

UCSF

UC San Francisco Previously Published Works

Title

A Modular Approach to Vocal Learning: Disentangling the Diversity of a Complex Behavioral Trait

Permalink

<https://escholarship.org/uc/item/3vw78863>

Journal

Neuron, 104(1)

ISSN

0896-6273

Authors

Wirthlin, Morgan
Chang, Edward F
Knörnschild, Mirjam
[et al.](#)

Publication Date

2019-10-01

DOI

10.1016/j.neuron.2019.09.036

Peer reviewed



Published in final edited form as:

Neuron. 2019 October 09; 104(1): 87–99. doi:10.1016/j.neuron.2019.09.036.

A Modular Approach to Vocal Learning: Disentangling the Diversity of a Complex Behavioral Trait

Morgan Wirthlin¹, Edward F. Chang², Mirjam Knörnschild³, Leah A. Krubitzer^{4,5}, Claudio V. Mello⁶, Cory T. Miller⁷, Andreas R. Pfenning¹, Sonja C. Vernes^{8,9}, Ofer Tchernichovski¹⁰, Michael M. Yartsev¹¹

¹Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, USA

²Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA, USA

³Animal Behavior Lab, Free University Berlin, Takustrasse 6, 14195, Berlin, Germany

⁴Center for Neuroscience, University of California, Davis, CA, USA

⁵Department of Psychology, University of California, Davis, CA, USA

⁶Department of Behavioral Neuroscience, OHSU, Portland, OR, USA

⁷Cortical Systems and Behavior Laboratory, Neurosciences Graduate Program, University of California San Diego, La Jolla, CA, USA

⁸Neurogenetics of Vocal Communication Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

⁹Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands

¹⁰Department of Psychology, Hunter College, The City University of New York, New York, NY, USA

¹¹Helen Wills Neuroscience Institute and Department of Bioengineering, UC Berkeley, Berkeley, CA USA

Abstract

Vocal learning is a behavioral trait in which the social and acoustic environment shapes the vocal repertoire of individuals. Over the past century, the study of vocal learning has progressed at the intersection of neuroscience, genomics, molecular biology, physiology, ecology, and evolution. Yet, despite the complexity of this trait, vocal learning is most frequently described as a binary trait, with species being classified as either vocal learners or vocal non-learners. As a result, studies have largely focused on a handful of vocal learner species. Recent studies, however, suggest a continuum in vocal learning capacity across taxa. Here, we further suggest that vocal learning is a multidimensional behavioral phenotype comprised of distinct, interconnected modules. Discretizing the vocal learning phenotype into its constituent components would facilitate integration of findings across a diversity of species, taking advantage of the ways in which each excels in a particular component, or in a specific combination of features. Such comparative studies can improve understanding of the mechanisms and evolutionary origins of vocal learning. We propose an initial set of vocal learning modules supported by behavioral

and neurobiological data, and highlight the need for diversifying the field of vocal learning to disentangle the complexity of the vocal learning phenotype.

Evolving definitions of vocal learning

The behavioral trait typically studied as “vocal learning” involves “the influence of auditory information, including feedback, on vocal development” [1]. Janik & Slater [2, 3] distinguished between (1) vocal usage learning, in which an individual can learn to produce an innate vocalization in a novel context (e.g., a dog can learn to bark when prompted); (2) vocal comprehension learning, in which an individual can learn to associate an acoustic signal with an action or an outcome (e.g., a dog can learn to sit after hearing the word “sit”); and (3) vocal production learning, in which an individual’s vocalizations are modified in some way as a result of social interaction with other individuals (as if a fictive dog could learn to repeat the word “speak,” as a parrot can, when in both cases the word is not a part of the species’ natural repertoire).

A lingering problem with the definition of vocal productive learning is that it is not entirely clear what level of vocal production plasticity qualifies a species to cross the ‘threshold’ to be considered a vocal production learner. How inclusive should be the set of species that possess the vocal learning trait (Figure 1)? In some taxa, all species are thought to be vocal production learners (songbirds & parrots [1]). In other taxa, conclusive evidence for vocal production learning exist in only a few species (hummingbirds, cetaceans, bats, and pinnipeds [2, 4]), and in others the evidence for vocal learning is weak (non-human primates, elephants, goats, rock hyraxes, and suboscine bellbirds [5–9]). In the inbred laboratory mouse, for example, some studies showed evidence for vocal plasticity [10], while other studies claim that vocalizations are entirely innate [11–13] and that variability in vocal production across behavioral states may be a byproduct of changes in locomotor activity [14]. This uneven evidential landscape presents a conundrum for researchers interested in performing comparative analyses of vocal behavior.

As a counterpoint to the view of vocal production learning as a binary trait, the Continuum Hypothesis was proposed by Arriaga, Petkov, & Jarvis [15, 16]. The Continuum Hypothesis states that species can be ranked according to their vocal learning ability on a linear, hierarchical scale ranging from vocal non-learners, through limited and moderate learners, to complex and high vocal learners. The most rudimentary or limited form of vocal learning is stated to be the ability to modify the amplitude and temporal structure of an innate call. Placed above this is the capacity for vocal improvisation, which may include spectral modifications, and at the top are species capable of spectra-temporal modification of their vocalizations in order to imitate (or mimic) an acquired acoustic target [16]. By allowing for the existence of a gray area between vocal learning and non-learning, the Continuum Hypothesis opened the field to investigation of a broader range of species. However, the Hypothesis rests under the assumption of a single dimension along which the evolution of vocal learning proceeds.

A Modular Hypothesis of vocal learning

We propose that rather than a binary or single continuous trait, vocal learning represents a multidimensional behavioral trait comprised of discrete sub-traits, which we term modules (Figure 2). Our use of the term module here is inspired by the mosaic model of brain evolution [17, 18], in which individual behavioral modules can be independently targeted by evolutionary selection. That is, it may be possible for a species to have a high degree of complexity in some modules but not others. We suggest that even if a species is a “poor” vocal learner on average (when considering all modules combined), it might still be a preferred system for studying the neurobiology of a single module in isolation. For example, from the perspective of studying mechanisms of human speech development, the transition from highly variable juvenile vocalizations (babbling) to stereotyped adult vocalizations can be observed both in excellent vocal learners such as songbirds (e.g., zebra finch [19]), but also in some vocal non-learners (e.g., the Japanese quail, [20]). Both species show a developmental transition from variable babbling to stereotyped calls. However, the quail does so without any imitative learning while the songbird acquires novel sounds learned from its tutors [20]. Here, the lack of other features in the quail that are traditionally attributed to the vocal learning trait could be an advantage, allowing studies to focus on a single component in isolation of other components. This is in contrast to a more complex vocal species, such as the songbird, where multiple features can all vary in conjunction and hence make it more difficult to identify the contributions of features independently. At the neural level, the quail does not possess the brain circuitry necessary for vocal mimicry and yet still retains the components necessary for vocal babbling. Thus, it can be used as an exquisite system for studying that specific module from the behavioral to the neural level.

More generally, our framework is aimed at facilitating integration between studies in songbird and non-songbird models in order to encourage a move towards a more holistic view of this complex trait. As with any behavior, vocal learning does not exist in a vacuum, but rather reflects the ways in which each species is uniquely adapted to its physical and social environment, and constrained by its particular evolutionary history [21]. To properly account for this context, comparison of a diverse set of vocal learning taxa is necessary to distinguish general patterns from lineage-specific features. As a starting point for further deliberation, we propose three initial modules: (1) vocal coordination, (2) vocal production variability & feedback, and (3) vocal versatility. Using these, we aim to highlight the utility of the modular approach in diversifying the study of the vocal learning trait.

Vocal coordination

The first proposed module refers to the ability of an animal to flexibly modify the temporal production of its vocal output. This may range from, at its most basic, reflexive vocal production in response to a stimulus, to predictive vocal production coordinated with the actions of a partner, to synchronized rhythmic entrainment in which the ongoing temporal vocal coordination becomes self-guided (Figure 3).

Gradient of Vocal Coordination Across Species

In its simplest expression, vocal coordination takes the form of a reflexive response to a stimulus, such as the call-and-response behavior exhibited by multiple species, that is likely gated by hypothalamic and amygdalar mechanisms with very little influence from learning [22]. Such behavior could reasonably be described as an example of vocal usage learning under Janik & Slater's definitions of vocal learning [3].

Other forms of vocal coordination can emerge as a response to different forms of “jamming” in the acoustic environment. In the example of call alternation, multiple vocalizing individuals may adjust the phase and tempo of their calls in response to neighboring callers in order to avoid signal interference. This has been extensively studied in anurans, whose breeding success depends on the precise timing of calls to maximize the chances of their own calls to be heard by a potential mate, or to actively interrupt the calls of their neighbors [23–25]. A similar ability to improve the signal-to-noise ratio of vocal signals in a noisy environment is also observed in group-foraging bats, where individuals can dynamically coordinate their echolocation calls during flight to avoid jamming interference [26]. This may be achieved through diverse mechanisms, including temporal vocal synchronization and shifting in the frequency of calls to avoid acoustic overlap [27, 28]. Here there are interesting mechanistic parallels with jamming avoidance in the weakly electric fish that present an exquisite example of non-vocal coordination and where the neurobiological underpinning have been extensively elucidated [29, 30]. Importantly, in both the electric fish, the frog, and the bat, the signals are innate, yet their temporal and spectral structure can be flexibly adjusted in response to the social and acoustic environment.

