to minimize overlap with competing transient sounds. Mechanisms for timing signals have evolved particular sophistication in animals when the signalers are competing for attention and the “noise” comprises competing transient signals from echoes or conspecific sounds (Greenfield 1994; Hall et al. 2006). However, modifying when one produces a call does not qualify as vocal learning and using auditory input just to decide when to call seems a particularly simple form of auditory-vocal feedback. If the noise level is not changing rapidly enough, or if the animal cannot wait to get a signal through, then it can modify the acoustic structure of calls to compensate for the noise. Several animal taxa have been shown to increase the length of their calls in the presence of prolonged noise. Brumm et al. (2004) showed that a New World monkey, the common marmoset, lengthens the syllables of a call when exposed to white noise, which Egnor and Hauser (2006) also showed in studies of the cotton-top tamarin (*Saguinus oedipus*).

One of the predictions of communication theory (Shannon and Weaver 1963) is that the redundancy of signaling should increase as the channel becomes noisier. Some bird species increase the number of syllables in their calls or the bout duration of their songs with increasing noise: Japanese quail (*Coturnix coturnix japonica*; Potash 1972) and king penguins (*Aptenodytes patagonicus*; Lengagne et al. 1999). Chaffinches (*Fringilla coelebs*) sing longer bouts of the same song in noisy areas, increasing the redundancy of their singing (Brumm and Slater 2006b). Among mammals, humpback whales increased repetitions of phrases in their songs when they were exposed to a low-frequency sonar (Miller et al. 2000; Fristrup et al. 2003). These responses have been interpreted as compensation to increase the ability of receivers to detect and classify signals in a noisy channel. Turnbull and Terhune (1993) have shown that a harbor seal (*Phoca vitulina*) can detect a regular series of calls at a lower signal to noise ratio than a single call alone, providing support on the receiver side for this interpretation.

One of the most widespread compensation mechanisms reported for animals is shifting frequency to avoid band-limited noise. This frequency shifting is important to avoid interference from conspecific vocalizations, which are particularly likely to overlap in frequency. For example, some bats shift their echolocation calls away from the frequencies of conspecifics nearby in what is called a jamming-avoidance response (Ulanovsky et al. 2004).

The propulsion noise of ships in marine environments and road traffic noise in terrestrial environments both tend to be most intense at low frequencies, and these low frequencies also propagate best. Shipping noise has been increasing globally over the past century, and these long-term changes in noise have led to long-term increases in the frequencies of the contact calls of right whales (*Eubalaena* sp.) in the Atlantic Ocean, apparently to compensate for increasing low-frequency shipping noise (Parks et al. 2007).

Road noise from traffic varies in different locations, providing contrasts in noise that enable tests for how animals compensate. Slabbekoorn and Peet (2003) showed that male great tits (*Parus major*) recorded singing in quiet and

noisy areas of cities showed a systematic correlation between amplitude of the noise and frequency of their songs. Halfwerk and Slabbekoorn (2009) conducted playback experiments to test whether great tits can rapidly switch the frequency of their songs outside of a noise band. When Halfwerk and Slabbekoorn (2009) played low-frequency noise, the subjects increased the frequencies of their low notes, and when they played high-frequency noise, the subjects decreased the frequencies of their high notes. This suggests that exposure to noise in a frequency band causes the birds to alter their singing pattern to emphasize energy outside of the noise band. There is evidence that some anurans (Parris et al. 2009; Cunningham and Fahrig 2010) and even an insect (bow-winged grasshoppers, *Chorthippus biguttulus*; Lampe et al. 2014) shift the frequency of their calls upward when in the presence of low-frequency noise. These results emphasize the taxonomic spread of mechanisms to compensate for noise, and how this selects for auditory-vocal feedback

One complication for these studies is that some species show a correlation between source level and frequency of calls: the harder they sing, the higher the frequency emphasis (Nemeth et al. 2013). These complications suggest the importance of measuring both level and frequency of calls as a function of noise. Potvin and Mulder (2013) set out explicitly to test whether birds elevate the frequency of their calls to avoid low-frequency noise or whether the rise in frequency is just a byproduct of calling more loudly. They exposed silvereyes (*Zosterops lateralis*) to sound playbacks of high- and low-frequency noise. When exposed to high- frequency noise, silvereyes reduced the minimum call frequency, but at the same time, they increased the average power of their calls. Cardoso and Atwell (2011) measured the intensity and frequency of songs of dark-eyed juncos (*Junco hyemalis*). They argue that oscine songbirds can to large measure control intensity and frequency independently, and that songbirds can adjust to low frequency noise either by singing louder, higher in frequency, or both. Understanding whether some frequency shifts are a byproduct of changes in source level must hinge on testing for taxon-specific linkages between acoustic features that result from the sound production mechanism, and on modeling the impact of changes in each feature on the active space of the calls in varying noise.

Reviewing the variety of noise-induced vocal modifications observed in wildlife, Hotchkin and Parks (2013, p 817) point out that “closely related species of mammals can exhibit very different vocal responses to noise.” Some of this variability may have to do with tactical responses to variation in the timing and frequency spectra of the noise, some with differences in the communication tasks the animals are conducting, and some with taxon-specific linkages between acoustic parameters such as level and frequency. When humans speak in noise, the Lombard effect not only influences the intensity of the voice, but it also involves simultaneous modification of timing and frequency content (Lane and Tranel 1971). These changes in Lombard speech are thought to be linked through the biomechanical properties of the speech production system. Evidence for mechanisms to compensate for noise is so varied and ubiquitous that it suggests a strong selection pressure to evolve mechanisms that cover the specific problems faced by each taxon.

The mechanisms used by animals to compensate for noise appear to match specific problems. When faced with intermittent interference, many species time their calls to fall in the intervals between transient noises. When faced with

band-limited noise, many species shift the frequency of their calls away from the dominant noise band. When faced with continuous wideband noise overlapping the call in frequency, many species call more loudly. Evidence that animals select which acoustic features to modify to most efficiently maintain effective communication in varying noise supports the view that compensation mechanisms are more complex than simple reflexes. Nuclei in the brainstem that coordinate respiration and larynx/syrinx could modulate the intensity of pressure, driving the intensity of vocalizations, leading to Lombard effect. Given correlation between level and frequency in some species, noise-stimulated changes in level could also lead to correlated changes in frequency. However, most of the other compensation mechanisms change the actual timing, repetition, or frequency of the signal. Modifying these features would appear to require modifying the actual pattern generated by a central pattern generator.

9.14 Modulation of Vocal Output Based on Auditory Input of Noise and Based on Balancing Benefits Against Costs and Risks of Different Effective Ranges of Communication

All animals that communicate must balance the costs and benefits of signaling. All signals involve some energetic cost, but a more significant cost in some settings is the risk that an unintended audience, such as a competitor, predator, or parasite, will detect your call. When male tungara frogs (*Physalaemus pustulosus*; Ryan et al. 1982) or field crickets (*Gryllus integer*; Cade 1975) produce calls to attract females, they also have a much higher risk of being killed by bat predators in the case of the frog or parasitic flies in the case of the cricket.

