



HHS Public Access

Author manuscript

Trends Neurosci. Author manuscript; available in PMC 2021 February 01.

Published in final edited form as:

Trends Neurosci. 2020 February ; 43(2): 115–126. doi:10.1016/j.tins.2019.12.006.

A Hierarchy of Autonomous Systems for Vocal Production

Yisi S. Zhang^{1,*}, Asif A. Ghazanfar^{1,2,3,*}

¹Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA

²Department of Psychology, Princeton University, Princeton, NJ 08544, USA

³Department of Ecology & Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

Abstract

Vocal production is hierarchical in the time domain. These hierarchies build upon biomechanical and neural dynamics across various timescales. We review studies in marmoset monkeys, songbirds and other vertebrates. To organize these data in an accessible and across-species framework, we interpret the different timescales of vocal production as belonging to different levels of an autonomous systems hierarchy. The first level accounts for vocal acoustics produced on short timescales; subsequent levels account for longer timescales of vocal output. The hierarchy of autonomous systems that we put forth accounts for vocal patterning, sequence generation, dyadic interactions and context-dependence by sequentially incorporating central pattern generators, intrinsic drives and sensory signals from the environment. We then show the framework's utility by providing an integrative explanation of infant vocal production learning in which social feedback modulates infant vocal acoustics through the tuning of a drive signal.

Keywords

vocalizations; marmoset monkey; birdsong; dynamical system; timescales; biomechanics; parental feedback; social reinforcement; Mayer wave; slow oscillations

A simplified framework for vocal production

Understanding the biology of vocal production—how all the mechanistic pieces are coordinated, and how this coordination is achieved over the course of development—is a formidable challenge. Here, we review what is known about the biomechanical and neural mechanisms of nonhuman animal vocal production. We focus on nonhuman species because we know so much more about their biology and development than we do for human speech production. Moreover, at the level of the central pattern generators (CPGs) and some

* Correspondence: yz9@princeton.edu, asifg@princeton.edu.

Authors' Contributions

Y.S.Z. and A.A.G wrote the manuscript.

Competing Interests

We have no competing interests.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

forebrain pathways, many circuits appear homologous or at the very least analogous across species [1, 2]. It will be apparent even in just the subset of studies that we describe that there is a problem of increasing complexity: As we learn more and more about the details of neural circuits and biomechanics in different species finding a common ground is difficult. We therefore put forth what we think is a plausible framework for the integrative biology of vocal production across species: a hierarchical set of autonomous systems with feedback [3]. In a system like this, high-level states are slower and provide the scaffolding for the fast, lower-level states [3, 4]. This approach has, for example, been used to model the fast acoustic features of speech as the result of the slower articulator movement [5, 6].

Our hope is that this framework will provide an integrative perspective on adaptive vocal production. It emphasizes that, first, vocal production should not be viewed as a one-to-one mapping from single-neuron activity to sound. The neural activity at each level of this system is under the influence of its upper level and in turn, determines (or at least strongly influences) how downstream neural activities unfold. The same set of neurons can be involved in various types of vocal output, and the aggregate activity of a large neural population across different structures is more relevant to the outcomes of the vocal production than any single neuron or small set of neurons. Second, the framework recognizes that vocal production is a consequence of the interplay between internal states and behavioral context [7]. The interactions between context, the internal state and the vocal output have often been invoked to explain vocal acoustic structure or even vocalization types, but much of the research is based largely on speculation as to what the animal's internal states may be and how it could influence vocal output [8]. Any context is comprised of multiple types of external sensory cues interacting with fluctuating internal states (which generate visceral cues) [9,10].

The scope of the problem

Vocal production is intrinsically hierarchical in the time domain. Human conversation, for example, consists of a sequence of vocal exchanges between two individuals; the duration of the conversation can vary greatly but must be composed of at least one utterance from each participant. These utterances are themselves sequences but of words. Words are, in turn, composed of consonant-vowel sounds. A similar temporal hierarchical structure is found in the vocal output of other animals, and these utterances can also be exchanged systematically between individuals. In many species of birds, mated pairs sing duets [11]; e.g., both sexes of the black-bellied wren (*Pheugopedius fasciatoventris*) initiate song production and both answer their mates' songs to form duets [12]. Among other vertebrates, there is a similar duetting among mated pairs of both Old and New World nonhuman primates [13, 14]. In frogs [15] and rodents [16], there are structured vocal exchanges between males during territorial defense. Beyond mating and territorial defense, animals also exhibit vocal turn-taking as affiliative gestures, a form of "grooming-at-a-distance" (marmoset monkeys: [17, 18]; macaque monkeys: [19]; lemurs: [20]; meerkats: [21]), much like humans seem to do [22]. Each individual's vocalizations often have multiple elements and these can be divided into smaller units based on their sequential structure and duration (e.g., motifs and syllables) (Figure 1A-C).

The neural correlates of vocal production also have a seemingly hierarchical structure. Across vertebrates, vocal production requires a source of air power and a sound-producing organ (e.g., the larynx, syrinx or swim bladder) [1, 23], and the 1st-order innervation of these peripheral structures are thought to arise from a homologous set of brainstem structures [24]. In primates and other mammals, the neural circuitry related to respiratory power includes groups of neurons within the pons and medulla that generate rhythmic patterns [25-27]; the circuitry influencing laryngeal tension arises from the nucleus ambiguus located in the medulla [27, 28]. Homologous circuitry is present in songbirds [29]. Figures 1D-F show only a portion of the neural pathways for vocal production in monkeys, birds and fish. Some nodes in these vocal-motor networks are tightly-correlated with the temporal patterns and frequency modulations of the vocalizations they produce; they are thus considered vocal central pattern generators (CPGs) [30-32]. The periaqueductal gray (PAG) in the midbrain provides descending control over these CPGs, and its activity influences the production of specific vocalizations [29, 33-36]. The PAG receives inputs from higher-order motor structures in the forebrain that are critical for the initiation of vocalizations and the integration of sensory cues (such as the vocalizations of conspecifics) with internal states (e.g., arousal levels). In the human lineage, this forebrain circuitry, and its relationship to brainstem structures have been elaborated upon considerably over the course of evolution [2, 37, 38].

Vocal production from a two-level hierarchical system

The minimal system that can autonomously generate motor behaviors consists of the appropriate effectors and self-generated neural activity that drives the movement of those biomechanics. Rhythmic motions, such as breathing, swallowing, walking, and swimming, are mainly driven by CPGs located within the spinal cord [39, 40] and the brainstem [41, 42]. The CPGs are intrinsically capable of producing rhythmic signals that directly influence the biomechanics of effectors by alternating the activation of flexor and extensor muscles. Vocal production is also driven by the coordinated activity of CPGs for multiple effectors (e.g., the diaphragm and larynx [30-32], and the result is the production of sound—or sound sequences—that often have a rhythmic structure [43, 44]. These vocalizations exhibit some stable features in the temporal and/or spectral domains and can be organized into distinct call types.

In primates and other mammals, the vocal apparatus consists of the lungs, the vocal folds of the larynx, and the vocal tract; the vocal tract consists of the oral and nasal cavities anterior to the larynx and whose shapes can be modified through articulations of the mouth and lips [23, 45]. Sound production is through self-sustaining vocal fold oscillations induced by increases in subglottal air pressure and vocal fold tension [46-49]. A simple mechanism to vary fundamental frequencies, as demonstrated in marmoset monkeys [50, 51] and zebra finches [52], is through manipulating two parameters, the air pressure and the vocal fold tension. As such, the spectral and temporal patterns emerge from the coordinated activity between respiratory and laryngeal (or syringeal for birds) muscles [50, 52, 53]. When coordinated by their respective CPGs, they define the spectral properties of the sound and provide the temporal structure for vocal output [49-51]. Here's a specific example: In a vocal fish, the plainfin midshipman, pacemaker neurons in the hindbrain control the

frequency of vocalizations, while “prepacemaker” neurons that innervate them control the overall duration of the vocalization [54].

