

Perspective

Convergent behavioral strategies and neural computations during vocal turn-taking across diverse species

Arkarup Banerjee^{1,2} and Daniela Vallentin^{3,4}



Abstract

Vocal exchanges between individuals are often coordinated in a temporally precise manner: one party is vocalizing while the other one is listening until the performance roles are switched. This vocal turn-taking behavior is widespread across the animal kingdom and thus provides an opportunity to study the neural circuit mechanisms from a comparative perspective. Although the physical prerequisites of the vocal tracts across animals can be different, the behavioral outcome of turn-taking is often similar with respect to vocal response timing and context-dependent adaptation. Here we review behavioral strategies of vocal turn-taking in diverse animals. Further, we highlight recent advances in studying the neural circuit mechanisms underlying vocal production and perception.

Addresses



¹ Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

² Cold Spring Harbor Laboratory School for Biological Sciences, Cold Spring Harbor, NY, USA

³ Max Planck Institute for Ornithology, 82319, Seewiesen, Germany

⁴ Max Planck Institute for Biological Intelligence (in Foundation), 82319, Seewiesen, Germany

Corresponding authors: Banerjee, Arkarup (abanerjee@cshl.edu); Vallentin, Daniela (daniela.vallentin@orn.mpg.de)

 (Banerjee A.),  (Vallentin D.)

neural circuit mechanisms requires experimentally tractable model systems and behavioral paradigms. For many species, including humans, one central adaptive behavior for social interactions is vocal communication. Such communication can arguably be unidirectional, but here we restrict ourselves to phenomena requiring bi-directional information flow between the sender and the receiver. Such vocal turn-taking requires the perception of auditory signals from the sender to generate appropriate vocal motor outputs making it ideally suited to study dynamic sensorimotor integration.

Using sounds to communicate is widespread in nature; from chirping birds [1] to duetting lemurs [2], a large number of species use vocalizations to cooperate as well as to compete. These vocal interactions often follow specific temporal patterns whereby two individuals exchange vocalizations in a coordinated manner avoiding overlaps and taking turns. Humans engaged in conversation, for example, take rapid turns to go back and forth (turn-taking) between listening and responding – a feat most of us tend to perform effortlessly, but which breaks down during communicative disorders [3].

In this review, we will first highlight behavioral evidence for vocal turn-taking across the animal kingdom, discussing its ubiquity, its ecological function, and how it may represent a common framework to study hitherto disparate phenomenon. Next, we will discuss recent developments in our understanding of neural circuit mechanisms for vocal communication taking examples from a few different species. Ultimately, we hope to convey that studying vocal turn-taking, at the intersection of ethology and systems neuroscience, represents a fruitful path forward in our quest to better understand the neurobiology of sensorimotor integration underlying adaptive behaviors.

Vocal turn-taking: a unifying framework to study diverse phenomenon

a) Vocal turn-taking is widespread across the animal kingdom

Spoken conversations between humans involve alternating exchange of vocalizations that tend to overlap

Current Opinion in Neurobiology 2022, 73:102529

This review comes from a themed issue on **Neurobiology of Behavior 2022**

Edited by **Tiago Branco** and **Mala Murthy**

For complete overview of the section, please refer the article collection - [Neurobiology of Behavior 2022](#)

Available online 31 March 2022

<https://doi.org/10.1016/j.conb.2022.102529>

0959-4388/© 2022 Elsevier Ltd. All rights reserved.