One underexplored aspect of noise compensation is the extent to which it helps a caller adjust its signal just to meet the detection requirements at the expected range of the audience, while reducing the risk that eavesdroppers will exploit the call to the caller's detriment. Brumm and Slater (2006a) have shown that zebra finches increase their song amplitude with increasing range to their intended receiver. And when male chickens (*Gallus gallus domesticus*) see a predator, they are less likely to produce an alarm call when not in the presence of an adult female (Evans and Marler 1992). However, few if any studies test whether animals respond to temporarily increased risk of eavesdropping by using noise compensation mechanisms to reduce the range at which their calls are detectable by the threat. Such a balance is implied in the presumed need to adjust calls to produce the correct range of detection. If animals were not responding to cost and risk, then why not just produce the loudest call possible? In fact, Zahavi and Zahavi (1997) argue in the context of reproductive advertisement displays that females may select a calling male based on his willingness to suffer the costs and risks of signaling. In this kind of setting, males may not adjust their signaling based upon noise, but just call as loudly as possible.

When the Lombard effect is induced in humans, it appears to be involuntary, and early descriptions sometimes called it a reflex. However Lane and Tranel (1971) argue that the Lombard effect is more complex than a reflex, and they emphasize that it is designed rather generally to adjust calling behavior

so that the caller can better communicate in the presence of varying noise. In humans, the extent of the Lombard effect depends on the importance of intelligibility as well as on the noise level (Lane and Tranel 1971). This differs from most reflexes that use feedback from internal sensory input to stabilize one simple form of motor output. The Lombard effect is much more complex in humans, who integrate information about noise and about the importance of getting the message across to modulate the level, length, and frequency of their vocalizations.

If animals evolved a mechanism to facilitate effective communication in varying noise, then it would function best by integrating all of the different kinds of information an animal needs to make the correct decision about adjusting the acoustic features of its signal. This view of noise compensation suggests that it is part of a broader set of mechanisms to modulate signals to reach the desired audience while reducing energetic costs and risks of detection by dangerous eavesdroppers. The signaler would have to integrate information from many different sources before making a decision about fine-tuning the acoustic properties of its call. If this integration is facilitated by information processing in the cortex, then the demonstration by Eliades and Wang (2012) that the Lombard effect in nonhuman primates also involves cortical circuits supports the idea that some animals as well as humans may integrate complex multimodal information to make decisions about how to adjust their vocalizations to balance the benefits, costs, and risks of vocalizing.

9.15 Neural Pathways for Learned and Unlearned Vocalizations in Birds and Mammals

The neural pathways for complex vocal learning have best been studied in oscine songbirds and humans. Comparisons of neural control of vocalization in these taxa vs. non-vocal learners emphasize the role of forebrain structures (telencephalon in birds (Jarvis 2007) and cortex in mammals (Jürgens 2009)) in producing learned vocalizations, while innate vocalizations are thought to be controlled by central pattern generators in the midbrain and/or brainstem (Wild 1997).

Over the last 40 years, neurobiologists have uncovered specialized areas of the songbird brain that process the information required for vocal learning. Different parts of the songbird brain are specialized to use auditory input to form templates, to use feedback from auditory input to improve the match between vocal output and the template, and to learn to produce a stable learned vocalization. Simpson and Vicario (1990, p 1541) “suggest that the learned features of oscine songbird vocalizations are controlled by a telencephalic pathway that acts in concert with other pathways responsible for simpler, unlearned vocalizations.”

Jürgens (2009) argues that there are two separate neural pathways for the control of innate vocalizations versus learned vocalizations in mammals. Jürgens (2009) summarizes data suggesting a pathway for innate vocalizations involving the anterior cingulate cortex, the periaqueductal gray (PAG) in the midbrain, and areas of the reticular formation in the brainstem that have direct connections with phonatory motor nuclei. He argues that voluntary control of these

vocalizations involves the anterior cingulate cortex, regulation of the initiation and intensity of innate vocalizations is performed in the midbrain PAG, and that most of the motor pattern generation involves the reticular formation of the lower brainstem. Many PAG neurons that correlate with vocalization fire before, not during, vocalization, and many respond to auditory, visual, or somatosensory input, suggesting that the PAG can mediate feedback to the vocal control system from other senses.

The muscles involved in sound production are also represented in the motor cortex, and Jürgens (2009) argues that learned vocal patterns are initiated by a second pathway involving the motor cortex. Jürgens (2002) argues that humans have a direct nerve fiber pathway from the facial motor cortical areas to the nucleus ambiguus in the brainstem, where motor neurons project to the intrinsic laryngeal muscles. He argues that this pathway is absent in nonhuman primates, and that it may be essential for cortical processes that enable vocal learning to provide fine motor control over the larynx. Birds with vocal learning also have similar projections directly from the telencephalon to motor neuron pools that innervate the syrinx, strengthening the evidence that direct connections from telencephalon are required for vocal learning to control fine details of vocal sound production.

Deacon (1998), by contrast, argues that many mammals including species thought to not be vocal learners may have had connections from the cortex to the nucleus ambiguus early in development, but that these weak cortical projections are not as important as the projection from PAG in species without complex vocal learning. Deacon (1998) argues that for human learned sounds to compete with the PAG pathway, these direct cortical pathways had to expand, leading to a larger, more distinct pathway for learned vocalizations that are driven under control from the premotor cortex rather than from PAG. Arriaga et al. (2012, p 107) reports “that mice have a cortical vocal premotor circuit that projects directly to vocal motoneurons in the brainstem.” This projection is much weaker than that reported for humans and songbirds, providing support for Deacon (1998)’s more nuanced hypothesis regarding projections from the cortex to vocal motor neurons.

Both Jürgens (2002) and Deacon (1998) hypothesize that a direct cortical pathway to phonatory motor nuclei is essential for complex vocal learning, and they both predict that complex vocal learners have a direct tract from motor cortex to the brainstem nucleus that innervates the sound production organ. As mentioned at the start of this section, this has been measured in humans and songbirds, but tests are limited among other species with or without complex vocal learning. The mammals demonstrated to be complex vocal learners (seals, cetaceans, and elephants) offer unique opportunities for testing these hypotheses about the role of the cortex and about the need for direct connections from cortex to phonatory brainstem nuclei. If this argument also holds for the other mammals capable of producing novel signals, then it would predict strong pathways linking cortical circuits involved in producing vocalizations and motoneurons in the nucleus ambiguus that innervate the laryngeal muscles (or in the case of toothed whales, other motoneurons that innervate their unique sound production organ).

9.16 Evolution of Complex Vocal Learning

The study of vocal learning typically focuses on the role of learning in the development of complex birdsongs or human speech. When this is contrasted with evidence that vocalizations of other taxa can develop without auditory input from conspecifics, it is often concluded that the ancestral state in mammals and birds was lacking the character of vocal learning, and that it independently originated in the taxa where it has so dramatically been demonstrated. For example, Nottebohm (1972) argued that vocal learning evolved independently in songbirds, parrots, and hummingbirds from ancestors that lacked this trait.