The emergent acoustic patterns, however, do not always slavishly reflect the output patterns of the CPGs; they are a joint consequence of the CPG patterns and the nonlinear dynamics of the tissue and biomechanical properties of the vocal apparatus. Typically, the production of different vocalizations is thought to occur through the differential assembly of neuronal populations that make up the vocal CPGs; i.e., it has been reasonably argued that there is a dedicated CPG assembly for each vocalization type [32]. Work in the domain of locomotion suggested, however, that this need not be the case. Distributed and coupled CPGs, instead of dedicated CPGs, yielded a successful model—with neurophysiological support—for switching between discrete modes of locomotion in the salamander: running and swimming [40], where the CPGs governing the limbs and body were coupled oscillators. The CPGs are usually modeled with low-dimensional nonlinear equations, a means to describe rich biologically relevant dynamics without the need to explicitly code the details of pattern formation [55, 56]. Models like this are used to simulate marmoset monkey calls and birdsongs [53, 57-59]. Together, the nonlinearities inherent in both the biomechanics of the vocal apparatus and the driving CPG activity are able to generate a set of acoustically discrete vocalizations with a repeated temporal pattern. They represent what we would refer to as a two-level hierarchical system (Figure 2A).

On some temporal scales, the rhythm of vocalizations seems to be in a similar range across mammals. In humans, regardless of languages and contexts, the amplitude modulation of the speech signal consists of a rhythm that ranges between 3 and 8 Hz and is correlated with production of syllables [60, 61]. The vocalizations and facial expressions of monkeys and apes also have this rhythmic structure [44, 62-66]. For example, in marmoset monkeys, experimental interruptions of their contact calls (by playing back noise) reveal that the seemingly continuous longer elements of this vocalization are actually made up of smaller elements: marmoset vocalizations can only be interrupted at periodic time points [44]. These periodic intervals occur at a frequency of ~ 7 Hz. These data suggest that, just as human speech is built up from the elemental units of syllables, marmoset monkey vocalizations (and likely the vocalizations of other species), are built of multiple sequentially uttered units on the same time scale [44](see [67] for review).

This suggests that humans and other mammals share a mechanistic substrate that produces 3 – 8 Hz vocal rhythms. We speculate that this substrate is at the two-level system we described (Figure 2A) and the set of CPGs involved are likely to be mostly overlapping with those involved in ingestive orofacial rhythms (like licking and chewing) for all mammals. These other orofacial movements are produced with a > 2 Hz rhythm. The CPGs for all orofacial rhythms (including vocalizations) are located in the brainstem, specifically the pons and medulla [31, 32, 68]. These regions contain the cranial nuclei for sensory processing in the head and face and final motor outputs to orofacial muscles. Vocal and ingestive rhythms occur at frequencies faster than the respiratory rhythm and both the slower respiratory rhythm and the faster orofacial rhythms can be linked directly to separable patterns of CPG activity [68]. Yet, neurons within a CPG network can still participate in multiple orofacial behaviors. For example, in macaque monkeys, neural activity in the

nucleus ambiguus of the medulla (which innervates laryngeal muscles) is modulated by vocalization, respiration, and/or swallowing [69, 70]. Likewise, neurons in the medullary ventral respiratory group are modulated not only by respiration but by swallowing and vocalizations as well [70]. Taken together, these data suggest that vocalizations and their rhythmicity are generated by a distributed network of CPGs that also participate in other rhythmic orofacial behaviors.

Complex vocalization from a three-level hierarchical system

Repetitively produced, rhythmic vocalizations driven by the two-level system do not necessarily represent ethologically relevant vocal behavior. A mechanism to produce different vocalizations at different times is needed. To enable that, one needs a third level. Let us consider the simplest situation regarding the timing of vocal output: when the animal is alone and spontaneously producing vocalizations (that is, without an external triggering event such as another individual or a threat). Rats, for example, produce spontaneous ultrasonic vocalization bouts at a rate of about once every 7 or 8 seconds (~ 0.1 Hz) [71]. Adult marmoset monkeys produce contact calls around every 10 s (~ 0.1 Hz) when alone and with no conspecifics responding to them [18, 72]. When in the hearing range of conspecifics, they exchange contact calls and their vocal production becomes anti-phase-locked to each other but with each marmoset maintaining a ~ 0.1 Hz rhythmic output [18]. Finally, the spontaneous speech of humans also has a similar slow temporal structure, whereby an individual will produce utterances of varying lengths roughly every 10 to 15 seconds (~ 0.1 Hz) [73] (though how this relates to turn-taking is not known). The temporal regularity of vocal production on a roughly similar timescale across species suggests that a common, slow oscillatory drive might exist within the mammalian vocal system.

However, what could be the biological basis for the much slower ~ 0.1 Hz rate of vocal production found in rats, marmoset monkeys, and humans? One link could be with the autonomic nervous system. In marmoset monkeys, spontaneous contact calls by adults are correlated with heart rate (a measure of arousal levels) [72]. This rhythmic arousal variation is the product of the “Mayer wave”, an oscillation of the autonomic nervous system (specifically, sympathetic vasomotor tone); the Mayer wave appears to be common to all mammals [74]. The neural activity of the PAG, a brain structure essential for vocal production [75], is modulated by this 0.1 Hz oscillation [76]. This could causally account for the spontaneous production of vocalizations every 10 seconds or so by adult marmoset monkeys, and possibly other mammals [18, 72]. Another possible source (and perhaps not unrelated) is the neocortex. The neocortex produces oscillations at various timescales [77, 78] and this includes infraslow oscillations on the order of 0.1 Hz [79-81]. These very slow neural oscillations are distinct from other neural oscillations (e.g., 1-4 Hz delta rhythm and the much faster gamma rhythm); they travel across the neocortical sheet with a stereotypical spatiotemporal trajectory and are state-dependent [82]. Moreover, there is at least one neurophysiological link between the Mayer wave (as measured by vasomotor tone) and neocortical activity: Infraslow fluctuations of neural activity occur as envelope over fast gamma band (30 – 80 Hz) activity, and this 0.1 Hz neural rhythm causes the 0.1 Hz oscillatory dilations and constrictions of arterioles [83]. This autonomic nervous system-related slow cortical rhythm represents the next level of our proposed hierarchical system

(Figure 2B). It initiates the two-level system to produce vocal output at particular times, in this case, roughly every 10 seconds.

Importantly, the 0.1 Hz rhythm not only initiates vocal production but, in infants at least, it also influences the sequential structure of vocalizations. When separated from adults, newborn infant marmosets spontaneously produce long sequences of both immature- and mature-sounding vocalizations [50, 84]. As is the case for human infants [85] and songbirds [43, 86, 87], the vocal output of marmoset infants is very rhythmic, tightly locked to the respiration rate around 1 Hz [84]. The time-varying spectral structure of infant marmoset vocal sequences also has a rhythm; however, it is at a rate that is an order of magnitude slower. Spectral entropy, a measure of the noisiness of the sound spectrum [50, 88], fluctuates during infant vocal sequences at a 0.1 Hz frequency [84]. This spectral rhythm is in tight coherence with the infant's Mayer oscillation. The different and discrete call types produced in these sequences are likely the result of this arousal-linked 0.1 Hz oscillation driving the CPG oscillators and the subsequent vocal biomechanics through different regimes. Results from a computational model of rhythmic vocal output lend support to this idea [57]. Thus, our three-level hierarchical system can also modulate (in addition to initiating their coordinated activity) the CPGs controlling the vocal apparatus.