Introduction

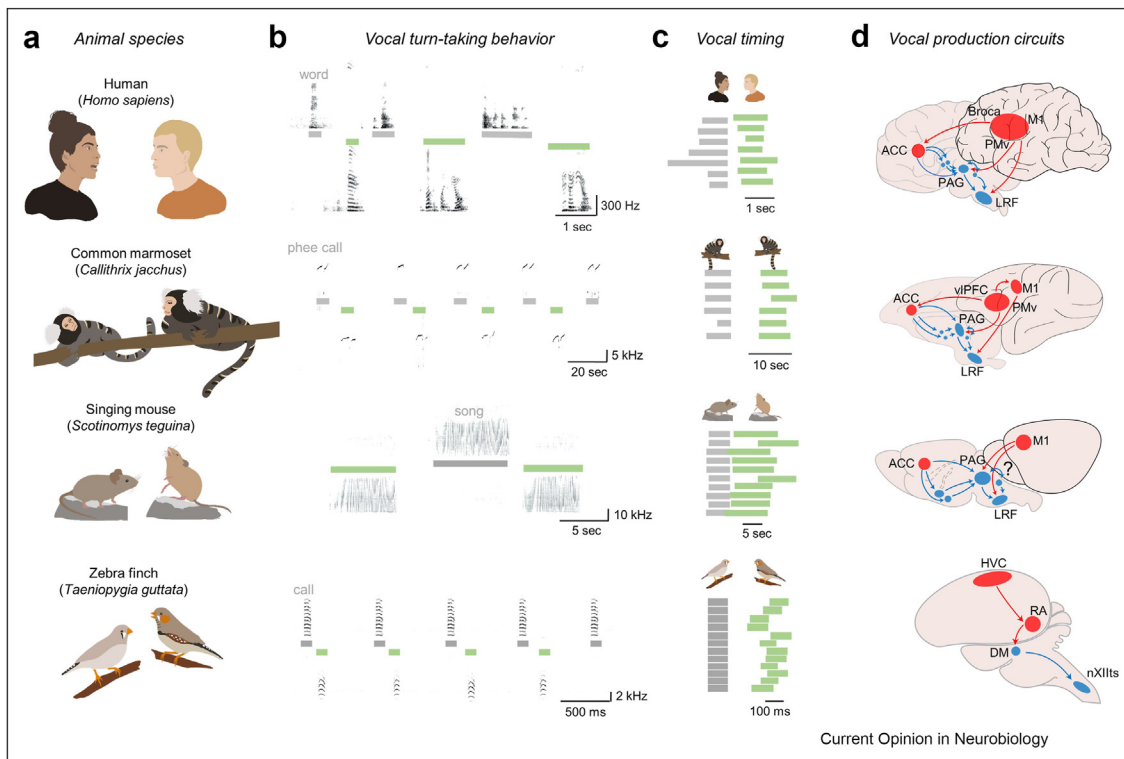
Adaptive behavior requires real-time adjustments in response to a rapidly changing environment. How the brain accomplishes this sensorimotor feat is one of the most fascinating and puzzling questions in neuroscience. Although examples of animals engaged in adaptive behavior abound in the natural world, understanding the

minimally [4]. This vocal turn-taking behavior not only requires fast sensory perception of the sender's vocal output but also the precise control of one's own vocal onset. During these interactions, participants simultaneously plan upcoming vocalizations while listening in order to respond as early as possible [5]. In humans, response times are typically about 250 ms (Figure 1 a), although average latency can vary across linguistic cultures [6]. Recent work has shown that speech planning activity can be localized to a frontotemporal brain network distinct from those involved in speech perception and production *per se* [7]. Due to the necessity of coordination in turn-based speech, and its early onset in ontogeny, vocal interaction has been proposed as a core component for language acquisition [8].

Vocal turn-taking in mammals is not restricted to humans; nonhuman primates as well as some rodent species have been shown to interact with each other by coordinating their vocalizations. Common marmosets (*Callithrix jacchus*)

live in large groups of 3–15 individuals and communicate with each other by performing 13 different calls which are used in specific contexts to indicate food, distance, or excitement [9]. Specifically, phee calls (example of a distance call) are preferentially used during vocal turn-taking [10,11]. In the family of mongooses, meerkats (*Suricata suricatta*), have been observed to communicate with each other using coordinated vocal interactions on sunny mornings during the cold season [12]. Timing of vocalizations could signal individual identity, internal state, or estimated physical distance (especially during phee calls in marmosets when individuals are out of site). Conveniently, marmosets engage in vocal interactions with playbacks in a laboratory setting [11], which is ideal to systematically explore how context affects vocal turn-taking. Lab mice or rats, however, have not been shown to vocally interact with conspecifics by taking turns. Notably, the Alston's singing mouse has recently been found to exhibit vocal turn-taking behavior [13]** (for details see Section below).

Figure 1



Vocal turn-taking behavior of animals. **a)** Display of pairs of different animals vocally interacting. **b)** Sonograms of vocalizations used by the animals shown in **A**. Although the spectral features of the vocalizations have a species-specific signature, all animals take turns while interacting. **c)** Temporal profile of vocal turn-taking. Vocalizations of animal 2 (green) aligned to the offset of the vocalizations of animal 1 (grey). Note the different timescales of these vocal exchanges. **d)** Neural circuits controlling vocal outputs: Broca: Broca's area; LRF: Laryngeal reticular formation; PMv: Premotor cortex (ventral); vPFC: ventral Pre-frontal cortex; PAG: Periaqueductal gray; M1: Primary motor cortex; ACC: Anterior cingulate cortex; nXIIIs: Hypoglossal nucleus; DM: Dorsomedial nucleus of the intercollicular complex; RA: Robust nucleus of the arcopallium; HVC: proper name; Human data credit: Giacomo Costalunga & Neetash Mysuru, Marmoset data credit: Steffen Hage, Singing mice data credit: Clifford Harpole, Zebra finch data credit: Jonathan Benichov.