Jarvis (2006, 2007) argues that all three avian vocal learning taxa share seven similar brain structures in the telencephalon. If the neural system controlling vocal learning evolved independently in the three vocal learning taxa of birds, then so many similarities in brain structures for vocal learning in the three avian taxa seems a remarkable coincidence. Jarvis (2007) compares vocal and auditory brain centers in the brains of humans and avian vocal learners, extending the argument for analogous brain structures for complex vocal learning to humans. He suggests that either these structures were used for some functions that made them particularly likely to be used for vocal learning or that a skill related to vocal learning was processed by these brain structures in a common ancestor.

The phylogeny of complex vocal learning suggests two or three independent origins in birds, and four or five among the mammals: humans, seals, cetaceans, elephants, and possibly bats (depending on whether vocal learning in bats is taken to be limited or complex; Fitch and Jarvis 2013). There may be similarities between the neural pathways for complex vocal learning in birds and mammals, but our lack of knowledge about these pathways in nonhuman mammals with complex vocal learning interferes with testing this idea. The data we have on the vocal learning abilities of different taxa are very spotty, and involve different criteria for learning. One critical component of a research program studying the evolution of vocal learning will be to identify a set of species likely to have or not to have each form of vocal learning, following explicit definitions. They will then need to be studied using comparable methods.

9.17 Did Complex Vocal Learning Evolve from Simpler Forms of Vocal Learning or Auditory-Vocal Feedback?

Much of the material reviewed in this chapter suggests that we need to question some assumptions about the evolution of vocal learning. There is something special about complex vocal learning systems that enable animals to learn new calls to form an open system of communication. However, rather than being an all-or-nothing capacity, there are a variety of ways that auditory input may alter vocal output, with varying taxonomic distributions and involving varying amounts of learning and flexibility. Petkov and Jarvis (2012) make a similar argument for distinguishing a range of vocal learning capabilities, which may have differing taxonomic distributions.

Research on vocal learning must recognize the variety of ways that auditory input can affect vocal output. Here, following Fitch and Jarvis (2013), this

chapter uses the term “complex vocal learning” for cases where animals can learn to imitate sounds that are not part of their species-specific vocal repertoire. This is essential for forming an open communication system, and may be fundamentally different from learning to fine-tune species-specific calls. The skills of learning to fine-tune species-specific calls, which Fitch and Jarvis (2013) call “limited vocal learning,” is more widely distributed among mammals. This chapter also defines a broader range of “auditory-vocal feedback,” which may not involve learning to develop a call based on an auditory template. If any of these forms of auditory-vocal feedback was shared among the ancestors of vocal learners and could form a substrate for the evolution of vocal learning, this could help to resolve the question of why the brain structures underlying independently evolved complex vocal learning appear to be so similar.

This chapter selected a definition for “vocal” that involves a sound production mechanism that may require auditory feedback to maintain stable acoustic features, and raises questions about the strength of evidence that hearing plays no role in the vocal development of many species identified as non-learners. The evidence reviewed on noise compensation suggests that most if not all mammals and birds have evolved the capability to modify acoustic parameters of their vocal output based on hearing the level and frequency range of ambient noise. Some bats have evolved similar mechanisms to modify their echolocation signals based on hearing echoes from previous signals. These phenomena appear to differ qualitatively from learning and memory mechanisms in which the learner forms or fine-tunes an auditory template and then learns to produce vocal sounds that match the template. Both limited and complex forms of vocal learning require auditory templates to be modifiable through audition, and require the ability to learn to match vocal motor output to a template. Thus it seems more likely that limited vocal learning is a better candidate than the simpler forms of auditory-vocal feedback for overlap in circuitry with complex learning. The distinction between complex versus limited vocal learning and the different forms of auditory-vocal feedback discussed in Sects. 9.11 – 9.13 suggests that they may involve neural circuits that do not fully overlap. However, it remains to be seen whether these neural circuits are completely independent, or whether some simpler forms may have served as precursors for the evolution of limited and/or complex vocal learning.

9.18 Selection Pressures for Auditory-Vocal Feedback and Vocal Learning

This chapter raises the question of whether simple forms of auditory-vocal feedback, which are widespread taxonomically, may have provided substrates for the evolution of more complex forms of vocal learning. This chapter does not propose a sharp definition separating auditory-vocal feedback from vocal learning. One critical distinction is that many of the simple forms of auditory-vocal feedback discussed here do not require memory, but vocal learning requires the formation of memories that enable the learning of auditory templates and the development of motor patterns producing vocalizations that match the template. As more evidence suggests a spectrum of complexity in vocal learning skills, more precise definitions will need to attend discussions of presence or

absence of different forms of vocal learning in different taxa. Here I discuss a range of selection pressures that may be important for the evolution of vocal learning.

9.18.1 Stabilizing the Acoustic Structure of Vocalizations

When sound is produced by stridulation or by sonic muscles exciting an air sac, a central motor program may be able to yield a predictable sound. It is less clear how a stereotyped vocalization can be produced by passing air through vocal folds without some form of auditory-vocal feedback. For pneumatic sound production, producing a stereotyped vocalization may involve a sophisticated mix of adjustments of air pressure, tension on vocal folds, and shape of the vocal tract. Slight variation in the tension or mass of the folds or shape of the vocal tract may lead to variation in the signal that would be hard to detect using any sense other than audition. Thus one function of auditory-vocal feedback may be to stabilize the acoustic structure of species-specific vocalizations.

9.18.2 Compensation for Noise

Another early selective pressure for modifying vocal output based on auditory input may stem from the ubiquitous problem of adjusting calls to maintain the receiver's ability to detect and classify the call. Although data on the exact evolutionary origin of the Lombard effect are incomplete, it appears that soon after vertebrates developed pneumatic mechanisms to produce sounds, they evolved mechanisms to modify sounds to improve communication in increased noise. This is a ubiquitous problem, especially important when sender and receiver are more than a few meters apart and when the sender must balance the benefits of delivering a signal to an intended receiver against costs if the signal is intercepted by another animal that may be a threat. The demonstration that invertebrates may modify calls produced by stridulation to compensate for noise shows that this selection pressure likely has a broader taxonomic scope than that for stabilizing pneumatically produced sounds.

9.18.3 Vocal Convergence

The broad taxonomic distribution of vocal convergence suggests that once the connections between auditory input and vocal output were established, a common function was to fine-tune production of specific call types to match those in one's group or population. Group-distinctive vocalizations have been argued to mediate cohesion of groups in a wide variety of taxa. The accommodation literature argues that signal matching can function as an affiliative signal. When an animal learns vocalizations from territorial neighbors it can better detect intruders and can use vocal matching to direct specific signals to specific recipients, often as a threat (Vehrencamp 2001). A variety of ecological and social settings could select for these diverse functions of vocal convergence. It is hard to imagine other mechanisms for developing group-distinctive calls among unrelated individuals, so this may be an important selective pressure for vocal learning among species whose ecology creates a benefit for group-distinctive calls.