Evidence from songbirds suggests that vocal coordination may represent a distinct module of vocal behavior even in more adept vocal learners. Although in zebra finches, only the male naturally learns to produce a song matching that of its tutors, both sexes produce several types of innate calls, including short calls that can be rhythmically coordinated [31]. In an acoustic jamming challenge in which the normal timing of a bird's short calls was interrupted by a vocal robot, the birds displayed rapid rhythmic adaptation to avoid their calls being disrupted. The birds were further able to avoid disruption presented by more complex rhythmic interruptions, demonstrating vocal coordination that is not simply responsive, but predictive of future calls [31]. Finally, lesions targeting to the posterior motor pathway (RA and HVC-to-RA projections) impaired the birds' ability to predictively coordinate the precise timing of these innate calls in both female and male zebra finches, suggesting a neuroanatomical locus for the precise timing of these calls [31].

Several animal species, including ‘vocal non-learners’, can synchronize the timing of vocal exchanges with a partner with greater degrees of precision, e.g. antiphonal calling, vocal turn-taking, and duetting [32, 33]. In mammals, neotropical singing mice (*Scotinomys* sp.) engage in agonistic countersinging behavior that is tightly temporally coordinated, comprised of simple notes that are identifiable from birth [34, 35]. As a further example, mate-paired gibbons sing elaborate and extensive duets comprised of complex sequences of individually simple calls [36, 37]. In pairs of siamang gibbons, the proper timing of their great-call duet sequence takes months to develop [38]. In one example, a newly formed pair performed frequent temporal sequencing errors, leading to premature termination of the duet

in all but 24% of these initial bout attempts. After 18 weeks, the pair achieved successful coordination of 79% of duet bouts, suggesting that learning the temporally coordinated structure of their duet was accomplished through extensive social feedback and practice [38]. Other species of gibbon have been shown to match the rate of singing of a simulated conspecific played via tape [39].

In birds, vocal duetting is observed in all three taxa traditionally considered vocal production learners—songbirds, parrots, and hummingbirds—but also in vocal non-learners such as swans, owls and woodpeckers [40, 41]. Extensive learned temporal vocal coordination is also observed in suboscine Passerines, the sister taxon of songbirds (i.e. oscine Passerines). The suboscine clade consists of over 1,000 species that are widely regarded as being vocal non-learners on the basis of deafening and hand-rearing experiments performed in just 3 closely related species of a single family, Tyrannidae [42–44]. In other suboscines, there is evidence for learning in the temporal domain: male long-tailed manakins (suboscine family Pipridae) develop coordinated duet songs over time as part of their cooperative lek-breeding displays [45]. Mate-paired horneros (family Furnariidae) sing a highly structured duet with a complex rhythmic time signature, in which the male and female must match and coordinate their tempo with that of their partner’s on a subsecond level [46, 47]. Pair-living Peruvian warbling antbirds (family Thamnophilidae) not only develop coordinated duets, but also display remarkable temporal vocal flexibility: females will respond to the signals of unpaired female sexual rivals by jamming their own male partner’s song, which the males will attempt to circumvent by adjusting the tempo of their own song in order to avoid jamming interference [48].

Duetting may exist on a continuum with the phenomenon of vocal turn-taking, where species’ vocal interactions display ongoing synchronized rhythmicity [32, 33, 49]. In marmosets, vocal turn-taking is learned during the first year of life, and may be facilitated by directed social feedback from the parents [50]. Turn-taking is not exclusive to the vocal domain, being evident also in human sign language as well as the gestural turn-taking observed in great apes [49]. Vocal and gestural coordination in primates has been widely proposed to be a fundamental precursor to human speech [22, 49].

Neural Mechanisms of Vocal coordination

Neurobiological evidence supports the existence of brain areas involved in temporal vocal coordination. In the case of human speech, strokes in various speech areas can result in “acquired arrhythmia,” the selective disruption of the ability to reproduce a rhythm or entrain to a beat, even in cases where rhythmic perception and melodic production are unaffected [51, 52]. Various cortical structures have been specifically implicated in the temporal coordination of vocal output in humans. Electrophysiological recordings during speech indicate that a high degree of temporal coordination is present within the laryngeal / orofacial sensorimotor cortex [53, 54]. It has been further proposed that this sensorimotor activity is guided primarily by a higher-order motor area for temporal processing, likely Broca’s area and the supplemental motor area (SMA), with additional roles for the basal ganglia and cerebellum [55, 56].

In songbirds, a very similar pattern is observed. As a zebra finch sings, neuronal activity in the primary motor output nucleus RA is tightly time-locked to specific song motifs [57]. The temporal encoding of this ensemble activity derives from descending projections from premotor nucleus HVC [58, 59]. The hypothesized role of HVC in temporal control over song has been suggested by cooling experiments, in which focal cooling of HVC leads to an overall slowing of song production, further implicating this nucleus in the temporal aspects of learned song [60]. Remarkably, in humans focal cooling of speech-associated premotor Broca's area leads to a specific slowdown in overall speech timing without significant loss in speech quality, while cooling of speech-associated ventral motor cortex results in a breakdown in the spectral quality of speech, with little effect on timing [61]. Finally, recordings in HVC of wild songbirds demonstrate that HVC encodes not only the temporal pattern for their own part in a coordinated duet, but that of their partner's as well [62, 63].

Taken together, these examples suggest that vocal coordination is a complex behavior involving varying degrees of learning, which is not imitative per se. Nevertheless, it has clear mechanistic overlaps with imitative vocal learning: both are guided by feedback, and both may be controlled by overlapping brain centers.

Vocal Production Variability & Feedback

The second modulate related to how vocal learning species dynamically change the acoustic variability of their vocal output, either during critical periods of development or seasonally in adulthood [19, 64, 65]. In particular, vocal learning typically begins with a highly variable babbling phase, which has been studied extensively in human infants and in songbirds [64]. The role of babbling is to promote vocal exploration, which allows the animal to learn how to use its vocal apparatus to produce a variety of vocal sounds, and then to match its vocal production to an external model in adulthood. This later consolidation of vocal sounds is guided by auditory and social feedback. Feedback can shape vocalization at multiple levels: most essential is the role of auditory feedback from the animal's own vocalization [66, 67]. Auditory feedback allows a self-evaluation of production errors [68] and may be influenced by an external auditory input, sometimes referred to as a sensory template [69]. Finally, vocal learning may be guided by higher level social selective pressures. For example, social feedback from a parent, or from adult conspecific, may reinforce the production of a call by its young in the correct context. (Figure 4).

Vocal variability with and without feedback

The ability to dynamically control variability during vocal production is not limited to vocal learners. For example, stages of vocal babbling resembling the dynamics by which subsong utterances are transformed into adult syllables and motifs in songbirds have been observed in quail [19, 20]. In its most rudimentary form, vocal babbling includes an initial phase of juvenile vocal variability, which largely disappears by adulthood without an apparent influence of any auditory or social feedback [2, 70, 71]. In juvenile Japanese quail, the precise spectral and temporal trajectory by which noisy initial calls develop into adult crows varies greatly between individuals, and yet the outcome of vocal development is nearly identical across individuals. Although we do not understand why vocal babbling

exists in some vocal non-learners, but not in others, it is a fascinating case of potential pre-adaptation for vocal learning. Importantly, it further provides an opportunity for studying the transition from vocal exploration to consolidation of structure in animal models that lack the complexity of an external feedback influences.

Similarly, in other species external sources influence vocal variability, but do not influence the vocalization ultimately learned. For example, parental interactions influence vocal production variability in the juvenile marmoset, who undergo a ‘babbling’-like phase of highly variable output [72, 73]. However, although these parental interactions affect the developmental rate of calls, they do not affect the final outcome of vocal development [74].

Juvenile begging calls present a further example where external feedback influences vocal variability. In many birds and mammals, infants engage in vocal begging to solicit parental food provisioning. Although these calls are innate, competing nestlings and littermates alter the variability of their calls, and at least in some cases, this variability is guided by social and auditory feedback. This includes established vocal learning taxa such as songbirds [75–78] and parrots [79], but also taxa not otherwise appreciated for their capacity for vocal learning, including cuckoos [80], pigeons [81], pigs [82, 83], meerkats [84], and pygmy marmosets [72]. In songbirds, nestlings are able to modify their begging calls to more closely match the acoustic parameters of their parents’ song, the capacity for which is reduced with lesions to the developing song system, suggesting that variable nest begging behavior could represent a prerequisite for the evolution of song learning [85].

On the other end of the developmental spectrum, some species retain vocal variability in adulthood. In many cases, this maintenance or reopening of vocal plasticity does not occur as a result of social learning, but rather as a result of changes in neuroendocrine factors, as in the case of the seasonal singing of many species of songbirds and vocalizing fish [65, 86]. In other cases, experimental disruption of auditory feedback leads to a dynamic increase in vocal variability of a previously fixed vocalization, revealing an ongoing role for this module in the experience-dependent maintenance of adult vocalizations [87].

Neural mechanisms of vocal variability

Interestingly, exploratory variability is not a passive feature of neural systems, but rather one that is dynamically regulated to facilitate motor learning [19, 88]. During juvenile song learning in zebra finches, the forebrain nucleus LMAN—a component of the cortico-basal ganglia-thalamo cortical forebrain loop required for song learning [89]—acts as a driver of vocal behavioral variability by injecting a variable spiking signal into the premotor output nucleus RA [90]. Over the course of learning, this variable output is refined by the influence of auditory feedback-based error correction, which ultimately enables the consolidation of the vocal motor pathway’s control of the learned adult song [68, 87, 91, 92]. In mammals, cortico-striatal-thalamo-cortical loops are known to play an essential role in flexible motor behavior [93–95], and likely play a similar role in the context of vocal production variability. However, the precise neuroanatomical substrate for the active regulation of vocal variability in human speech learning and other mammalian vocal learning systems has not been well elucidated.