A spectacular elaboration of turn-taking behavior is observed in avian species. Songbirds have been extensively studied for their exquisite vocal turn-taking performances. Their behavior ranges from the exchange of simple innate calls in zebra finches (*Taeniopygia guttata*) [14] to complex interactions with hundreds of learned songs in nightingales (*Luscinia megarynchos*) [15]. Many tropical bird species perform duets in opposite-sex pairs during which they sing cooperatively to defend a territory or reinforce the pair bond [16]. Some duetting styles (i.e., antiphonal duets) are characterized by synchronized time- and pattern-specific responses between partners, in which each bird rapidly adjusts its vocalizations over the course of an interaction depending on the partner. During duets, plain-tailed wrens, for instance, time their song more accurately compared to solo singing. Moreover, males sing louder with a female suggesting context-dependent song adjustments [17,18]. Recent studies on white-browed sparrow weavers further indicate that duetting birds accurately synchronize their vocal output to avoid overlaps and optimize their alternating turn-taking behavior, which sometimes even follows pair-specific rules [19–21]**.

Notably, songbirds with a very limited vocal repertoire also exhibit precise vocal turn-taking behavior. Throughout the day, a male zebra finch switches between singing its stereotyped courtship song and much shorter contact calls [14]. While contact calls are used to counter-sing with vocal partners at timescales comparable to humans, songs are not [22]**. In addition, zebra finches are capable of adapting their response strategy depending on the social context by flexibly adjusting the timing of their calls in relation to a vocal partner. This ability might be relevant for other contexts that go beyond mate attraction or territorial defense; during group flights, for example, individual birds may potentially coordinate their calls indicating their position, velocity, or directional changes.

The phenomenon of turn-taking is truly widespread in the animal kingdom. Mammalian lineages, such as bats and dolphins, which inhabit terrestrial and aquatic niches respectively, participate in vocal turn-taking as well. For example, adult white-winged Vampire bats [23] use antiphonal calling to communicate outside of their roost. Accumulating evidence suggests that multiple bat species on the one hand learn their vocal repertoire [24,25] and on the other hand use it in a more complex manner than previously thought [26]. Dolphins produce characteristic whistles during coordinated vocal interactions, which seem to facilitate individual recognition and maintenance of group cohesion [27]. Going even beyond vertebrates, turn-taking behavior is also prevalent in insects [28]. One prominent example is *Drosophila virilis*, a species of fruit fly that engages in acoustic duetting [29]. Further exploration of vocal turn-taking behavior in other species (e.g. the fish *Danionella cerebrum* which can

vocalize [30]) will reveal whether this phenomenon is even more ubiquitous than previously appreciated.

b) The potential function of vocal turn-taking

The ecological significance of vocal turn-taking behavior ranges from reproductive competition and conflict to sexual selection strategies. To avoid overlap with a communication partner implies that vocal signal should not be masked but clearly conveyed. This is useful in the framework of territorial defense when the aim is to defeat a vocal sparring partner or, alternatively during mating when females are supposed to be attracted by two individuals. Targeted overlap of a communication signal can also be interpreted as a sign of aggression to perturb the performance of the vocal partner. However, whether vocal overlap is a sign of aggression is debated [31,32] and further studies are needed to clarify this issue.

The ability to coordinate vocalizations in an interspersed manner precedes spoken language developmentally and evolutionarily, extending to other species ranging from nonhuman primates to birds and anurans [33]. Antiphonal advertisement calls of frogs have been well studied and we would like to refer the reader to Wells 1977 [34] for an extensive review. In all cases, vocal interactions generally involve perceiving relevant acoustic signals and initiating exact motor commands to generate an appropriate vocal reply. This social form of sensorimotor coordination reduces acoustic overlap, thereby ensuring that signals are accurately transmitted and can be correctly detected. Signaler and listener can respond to each other's vocalizations with variable delay, together with other adjusted structural vocal parameters for communicating their current state [35]. The fundamental principles of vocal turn-taking can only be fully understood when both participants and their associated internal states are taken into account.

All vocal turn-taking species partition acoustic space to maximize information transfer. For example, many anurans are highly vocal and communal social interactions are characterized by nonoverlapping vocalizations of multiple individuals. However, this strategy can also be changed to achieve maximal overlap which results in synchronous vocal outputs. It is hypothesized that the switch between these strategies might be an epiphenomenon to reset an individual's signal output upon hearing a neighbor's signal [36]. Interestingly, when multiple species inhabit the same niche, they vocalize at different times of the day to minimize acoustic overlap. For instance, frog species within a community take turns with other species by producing their advertisement calls only at specific times during the day while the other species remain quiet [37]. Although not strictly vocal turn-taking between two conspecifics as defined

above, here multiple species are taking vocal turns sequentially. This provides further evidence that communication evolves to minimize acoustic overlap.