9.18.4 Echolocation

The evolution of echolocation may also provide selective pressures for specialized mechanisms that modify vocal output of echolocation clicks based on information received from incoming echoes. The use of echolocation requires very rapid feedback between auditory processing of echoes to regulate the timing and acoustic features of the next sonar pulses (Moss and Sinha 2003). As bats or dolphins approach a target, they may reduce the intensity of their outgoing pulse to stabilize the level of the echo (Madsen and Surlykke 2013). Bats with frequency modulated echolocation signals can detect differences in the delay between click and echo of as little as 60 microseconds, and detection of these delays is used to time outgoing clicks (Moss and Sinha 2003). Bats that analyze the Doppler shift of constant frequency echolocation signals are able to change the frequency of the outgoing signal so that the Doppler-shifted echo matches the best frequency for hearing (Metzner et al. 2002). The neurobiology of auditory processing and vocal motor control has been studied in echolocating bats, which are one of the nonhuman mammalian taxa known for vocal learning (Metzner and Schuller 2010). Bats have a mammalian auditory system, and the vocal motor pathway controlling the larynx is shared with many mammals, but they also have many specializations for rapid sensorimotor integration to provide rapid feedback between auditory processing of one click and producing the next. Echolocation has clearly thus provided a selection pressure in bats for specialized and rapid forms of modifying vocal output based on auditory input. The presence of vocal learning abilities in the two taxa specialized for echolocation, bats and toothed whales, suggests that echolocation may have selected for more sophisticated and rapid vocal learning skills in these taxa.

9.18.5 Sexual Selection

Sexual selection also can create selection pressures for using vocal learning to elaborate the acoustic structure of signals (e.g., increasing the song repertoire: Catchpole 1980, 1986), for males to copy the signals of more successful males or to match calls of neighbors to direct a threat (Vehrencamp 2001), or for males to copy acoustic features preferred by females (West and King 1988). As with nonacoustic displays, sexual selection can lead to the evolution of elaborate and striking advertisement displays. This elaboration can make the role of vocal learning in development of sexually selected songs particularly obvious. Many of the species shown to be capable of vocal learning – songbirds, seals, and whales – produce songs that are sexually selected advertisement displays. The bias toward males producing these songs, and the evidence for onset of learned signals by males at the time of sexual maturity suggest that sexual selection has influenced the evolution of vocal learning in songbirds, seals, whales, and perhaps elephants. But this does not necessarily imply that sexual selection provided the original selection pressure for links between auditory input and vocal motor output in the first place.

A contrasting view is suggested by the much broader distribution of other forms of auditory-vocal feedback, which suggests that ubiquitous problems of stabilizing vocalizations or of communication in noise may have provided earlier selection pressures for modifying vocal output based on auditory input. Vocal convergence appears to have a more limited taxonomic scope, but may also pro-

vide an early selection pressure for fine-tuning acoustic features of production-first vocalizations. Sexual selection in this case may have used some of these building blocks for the evolution of vocal learning mechanisms to generate more complex or more precisely matched vocal repertoires. However, the evidence cited in Sect. 9.15 which locates neural pathways for some forms of auditory-vocal feedback in the midbrain and brainstem, and those critical for vocal learning in the telencephalon are at odds with this hypothesis. Resolution of these contrasting views requires a broader taxonomic scope studying capabilities of noise compensation, vocal convergence, and vocal learning and their neural mechanisms.

9.18.6 Functions of Vocal Learning in Highly Social Taxa with Prolonged Periods of Dependency and Reliance on Knowledge of Older Members of the Group

Much has been written about similarities between birdsong and human speech (Doupe and Kuhl 1999; Fitch and Jarvis 2013). But there are many differences between sexually selected advertisement displays and speech. Although both communication systems may have sensitive periods with predispositions for learning the correct sounds, in many songbird species it is just the male that sings, and there is much less evidence for songs than speech having flexible associations between specific learned vocalizations and external referents.

Some of the vocal learning species that do not sing provide better evidence of these latter features. An African gray parrot trained to produce human speech sounds not only learned to produce words, but could also use these words to report on features of objects such as color, shape, and number (Pepperberg 1999). Similarly a bottlenose dolphin (*Tursiops truncatus*), trained to mimic synthetic computer-generated sounds, learned to associate each sound with an arbitrary object such as a pipe, ball, or a frisbee. The dolphin then could produce the correct sound when shown one of these objects, so was able to vocally label an arbitrary object with a learned sound (Richards et al. 1984). Birdsong has provided fascinating parallels with speech in terms of sensitive periods, predispositions to learn, and the neurobiology of vocal development, but the abilities of species such as parrots and dolphins to learn arbitrary associations between learned vocalization and external referents provide a less explored and equally fascinating set of parallels with other aspects of speech.

Charvet and Striedter (2011) point out that oscine songbirds and parrots develop an unusually enlarged telencephalon through delayed maturation of this part of the brain. This requires a prolonged period of dependency, which is enabled by parental care. Charvet and Striedter (2011) argue that delaying telencephalic neurogenesis fosters the evolution of learned vocalization in humans, songbirds, and parrots. Following this argument, it is worth noting that among the most sophisticated vocal learners, there are a set of long-lived highly social animals with slow maturation and a prolonged period of dependence. Elephants and some toothed whales rival humans in having young dependent until their teenage years and with many adult females having a postreproductive period during which their reproductive effort is thought to be devoted to parental care (Marsh and Kasuya 1986 ; Tyack 1986; Lahdenperä et al. 2014). These older females retain knowledge of great importance to their groups (McComb et al.

2001; Brent et al. 2015). This social setting in which the young are dependent for years during which they can learn valuable information from caregivers may provide selective pressures for the use of vocal learning to develop more open forms of communication. The selection of songbirds as a model organism for vocal learning has enabled great progress in our understanding of the neural basis of the development of learned vocalizations, but I would argue that focus on communication in these other taxa may provide similarly valuable insights into the evolution of complex vocal learning in our own mammalian ancestors and into features of human communication for which we have few animal models.

9.19 Summary

The classic works on vocal learning have studied songbirds and humans to show how some species listen to an acoustic model and form an auditory template for the model. Then, often at a later date, they are able to listen to their own vocalizations, and slowly learn to produce sounds that match the template. If deprived of acoustic model to copy, these species learn to match their vocalizations to an inherited auditory template. This remarkable ability is usually contrasted with evidence that many other species can develop normal vocalizations with no auditory feedback at all.

However, there are a variety of ways that auditory input influences the acoustic structure of vocalizations that are more subtle than classic vocal learning. Some neurobiologists have argued that simple reflex-like auditory-vocal feedback taps completely different neural systems from vocal learning. But the last decades have revealed many different ways by which auditory input influences vocalizations. Some do not appear to involve learning and memory, some involve very rapid matching, and others involve matching on slow time scales similar to classic vocal learning. This chapter reviews these different forms of auditory-vocal feedback and asks whether some may use some of the same neural circuits and/or whether they may have provided neural substrates for the evolution of vocal learning.