Integrating these findings, we suggest that vocal production variability represents a distinct behavioral module supported at least in songbirds by a distinct neurobiological substrate. The observation that species dynamically regulate vocal production variability through a diversity of mechanisms motivates a comparative approach to identify its diverse underlying mechanisms as well as their influence on vocal plasticity.

Vocal Versatility

Lastly, we propose a two-axis module which we refer to vocal versatility. This module is defined as the combination of two related traits: vocal plasticity, the degree to which a species repertoire can be altered as a result of experience and, vocal diversity which is the repertoire size of vocal sounds a species is capable of producing (Figure 5). Vocal versatility thus ranges from species that produce only a few innate calls, such as humming fish or crocodilians, to species with advanced interspecific mimicry and the ability to imitate novel sounds not part of the species' natural repertoire, as exemplified by many songbirds and parrots. However, this two-dimensional framework also includes species with diverse vocal repertoires that are largely innate, such as in the domesticated chicken and the macaque monkey, as well as species with fairly limited natural diversity of vocal sounds that are nevertheless amenable to learned modification, such as harbor seals. Such uncoupling between vocal plasticity and diversity is in turn linked to the evolution of peripheral structures for sound production, which expand the vocal repertoire, versus the evolution of the central nervous system, which expands vocal plasticity.

Peripheral Components of Vocal Versatility: Source and Filter

Broadly across vocalizing vertebrates, a primary structure serves as the 'source' of vocal output (the larynx, and additionally the syrinx in birds), which is 'filtered' spectrally through the vocal tract (including the trachea, pharynx, and nasal cavity) with a set of supralaryngeal structures that can dynamically change the length of the tract and its resonance frequencies. The set of articulatory filtering mechanisms exhibit wide phenotypic diversity across taxa, and include the tongue, lips, jaw, palate, beak, and trunk. The source-filter theory distinguishes these physical structures based on their role in shaping acoustic output as well as by their anatomical separation [96, 97]. Importantly, it provides a helpful framework for conceptualizing how different species achieve different levels of vocal versatility. For example, whereas in songbirds vocal complexity is achieved largely by controlling the source (i.e. their syrinxes), in human speech, acoustic output is extensively shaped by filtering (e.g., using our tongue). We organize our description of this module to focus separately on the source and filter structures and their related neurobiological mechanism in each section.

Source: Evolution of Structure Complexity and Forebrain Control

At the source level, several species have evolved unique adaptations in their primary vocal organ to enable a more flexible and elaborate forms of vocal output. The syringeal anatomy of parrots and songbirds is more complex compared non-learning birds such as ducks and ibises [98, 99] and songbirds have further evolved "superfast" syringeal musculature as well as the ability to independently control sound production in each of the two branches of

the syrinx. This in turn, allows for the production of more complex vocalizations, with the degree of anatomical complexity possibly also correlating with the species' capacity for vocal imitation [98–102]. Similarly, great diversity in source complexity are evident across mammals. The laryngeal anatomy of some bats has also been found to contain superfast muscles [103] while in other vocal learning mammals, however, there is little evidence for increasing complexity at the source level. Yet, many animals have also found creative ways of achieving vocal diversity by utilizing different features of the source. Such is the case for the screams of macaques and pant-hoots of chimpanzees, where harmonic calls are transformed into a graded series of increasingly complex and noisy calls through relatively simple additive application of respiratory pressure [104, 105].

However, the complexity of the vocal organ alone may not be sufficient to enable an advanced form of vocal versatility. Recent support for this notion comes from anatomical studies in non-human primates. While in the past it has been claimed that the human 'descended larynx' is uniquely evolved to enable complex speech [106], more recent work has disputed this claim, suggesting that even the vocal tract of macaques may be 'speech-ready,' missing only the necessary neural structures for learning rather than the physiological capacity for more complex vocal production [107]. Thus, in humans as in other vocal learners, it is believed that at the level of the central nervous system, higher brain structures must gain control over the brainstem in order to facilitate higher vocal versatility. Experimental work in humans and songbirds has illuminated many of the neurobiological mechanisms underlying the flexible cortical (or pallial in the case of avians) control over sound production [54, 108, 109]. Nevertheless, further work is needed to bridge the gap between findings in the avian and mammalian brains. One potential approach is to search for shared mechanisms underlying the capacity for modifying vocal pitch, a trait shared by several avian and mammalian species including songbirds [110], parrots [111], cetaceans [112], pinnipeds [113], bats [114, 115], and New World primates [116, 117]. Identifying common neural mechanisms facilitating this capacity would first require detailed delineation of forebrain structures that enable the adaptive modification of pitch in the mammalian brain. However, this challenge is becoming more tractable with novel genetic and anatomical tracing approaches for identifying homologous brain structures and cell types across taxa [118, 119].

Filter: Evolution of Structure Complexity and Forebrain Control

While major research focus has been devoted to laryngeal and syringeal control, we know relatively little about the diverse mechanisms mediating articulatory filtering in vocalizing animals. Considering the crucial role of supralaryngeal structures in enabling more refined spectral shaping during vocal production (including in human speech), this topic deserved far more attention. For example, parrots display high dynamic control over their tongue, which plays a prominent role in the production of their extensive natural vocal repertoire as well as their expert imitative abilities [120, 121]. Parrots have been shown to have more extensive forebrain circuitry for learned vocal behavior than songbirds [122] but the extent to which this circuitry plays a role in the control of vocal articulatory structures beyond the syrinx remains unexplored. Among primates, control of the tongue via direct motor cortex projections to the hypoglossal nucleus appears to be absent in saddle-back tamarin,

present to a degree in squirrel monkey and rhesus macaque, and most densely innervated in humans, suggesting that the density of this representation may be correlated with the degree of tongue-mediated vocal behavior in these species [123]. Among bats, the Egyptian fruit bat (*Rousettus aegypticus*) similarly displays an exceptional degree of control over the tongue: whereas most bats' echolocation calls are produced from the larynx, in *R. aegypticus* they are exclusively tongue-based [124, 125]. More recently, these bats have been demonstrated to exhibit a capacity for long term and persistent vocal plasticity of their social communication calls across multiple spectral parameters [115], further supporting a potential link between the degree of control over the articulatory filtering structures and the capacity for vocal plasticity.

Coordination of Source and Filter

Lastly, nearly all animal vocalizations involve some degree of coordination between these various source and articulatory filtering mechanisms (as well as respiratory output), which is reflected in their increasingly integrated representation at higher levels of vocal control circuits [126–128]. However, as the precise mechanisms for the neural control of source and filter structures are seldom examined in combination, the possible mechanisms for integrating and controlling them remain largely unexplored. In humans, electrophysiological recordings during speech production have demonstrated that the diverse articulatory structures as well as the larynx each have distinct somatotopic representations and temporally coordinated activity in the sensorimotor cortex [53]. Based on the vocal learner convergences observed for laryngeal / syringeal control, it is likely that other vocal learners could possess similar shared specializations for the coordinated cortical control of the articulatory structures that enable vocal versatility. We suggest that comparative work to address this possibility represents a rich domain for future neurobiological inquiry.

Conclusions

We formulate a framework for the study of vocal learning, in which the behavior is proposed to be a multidimensional trait comprised of distinct but interdependent subcomponents, which we term modules. Within this framework, we suggest three potential modules—vocal coordination, vocal production variability & feedback, and vocal versatility. The modular approach proposes that incorporating species along the different axis of distinct modules will facilitate a better mechanistic and evolutionary understanding of the building blocks which together comprise the vocal learning trait. The three modules discussed here benefited from positioning species along these individual axes, starting from the origin (or most rudimentary level) and advancing towards the champion species that best exemplify each module. The broadly comparative work that will be necessary to assess species across these dimensions is greatly facilitated by rapid advancement in modern genomic tools and computational approaches [129–134] that can enable linking behavioral and neurobiological components along each axis. Ultimately, we believe a modular, multidimensional framework will allow for a more efficient inclusion and comparisons of traits between human and animal models, as we may ask not which species represents the singular best model for vocal learning, but rather which particular aspects of vocal learning can best be studied across a diverse plurality of model systems.