Interestingly, in most animals, only certain vocal types are reserved for turn-taking, whereas the rest are used in the solo context. For example, in zebra finches, vocal usage depends upon specific contexts. During social bonding, they use short calls but switch to producing stereotyped songs (in a unidirectional manner) while attracting females. Similarly, marmosets specifically use phee calls for antiphonal interactions but have a larger repertoire of vocalizations for signaling food or emotional state. One potential benefit of having specialized call types for vocal turn-taking might be to unambiguously signal a conspecific with an explicit expectation of a response in return.

Vertebrates have developed numerous ways of producing sounds for communication. For example, some fish drum on their swim bladder [38] or stridulate their pectoral fins [39]; amphibians, reptiles, and mammals have a larynx that functions as a voice box during vocalizations [40]; and birds have a syrinx with membranes that vibrate during singing [41]. The diversity of vocal organs in vertebrates is partly mirrored by different neural structures controlling the vocal organ, but there are also similarities among species. Recent findings suggest that physiological and functional properties of neuronal circuits in birds and mammals are comparable, and analogous circuits can be identified in endbrain structures [42–46] as well as at the brainstem level in all vertebrates. Taken together, all vocal turn-taking species achieve the same behavior with a slightly different toolkit. In addition, since different species can either share communication strategies for vocal interactions or use distinct strategies, a comparative approach to study this behavior will provide insights into the canonical circuit functions and identify specialized ones underlying vocal control and production, as well as audio-vocal integration processes.

What are the neural principles underlying vocal turn-taking in mammals?

The neural processes underlying vocal communication behaviors are quite complex. It requires the integration of auditory inputs, planning, and generation of appropriate motor commands to move the vocal muscles, all within a fraction of a second. Since, lesions in different brain regions can cause specific deficits (e.g., sensory or motor aphasia), we have a parts-list of different brain regions mediating vocal communication. However, we do not understand the neural circuit-level mechanisms of vocal communication, despite its relevance to many communication disorders.

In mammals, species-typical sounds are generated by the primary “vocal motor network,” consisting of

evolutionarily conserved brain areas in the midbrain and brainstem [45]. Based on human brain lesions, gene expression profiles, and neurophysiology evidence in primates, a second frontal/motor “volitional control system” has been proposed [47]. Such cortical control over vocal production circuits is thought to render flexibility to vocal behaviors, such as during conversations or for modifying vocal usage based on context [48]. A key player of the primary “vocal motor network” is the periaqueductal gray (PAG). PAG lesions result in mutism and stimulation, either electrical or chemical, generates species-typical vocalizations in monkeys [49]. A recent study in laboratory mice has revealed the central importance of PAG in controlling ultrasonic vocalizations (USV). The first step was to use a clever genetic strategy to selectively label PAG neurons that are transiently activated in relation to USVs, referred to as the PAG-USV neurons [50]**. Blocking neurotransmission of the PAG-USV neurons using tetanus toxin light chain reduced the USV production and conversely, chemical or optogenetic activation led to more USVs in the absence of social cues. Importantly, optogenetic stimulation evoked USVs were produced during the exhalation phase of the respiratory cycle and did not differ from natural USVs in acoustic features. Finally, the authors showed that selective activation of PAG-USV neurons that project to the nucleus retroambiguus in the brain stem is sufficient to elicit USVs [50]**.

But what triggers or suppresses such vocal production circuits, especially in a context-dependent manner? In monkeys, stimulation of the anterior cingulate cortex (ACC) that projects directly to PAG, elicits species-specific vocalizations [51]. Other brain areas such as the hypothalamus and amygdala, when electrically stimulated also result in the production of vocalizations, suggesting that many regions upstream of PAG could influence context-specific vocalizations. Recent studies in laboratory mice have revealed the detailed circuit mechanism by which hypothalamus and amygdala control vocal production via the PAG [52,53]. Optogenetic activation of PAG projecting amygdala neurons can transiently suppress ongoing USVs. In contrast, activation of GABAergic hypothalamic neurons (medial preoptic area) that project to PAG elicits USVs in the absence of social cues [52]. It has been shown that increased activity of the hypothalamic input to PAG (e.g., ESR1+ve neurons in the lateral preoptic area) can flexibly scale the duration and amplitude of the USVs [53]**. Together, these studies begin to define the core circuit elements of the “primary vocal motor network” wherein PAG-USV neurons control USV production via pattern generators in the brainstem and in turn, hypothalamic or amygdala inputs to PAG provide context-specific behavioral gating (Figure 1 d).