Context-dependent vocalization from a four-level hierarchical system

Vocal output is context-dependent. Spontaneous vocalizations described above are a special case of undirected contexts (no conspecifics are present) and thus reflect, in large part, the internal state of the animal. Most of the time, however, vocalizations involve another individual who responds in kind. In human vocal turn-taking, individuals become anti-phase-locked and entrained to each other through their speech [89]. Similar "coupled oscillator" turn-taking behavior is observed in adult marmoset monkeys [18], macaque monkeys [90], meerkats [21], and potentially other species [67, 90-94]. In both human and marmoset monkey turn-taking, vocal exchanges are separated by intervals within a small range, albeit at different timescales [18, 95]. The coupling between individuals is modulated by the exchange of visual, auditory and potentially other sensory signals. The processing of these sensory signals in the brain comprises the fourth level of the hierarchical system proposed here (Figure 2C).

Consider the following scenario: two marmoset monkeys can hear each other but cannot see each other. In this context, the two marmosets exchange contact "phee" calls with anti-phase-locking, and mutually entrain to each other's call timing [18]. Manipulating the amplitude of contact call playback – a simulation of the physical distance of a caller out of sight – revealed that the marmoset vocal responses to nearby conspecifics were delayed and quieter; the opposite was true for far-away conspecifics [17]. A simple model can account for this behavior whereby the drive to vocalize is inhibited by auditory feedback [17, 96]. Of course, varying inter-individual distances is not unique to marmoset monkeys. Zebra finches will adjust the intensity of their courtship songs according to their physical distance from a conspecific female [97]. Even human speakers are tacitly aware that as the distance between themselves and listeners is increased, vocal intensity must also be increased to maintain effective communication [98, 99]. As a result, low amplitude speech signals elicit high

amplitude responses from the listener. Adaptively adjusting speech amplitude is interpreted as a cooperative act requiring a high-level social skill [98, 99]. However, a simple model incorporating auditory feedback can account for the behavior in humans as well as marmoset monkeys [17].

Changing the social distance between two individuals can also result in qualitative acoustic changes—the production of different call types [100]. In one study, an adult marmoset monkey was placed in four different contexts in which the marmoset could be (1) by itself, (2) with a partner at the opposite corner of the room behind a curtain, (3) with a visible partner at the opposite corner of the room, and (4) with a partner close by. As the distance between two individuals decreased, marmoset monkeys switched from producing loud contact calls to producing softer, shorter and noisier vocalizations [100]. Decreasing the distance between two individuals increased the robustness of auditory and visual signals from the conspecific. Together, the effect seems to be to increase the inhibition on the drive to produce vocalizations, which in turn reduces the overall time spent on calling and reduces the power to produce vocalizations. This is similar to what is observed for the spontaneous vocalizations of infants—the level of the drive changes what vocalizations are produced. The presence of increasing sensory cues coming from conspecifics seems to gradually diminish the dependence on the internal state fluctuation (i.e., the 0.1 Hz Mayer wave). This is supported by the decreasing coherence between vocal output and heart rate as a function of decreasing social distance [100]. In real life, the environment, including the proximity of conspecifics, slowly changes and continuously modulates the dynamics of the drive to vocalize. Relevant variables such as social distance represented by multiple sensory cues constitute the fourth (and final) level of the hierarchical set of autonomous systems (Figure 2C).

Using the autonomous systems framework to understand vocal development

How do infants, capable from postnatal day one of producing spontaneous vocalizations, develop adaptive vocal behavior? Sensory input from the social environment influences the structure of vocalizations very early in life. In humans, the acoustic structure of newborn crying reflects the acoustic structure of the ambient language environment [101]. In fairy wrens, the acoustic structure of parental nesting calls is learned prenatally by their offspring and partially reproduced postnatally in the offsprings' begging calls [102]. In marmoset monkeys, it is suggested that sensory experiences as early as in the womb contribute to shaping the temporal structure of the vocal output in early postnatal life [84]. The early vocal sequences of infant marmoset monkeys (within the first postnatal week), as well as their heart rate profiles, exhibited greater similarities in their sequential structure between dizygotic twins than with their non-twin siblings and non-relatives [84]. The similar autonomic modulation of the twins is likely through the interactions with their mothers both pre- and post-natally. In humans, changes in maternal arousal levels can influence offspring both in the womb and through physical contact postnatally. For example, the cardiorespiratory dynamics of human infants will entrain to their mother's dynamics when they are laying on her body [103] and even during face-to-face communication [104]. These

inter-individual autonomic system influences can affect the vocal behavior of those infants [105]. Taken together, these studies suggest that the set of autonomous systems proposed here are working in concert very early in life, that vocal development does not proceed in discrete stages seemingly represented by the levels of the hierarchy (Figure 2).

Vocalizations continue to undergo changes over the course of postnatal development in many species. These changes include the improved ability to produce the correct sound in the right context, and this improvement is influenced by social factors. This is true for humans [106-108], birds [109-111], bats [112, 113] and marmoset monkeys [50, 114-116]. For example, newborn marmosets spontaneously produce many different call types when alone, some of which are not appropriate for the context or immature-sounding [50]. Through development, marmosets gradually switch to producing only mature-sounding, high-energy contact calls when they are out of sight or far away from conspecifics [100]. In order to produce such calls, laryngeal muscles providing the vocal fold tension and expiratory muscles providing the subglottal pressure have to be strongly contracted. How are the neural dynamics shaped to direct a strong input to the muscles? This could be achieved through tuning the drive signal (arousal state) via infant-parent interactions.

The effect of parental feedback on changing vocal behavior is both immediate and cumulative. In one experiment, an infant marmoset and its parent were placed in the opposite corners of a room with an acoustically-transparent curtain in between. Instead of turn-taking using contact calls as adult marmosets do in this scenario [18], infant marmosets produce various types of vocalizations. However, they do immediately respond with a more adult-like contact call following the parental call [117]. It was found that parental calls were contingent upon the dynamics of the acoustic change in infant calls: they are produced at the phase of the 0.1 Hz acoustic oscillation when the infant calls were at the transition point from less to more adult-like [117]. Thus, on a moment-to-moment basis, the contingent parental calls are affecting the dynamics of the drive to the CPGs in the infants; it is providing a reinforcement signal to the infants (Figure 3). Parent-infant interactions can lead to a long-term change in infant vocal behavior. In humans [106-108] and songbirds [109-111], contingent parental responses accelerate the development of infant vocalizations, making them sound more mature more quickly. Similarly, in marmoset monkeys, the timing of the transition from immature to mature-sounding contact calls is correlated with the amount of contingent parental vocal feedback [50].

A study in which the amount of contingent parental responses was experimentally manipulated revealed that such feedback has a causal influence on vocal learning in this primate species as it does in humans [116]. Other studies showed that infant marmosets with limited parental feedback continued producing a large number of immature calls [114, 115]. Together, these results suggest that the contingent vocal feedback and parental interactions are not merely a perturbation to the infant vocal system but rather is changing the stable state of the drive oscillator in order to keep producing higher energy (mature sounding) calls (Figure 3). What could be the biological basis for the sustained change of drive oscillations? The dopamine (DA) neurons of mammalian PAG represent the social context and arousal in response to social cues [118, 119]. Analogously, a recent study in songbirds has suggested that the DA neurons in the PAG of juveniles play a role in initiating song learning through

the PAG-HVC pathway in the presence of a singing tutor [120]. In mammals, the PAG is also part of the pathway of vocal control. We thus hypothesize that dopamine signaling in the PAG modulated by cortical inputs is involved in vocal learning by marmoset monkeys (and perhaps other vocal learning animals) through contingent parental feedback.