This chapter defines vocalizations as sounds generated by passing air through vibrating lips, which then may be filtered by an upper respiratory path. The acoustic features of these kinds of vocal sounds are influenced by a complex combination of air pressure, tension, mass of the lips, and configuration of the vocal tract. A vocal motor program may be able to generate basic vocalizations such as crying or laughter, but auditory feedback may be required for matching more precise acoustic features of vocalizations.

This chapter discriminates limited vocal learning, in which an animal uses auditory input to fine-tune acoustic features of an inherited auditory template, from complex vocal learning, in which an animal can learn to produce a novel sound by learning a novel auditory template. The best evidence for complex vocal learning comes from birds and mammals that learn to imitate novel human-made sounds. Songbirds, parrots, seals, elephants, and dolphins all stand out for remarkable abilities of this kind of vocalization. Some bats and whales vary enough in their natural vocalizations, that one can track how conspecifics learn to match these variations.

A much broader range of mammals have been shown to converge in the acoustic structure of vocalizations when they are in a socially interacting group

of conspecifics. If this convergence involves fine-tuning the acoustic structure of an inherited call based on hearing others, then it qualifies as limited vocal learning. It is not known whether this limited form of vocal learning uses different neural mechanisms than complex vocal learning.

Echolocating bats use rapid auditory-vocal feedback to modify outgoing echolocation signals based upon incoming echoes. They use midbrain and brainstem for a system that can rapidly modulate vocalizations in tens of milliseconds. These systems are capable of very precise matching, but need not involve learning or memory. This suggests that complexity and precision of matching does not necessarily imply higher development of vocal learning. On the other hand, these systems may facilitate the evolution of vocal learning, as the two echolocating taxa (bats and toothed whales) contain vocal learners.

All birds and mammals tested use auditory-vocal feedback to adjust their vocalizations to compensate for the effects of noise, and many species compensate for varying costs and benefits of communicating. Neurobiologists have shown that some of these mechanisms can be produced in the brainstem, and they often treat these mechanisms as reflex-like. However, many different options for compensation are available to animals. To make the correct decision about which to use, they must integrate information about noise, about their audience, and about risks that others may detect their signals to their disadvantage. In some nonhuman primates, these mechanisms involve cortical pathways as well as the brainstem, and it remains to be seen whether compensation mechanisms are completely independent of those used in vocal learning, or whether this ubiquitous capability might provide some substrates for the evolution of vocal learning.

Most work on the evolution of vocal learning assumes that the ancestral state involved no vocal learning, and therefore that taxonomically remote vocal learners must represent independent evolution of this trait. This interpretation is supported by the demonstration of specialized telencephalic nuclei in the brains of avian vocal learners, and of their absence in non-learners. However, the similarities in neural circuits used by the three avian taxa with vocal learning present a puzzle. If vocal learning originated independently, is it likely that all three taxa would use the same nuclei connected in the same way? One answer could be that vocal learning is not an all-or-none phenomenon, and that some of the less obvious forms of auditory-vocal feedback may provide shared substrates for the evolution of vocal learning.

This chapter has explored recent evidence for auditory-vocal feedback and for limited vocal learning to broaden the scope of how we think about vocal learning and to suggest new approaches to studying the evolution of vocal learning. Progress will require more precise definitions of different forms of vocal learning, broad comparative review of their presence and absence, and behavioral and neurobiological investigations into the mechanisms underlying the skills.

Acknowledgments

P. L. T. thanks Henrik Brumm, Bjorn Merker, Arthur Popper, Peter Slater, and Rod Suthers for very helpful comments on an earlier draft of this chapter. This work received funding from the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland) and their support is grate-

fully acknowledged. MASTS is funded by the Scottish Funding Council (grant reference HR09011) and contributing institutions.

References

Arriaga, G., & Jarvis, E. D. (2013). Mouse vocal communication system: are ultrasounds learned or innate? *Brain and Language*, 124, 96–116.

Arriaga, G., Zhou, E. P., & Jarvis, E. D. (2012). Of mice, birds and men: the mouse ultrasonic sound system has some features similar to humans and song-learning birds. *PLoS ONE*, 7(10), 1–15.

Au, W. W. L., Pawloski, J. L., Nachtigall, P. E., Blonz, M., & Gisner, R. C. (1995). Echolocation signals and transmission beam pattern of a false killer whale (*Pseudorca crassidens*). *The Journal of the Acoustical Society of America*, 98(1), 51–59.

Baptista, L. F., & Morton, M. L. (1981). Interspecific song acquisition by a white-crowned sparrow. *The Auk*, 98(2), 383–385.

Baptista, L. F., & Schuchmann, K.-L. (1990). Song learning in the Anna hummingbird (*Calypte anna*). *Ethology*, 84, 15–26.

Baylis, J. (1982). Avian vocal mimicry: Its function and evolution. In D. E. Kroodsma & E. H. Miller (Eds.), *Acoustic communication in birds* (pp. 51–83). New York: Academic.

Beecher, M. D., & Brenowitz, E. A. (2005). Functional aspects of song learning in songbirds. *Trends in Ecology & Evolution*, 20(3), 143–149.

Boughman, J. W. (1998). Vocal learning by greater spear-nosed bats. *Proceedings of the Royal Society B: Biological Sciences*, 265, 227–233.

Brenowitz, E. A., & Kroodsma, D. E. (1996). The neuroethology of birdsong. In D. E. Kroodsma & E. H. Miller (Eds.), *Ecology and evolution of acoustic communication in birds* (pp. 269–281). Ithaca, NY: Cornell University Press.

Brent, L. J., Franks, D. W., Foster, E. A., Balcomb, K. C., Cant, M. A., & Croft, D. P. (2015). Ecological knowledge, leadership, and the evolution of menopause in killer whales. *Current Biology*, 25, 1–5.

Briefer, E. F., & McElligott, A. G. (2012). Social effects on vocal ontogeny in an ungulate, the goat, *Capra hircus*. *Animal Behaviour*, 83(4), 991–1000.

Brumm, H., Schmidt, R., & Schrader, L. (2009). Noise-dependent vocal plasticity in domestic fowl. *Animal Behaviour*, 78(3), 741–746.

Brumm, H., & Slater, P. J. B. (2006a). Animals can vary signal amplitude with receiver distance: evidence from zebra finch song. *Animal Behavior*, 71, 699–705.

Brumm, H., & Slater, P. J. B. (2006b). Ambient noise, motor fatigue, and serial redundancy in chaffinch song. *Behavioral Ecology and Sociobiology*, 60, 475–481.

Brumm, H., Voss, K., Köllmer, I., & Todt, D. (2004). Acoustic communication in noise: Regulation of call characteristics in a New World monkey. *Journal of Experimental Biology*, 207(3), 443–448.

Brumm, H., & Zollinger, S. A. (2011). The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour*, 148, 1173–1198.

Cade, W. (1975). Acoustically orienting parasitoids: Fly phonotaxis to cricket song. *Science*, 190, 1312–1313.