The ability to exert voluntary control over phonation is a crucial element of human speech. Does this have a

phylogenetic precursor in other species? Behavioral evidence in favor of this idea comes from experiments demonstrating that macaque monkeys (*Macaca mulatta*) can be successfully trained to emit vocalizations (coo calls) in response to food [54] or to arbitrary visual stimuli [55]. Insights into neural circuit mechanisms for such voluntary control come from neural (extracellular) recordings during volitional control of vocalizations. For example, the macaque premotor cortex (PMv) contains vocalization-specific population of neurons that discharge either immediately before or during vocal onset, showing a preference for conditioned (voluntary) vocalizations compared to spontaneous ones [55]. Similar results have been observed in the ventrolateral prefrontal cortex (vlPFC), the putative Broca's area analog in monkeys. Macaque monkeys were successfully trained on a Go No-Go task to vocalize in response to a Go visual stimulus or withhold their vocalizations in response to the No-Go visual stimulus [55]. Single-unit electrophysiology recordings during this paradigm revealed that vlPFC neurons discharged preferentially during cued vocalizations and not during arousal outside of training. Moreover, neural activity of specific neurons immediately before vocal onset was correlated with acoustic features such as the call duration. Taken together, these findings provide evidence for a dual network model of vocalizations, wherein a frontal/motor "volitional control system" involving multiple cortical areas (e.g., PFC and PMv) is capable of exerting voluntary control over the phylogenetically older "primary vocal motor network" described above (Figure 1 d).

For turn-taking such as during a conversation, vocal production needs to be flexibly controlled by the auditory inputs from the vocal partner. What are the neural circuit mechanisms underlying such flexible auditory-motor coupling in the brain? For example, the New World marmoset monkeys participate in vocal exchanges. Consequently, marmosets have emerged as a prominent primate model to study vocal communication [10,11,56]. Marmosets have a large repertoire of vocalizations, including trills, twitters, and phee calls that are used in distinct social settings [57]. Among these, the most extensively studied is the phee call – a long-range contact call, which the common marmoset (*C. jacchus*) uses to participate in vocal turn-taking during cooperative vocal communication [10,58] (Figure 1 b). This turn-taking behavior is on the order of seconds (Figure 1 c) and thus, follows a slower timescale compared to human conversation. Phee calls are thought to encode a variety of social information such as caller sex and individual identity [59–61]. Recent studies have shown that marmosets are capable of interrupting and modulating ongoing vocalizations based on auditory inputs [44]. For example, auditory perturbation during ongoing phee calls showed that animals interrupt vocalizations only at discreet time points, suggesting that each phee call is in fact composed of multiple vocal units [62]**.

Further, it has been observed that marmosets rarely initiate vocalizations during the presentation of calls and therefore seem to inhibit calling to avoid interference [10,11].

A few studies have measured neural activity in different cortical regions during perception and production of phee calls. Neurons in the marmoset primary auditory cortex respond to spectrotemporal acoustic patterns of the species-typical vocalizations [63]. In addition to these purely sensory responses, auditory cortex neurons show significant vocalization-induced modulation of firing rates during self-generated phee calls [64]. Most of these auditory cortical neurons show a suppression of neural activity that starts prior to the onset of vocalizations, consistent with an efference copy mechanism thought to convey the predicted sensory consequences of self-generated movements [65]. In line with this model, these vocalization-suppressed neurons in auditory cortex were subsequently found to be sensitive to altered self-generated vocal feedback [66]. Moreover, neurons in auditory cortex are sensitive to compensatory vocal control in response to auditory feedback. Electrical stimulation of the same sites evokes rapid changes in vocal production [67]. Therefore, behavioral and physiological evidence suggests that auditory cortical activity plays a causal role in feedback-dependent vocal control [67].

What are the brain areas that control audio-vocal interactions in Marmosets? Ventral premotor cortex (vPMC) and prefrontal cortex (PFC) showed heightened immediate early gene expression during antiphonal conversations (Figure 1 d) [60,68]. Subsequent electrophysiological recordings identified neurons in PFC and vPMC that show increased neural activity during vocal production [69]. Interestingly, these neurons are also active during spontaneously generated phee calls in contrast to the vlPFC neurons reported in the macaques that are modulated only by cued vocalizations [54,55]. Further studies are needed to resolve whether this discrepancy arises due to operant-conditioning in the case of the macaque experiments or it reflects species-specific differences.

While these pioneering studies have identified the brain regions involved in primate turn-taking behaviors, our understanding of neural circuit mechanisms underlying vocal interactions, especially in mammals, remains quite rudimentary. While lab mice (*Mus musculus*) or rats (*Rattus rattus*) produce ample ultrasonic vocalizations, they do not seem to take turns. Recently, another mouse species, the Alston's singing mouse (*Scotinomys teguina*) has been found to not only vocalize in the audible range for humans but also to exhibit vocal turn-taking with conspecifics [70]. These rodents sing both spontaneous "solo" songs, as well as "duets." This counter-singing behavior requires sub-second modification of motor

outputs in response to auditory cues with a temporal precision that resembles human conversation [13]. Importantly, vocalizations that are relatively stable during solo singing become highly variable when the individual participates in vocal turn-taking, demonstrating strong context dependence – a hallmark of social behaviors. Furthermore, as a small rodent, it is readily amenable to many behavioral, electrophysiological, and viral tools developed in recent decades. Therefore, a more detailed analysis of the underlying turn-taking circuitry in mammals was conducted with the Alston's singing mice.