References

1. Nottebohm F (1972). The Origins of Vocal Learning. *The American Naturalist* 106, 116–140.
2. Janik VM, and Slater PJB (1997). Vocal Learning in Mammals. *Advances in the Study of Behavior* 26, 59–100.
3. Janik VM, and Slater PJB (2000). The different roles of social learning in vocal communication. *Animal Behaviour* 60, 1–11. [PubMed: 10924198]
4. Jarvis ED (2004). Learned Birdsong and the Neurobiology of Human Language. *Annals of the New York Academy of Sciences* 1016, 749–777. [PubMed: 15313804]
5. Fischer J (2017). Primate vocal production and the riddle of language evolution. *Psychonomic Bulletin & Review* 24, 72–78. [PubMed: 27368619]
6. Stoeger AS, and Manger P (2014). Vocal learning in elephants: neural bases and adaptive context. *Current Opinion in Neurobiology* 28, 101–107. [PubMed: 25062469]
7. Briefer EF, and McElligott AG (2012). Social effects on vocal ontogeny in an ungulate, the goat, *Capra hircus*. *Animal Behaviour* 83, 991–1000.
8. Kershenbaum A, Ilany A, Blaustein L, and Geffen E (2012). Syntactic structure and geographical dialects in the songs of male rock hyraxes. *Proceedings of the Royal Society B: Biological Sciences* 279, 2974–2981.
9. Saranathan V, Hamilton D, Powell GVN, Kroodsmas D, and Prum RO (2007). Genetic evidence supports song learning in the three-wattled bellbird *Procnias tricarunculata* (Cotingidae). *Molecular Ecology* 16, 3689–3702. [PubMed: 17845441]
10. Arriaga G, Zhou EP, and Jarvis ED (2012). Of Mice, Birds, and Men: The Mouse Ultrasonic Song System Has Some Features Similar to Humans and Song-Learning Birds. *PLOS ONE* 7, e46610. [PubMed: 23071596]
11. Hammerschmidt K, Reisinger E, Westekemper K, Ehrenreich L, Strenzke N, and Fischer J (2012). Mice do not require auditory input for the normal development of their ultrasonic vocalizations. *BMC Neuroscience* 13, 40. [PubMed: 22533376]
12. Hammerschmidt K, Whelan G, Eichele G, and Fischer J (2015). Mice lacking the cerebral cortex develop normal song: Insights into the foundations of vocal learning. *Scientific Reports* 5, 8808. [PubMed: 25744204]
13. Mahrt EJ, Perkel DJ, Tong L, Rubel EW, and Portfors CV (2013). Engineered Deafness Reveals That Mouse Courtship Vocalizations Do Not Require Auditory Experience. *The Journal of Neuroscience* 33, 5573–5583. [PubMed: 23536072]
14. Blumberg MS (1992). Rodent ultrasonic short calls: Locomotion, biomechanics, and communication. *Journal of Comparative Psychology* 106, 360–365. [PubMed: 1451418]
15. Petkov CI, and Jarvis ED (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. In *Frontiers in evolutionary neuroscience*, Volume 4. p. 12. [PubMed: 22912615]
16. Arriaga G, and Jarvis ED (2013). Mouse vocal communication system: Are ultrasounds learned or innate? *Brain and Language* 124, 96–116. [PubMed: 23295209]
17. Barton RA, and Harvey PH (2000). Mosaic evolution of brain structure in mammals. *Nature* 405, 1055–1058. [PubMed: 10890446]
18. Moore JM, and DeVoogd TJ (2017). Concerted and mosaic evolution of functional modules in songbird brains. *Proceedings of the Royal Society B: Biological Sciences* 284, 20170469.
19. Tchernichovski O, Mitra PP, Lints T, and Nottebohm F (2001). Dynamics of the Vocal Imitation Process: How a Zebra Finch Learns Its Song. *Science* 291, 2564–2569. [PubMed: 11283361]
20. Derégnaucourt S, Saar S, and Gahr M (2009). Dynamics of crowing development in the domestic Japanese quail (*Coturnix coturnix japonica*). *Proceedings of the Royal Society B: Biological Sciences* 276, 2153–2162.
21. Krubitzer LA, and Prescott TJ (2018). The Combinatorial Creature: Cortical Phenotypes within and across Lifetimes. *Trends in Neurosciences* 41, 744–762. [PubMed: 30274608]
22. Ghazanfar AA, and Takahashi DY (2014). The evolution of speech: vision, rhythm, cooperation. *Trends in Cognitive Sciences* 18, 543–553. [PubMed: 25048821]

23. Awbrey FT (1978). Social Interaction among Chorusing Pacific Tree Frogs, *Hyla regilla*. *Copeia* 1978, 208–214.
24. Zelick RD, and Narins PM (1983). Intensity discrimination and the precision of call timing in two species of neotropical treefrogs. *Journal of comparative physiology* 153, 403–412.
25. Grafe TU (1996). The function of call alternation in the African reed frog (*Hyperolius marmoratus*): precise call timing prevents auditory masking. *Behavioral Ecology and Sociobiology* 38, 149–158.
26. Ulanovsky N, and Moss CF (2008). What the bat's voice tells the bat's brain. *Proceedings of the National Academy of Sciences* 105, 8491.
27. Obrist MK (1995). Flexible bat echolocation: the influence of individual, habitat and conspecifics on sonar signal design. *Behavioral Ecology and Sociobiology* 36, 207–219.
28. Ulanovsky N, Fenton MB, Tsoar A, and Korine C (2004). Dynamics of jamming avoidance in echolocating bats. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271, 1467–1475.
29. Heiligenberg W (1977). Principles of electrolocation and jamming avoidance in electric fish: a neuroethological approach, (New York: Springer-Verlag).
30. Bell CC, and Maler L (2005). Central Neuroanatomy of Electrosensory Systems in Fish. In *Electroreception*, Bullock TH, Hopkins CD, Popper AN and Fay RR, eds. (New York, NY: Springer New York), pp. 68–111.
31. Benichov Jonathan I., Benezra Sam E., Vallentin D, Globerson E, Long Michael A., and Tchernichovski O (2016). The Forebrain Song System Mediates Predictive Call Timing in Female and Male Zebra Finches. *Current Biology* 26, 309–318. [PubMed: 26774786]
32. Benichov JI, Globerson E, and Tchernichovski O (2016). Finding the Beat: From Socially Coordinated Vocalizations in Songbirds to Rhythmic Entrainment in Humans. *Frontiers in Human Neuroscience* 10, 255. [PubMed: 27375455]
33. Pika S, Wilkinson R, Kendrick Kobin H, and Vernes Sonja C (2018). Taking turns: bridging the gap between human and animal communication. *Proceedings of the Royal Society B: Biological Sciences* 285, 20180598. [PubMed: 29875303]
34. Campbell P, Pasch B, Warren AL, and Phelps SM (2014). Vocal Ontogeny in Neotropical Singing Mice (*Scotinomys*). *PLOS ONE* 9, e113628. [PubMed: 25469986]
35. Okobi DE, Banerjee A, Matheson AMM, Phelps SM, and Long MA (2019). Motor cortical control of vocal interaction in neotropical singing mice. *Science* 363, 983. [PubMed: 30819963]
36. Elliott HH (1981). Video Analysis of Siamang (*Hylobates Syndactylus*) Songs. *Behaviour* 76, 128–151.
37. Clarke E, Reichard UH, and Zuberbühler K (2006). The Syntax and Meaning of Wild Gibbon Songs. *PLoS ONE* 1, e73. [PubMed: 17183705]
38. Maples EG Jr, Haraway MM, and Hutto CW (1989). Development of coordinated singing in a newly formed siamang pair (*Hylobates syndactylus*). *Zoo Biology* 8, 367–378.
39. Maples EG Jr, Haraway MM, and Collie L (1988). Interactive singing of a male Mueller's gibbon with a simulated neighbor. *Zoo Biology* 7, 115–122.
40. Hall ML (2009). A Review of Vocal Duetting in Birds. In *Advances in the Study of Behavior*, Volume 40, Naguib M, Zuberbuehler K, Clayton NS and Janik VM, eds. (Academic Press), pp. 67–121.
41. Farabaugh S (1982). The ecological and social significance of duetting. In *Acoustic Communication in Birds: Song Learning & Its Consequences*, Volume 2, Kroodsma DE and Miller EH, eds. (New York: Academic Press), pp. 85–124.
42. Kroodsma DE (1984). Songs of the Alder Flycatcher (*Empidonax alnorum*) and Willow Flycatcher (*Empidonax traillii*) Are Innate. *The Auk* 101, 13–24.
43. Kroodsma DE (1989). Male eastern phoebes (*Sayornis phoebe*; Tyrannidae, Passeriformes) fail to imitate songs. *Journal of Comparative Psychology* 103, 227–232.
44. Kroodsma DE, and Konishi M (1991). A suboscine bird (eastern phoebe, *Sayornis phoebe*) develops normal song without auditory feedback. *Animal Behaviour* 42, 477–487.