Using the hierarchy of autonomous systems as a hypothesis-testing framework

The hierarchical autonomous systems framework for vocal production by its very nature suggests that components of vocal production can be, in some sense, isolated from each other. For example, the processes related to the production of a vocal sound (2-level system; Figure 2A) can be separated from the process that determines the initiation of a vocalization (3-level system; Figure 2B). Along the same lines, the type of vocalization produced (independent of its initiation) is determined by what is called for by the context (4-level system; Figure 2C). Neurophysiologic data from macaque monkeys trained to produce vocalizations on cue provide some support for this notion of separability, showing that sensory cues set up the decision to produce a vocalization in the ventrolateral prefrontal cortex while the anterior cingulate cortex reflects the degree to which the animal is motivated to produce a vocalization [121]. More specific predictions of the proposed multilevel framework could include testing the degree to which each level of the hierarchy functions as an autonomous system. For example, if the two-level system is autonomous, then experimental changes in laryngeal tension or the movement of other parts of the vocal apparatus should naturally affect the acoustic structure of vocalizations but it should not affect the rate or probability of producing a vocalization. Along the same lines, pharmacological manipulations of the autonomic nervous system (the three-level system) that affect arousal levels (e.g., beta blockers or beta agonists) should influence the timing of vocalizations independent of vocalization type. Finally, microstimulation, inactivation or other manipulations of motor structures in the forebrain (the four-level system) should affect the probability of producing a particular type of vocalization without affecting its acoustic structure (i.e., novel vocal sounds will not be produced). Certainly, we are well aware that the biology of vocal production is much more complex than we presented it [7] (for example, we completely neglected the important role of hormones [122-124]), but despite these limitations, we would argue for the utility of the proposed framework, particularly for comparative studies.

Concluding Remarks

The hierarchical structure of the motor system has long been adopted to facilitate our understanding of motor planning and motor control. Vocal production is a concrete example of a motor behavior that can be well quantified due to the relatively low-dimensional nature of acoustic signals. In this review, we summarize findings in animal vocal production occurring at different timescales into a hierarchical autonomous systems framework to understand vocal production. Each hierarchy of this framework operates at a different temporal scale, starting with acoustic features within vocalizations, and proceeding to sequences of vocalizations, and finally to vocal interactions and context-dependent vocal

production. Following this deduction, we gradually reveal the dynamics of higher-level brain regions that provide the manifolds for the dynamics of the lower level to unfold. Essentially, we propose that at least four levels are needed to fully understand the vocal production system, i.e., from the lowest to the highest: the vocal biomechanics, the CPGs, the drive, and the sensory representation. With the help of the autonomous systems perspective and findings in recent experiments, non-trivial transitions in vocal behavior through development can be explained including those induced by social experience.

Acknowledgments

We thank Daniel Takahashi and Diana Liao for their influence in shaping the ideas presented in this manuscript. Midshipman grunts courtesy of Andy Bass. This work was supported by the National Institutes of Health-NINDS (R01NS054898 to AAG).

References

1. Bass AH (2014) Central pattern generator for vocalization: is there a vertebrate morphotype? *Current opinion In neurobiology* 28, 94–100 [PubMed: 25050813]
2. Hage SR and Nieder A (2016) Dual Neural Network Model for the Evolution of Speech and Language. *Trends in Neurosciences* 39, 813–829 [PubMed: 27884462]
3. Kiebel SJ, et al. (2008) A Hierarchy of Time-Scales and the Brain. *Plos Comput Biol* 4
4. Flack JC, et al. (2013) Timescales, symmetry, and uncertainty reduction in the origins of hierarchy in biological systems. *Evolution cooperation and complexity*, 45–74
5. Deng L, et al. (2006) Structured speech modeling. *IEEE Transactions on Audio, Speech, and Language Processing* 14, 1492–1504
6. Dunbar RI and Shultz S (2007) Evolution in the social brain. *Science* 317, 1344–1347 [PubMed: 17823343]
7. Gomez-Marin A and Ghazanfar AA (2019) The life of behavior. *Neuron* 104, 25–36 [PubMed: 31600513]
8. Fischer J and Price T (2017) Meaning, intention, and inference in primate vocal communication. *Neuroscience & Biobehavioral Reviews* 82, 22–31 [PubMed: 27773691]
9. Barrett LF and Simmons WK (2015) Interoceptive predictions in the brain. *Nature Reviews Neuroscience* 16, 419–429 [PubMed: 26016744]
10. Azzalini D, et al. (2019) Visceral signals shape brain dynamics and cognition. *Trends Cogn Sci*
11. Hall ML (2009) A review of vocal duetting in birds. *Advances in the Study of Behavior* 40, 67–121
12. Logue DM and Gammon DE (2004) Duet song and sex roles during territory defence in a tropical bird, the black-bellied wren, *Thryothorus fasciatoventris*. *Animal Behaviour* 68, 721–731
13. Haimoff EH (1986) Convergence in the duetting of monogamous Old World primates. *Journal of Human Evolution* 15, 51–59
14. Müller AE and Anzenberger G (2002) Duetting in the titi monkey *Callicebus cupreus*: structure, pair specificity and development of duets. *Folia Primatologica* 73, 104–115
15. Greenfield MD and Rand AS (2000) Frogs have rules: selective attention algorithms regulate chorusing in *Physalaemus pustulosus* (Leptodactylidae). *Ethology* 106, 331–347
16. Okobi DE, et al. (2019) Motor cortical control of vocal interaction in neotropical singing mice. *Science* 363, 983–988 [PubMed: 30819963]
17. Choi JY, et al. (2015) Cooperative vocal control in marmoset monkeys via vocal feedback. *J Neurophysiol* 114, 274–283 [PubMed: 25925323]
18. Takahashi DY, et al. (2013) Coupled Oscillator Dynamics of Vocal Turn-Taking in Monkeys. *Curr Biol* 23, 2162–2168 [PubMed: 24139740]
19. Arlet M, et al. (2015) Grooming-at-a-distance by exchanging calls in non-human primates. *Biology letters* 11, 20150711 [PubMed: 26510675]