- Cardoso, G. C., & Atwell, J. W. (2011). On the relation between loudness and the increased song frequency of urban birds. *Animal Behaviour*, 82(4), 831–836.
- Catchpole, C. K. (1980). Sexual selection and the evolution of complex songs among warblers of the genus *Acrocephalus*. *Behaviour*, 74, 149–166.
- Catchpole, C. K. (1986). Song repertoires and reproductive success in the great reed warbler *Acrocephalus arundinaceus*. *Behavioral Ecology and Sociobiology*, 19, 439–445.
- Catchpole, C. K., & Slater, P. J. B. (2008). *Bird song: Biological themes and variations*. Cambridge, England: Cambridge University Press.
- Charvet, C. J., & Striedter, G. F. (2011). Developmental modes and developmental mechanisms can channel brain evolution. *Frontiers in Neuroanatomy*, 5, 1–5.
- Cranford, T. W., Amundin, M., & Norris, K. S. (1996). Functional morphology and homology in the odontocete nasal complex. *Journal of Morphology*, 228, 223–285.
- Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild chimpanzees produce group specific calls: A case for vocal learning? *Ethology*, 110, 221–243.
- Cunnington, G. M., & Fahrig, L. (2010). Plasticity in the vocalizations of anurans in response to traffic noise. *Acta Oecologica*, 36(5), 463–470.
- Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain*. New York: W. W. Norton.
- Delarue, J., Laurinolli, M., & Martin, B. (2009). Bowhead whale (*Balaena mysticetus*) songs in the Chukchi Sea between October 2007 and May 2008. *The Journal of the Acoustical Society of America*, 126(6), 3319–3328.
- Derégnaucourt, S., Mitra, P. P., Fehér, O., Pytte, C., & Tchernichovski, O. (2005). How sleep affects the developmental learning of bird song. *Nature*, 433(7027), 710–716.
- Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, 22, 567–631.
- Egnor, S. E. R., & Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. *Trends in Neurosciences*, 27, 649–654.
- Egnor, S. E. R., & Hauser, M. D. (2006). Noise-induced vocal modulation in cotton-top tamarins (*Saguinus oedipus*). *American Journal of Primatology*, 68 (12), 1183–1190.
- Eliades, S. J., & Wang, X. (2012). Neural correlates of the Lombard effect in primate auditory cortex. *The Journal of Neuroscience*, 32(31), 10737–10748.
- Evans, C. S., & Marler, P. (1992). Female appearance as a factor in the responsiveness of male chickens during anti-predator behaviour and courtship. *Animal Behaviour*, 43(1), 137–145.
- Farabaugh, S. M., Linzenbold, A., & Dooling, R. J. (1994). Vocal plasticity in budgerigars (*Melopsittacus undulatus*): Evidence for social factors in the learning of contact calls. *Journal of Comparative Psychology*, 108, 81–92.
- Fitch, W., & Jarvis, E. (2013). Birdsong and other animal models for human speech, song, and vocal learning. In M. A. Arbib (Ed.), *Language, music, and the brain (Strüngmann forum reports, Vol. 10, pp. 499–539)*. Cambridge, MA: MIT Press.
- Fripp, D., Owen, C., Quintana-Rizzo, E., Shapiro, A., Buckstaff, K., Jankowski, K., et al. (2004). Bottlenose dolphin calves model their signature whistles on

the whistles of community members they rarely hear. *Animal Cognition*, 8, 17–26.

Fristrup, K. M., Hatch, L. T., & Clark, C. W. (2003). Variation in humpback whale (*Megaptera novaeangliae*) song length in relation to low-frequency sound broadcasts. *The Journal of the Acoustical Society of America*, 113(6), 3411–3424.

Gans, C. (1973). Sound production in the Salientia: Mechanism and evolution of the emitter. *American Zoologist*, 13, 1179–1194.

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

Giles, H. (1984). The dynamics of speech accommodation. *International Journal of the Sociology of Language*, 46, 1–155.

Greenfield, M. D. (1994). Cooperation and conflict in the evolution of signal interactions. *Annual Review of Ecology and Systematics*, 25, 97–126.

Guinee, L. N., Chu, K., & Dorsey, E. M. (1983). Changes over time in the songs of known individual humpback whales (*Megaptera novaeangliae*). In R. Payne (Ed.), *Communication and behavior of whales* (pp. 59–80). Boulder, CO: Westview Press.

Hage, S. R., Jürgens, U., & Ehret, G. (2006). Audio-vocal interaction in the pontine brainstem during self-initiated vocalization in the squirrel monkey. *European Journal of Neuroscience*, 23, 3297–3308.

Halfwerk, W., & Slabbekoorn, H. (2009). A behavioural mechanism explaining noise-dependent frequency use in urban birdsong. *Animal Behaviour*, 78(6), 1301–1307.

Hall, M. L., Illes, A., & Vehrencamp, S. L. (2006). Overlapping signals in banded wrens: Longterm effects of prior experience on males and females. *Behavioral Ecology*, 17(2), 260–269.

Hammerschmidt, K., Reisinger, E., Westekemper, K., Ehrenreich, L., Strenzke, N., & Fischer, J. (2012). Mice do not require auditory input for the normal development of their ultrasonic vocalizations. *BMC Neuroscience*, 13(1), 40.

Hardus, M. E., Lameira, A. R., Van Schaik, C. P., & Wich, S. A. (2009). Tool use in wild orangutans modifies sound production: A functionally deceptive innovation? *Proceedings of the Royal Society of London B: Biological Sciences*, 276(1673), 3689–3694. doi: 10.1098/rspb.2009.1027.

Hayes, C. (1951). *The ape in our house*. New York: Harper and Brothers.

Hayes, K. J., & Hayes, C. (1952). Imitation in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology*, 45, 450–459.

Hiss, A. (1983). Hoover. *The New Yorker*, 3, 25–27.

Holy, T. E., & Guo, Z. (2005). Ultrasonic songs of male mice. *PLoS Biology*, 3(12), 2177–2186.

Horner, K., Serviere, J., & Granier-Deferre, C. (1987). Deoxyglucose demonstration of in-utero hearing in the guinea pig foetus. *Hearing Research*, 26(3), 327–333.

Hotchkin, C., & Parks, S. (2013). The Lombard effect and other noise-induced vocal modifications: Insight from mammalian communication systems. *Biological Reviews*, 88(4), 809–824.

Janik, V. M. (2000). Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*). *Science*, 289(5483), 1355–1357.