Using four complementary lines of evidence, Okobi, Banerjee et al. defined a region of orofacial motor cortex (OMC) that mediates flexible vocal interactions in the singing mouse. Intracortical microstimulation (ICMS) was first used to map a functional hotspot located on the anterolateral aspect of motor cortex, which was referred to as the OMC. Electrical stimulation of OMC disrupted ongoing singing behavior. Additionally, mild focal cooling of OMC slowed down the progression of the motor sequence by incorporating additional notes, resulting in considerably longer songs. Furthermore, silencing OMC activity (via GABA-A agonist muscimol), the authors showed that the animals could still sing their songs but could no longer participate in vocal interactions. While previous studies have used immediate early gene expression or electrophysiology to suggest cortical involvement in nonmammalian communication, these results represent the first direct demonstration of motor cortical dependence of vocal interactions in a non-primate mammalian species. These experiments demonstrated hierarchical vocal motor control in a rodent [13], and extends the scope of the dual network model of vocal production [47] that was previously thought to be evolutionarily restricted to the primate lineage.

In summary, evidence from primate and rodents suggest that in all mammals, species-typical sounds, like laughing or crying, which tend to be stereotyped and relatively inflexible, are generated by the primary “vocal motor network.” This network consists of vocal pattern generating regions in the brainstem and operates under limbic control. There exists a second frontal/motor “volitional control system,” which exerts *cortical control* over this primary vocal motor network [47]. The volitional control system effectively renders certain types of vocalizations (i.e., speech) more flexible.

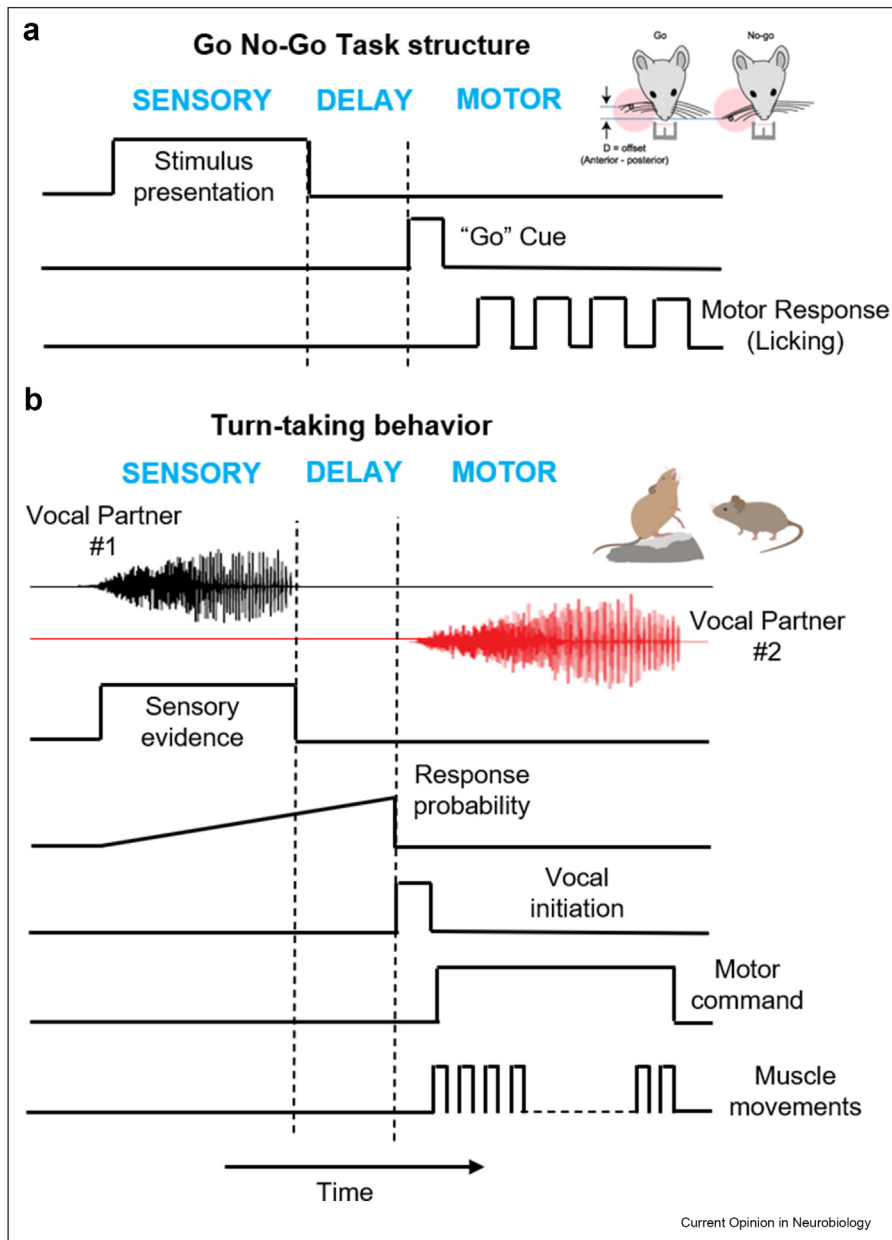
Neural control of vocal turn-taking in birds