20. Kulahci IG, et al. (2015) Lemurs groom-at-a-distance through vocal networks. *Animal Behaviour* 110, 179–186
21. Demartsev V, et al. (2018) Vocal turn-taking in meerkat group calling sessions. *Curr Biol* 28, 3661–3666. e3663 [PubMed: 30416063]
22. Dunbar RI (1993) Coevolution of neocortical size, group size and language in humans. *Behavioral and brain sciences* 16, 681–694
23. Ghazanfar AA and Rendall D (2008) Evolution of human vocal production. *Curr Biol* 18, R457–R460 [PubMed: 18522811]
24. Bass AH, et al. (2008) Evolutionary origins for social vocalization in a vertebrate hindbrain–spinal compartment. *Science* 321, 417–421 [PubMed: 18635807]
25. Luthé L, et al. (2000) Neuronal activity in the medulla oblongata during vocalization. A single-unit recording study in the squirrel monkey. *Behav Brain Res* 116, 197–210 [PubMed: 11080551]
26. Hage SR and Jürgens U (2006) Localization of a vocal pattern generator in the pontine brainstem of the squirrel monkey. *The European journal of neuroscience* 23, 840–844 [PubMed: 16487165]
27. Jürgens U (2002) Neural pathways underlying vocal control. *Neuroscience and biobehavioral reviews* 26, 235–258 [PubMed: 11856561]
28. Hage SR and Jürgens U (2006) On the role of the pontine brainstem in vocal pattern generation: a telemetric single-unit recording study in the squirrel monkey. *J Neurosci* 26, 7105–7115 [PubMed: 16807339]
29. Schmidt MF, et al. (2012) Breathing and vocal control: The respiratory system as both a driver and a target of telencephalic vocal motor circuits in songbirds. *Experimental physiology* 97, 455–461 [PubMed: 21984733]
30. Bass AH, et al. (2008) Evolutionary Origins for Social Vocalization in a Vertebrate Hindbrain–Spinal Compartment. *Science* 321, 417–421 [PubMed: 18635807]
31. Barlow SM, et al. (2010) Central pattern generators for orofacial movements and speech In *Handbook of mammalian vocalization* (Brudzynski SM ed), pp. 351–369, Academic Press
32. Hage SR (2010) Localization of the central pattern generator for vocalization. *Handbook of Behavioral Neuroscience* 19, 329–337
33. Kittelberger JM, et al. (2006) The midbrain periaqueductal gray and vocal patterning in a teleost fish. *J Neurophysiol*
34. Esposito A, et al. (1999) Complete mutism after midbrain periaqueductal gray lesion. *Neuroreport* 10, 681–685 [PubMed: 10208530]
35. Jürgens U (1994) The role of the periaqueductal grey in vocal behaviour. *Behavioural brain research* 62, 107–117 [PubMed: 7945960]
36. Tschida K, et al. (2019) A Specialized Neural Circuit Gates Social Vocalizations in the Mouse. *Neuron*
37. Holstege G and Subramanian HH (2016) Two different motor systems are needed to generate human speech. *J Comp Neurol* 524, 1558–1577 [PubMed: 26355872]
38. Owren MJ, et al. (2011) Two organizing principles of vocal production: Implications for nonhuman and human primates. *American journal of primatology* 73, 530–544 [PubMed: 21509789]
39. Grillner S and Wallen P (1985) Central pattern generators for locomotion, with special reference to vertebrates. *Annu Rev Neurosci* 8, 233–261 [PubMed: 2984978]
40. Ijspeert AJ, et al. (2007) From swimming to walking with a salamander robot driven by a spinal cord model. *Science* 315, 1416–1420 [PubMed: 17347441]
41. Bianchi AL, et al. (1995) Central control of breathing in mammals: neuronal circuitry, membrane properties, and neurotransmitters. *Physiological Reviews* 75, 1–45 [PubMed: 7831394]
42. Jean A (2001) Brain stem control of swallowing: neuronal network and cellular mechanisms. *Physiol Rev* 81, 929–969 [PubMed: 11274347]
43. Sasahara K, et al. (2015) A rhythm landscape approach to the developmental dynamics of birdsong. *Journal of the Royal Society Interface* 12, 20150802
44. Pomberger T, et al. (2018) Precise Motor Control Enables Rapid Flexibility in Vocal Behavior of Marmoset Monkeys. *Curr Biol* 28, 788–794 e783 [PubMed: 29478857]

45. Fitch WT and Hauser MD (1995) Vocal production in nonhuman primates - acoustics, physiology, and functional constraints on honest advertisement. *American Journal of Primatology* 37, 191–219 [PubMed: 31936952]
46. van den Berg J (1958) Myoelastic-Aerodynamic Theory of Voice Production. *Journal of Speech, Language, and Hearing Research* 1, 227–244
47. Mindlin GB and Laje R (2006) *The physics of birdsong*. Springer
48. Titze IR and Martin DW (1998) *Principles of voice production*. ASA
49. Elemans CPH, et al. (2015) Universal mechanisms of sound production and control in birds and mammals. *Nat Commun* 6
50. Takahashi DY, et al. (2015) The developmental dynamics of marmoset monkey vocal production. *Science* 349, 734–738 [PubMed: 26273055]
51. Teramoto Y, et al. (2017) Vocal development in a Waddington landscape. *eLife* 6, e20782 [PubMed: 28092262]
52. Amador A, et al. (2013) Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature* 495, 59–64 [PubMed: 23446354]
53. Zhang YS, et al. (2019) Vocal state transition through laryngeal development. *Nat Comm* In press.
54. Chagnaud BP, et al. (2011) Vocalization frequency and duration are coded in separate hindbrain nuclei. *Nat Commun* 2, 346 [PubMed: 21673667]
55. Kelso JS (1997) *Dynamic patterns: The self-organization of brain and behavior*. MIT press
56. Thelen E and Smith LB (1996) *A dynamic systems approach to the development of cognition and action*. MIT press
57. Zhang YS and Ghazanfar AA (2018) Vocal development through morphological computation. *PLoS biology* 16, e2003933 [PubMed: 29462148]
58. Laje R, et al. (2002) Neuromuscular control of vocalizations in birdsong: a model. *Physical review. E, Statistical, nonlinear, and soft matter physics* 65, 051921
59. Gardner T, et al. (2001) Simple motor gestures for birdsongs. *Physical review letters* 87, 208101 [PubMed: 11690514]
60. Chandrasekaran C, et al. (2009) The natural statistics of audiovisual speech. *Plos Comput Biol* 5, e1000436 [PubMed: 19609344]
61. Greenberg S, et al. (2003) Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics* 31, 465–485
62. Lameira AR, et al. (2015) Speech-like rhythm in a voiced and voiceless orangutan call. *PloS one* 10, e116136 [PubMed: 25569211]
63. Toyoda A, et al. (2017) Speech - like orofacial oscillations in stump - tailed macaque (*Macaca arctoides*) facial and vocal signals. *American journal of physical anthropology* 164, 435–439 [PubMed: 28681947]
64. Terleph TA, et al. (2018) An analysis of white - handed gibbon male song reveals speech - like phrases. *American journal of physical anthropology* 166, 649–660 [PubMed: 29508909]
65. Bergman TJ (2013) Speech-like vocalized lip-smacking in geladas. *Curr Biol* 23, R268–R269 [PubMed: 23578870]
66. Ghazanfar AA, et al. (2012) Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Curr Biol* 22, 1176–1182 [PubMed: 22658603]
67. Ghazanfar AA and Takahashi DY (2014) The evolution of speech: vision, rhythm, cooperation. *Trends Cogn Sci* 18, 543–553 [PubMed: 25048821]
68. Moore JD, et al. (2014) How the brainstem controls orofacial behaviors comprised of rhythmic actions. *Trends In neurosciences* 37, 370–380 [PubMed: 24890196]
69. Chiao G, et al. (1994) Neuronal activity in nucleus ambiguus during deglutition and vocalization in conscious monkeys. *Experimental brain research* 100, 29–38 [PubMed: 7813650]
70. Larson CR, et al. (1994) Modification in activity of medullary respiratory-related neurons for vocalization and swallowing. *J Neurophysiol* 71, 2294–2304 [PubMed: 7931518]
71. Wright JM, et al. (2012) α - and β -Adrenergic receptors differentially modulate the emission of spontaneous and amphetamine-induced 50-kHz ultrasonic vocalizations in adult rats. *Neuropsychopharmacology* 37, 808 [PubMed: 22030713]