- Janik, V. M., & Sayigh, L. S. (2013). Communication in bottlenose dolphins: 50 years of signature whistle research. *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology*, 199(6), 479–489.
- Janik, V. M., & Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behavior*, 26, 59–99.
- Jarvis, E. D. (2006). Selection for and against vocal learning in birds and mammals. *Ornithological Science*, 5, 5–14.
- Jarvis, E. D. (2007). Neural systems for vocal learning in birds and humans: A synopsis. *Journal of Ornithology*, 148, S35–S44.
- Jones, G., & Ransome, R. D. (1993). Echolocation calls of bats are influenced by maternal effects and changes over a lifetime. *Proceedings of the Royal Society B: Biological Sciences*, 252, 125–128.
- Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews*, 26(2), 235–258.
- Jürgens, U. (2009). The neural control of vocalization in mammals: A review. *Journal of Voice*, 23(1), 1–10.
- Kelley, L. A., & Healy, S. D. (2011). Vocal mimicry. *Current Biology*, 21, R9–R10.
- Kellogg, W. N., & Kellogg, L. A. (1933). *The ape and the child*. New York: McGraw-Hill.
- Kikusui, T., Nakanishi, K., Nakagawa, R., Nagasawa, M., Mogi, K., & Okanoya, K. (2011). Cross fostering experiments suggest that mice songs are innate. *PloS ONE*, 6(3), 1–10.
- King, S. L., Sayigh, L. S., Wells, R. S., Fellner, W., & Janik, V. M. (2013). Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proceedings of the Royal Society of London B: Biological Sciences*, 280(1757), 20130053.
- Klatt, D. H., & Stefanski, R. A. (1974). How does a mynah bird imitate human speech? *The Journal of the Acoustical Society of America*, 55, 822–832.
- Konishi, M. (1963). The role of auditory feedback in the vocal behavior of the domestic fowl. *Zeitschrift für Tierpsychologie*, 20, 349–367.
- Konishi, M. (1965). The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Zeitschrift für Tierpsychologie*, 22, 770–783.
- Konishi, M. (2004). The role of auditory feedback in birdsong. *Annals of the New York Academy of Sciences*, 1016, 463–475.
- Kroodsma, D. E., & Konishi, M. (1991). A suboscine songbird (eastern phoebe, *Sayornis phoebe*) develops normal song without auditory feedback. *Animal Behavior*, 44, 477–487.
- Lahdenperä, M., Mar, K. U., & Lummaa, V. (2014). Reproductive cessation and post-reproductive lifespan in Asian elephants and pre-industrial humans. *Frontiers in Zoology*, 11(1), 54.
- Lampe, U., Reinhold, K., & Schmoll, T. (2014). How grasshoppers respond to road noise: Developmental plasticity and population differentiation in acoustic signalling. *Functional Ecology*, 28(3), 660–668.
- Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, 14, 677–709.
- Lengagne, T., Aubin, T., Lauga, J., & Jouventin, P. (1999). How do king penguins (*Aptenodytes patagonicus*) apply the mathematical theory of information to communicate in windy conditions? *Proceedings of the Royal Society of London B: Biological Sciences*, 266, 1623–1628.

- Lieberman, P. (1984). *The biology and evolution of language*. Cambridge, MA: Harvard University Press.
- Madsen, P. T., Jensen, F. H., Carder, D., & Ridgway, S. (2011). Dolphin whistles: A functional misnomer revealed by heliox breathing. *Biology Letters*, 8(2), 211–213.
- Madsen, P. T., & Surlykke, A. (2013). Functional convergence in bat and toothed whale biosonars. *Physiology*, 28(5), 276–283.
- Mammen, D. L., & Nowicki, S. (1981). Individual differences and within-flock convergence in chickadee calls. *Behavioral Ecology and Sociobiology*, 9, 179–186.
- Marsh, H., & Kasuya, T. (1986). Evidence for reproductive senescence in female cetaceans. *Reports of the International Whaling Commission*, 8, 57–74.
- Marshall, A. J., Wrangham, R. W., & Arcadi, A. C. (1999). Does learning affect the structure of vocalizations in chimpanzees? *Animal Behaviour*, 58, 825–830.
- McComb, K., Moss, C., Durant, S. M., Baker, L., & Sayialel, S. (2001). Matriarchs as repositories of social knowledge in African elephants. *Science*, 292(5516), 491–494.
- Metzner, W. (1989). A possible neuronal basis for Doppler-shift compensation in echo-locating horseshoe bats. *Nature*, 341 (6242), 529–532.
- Metzner, W. (1993). An audio-vocal interface in echolocating horseshoe bats. *The Journal of Neuroscience*, 13(5), 1899–1915.
- Metzner, W. (1996). Anatomical basis for audio-vocal integration in echolocating horseshoe bats. *Journal of Comparative Neurology*, 368 (2), 252–269.
- Metzner, W., & Schuller, G. (2010). Vocal control in echolocating bats. In M. B. Stefan (Ed.), *Handbook of behavioral neuroscience* (Vol. 19, pp. 403–415). London: Elsevier.
- Metzner, W., Zhang, S., & Smotherman, M. (2002). Doppler-shift compensation behavior in horseshoe bats revisited: Auditory feedback controls both a decrease and an increase in call frequency. *Journal of Experimental Biology*, 205, 1607–1616.
- Miksis, J., Tyack, P., & Buck, J. (2002). Captive dolphins, *Tursiops truncatus*, develop signature whistles that match acoustic features of human-made model sounds. *The Journal of the Acoustical Society of America*, 112(2), 728–739.
- Miller, P., Biassoni, N., Samuels, A., & Tyack, P. (2000). Whale songs lengthen in response to sonar. *Nature*, 405 (6789), 903.
- Mitani, J. C., & Gros-Louis, J. (1998). Chorusing and call convergence in chimpanzees: Tests of three hypotheses. *Behaviour*, 135, 1041–1064.
- Moss, C. F., & Sinha, S. R. (2003). Neurobiology of echolocation in bats. *Current Opinion in Neurobiology*, 13(6), 751–758.
- Nemeth, E., Pieretti, N., Zollinger, S. A., Geberzahn, N., Partecke, J., Miranda, A. C., et al. (2013). Bird song and anthropogenic noise: Vocal constraints may explain why birds sing higher frequency songs in cities. *Proceedings of the Royal Society of London B: Biological Sciences*, 280, 20122798.
- Nonaka, S., Takahashi, R., Enomoto, K., Katada, A., & Unno, T. (1997). Lombard reflex during PAG-induced vocalization in decerebrate cats. *Neuroscience Research*, 29(4), 283–289.
- Nottebohm, F. (1970). Ontogeny of bird song. *Science*, 167(3920), 950–956.