How is this “vocal motor network” and the “volitional control system” implemented in other animal species with a different brain architecture and network connectivity? Songbirds are among the most vocal animals and their brain is comprised of dedicated and well-defined pathways for auditory perception and vocal

production [71] for which molecular and functional homologies to the mammalian cortex have been described [72]. Due to its anatomical location as an interface between the auditory pathway and the vocal motor pathway, the premotor nucleus HVC (proper name) is well positioned to play an important role during sensorimotor integration. Functionally, it has been shown that HVC shows stereotyped activity while a bird is singing [73], thus suggesting a role in patterned vocal signals, and likely in vocal turn-taking. Recent research by Benichov & Vallentin (2020) [22]** found that zebra finches tend to respond with preferred latencies to specific social partners. In the experiment, isochronous stack call playbacks were presented to an isolated bird. Different individuals responded with a specific stereotyped latency (198–322 ms). Besides, adding another bird to the setup resulted in one of the individuals responding later in time, showing that zebra finches might anticipate the calls of a vocal conspecific in order to adjust their own call timing, and ultimately avoid overlapping. The initial findings highlight the flexibility of the vocal system and its dependence on the social context. Following the behavioral characterization, HVC was pharmacologically inactivated bilaterally. The response timing got reversibly impaired, the precision was lost, with birds no longer showing a preferred response time. Benichov et al. (2016) [74] observed similar results after RA lesion, as expected from being immediately downstream of HVC in the descending motor pathway. Finally, to identify the neural circuit mechanism within HVC related to call timing, Benichov & Vallentin (2020) performed intracellular recordings from HVC neurons while the birds were calling. Results showed cells with call-related premotor activity, call-related inhibition that preceded the premotor activity, and inhibitory interneurons transiently increasing their firing rate in relation to calling and then reducing it. Interestingly, the rise in interneuron activity preceded the premotor signal, suggesting a role for this type of neurons in regulating premotor cells in HVC, and therefore call timing. To test this hypothesis, inhibition within HVC was temporarily lifted by means of Gabazine application. This pharmacological manipulation resulted in a change in call timing i.e., birds were calling faster in response to the call playback by up to 100 ms. Thus, premotor neurons provide accuracy to vocal onsets whereas interneurons ensure that vocalizations are triggered at appropriate times. All in all, the findings underpin that HVC is necessary for call timing precision, thus clarifying the forebrain role in vocal turn-taking.

A similar role for inhibition during vocal turn-taking has been described in HVC of duetting wrens [75]. These animals counter sing in a rapid manner with each other. Just like in other songbirds, HVC neurons are active during song production but not during auditory input when these birds are engaged in song production. To

Figure 2



| Vocal turn-taking as paradigms for studying ethologically relevant sensorimotor transformations. a) Typical design of a Go No-Go task wherein a subject is presented with sensory stimuli followed by a delay period leading up to the motor response. Cartoon adapted from O'Conner et al., 2010 **b)** Computations and their neural signatures underlying sensorimotor transformation during vocal turn-taking – from sensory evidence accumulation to decision making and finally leading up to motor commands and muscle movements. This example highlights vocal turn-taking behavior in the singing mice – a novel mammalian model system to study neural circuits for vocal communication. The minimal temporal overlap between the songs of the two mice conveniently delineates the behavior into sensory, delay, and motor epochs, a feature typically engineered in many tasks.

test the hypothesis whether inhibition within HVC suppresses auditory information during vocal interactions, the authors lifted inhibition by means of urethane anesthesia and showed that auditory-related activity reemerged. Taken together, inhibition might play a crucial role in withholding vocal production while listening to a communication partner.

Remarkably, in spite of millions of years of divergence between birds and the singing mice, similar neural control circuitry underlies vocal turn-taking behavior. This underscores the importance of the comparative neuroethology approach to discover the canonical neural principles of vocal turn-taking behavior in diverse species.