72. Borjon JI, et al. (2016) Arousal dynamics drive vocal production in marmoset monkeys. *J Neurophysiol* 116, 753–764 [PubMed: 27250909]
73. Henderson A, et al. (1966) Sequential temporal patterns in spontaneous speech. *Language and Speech* 9, 207–216
74. Julien C (2006) The enigma of Mayer waves: Facts and models. *Cardiovascular research* 70, 12–21 [PubMed: 16360130]
75. Lu CL and Jürgens U (1993) Effects of chemical stimulation in the periaqueductal gray on vocalization in the squirrel monkey. *Brain Research Bulletin* 32, 143–151 [PubMed: 8102315]
76. Morris KF, et al. (2010) Respiratory and Mayer wave-related discharge patterns of raphé and pontine neurons change with vagotomy. *Journal of Applied Physiology* 109, 189–202 [PubMed: 20360432]
77. Buzsáki G (2004) Neuronal Oscillations in Cortical Networks. *Science* 304, 1926–1929 [PubMed: 15218136]
78. Palva S and Palva JM (2018) Roles of brain criticality and multiscale oscillations in temporal predictions for sensorimotor processing. *Trends In neurosciences* 41, 729–743 [PubMed: 30274607]
79. Fox MD and Raichle ME (2007) Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature reviews neuroscience* 8, 700 [PubMed: 17704812]
80. Chan AW, et al. (2015) Mesoscale infraslow spontaneous membrane potential fluctuations recapitulate high-frequency activity cortical motifs. *Nat Commun* 6, 7738 [PubMed: 26190168]
81. Leopold DA, et al. (2003) Very slow activity fluctuations in monkey visual cortex: implications for functional brain imaging. *Cerebral cortex* 13, 422–433 [PubMed: 12631571]
82. Mitra A, et al. (2018) Spontaneous infra-slow brain activity has unique spatiotemporal dynamics and laminar structure. *Neuron* 98, 297–305. e296 [PubMed: 29606579]
83. Mateo C, et al. (2017) Entrainment of arteriole vasomotor fluctuations by neural activity is a basis of blood-oxygenation-level-dependent “resting-state” connectivity. *Neuron* 96, 936–948. e933 [PubMed: 29107517]
84. Zhang YS and Ghazanfar AA (2016) Perinatally influenced autonomic nervous system fluctuations drive infant vocal sequences. *Curr Biol* 26, 1249–1260 [PubMed: 27068420]
85. MacNeilage PF (2008) *The origin of speech*. Oxford University Press
86. Tchernichovski O, et al. (2001) Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* 291, 2564–2569 [PubMed: 11283361]
87. Veit L, et al. (2011) Learning to breathe and sing: development of respiratory-vocal coordination in young songbirds. *J Neurophysiol* 106, 1747–1765 [PubMed: 21697438]
88. Tchernichovski O, et al. (2000) A procedure for an automated measurement of song similarity. *Anim Behav* 59, 1167–1176 [PubMed: 10877896]
89. Sacks H, et al. (1978) *A Simplest Systematics for the Organization of Turn Taking for Conversation***This chapter is a variant version of “A Simplest Systematics for the Organization of Turn-Taking for Conversation,” which was printed in *Language*, 50, 4 (1974), pp. 696–735. An earlier version of this paper was presented at the conference on “Sociology of Language and Theory of Speech Acts,” held at the Centre for Interdisciplinary Research of the University of Bielefeld, Germany. We thank Dr. Anita Pomerantz and Mr. Richard Faumann for pointing out to us a number of errors in the text. 7–55
90. Katsu N, et al. (2019) Temporal adjustment of short calls according to a partner during vocal turn-taking in Japanese macaques. *Current Zoology* 65, 99–105 [PubMed: 30697245]
91. Pika S, et al. (2018) Taking turns: bridging the gap between human and animal communication. *Proceedings of the Royal Society B: Biological Sciences* 285, 20180598
92. Fröhlich M (2017) Taking turns across channels: Conversation-analytic tools in animal communication. *Neuroscience & Biobehavioral Reviews* 80, 201–209 [PubMed: 28501552]
93. Logue DM and Krupp DB (2016) Duetting as a collective behavior. *Frontiers in Ecology and Evolution* 4, 7
94. Levinson SC (2016) Turn-taking in human communication--origins and implications for language processing. *Trends Cogn Sci* In press

95. Kuriki S, et al. (1999) Motor planning center for speech articulation in the normal human brain. *Neuroreport* 10, 765–769 [PubMed: 10208545]
96. Takahashi DY, et al. (2012) A computational model for vocal exchange dynamics and their development in marmoset monkeys. 1–2
97. Brumm H and Slater PJ (2006) Animals can vary signal amplitude with receiver distance: evidence from zebra finch song. *Animal Behaviour* 72, 699–705
98. Johnson CJ, et al. (1981) Effects of interpersonal distance on children's vocal intensity. *Child Development* 52, 721–723
99. Pelegrín-García D, et al. (2011) Vocal effort with changing talker-to-listener distance in different acoustic environments. *Journal of the Acoustical Society of America* 129, 1981–1990 [PubMed: 21476654]
100. Liao DA, et al. (2018) Internal states and extrinsic factors both determine monkey vocal production. *Proceedings of the National Academy of Sciences* 115, 3978–3983
101. Mampe B, et al. (2009) Newborns' cry melody is shaped by their native language. *Curr Biol* 19, 1994–1997 [PubMed: 19896378]
102. Colombelli-Négrel D, et al. (2012) Embryonic learning of vocal passwords in superb fairy-wrens reveals intruder cuckoo nestlings. *Curr Biol* 22, 2155–2160 [PubMed: 23142041]
103. Van Puyvelde M, et al. (2015) Whose clock makes yours tick? How maternal cardiorespiratory physiology influences newborns' heart rate variability. *Biological psychology* 108, 132–141 [PubMed: 25869175]
104. Feldman R, et al. (2011) Mother and infant coordinate heart rhythms through episodes of interaction synchrony. *Infant Behavior and Development* 34, 569–577 [PubMed: 21767879]
105. Wass SV, et al. (2019) Parents Mimic and Influence Their Infant's Autonomic State through Dynamic Affective State Matching. *Curr Biol* 29, 2415–2422. e2414 [PubMed: 31303488]
106. Kuhl PK, et al. (2003) Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences* 100, 9096–9101
107. Goldstein MH and Schwade JA (2008) Social Feedback to Infants' Babbling Facilitates Rapid Phonological Learning. *Psychological Science* 19, 515–523 [PubMed: 18466414]
108. Goldstein MH, et al. (2003) Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences* 100, 8030–8035
109. West MJ and King AP (1988) Female visual displays affect the development of male song in the cowbird. *Nature* 334, 244–246 [PubMed: 3398921]
110. Carouso-Peck S and Goldstein MH (2019) Female Social Feedback Reveals Non-imitative Mechanisms of Vocal Learning in Zebra Finches. *Curr Biol*
111. Chen Y, et al. (2016) Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proceedings of the National Academy of Sciences* 113, 6641–6646
112. Prat Y, et al. (2017) Crowd vocal learning induces vocal dialects in bats: Playback of conspecifics shapes fundamental frequency usage by pups. *PLoS biology* 15, e2002556 [PubMed: 29088225]
113. Prat Y, et al. (2015) Vocal learning in a social mammal: Demonstrated by isolation and playback experiments in bats. *Science Advances* 1, e1500019 [PubMed: 26601149]
114. Gultekin YB and Hage SR (2017) Limiting parental feedback disrupts vocal development in marmoset monkeys. *Nat Commun* 8, 14046 [PubMed: 28090084]
115. Gultekin YB and Hage SR (2018) Limiting parental interaction during vocal development affects acoustic call structure in marmoset monkeys. *Science Advances* 4, eaar4012 [PubMed: 29651461]
116. Takahashi DY, et al. (2017) Vocal Learning via Social Reinforcement by Infant Marmoset Monkeys. *Curr Biol* 27, 1844–1852. e1846 [PubMed: 28552359]
117. Takahashi DY, et al. (2016) Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philos T R Soc B* 371
118. Cho JR, et al. (2017) Dorsal Raphe Dopamine Neurons Modulate Arousal and Promote Wakefulness by Salient Stimuli. *Neuron* 94, 1205–1219. e1208 [PubMed: 28602690]