- Nottebohm, F. (1972). The origins of vocal learning. *American Naturalist*, 106, 116–140.
- Nottebohm, F., & Liu, W.-C. (2010). The origins of vocal learning: New sounds, new circuits, new cells. *Brain and Language*, 115, 3–17.
- Nottebohm, F., & Nottebohm, M. E. (1971). Vocalizations and breeding behavior of surgically deafened ring doves (*Streptopelia risoria*). *Animal Behavior*, 19, 313–327.
- Nowicki, S. (1989). Vocal plasticity in captive black-capped chickadees: the acoustic basis and rate of call convergence. *Animal Behaviour*, 37, 64–73.
- Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology*, 73(6), 530–544.
- Parks, S. E., Clark, C. W., & Tyack, P. L. (2007). Short- and long-term changes in right whale calling behavior: The potential effects of noise on acoustic communication. *The Journal of the Acoustical Society of America*, 122(6), 3725–3731.
- Parris, K. M., Velik-Lord, M., & North, J. M. (2009). Frogs call at a higher pitch in traffic noise. *Ecology and Society*, 14(1), 25.
- Payne, K., & Payne, R. (1985). Large scale changes over 19 years in songs of humpback whales in Bermuda. *Zeitschrift für Tierpsychologie*, 68, 89–114.
- Payne, K., Tyack, P., & Payne, R. (1983). Progressive changes in the songs of humpback whales (*Megaptera novaeangliae*): A detailed analysis of two seasons in Hawaii. In R. Payne (Ed.), *Communication and behavior of whales* (pp. 9–57). Boulder, CO: Westview Press.
- Pepperberg, I. M. (1999). *The Alex studies*. Cambridge, MA: Harvard University Press.
- Pepperberg, I. M. (2010). Vocal learning in grey parrots: A brief review of perception, production, and cross-species comparisons. *Brain and Language*, 115, 81–91.
- Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*, 4 (12), 1–22.
- Poole, J., Tyack, P., Stoeger-Horwath, A., & Watwood, S. (2005). Elephants are capable of vocal learning. *Nature*, 434 (7032), 455–456.
- Potash, L. M. (1972). Noise-induced changes in calls of the Japanese quail. *Psychonomic Science*, 26, 252–254.
- Potvin, D. A., & Mulder, R. A. (2013). Immediate, independent adjustment of call pitch and amplitude in response to varying background noise by silvereyes (*Zosterops lateralis*). *Behavioral Ecology*, 24(6), 1363–1368.
- Ralls, K., Fiorelli, P., & Gish, S. (1985). Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*. *Canadian Journal of Zoology*, 63 (5), 1050–1056.
- Reiss, D., & McCowan, B. (1993). Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): Evidence for vocal learning. *Journal of Comparative Psychology*, 107(3), 301–312.
- Rendell, L., Mesnick, S. L., Dalebout, M. L., Burtenshaw, J., & Whitehead, H. (2012). Can genetic differences explain vocal dialect variation in sperm whales, *Physeter macrocephalus*? *Behavior Genetics*, 42(2), 332–343.
- Rendell, L., & Whitehead, H. (2005). Spatial and temporal variation in

- sperm whale coda vocalizations: Stable usage and local dialects. *Animal Behaviour*, 70(1), 191–198.
- Richards, D. G., Wolz, J. P., & Herman, L. M. (1984). Vocal mimicry of computer-generated sounds and vocal labelling of objects by a bottlenosed dolphin, *Tursiops truncatus*. *Journal of Comparative Psychology*, 98, 10–28.
- Romand, R., & Ehret, G. (2004). Development of sound production in normal, isolated, and deafened kittens during the first postnatal months. *Developmental Psychobiology*, 17, 629–649.
- Ryan, M. J., Tuttle, M. D., & Rand, A. S. (1982). Bat predation and sexual advertisement in a neotropical frog. *American Naturalist*, 119, 136–139.
- Sanvito, S., Galimberti, F., & Miller, E. H. (2007). Observational evidences of vocal learning in southern elephant seals: a longitudinal study. *Ethology*, 113, 137–146.
- Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2006). Vocal expression of emotions in normally hearing and hearing-impaired infants. *Journal of Voice*, 20(4), 585–604.
- Schleidt, W. (1961). Operative Entfernung des Gehörorgans ohne Schädigung angrenzender Labyrinthteile bei Putenküken. *Experientia*, 17, 464–465.
- Shannon, C., & Weaver, W. (1963). *Mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- Simpson, H. B., & Vicario, D. S. (1990). Brain pathways for learned and unlearned vocalizations differ in zebra finches. *The Journal of Neuroscience*, 10(5), 1541–1556.
- Slabbekoorn, H., & Peet, M. (2003). Birds sing at a higher pitch in urban noise. *Nature*, 424, 267.
- Smolker, R., & Pepper, J. W. (1999). Whistle convergence among allied male bottlenose dolphins (Delphinidae, *Tursiops* sp.). *Ethology*, 105, 595–617.
- Snowdon, C. T., & Elowson, A. M. (1999). Pygmy marmosets modify call structure when paired. *Ethology*, 105, 893–908.
- Stoeger, A. S., Mietchen, D., Oh, S., de Silva, S., Herbst, C. T., Kwon, S., et al. (2012). An Asian elephant imitates human speech. *Current Biology*, 22(22), 2144–2148.
- Street, R. L., & Giles, H. (1982). Speech accommodation theory: A social cognitive approach to language and speech behavior. In M. Roloff & C. R. Berger (Eds.), *Social cognition and communication* (pp. 193–226). Beverly Hills, CA: Sage.
- Sugiura, H. (1998). Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. *Animal Behaviour*, 55, 673–687.
- Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.
- Todt, D. (1975). Social learning of vocal patterns and modes of their application in grey parrots (*Psittacus erithacus*). *Zeitschrift für Tierpsychologie*, 39, 178–188.
- Trainer, J. M. (1989). Cultural evolution in song dialects of yellow-rumped caciques in Panama. *Ethology*, 80, 190–204.
- Turnbull, S., & Terhune, J. (1993). Repetition enhances hearing detection thresholds in a harbor seal (*Phoca vitulina*). *Canadian Journal of Zoology*, 71(5), 926–932.
- Tyack, P. L. (1986). Population biology, social behavior, and communication in whales and dolphins. *Trends in Ecology and Evolution*, 1, 144–150.

- Ulanovsky, N., Fenton, M. B., Tsoar, A., & Korine, C. (2004). Dynamics of jamming avoidance in echolocating bats. *Proceedings of the Royal Society of London B: Biological Sciences*, 271 (1547), 1467–1475.
- Vehrencamp, S. L. (2001). Is song-type matching a conventional signal of aggressive intentions? *Proceedings of the Royal Society of London B: Biological Sciences*, 268, 1637–1642.
- Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., et al. (2015). Vocal learning in the functionally referential food grunts of chimpanzees. *Current Biology*, 25 (4), 495–499.
- Watwood, S., Tyack, P., & Wells, R. (2004). Whistle sharing in paired male bottlenose dolphins, *Tursiops truncatus*. *Behavioral Ecology and Sociobiology*, 55(6), 531–543.
- Weiss, D. J., Garibaldi, B. T., & Hauser, M. D. (2001). The production and perception of long calls by cotton-top tamarins (*Saguinus oedipus*): Acoustic analyses and playback experiments. *Journal of Comparative Psychology*, 115, 258–271.
- Wemmer, C., & Mishra, H. R. (1982). Observational learning by an Asiatic elephant of an unusual sound production method. *Mammalia*, 46, 556–557.
- West, M. J., & King, A. P. (1988). Female visual displays affect the development of male song in the cowbird. *Nature*, 334, 244–246.
- Wich, S. A., Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., & Shumaker, R. W. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates*, 50(1), 56–64.
- Wild, J. M. (1997). Neural pathways for the control of birdsong production. *Journal of Neurobiology*, 33 (5), 653–670.
- Winter, P., Handley, P., Ploog, D., & Schott, D. (1973). Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour*, 47, 230–239.
- Zahavi, A., & Zahavi, A. (1997). *The handicap principle*. Oxford, England: Oxford University Press.