45. Trainer JM, McDonald DB, and Learn WA (2002). The development of coordinated singing in cooperatively displaying long-tailed manakins. *Behav. Ecol* 13, 65–69.
46. Laje R, and Mindlin GB (2003). Highly Structured Duets in the Song of the South American Hornero. *Physical Review Letters* 91, 258104. [PubMed: 14754163]
47. Diniz P, da Silva Júnior EF, Webster MS, and Macedo RH (2018). Duetting behavior in a Neotropical ovenbird: sexual and seasonal variation and adaptive signaling functions. *Journal of Avian Biology* 49, jav-01637.
48. Tobias JA, and Seddon N (2009). Signal Jamming Mediates Sexual Conflict in a Duetting Bird. *Current Biology* 19, 577–582. [PubMed: 19285404]
49. Levinson SC (2016). Turn-taking in Human Communication – Origins and Implications for Language Processing. *Trends in Cognitive Sciences* 20, 6–14. [PubMed: 26651245]
50. Chow CP, Mitchell JF, and Miller CT (2015). Vocal turn-taking in a non-human primate is learned during ontogeny. *Proceedings of the Royal Society B: Biological Sciences* 282, 20150069.
51. Fries W, and Swihart AA (1990). Disturbance of rhythm sense following right hemisphere damage. *Neuropsychologia* 28, 1317–1323. [PubMed: 2280838]
52. Mavlov L (1980). Amusia Due to Rhythm Agnosia in a Musician with Left Hemisphere Damage: A Non-Auditory Supramodal Defect. *Cortex* 16, 331–338. [PubMed: 7471774]
53. Bouchard KE, Mesgarani N, Johnson K, and Chang EF (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495, 327. [PubMed: 23426266]
54. Dichter BK, Breshears JD, Leonard MK, and Chang EF (2018). The Control of Vocal Pitch in Human Laryngeal Motor Cortex. *Cell* 174, 21–31.e29. [PubMed: 29958109]
55. Kotz SA, and Schwartze M (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in Cognitive Sciences* 14, 392–399. [PubMed: 20655802]
56. Kotz SAE, and Schwartze M (2011). Differential input of the supplementary motor area to a dedicated temporal processing network: functional and clinical implications. *Frontiers in Integrative Neuroscience* 5, 86. [PubMed: 22363269]
57. Leonardo A, and Fee MS (2005). Ensemble Coding of Vocal Control in Birdsong. *J. Neurosci* 25, 652–661. [PubMed: 15659602]
58. Nottebohm F, Paton JA, and Kelley DB (1982). Connections of vocal control nuclei in the canary telencephalon. *The Journal of Comparative Neurology* 207, 344–357. [PubMed: 7119147]
59. Hahnloser RHR, Kozhevnikov AA, and Fee MS (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419, 65–70. [PubMed: 12214232]
60. Long MA, and Fee MS (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* 456, 189–194. [PubMed: 19005546]
61. Long Michael A., Katlowitz Kalman A., Svirsky Mario A., Clary Rachel C., Byun Tara M., Majaj N, Oya H, Howard Matthew A. III, and Greenlee Jeremy D.W. (2016). Functional Segregation of Cortical Regions Underlying Speech Timing and Articulation. *Neuron* 89, 1187–1193. [PubMed: 26924439]
62. Fortune ES, Rodríguez C, Li D, Ball GF, and Coleman MJ (2011). Neural Mechanisms for the Coordination of Duet Singing in Wrens. *Science* 334, 666. [PubMed: 22053048]
63. Hoffmann S, Trost L, Voigt C, Leitner S, Lemazina A, Sagunsky H, Abels M, Kollmansperger S, Maat AT, and Gahr M (2019). Duets recorded in the wild reveal that interindividually coordinated motor control enables cooperative behavior. *Nature Communications* 10, 2577.
64. Doupe AJ, and Kuhl PK (1999). Birdsong and Human Speech: Common Themes and Mechanisms. *Annual Review of Neuroscience* 22, 567–631.
65. Tramontin AD, and Brenowitz EA (2000). Seasonal plasticity in the adult brain. *Trends in Neurosciences* 23, 251–258. [PubMed: 10838594]
66. Konishi M (1965). The role of auditory feedback in the control of vocalization in the White-crowned Sparrow. *Zeitschrift für Tierpsychologie* 22, 770–783. [PubMed: 5874921]
67. Fehér O, Wang H, Saar S, Mitra PP, and Tchernichovski O (2009). *De novo* establishment of wild-type song culture in the zebra finch. *Nature* 459, 564–568. [PubMed: 19412161]

68. Gadagkar V, Puzerey PA, Chen R, Baird-Daniel E, Farhang AR, and Goldberg JH (2016). Dopamine neurons encode performance error in singing birds. *Science* 354, 1278. [PubMed: 27940871]
69. Margoliash D (2002). Evaluating theories of bird song learning: implications for future directions. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* 188, 851–866. [PubMed: 12471486]
70. Kroodsma DE (1982). Learning and the ontogeny of sound signals in birds. In *Acoustic communication in birds*, Volume 2, Kroodsma DE, Miller EH and Ouellet H, eds. (New York: Academic Press), pp. 1–23.
71. Egnor SER, and Hauser MD (2004). A paradox in the evolution of primate vocal learning. *Trends in Neurosciences* 27, 649–654. [PubMed: 15474164]
72. Elowson AM, Charles S, and Cristina L-P (1998). Infant ‘Babbling’ in a Nonhuman Primate: Complex Vocal Sequences with Repeated Call Types. *Behaviour* 135, 643–664.
73. Pistorio AL, Vintch B, and Wang X (2006). Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). *The Journal of the Acoustical Society of America* 120, 1655–1670. [PubMed: 17004487]
74. Gultekin YB, and Hage SR (2017). Limiting parental feedback disrupts vocal development in marmoset monkeys. *Nature Communications* 8, 14046.
75. David JTH (1988). Supply and Demand in Tree Swallow Broods: A Model of Parent-Offspring Food-Provisioning Interactions in Birds. *The American Naturalist* 131, 175–202.
76. Price K, Harvey H, and Ydenberg RON (1996). Begging tactics of nestling yellow-headed blackbirds, *Xanthocephalus xanthocephalus*, in relation to need. *Animal Behaviour* 51, 421–435.
77. Tuero DT, Gloag R, and Reboreda JC (2015). Nest environment modulates begging behavior of a generalist brood parasite. *Behavioral Ecology* 27, 204–210.
78. Ligout S, Dentressangle F, Mathevon N, and Vignal C (2016). Not for Parents Only: Begging Calls Allow Nest-Mate Discrimination in Juvenile Zebra Finches. *Ethology* 122, 193–206.
79. Stamps J, Clark A, Arrowood P, and Kus B (1989). Begging Behavior in Budgerigars. *Ethology* 81, 177–192.
80. Davies NB, Kilner RM, and Noble DG (1998). Nestling cuckoos, *Cuculus canorus*, exploit hosts with begging calls that mimic a brood. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265, 673–678.
81. Mondloch CJ (1995). Chick hunger and begging affect parental allocation of feedings in pigeons. *Animal Behaviour* 49, 601–613.
82. Weary DM, and Fraser D (1995). Calling by domestic piglets: reliable signals of need? *Animal Behaviour* 50, 1047–1055.
83. Iacobucci P, Colonnello V, D’Antuono L, Cloutier S, and Newberry RC (2015). Piglets call for maternal attention: Vocal behaviour in *Sus scrofa domestica* is modulated by mother’s proximity. *Applied Animal Behaviour Science* 171, 88–93.
84. Manser MB, and Avey G (2000). The effect of pup vocalisations on food allocation in a cooperative mammal, the meerkat (*Suricata suricatta*). *Behavioral Ecology and Sociobiology* 48, 429–437.
85. Liu W. c., Wada K, and Nottebohm F (2009). Variable Food Begging Calls Are Harbingers of Vocal Learning. *PLoS ONE* 4, e5929. [PubMed: 19529766]
86. Forlano PM, Sisneros JA, Rohmann KN, and Bass AH (2015). Neuroendocrine control of seasonal plasticity in the auditory and vocal systems of fish. *Frontiers in Neuroendocrinology* 37, 129–145. [PubMed: 25168757]
87. Tschida K, and Mooney R (2012). The role of auditory feedback in vocal learning and maintenance. *Current Opinion in Neurobiology* 22, 320–327. [PubMed: 22137567]
88. Wu HG, Miyamoto YR, Castro LNG, Olveczky BP, and Smith MA (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nat Neurosci* 17, 312–321. [PubMed: 24413700]
89. Ölveczky BP, Andalman AS, and Fee MS (2005). Vocal Experimentation in the Juvenile Songbird Requires a Basal Ganglia Circuit. *PLoS Biol* 3, e153. [PubMed: 15826219]

90. Kao MH, Doupe AJ, and Brainard MS (2005). Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* 433, 638–643. [PubMed: 15703748]
91. Andalman AS, and Fee MS (2009). A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proceedings of the National Academy of Sciences* 106, 12518–12523.
92. Warren TL, Tumer EC, Charlesworth JD, and Brainard MS (2011). Mechanisms and time course of vocal learning and consolidation in the adult songbird. *Journal of Neurophysiology* 106, 1806–1821. [PubMed: 21734110]
93. Wise SP, Murray EA, and Gerfen CR (1996). The Frontal Cortex-Basal Ganglia System in Primates. *10*, 317–356.
94. Kreitzer AC, and Malenka RC (2008). Striatal Plasticity and Basal Ganglia Circuit Function. *Neuron* 60, 543–554. [PubMed: 19038213]
95. Dhawale AK, Smith MA, and Ölveczky BP (2017). The Role of Variability in Motor Learning. *Annual Review of Neuroscience* 40, 479–498.
96. Fant G (1960). *Acoustic theory of speech production*, (The Hague: Mouton).
97. Taylor AM, and Reby D (2010). The contribution of source-filter theory to mammal vocal communication research. *Journal of Zoology* 280, 221–236.
98. Ames PL (1971). *The morphology of the syrinx in passerine birds*, Volume 37, (Peabody Museum of Natural History, Yale University New Haven, CT).
99. Warner RW (1972). The anatomy of the syrinx in passerine birds. *Journal of Zoology* 168, 381–393.
100. Suthers RA (1990). Contributions to birdsong from the left and right sides of the intact syrinx. *Nature* 347, 473–477.
101. Zollinger SA, and Suthers RA (2004). Motor mechanisms of a vocal mimic: implications for birdsong production. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271, 483–491.
102. Uchida AM, Meyers RA, Cooper BG, and Goller F (2010). Fibre architecture and song activation rates of syringeal muscles are not lateralized in the European starling. *J Exp Biol* 213, 1069–1078. [PubMed: 20228343]
103. Elemans CPH, Mead AF, Jakobsen L, and Ratcliffe JM (2011). Superfast Muscles Set Maximum Call Rate in Echolocating Bats. *Science* 333, 1885. [PubMed: 21960635]
104. Fitch WT, Neubauer J, and Herzog H (2002). Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour* 63, 407–418.
105. Riede T, Owren MJ, and Arcadi AC (2004). Nonlinear acoustics in pant hoots of common chimpanzees (*Pan troglodytes*): Frequency jumps, subharmonics, biphonation, and deterministic chaos. *American Journal of Primatology* 64, 277–291. [PubMed: 15538766]
106. Lieberman P (1984). *The biology and evolution of language*, (Cambridge, MA: Harvard University Press).
107. Fitch WT, de Boer B, Mathur N, and Ghazanfar AA (2016). Monkey vocal tracts are speech-ready. *Science Advances* 2, e1600723. [PubMed: 27957536]
108. Sober SJ, Wohlgenuth MJ, and Brainard MS (2008). Central Contributions to Acoustic Variation in Birdsong. *The Journal of Neuroscience* 28, 10370. [PubMed: 18842896]
109. Simonyan K, and Horwitz B (2011). Laryngeal Motor Cortex and Control of Speech in Humans. *The Neuroscientist* 17, 197–208. [PubMed: 21362688]
110. Sober SJ, and Brainard MS (2012). Vocal learning is constrained by the statistics of sensorimotor experience. *Proceedings of the National Academy of Sciences* 109, 21099.
111. Osmanski MS, and Dooling RJ (2009). The effect of altered auditory feedback on control of vocal production in budgerigars (*Melopsittacus undulatus*). *The Journal of the Acoustical Society of America* 126, 911–919. [PubMed: 19640055]
112. Moore PWB, and Pawloski DA (1990). Investigations on the Control of Echolocation Pulses in the Dolphin (*Tursiops truncatus*). In *Sensory Abilities of Cetaceans: Laboratory and Field Evidence*, Thomas JA and Kastelein RA, eds. (Boston, MA: Springer US), pp. 305–316.
113. Stansbury AL, and Janik VM (2019). Formant Modification through Vocal Production Learning in Gray Seals. *Current Biology* 29, 2244–2249.e2244. [PubMed: 31231051]