119. Matthews Gillian A., et al. (2016) Dorsal Raphe Dopamine Neurons Represent the Experience of Social Isolation. *Cell* 164, 617–631 [PubMed: 26871628]
120. Tanaka M, et al. (2018) A mesocortical dopamine circuit enables the cultural transmission of vocal behaviour. *Nature* 563, 117–120 [PubMed: 30333629]
121. Gavrilov N, et al. (2017) Functional specialization of the primate frontal lobe during cognitive control of vocalizations. *Cell reports* 21, 2393–2406 [PubMed: 29186679]
122. Kelley DB (1986) Neuroeffectors for vocalization in *Xenopus laevis*: hormonal regulation of sexual dimorphism. *Journal of neurobiology* 17, 231–248 [PubMed: 3519865]
123. Marler P, et al. (1988) The role of sex steroids in the acquisition and production of birdsong. *Nature* 336, 770 [PubMed: 3205304]
124. Bercovitch FB, et al. (1995) The endocrine stress response and alarm vocalizations in rhesus macaques. *Animal Behaviour*
125. Hage SR (2018) Dual neural network model of speech and language evolution: new insights on flexibility of vocal production systems and involvement of frontal cortex. *Current opinion In behavioral sciences* 21, 80–87
126. Ackermann H, et al. (2014) Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences* 37, 529–546 [PubMed: 24827156]
127. Zhao W, et al. (2019) Inception of memories that guide vocal learning in the songbird. *Science* 366, 83–89 [PubMed: 31604306]
128. Feng NY and Bass AH (2016) “Singing” fish rely on circadian rhythm and melatonin for the timing of nocturnal courtship vocalization. *Curr Biol* 26, 2681–2689 [PubMed: 27666972]

Outstanding questions

Can the idea of distributed CPGs for vocal production be empirically validated, and if so, what is the neural basis for the versatile patterning of the vocal CPG circuit?

How do global fluctuations generated by the autonomic nervous system influence the dynamics of central pattern generators and other neural structures related to vocal production?

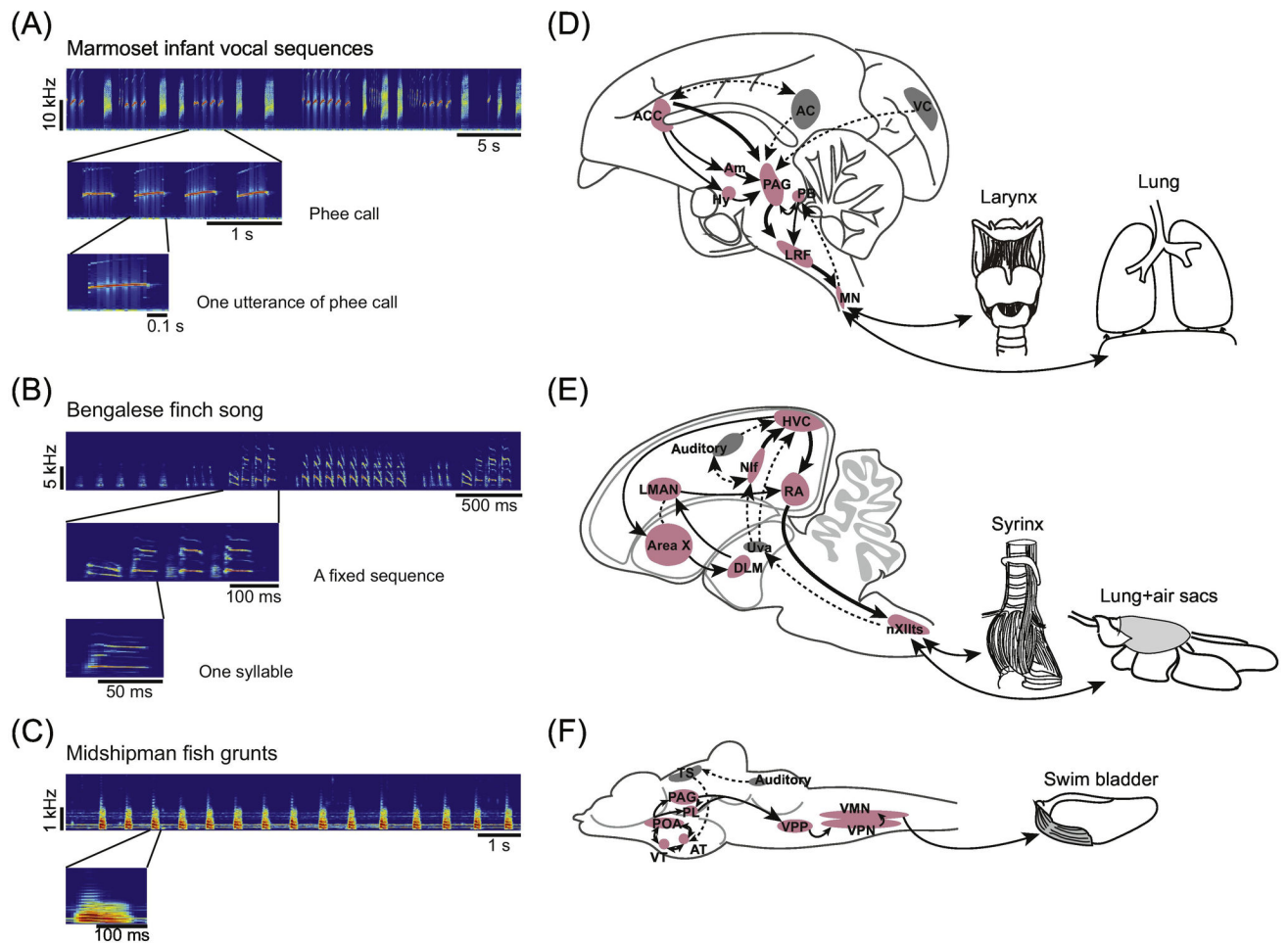
In what manner, and at what levels, does social reinforcement influence the neural dynamics and sensorimotor integration required for vocal production?

Highlights

Vocal production can be understood in terms of a timescale-based hierarchy of autonomous dynamical systems representing biomechanics and central pattern generators, internal drives, and the environment.

Each level of the vocal system generates a dynamical aspect of vocal output at a specific timescale.

This proposed autonomous systems framework can be used to illuminate how social reinforcement shapes the drive to generate mature sounding vocalizations in developing animals.



Trends in Neurosciences

Figure 1.

Temporal hierarchies of vocalizations, and the corresponding brain structure and vocal apparatus. (A) A segment of infant marmoset monkey vocalization comprised of utterances of different types of calls (contributed by Ghazanfar lab). (B) A segment of Bengalese finch song comprised of fixed and variable syllable sequences (contributed by Yisi Zhang). (C) A segment of midshipman fish grunts (courtesy of Dr. Andrew Bass). (D) Vocal communication system and vocal apparatus of nonhuman primates. Abbreviations: ACC, anterior cingulate cortex; AC, auditory cortex; Am, amygdala; Hy, hypothalamus; PAG, periaqueductal gray; LRF, lateral reticular formation; PB, parabrachial nucleus; MN, motor nuclei; VC, visual cortex (adapted from [125] and [126]). (E) Song system and vocal apparatus of songbirds. Abbreviations: HVC (as a proper name); RA, robust nucleus of the arcopallium; NIF, nucleus interfascialis of the nidopallium; LMAN, lateral magnocellular nucleus of the anterior nidopallium; Area X, striato-pallidum basal ganglia nucleus; DLM, dorsolateral thalamic nucleus; Uva, nucleus uvulaeformis; nXIIts, tracheosyringeal part of the hypoglossal motor nucleus (adapted from [127]). (F) Vocal network and vocal apparatus of fish. Abbreviations: POA, preoptic area; AT, anterior tuberal nucleus; VT, ventral tuberal nucleus; PL, paralemnisal midbrain tegmentum; TS, torus semicircularis; VPP, vocal

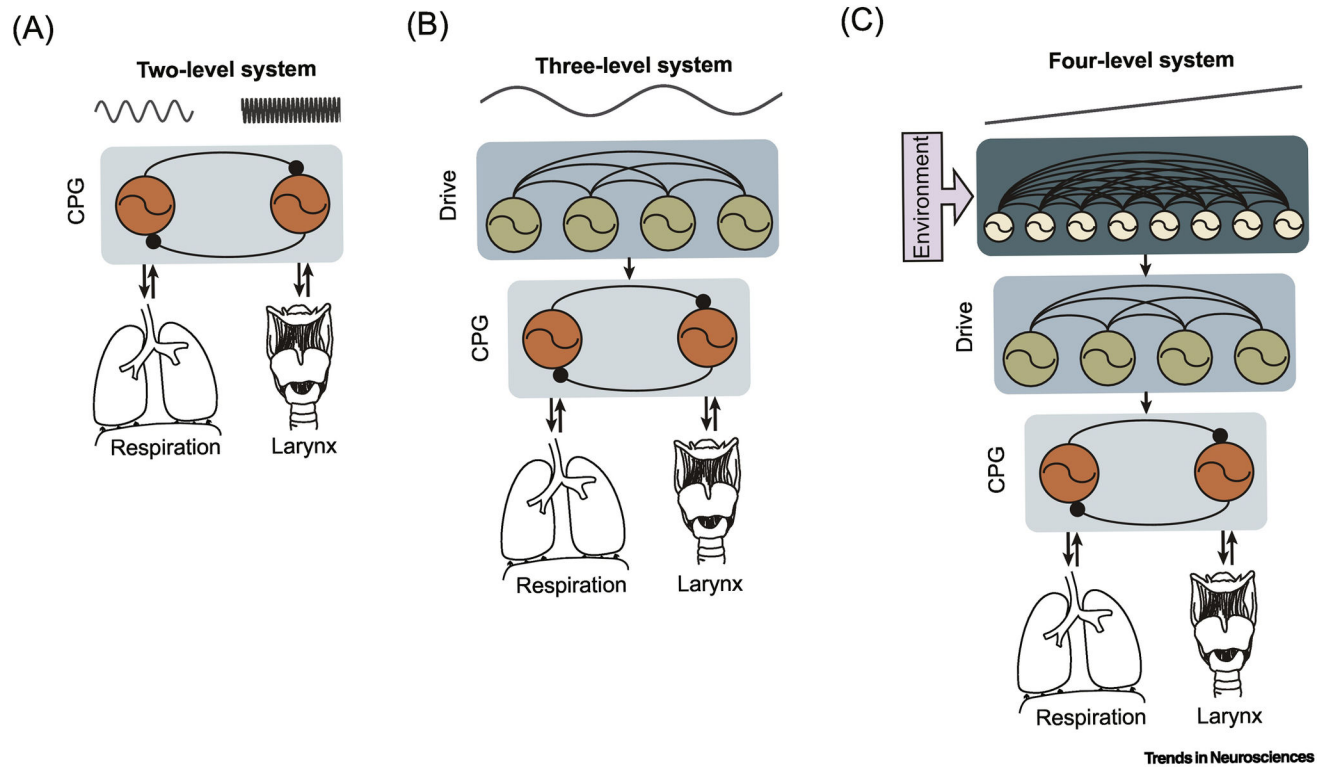
prepacemaker nucleus; VPN, vocal pacemaker nucleus; VMN vocal motor nucleus (adapted from [128]). Solid lines indicate vocal pathways and dotted lines indicate sensory pathways.

Author Manuscript

Author Manuscript

Author Manuscript

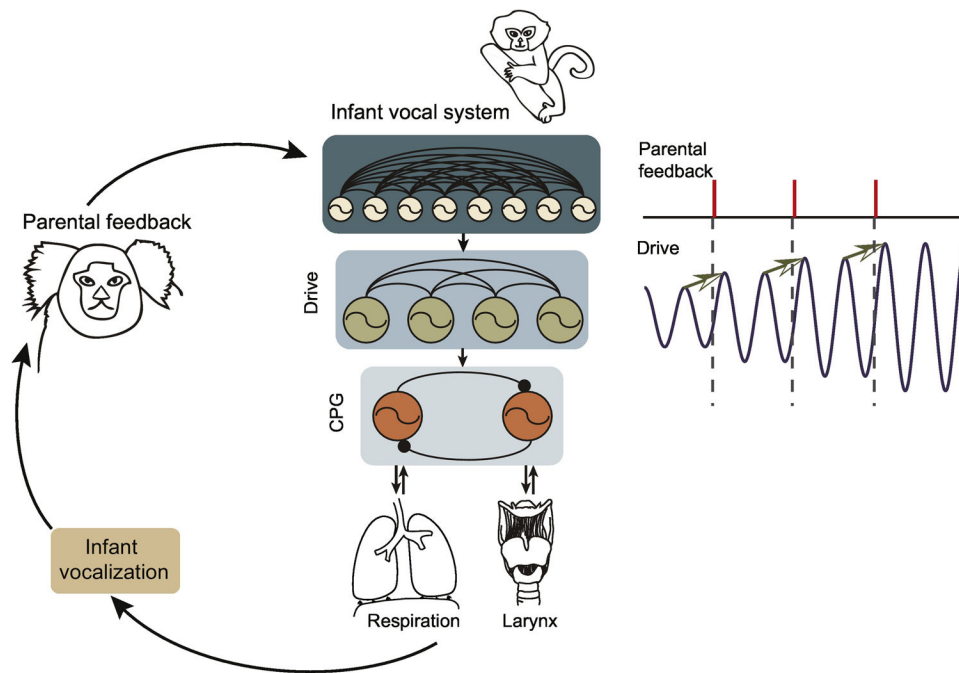
Author Manuscript



Trends in Neurosciences

Figure 2.

A hierarchical structure of the vocal production system based on time-scales. (A) Two-level system: the vocal biomechanics (lungs and larynx) and two coupled central pattern generators (CPGs). (B) Three-level system: adding a drive signal on top of the CPGs enables the continuous production of various calls. (C) Four-level system: a fourth layer provides the modulation to the drive and allows animals to adjust the vocal output with respect to the environment.



Trends in Neurosciences

Figure 3.

Vocal development through social feedback. Here we illustrate vocal development through marmoset monkey parent-infant vocal interaction. The parent responds at the transitions where the infant starts producing more mature-sounding vocalizations [117]. As more mature-sounding vocalizations indicate a greater underlying drive signal [53], parental responses occur in the rising phase of the oscillatory drive. The contingent parental calls have a cumulative effect on the infant vocal production towards more high-energy calls, accelerating vocal development on the timescale of days [116]. This process can be a consequence of shaping the drive signal of infant vocal production. The social feedback process described here compactly illustrates its cumulative influence on cumulative changes in the drive signal.