114. Prat Y, Azoulay L, Dor R, and Yovel Y (2017). Crowd vocal learning induces vocal dialects in bats: Playback of conspecifics shapes fundamental frequency usage by pups. *PLOS Biology* 15, e2002556. [PubMed: 29088225]
115. Genzel D, Desai J, Paras E, and Yartsev MM (2019). Long-term and persistent vocal plasticity in adult bats. *Nature Communications* 10, 3372.
116. Zhao L, Rad Bahar B, and Wang X (2019). Long-lasting vocal plasticity in adult marmoset monkeys. *Proceedings of the Royal Society B: Biological Sciences* 286, 20190817. [PubMed: 31238841]
117. Hotchkin CF, Parks SE, and Weiss DJ (2015). Noise-Induced Frequency Modifications of Tamarin Vocalizations: Implications for Noise Compensation in Nonhuman Primates. *PLOS ONE* 10, e0130211. [PubMed: 26107515]
118. Pfenning AR, Hara E, Whitney O, Rivas MV, Wang R, Roulhac PL, Howard JT, Wirthlin M, Lovell PV, Ganapathy G, et al. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* 346, 1333.
119. Briscoe SD, and Ragsdale CW (2018). Homology, neocortex, and the evolution of developmental mechanisms. *Science* 362, 190. [PubMed: 30309947]
120. Patterson DK, and Pepperberg IM (1994). A comparative study of human and parrot phonation: Acoustic and articulatory correlates of vowels. *The Journal of the Acoustical Society of America* 96, 634–648. [PubMed: 7930064]
121. Ohms VR, Beckers GJL, ten Cate C, and Suthers RA (2012). Vocal tract articulation revisited: the case of the monk parakeet. *The Journal of Experimental Biology* 215, 85. [PubMed: 22162856]
122. Chakraborty M, Walløe S, Nedergaard S, Fridel EE, Dabelsteen T, Pakkenberg B, Bertelsen MF, Dorrestein GM, Brauth SE, Durand SE, et al. (2015). Core and Shell Song Systems Unique to the Parrot Brain. *PLoS ONE* 10, e0118496. [PubMed: 26107173]
123. Jürgens U, and Alipour M (2002). A comparative study on the cortico-hypoglossal connections in primates, using biotin dextranamine. *Neuroscience Letters* 328, 245–248. [PubMed: 12147317]
124. Yovel Y, Geva-Sagiv M, and Ulanovsky N (2011). Click-based echolocation in bats: not so primitive after all. *Journal of Comparative Physiology A* 197, 515–530.
125. Lee W-J, Falk B, Chiu C, Krishnan A, Arbour JH, and Moss CF (2017). Tongue-driven sonar beam steering by a lingual-echolocating fruit bat. *PLOS Biology* 15, e2003148. [PubMed: 29244805]
126. Jürgens U (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews* 26, 235–258. [PubMed: 11856561]
127. Vicario DS, and Nottebohm F (1988). Organization of the zebra finch song control system: I. Representation of syringeal muscles in the hypoglossal nucleus. *The Journal of Comparative Neurology* 271, 346–354. [PubMed: 3385013]
128. Vicario DS (1991). Organization of the Zebra Finch Song Control System: II. Functional Organization of Outputs From Nucleus *Robustus Archistriatalis*. *The Journal of Comparative Neurology* 309, 486–494. [PubMed: 1655832]
129. Buenostro JD, Wu B, Chang HY, and Greenleaf WJ (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current protocols in molecular biology* 109, 21.29.21–29.
130. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. *Nature Biotechnology* 28, 1045.
131. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD, et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515, 355–364. [PubMed: 25409824]
132. Zhu Y, Sousa AMM, Gao T, Skarica M, Li M, Santpere G, Esteller-Cucala P, Juan D, Ferrández-Peral L, Gulden FO, et al. (2018). Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* 362, eaat8077.
133. Kowalczyk A, Meyer WK, Partha R, Mao W, Clark NL, and Chikina M (2019). RERconverge: an R package for associating evolutionary rates with convergent traits. *Bioinformatics*.

134. Berger MJ, Wenger AM, Guturu H, and Bejerano G (2018). Independent erosion of conserved transcription factor binding sites points to shared hindlimb, vision and external testes loss in different mammals. *Nucleic Acids Research* 46, 9299–9308. [PubMed: 30137416]
135. Takahashi DY, Liao DA, and Ghazanfar AA (2017). Vocal Learning via Social Reinforcement by Infant Marmoset Monkeys. *Current Biology* 27, 1844–1852.e1846. [PubMed: 28552359]
136. Watson Stuart K., Townsend Simon W., Schel Anne M., Wilke C, Wallace Emma K., Cheng L, West V, and Slocombe Katie E. (2015). Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees. *Current Biology* 25, 495–499. [PubMed: 25660548]
137. Fischer J, Wheeler BC, and Higham JP (2015). Is there any evidence for vocal learning in chimpanzee food calls? *Current Biology* 25, R1028–R1029. [PubMed: 26528740]
138. Miller CT, and Thomas AW (2012). Individual recognition during bouts of antiphonal calling in common marmosets. *Journal of Comparative Physiology A* 198, 337–346.
139. Fee MS, and Goldberg JH (2011). A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience* 198, 152–170. [PubMed: 22015923]
140. Riede T, Tokuda IT, and Farmer CG (2011). Subglottal pressure and fundamental frequency control in contact calls of juvenile *Alligator mississippiensis*. *The Journal of Experimental Biology* 214, 3082. [PubMed: 21865521]
141. Hauser MD, Evans CS, and Marler P (1993). The role of articulation in the production of rhesus monkey, *Macaca mulatta*, vocalizations. *Animal Behaviour* 45, 423–433.
142. Ball G, Sockman K, Duffy D, and Gentner T (2006). A neuroethological approach to song behavior and perception in European starlings: interrelationships among testosterone, neuroanatomy, immediate early gene expression, and immune function. *Advances in the Study of Behavior* 36, 59–122.

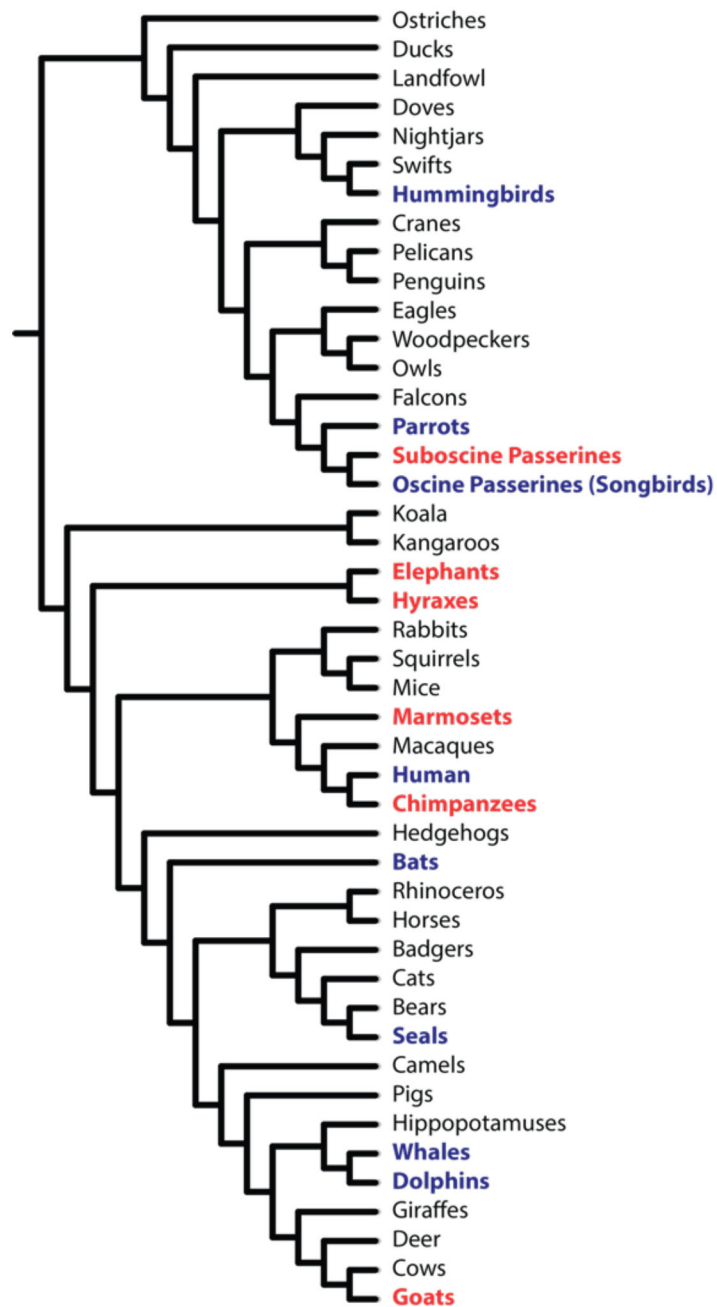


Figure 1. A phylogeny of vocal learners.

As vocalizations are thought to be innate in vocalizing fish, amphibians, and reptiles, most studies of vocal learning focus on birds and mammals. The canonical set of taxa widely cited as being vocal production learners [2–4] is fairly restricted (in blue). However, some limited evidence has led to claims that a much broader range of taxa may possess some capacity for vocal learning (in red) [6–9, 135, 136]. In many cases these examples rely on highly anecdotal examples, in others the evidence has been contested [137]. The remaining taxa are largely assumed to be vocal non-learners, although this has been conclusively demonstrated

(classically, through deafening, hand-rearing, or cross-fostering experiments) in remarkably few taxa.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

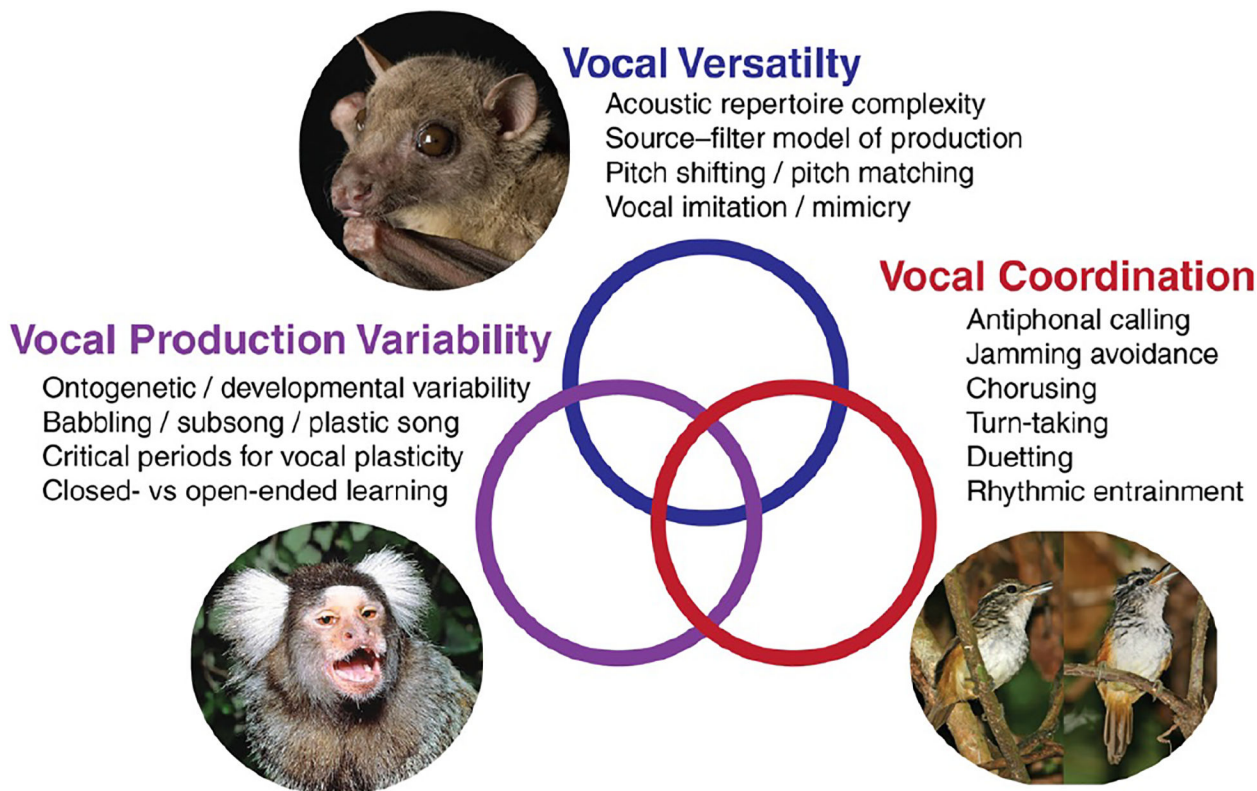


Figure 2. Vocal learning as the intersection of several component sub-traits.

Vocal production learning has traditionally been viewed, explicitly or implicitly, as a binary trait that is either present or absent. More recently it has been proposed to be a continuum, with species increasing linearly in ability from ‘low’ to ‘high’ vocal learners [16]. An alternate view, presented here, suggests that vocal learning may be a multidimensional trait, in which species exhibit varying capacities for learning across a set of interrelated behavioral trait modules, which do not necessarily vary co-linearly. This view enables a neuroethological approach, in which species that excel in a particular module could represent ideal model to disentangle particular aspects of the vocal learning phenotype in isolation. As a starting point, we present three example modules, listed here with (in black) related concepts from the broader vocal learning literature. Species discussed in relation to each module in this review include, for vocal production variability, common marmoset (*Callithrix jacchus*); for vocal coordination, warbling antbirds (*Hypocnemis sp.*); and for vocal versatility, Egyptian fruit bat (*Rousettus aegypticus*).

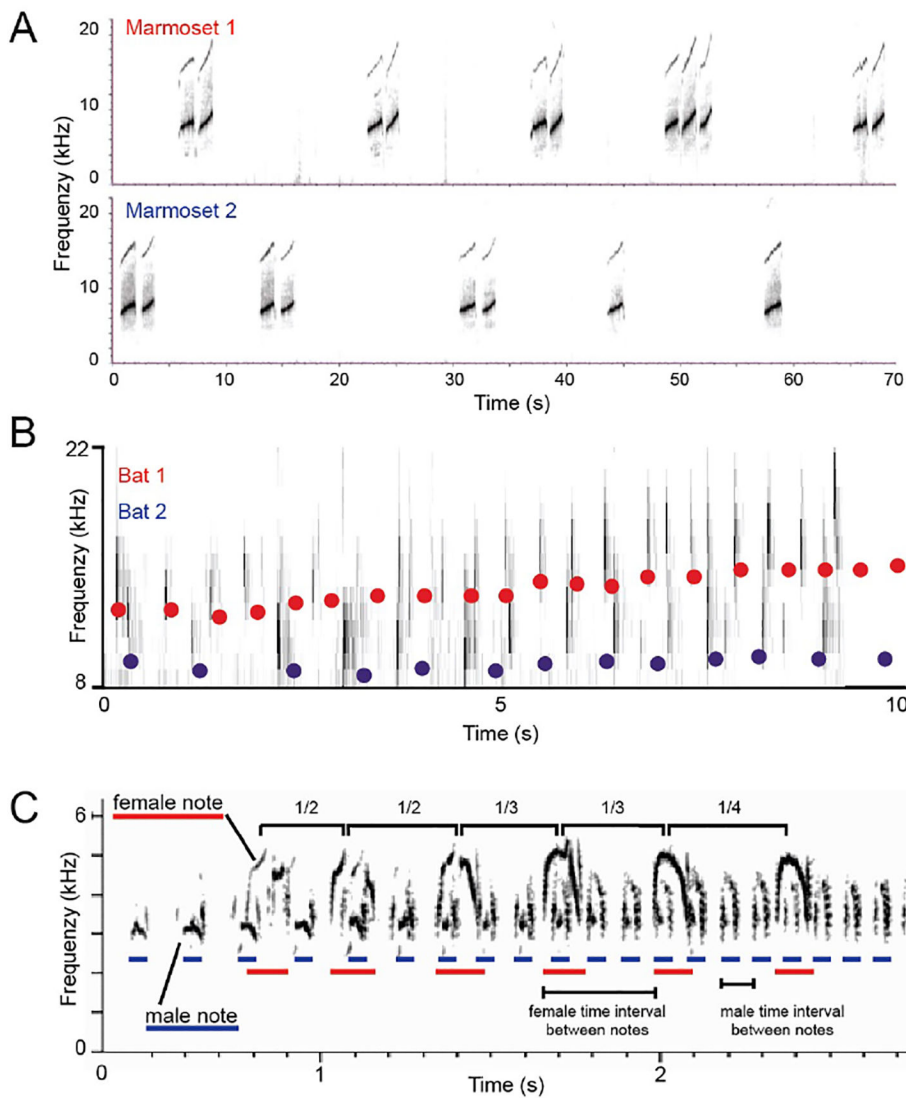


Figure 3. Vocal coordination.

As with any social signal, vocal communication signals must be appropriately timed and coordinated in order to produce their intended effect on their recipient(s). (A) In marmoset antiphonal calling (adapted from [138]), adults engage in ongoing bouts of rhythmically coordinated calling, a phenomenon known as turn-taking that is also exhibited in human speech. (B) Bats use the acoustic reflections of echolocation calls to navigate and detect prey. While flying and foraging in a group context, the echolocation of conspecifics then threatens to “jam” the signal of a caller, a possibility that much be avoided. One of several solutions to this problem, exhibited by the European free-tailed bat (*Tadarida teniotis*), is for individuals to dynamically shift the acoustic frequency of their calls to avoid overlap (blue and red dots indicate the minimum frequency of echolocation calls of two bats, adapted from [28]). (C) During vocal duetting, conspecifics tightly coordinate their vocal production as means of defending territories or maintain social bonds, among other possibilities. In some cases, as in the Rufous hornero (*Furnarius rufus*), these duets are refined over time between

partners to match a highly-structured pattern with subsecond temporal precision (adapted from [46]).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

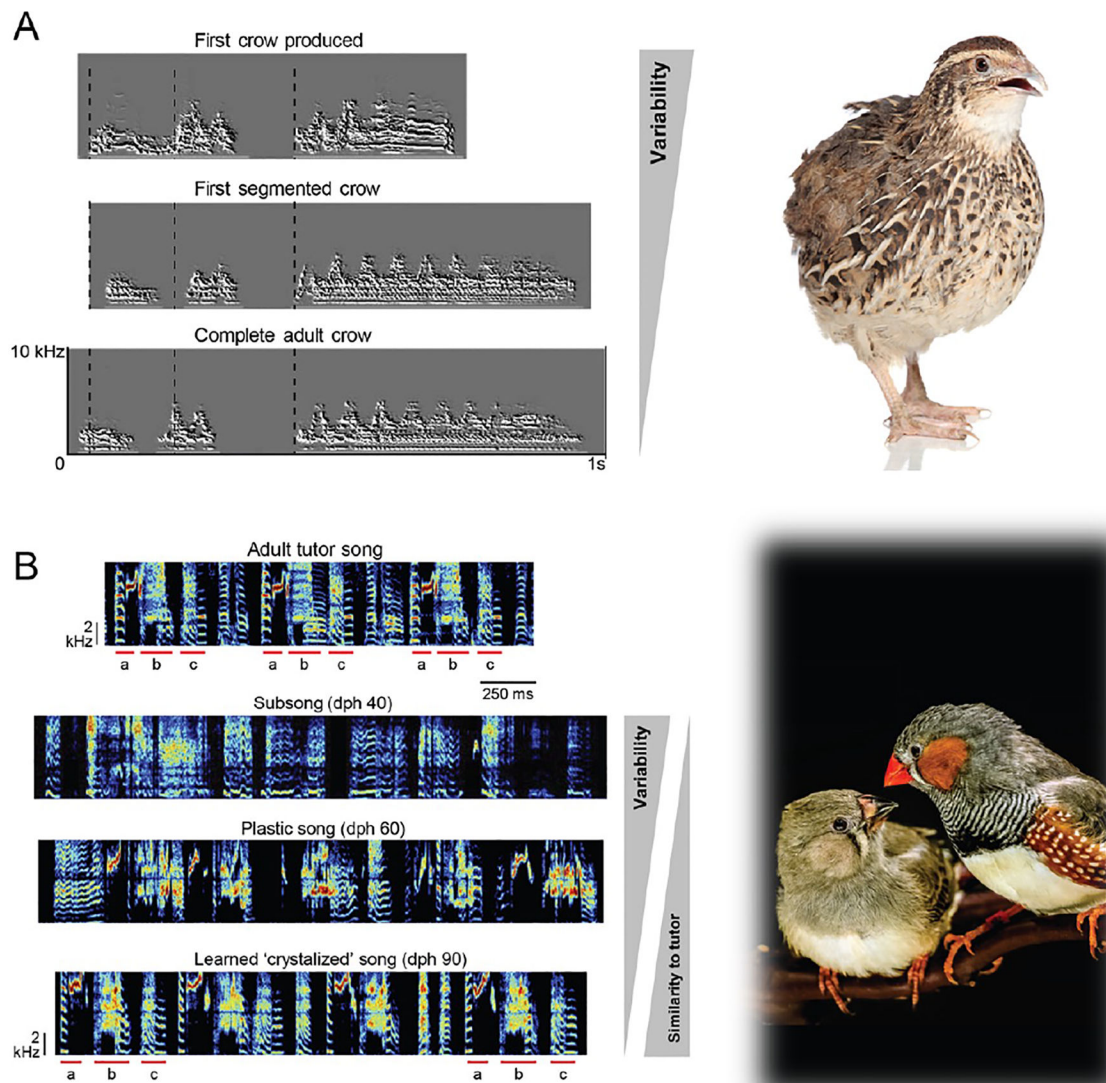


Figure 4. Vocal production variability & feedback.

During the plastic period of vocal development, diverse sources contribute to variability in vocal output, and the neural motor signal that underlies it. This exploratory variability is dynamically regulated to facilitate motor learning. (A) Juvenile Japanese quail exhibit an initial period of vocal variability that wanes over the course of the development of their adult crow [20]. (B) Juvenile zebra finches similarly experience an initial period of high vocal variability (subsong), which gradually (via plastic song) crystallizes into the adult song type. Uniquely in the finch relative to the quail, this decrease in vocal variability coincides with increasing similarity to the song of a tutor it was exposed to during this critical period [139]. Vocal learning is then in some respects the process by which internally guided developmental processes (as in the quail) are exposed to external influences (as in the finch).

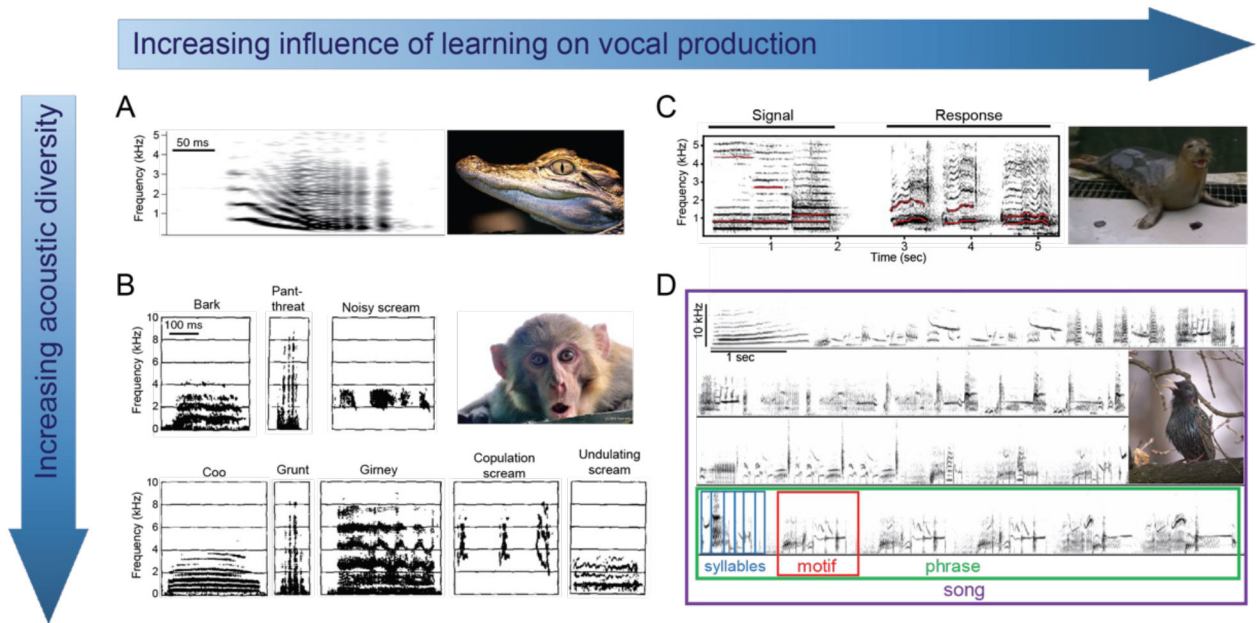


Figure 5. Vocal versatility.

A fundamental concept in the study of animal vocalizations is the complexity of its vocal repertoire. In some cases, as in vocalizing fish and crocodilians (A, adapted from [140]), a species may present a very small set of calls used under a limited set of circumstances. Other species may present a more diverse acoustic repertoire with a range of categorically discrete call types, as in the case of the macaque monkey (B, adapted from [141]). Expanding the complexity of the vocal repertoire can be accomplished through various means, including increasing the dynamic motility of the vocal apparatus and taking advantage of nonlinear vocal phenomena. Vocal repertoires may also be made more versatile through the influence of learning. At its most basic, this may involve small modifications to pre-existing vocal signals to match a template, as in the case of a harbor seal trained to imitate a sequence of human vowel formants (C, adapted from [113]). In its most advanced instantiation, species may present the ability to mimic other species, novel sounds, and complex series of syllables, motifs, and phrases, as in the common starling (D, adapted from [142]).