Conclusions and outlook

In systems neuroscience, typical paradigms to study sensorimotor transformations involve training animals to associate arbitrary stimuli to specific actions. For example, in a Go No-Go task, a rodent may be trained to associate a sensory stimulus with a specific movement and also to withhold the movement in response to a different sensory stimulus (Figure 2 a). Such paradigms allow experimenters to precisely control sensory stimuli as well as measure behavioral outputs with great precision and have taught us a lot about neural computations and circuit mechanisms underlying such sensorimotor behaviors.

There is a growing emphasis on using naturalistic behaviors in neuroscience [76–78]. Vocal communication, especially turn-taking behaviors described above goes beyond “naturalistic” – they represent tractable natural and self-guided behaviors ideally suited to study neural computations underlying sensorimotor integration. Since vocal partners take turns and avoid jamming each other as described above, this behavior can be conveniently delineated into sensory, delay, and motor epochs, a feature typically engineered in many operant conditioning tasks (Figure 2). Therefore, understanding the neural basis of vocal turn-taking behaviors in diverse species offers a potentially paradigm-changing approach for investigating the substrates of ethologically relevant perception, cognition, and action in the laboratory (Figure 2 b).

Finally, studying natural behaviors such as vocal turn-taking enables one to go beyond the “proximate” question of how neural circuits function toward the “ultimate” question of how neural circuits evolve (specified by genes and modified by learning). Comparing and contrasting neural circuit mechanisms across diverse species can help us to discover canonical algorithms underlying vocal turn-taking. Going forward, out of the plethora of questions that can be addressed while studying vocal turn-taking behavior, we would like to highlight the following three:

- Can a unifying signature of vocal turn-taking behavior be identified that holds true across animals? Or did different strategies evolve due to ecological needs?
- Are there common neural circuit motifs across animal species that govern the temporal basis of vocal turn-taking?
- What are the relative contributions of cortical and subcortical areas to the execution of vocal turn-taking during different contexts?

The spectacular progress in neural measurement and manipulation technologies in the last decade, coupled with gene editing platforms such as CRISPR, can

hopefully allow us to leverage the diversity of natural behaviors in pursuit of generalizable principles in neuroscience.

Conflict of interest statement

Nothing declared.

Acknowledgments

We would like to thank Adele and Giacomo Costalunga for their help with the Figure design. AB would like to acknowledge funding by the Simons foundation, NY. DV received funding from Emmy Noether grant (Project number VA742/2-1) by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), European Starting grant (ERC-2017-StG - 757459 MIDNIGHT) and a CRC grant by DFG – Project number 327654276.

References

1. Catchpole CK, Slater PJB: *Bird song: biological themes and variations*. Cambridge: Cambridge University Press; 2008.
2. De Gregorio C, et al.: **Categorical rhythms in a singing primate**. *Curr Biol* 2021, **31**:R1379–R1380.
3. Konopka G, Roberts TF: **Animal models of speech and vocal communication deficits associated with psychiatric disorders**. *Biol Psychiatr* 2016, **79**:53–61.
4. Levinson SC: **Turn-taking in human communication—origins and implications for language processing**. *Trends Cognit Sci* 2016, **20**:6–14.
5. Levinson SC, Torreira F: **Timing in turn-taking and its implications for processing models of language**. *Front Psychol* 2015, **6**:731.
6. Stivers T, et al.: **Universals and cultural variation in turn-taking in conversation**. *Proc Natl Acad Sci U S A* 2009, **106**:10587–10592.
7. Castellucci GA, et al.: **A speech planning network for interactive language use**. *Nature* 2022, **602**:117–122.
8. Donnelly S, Kidd E: **The longitudinal relationship between conversational turn-taking and vocabulary growth in early language development**. *Child Dev* 2021, **92**:609–625.
9. Bezerra BM, Souto A: **Structure and usage of the vocal repertoire of callithrix jacchus**. *Int J Primatol* 2008, **29**:671–701.
10. Takahashi DY, Narayanan DZ, Ghazanfar AA: **Coupled oscillator dynamics of vocal turn-taking in monkeys**. *Curr Biol* 2013, **23**:2162–2168.
11. Dohmen D, Hage SR: **Limited capabilities for condition-dependent modulation of vocal turn-taking behavior in marmoset monkeys**. *Behav Neurosci* 2019, **133**:320–328.
12. Demartsev V, et al.: **Vocal turn-taking in meerkat group calling sessions**. *Curr Biol* 2018, **28**:3661–3666. e3.
13. Okobi Jr DE, et al.: **Motor cortical control of vocal interaction in neotropical singing mice**. *Science* 2019, **363**:983–988.
14. Beckers GJ, Gahr M: **Neural processing of short-term recurrence in songbird vocal communication**. *PLoS One* 2010, **5**:e11129.
15. Hultsch H, Todt D: **Repertoire sharing and song-post distance in nightingales (*Luscinia-Megarhynchos B*)**. *Behav Ecol Sociobiol* 1981, **8**:183–188.
16. Hall ML: **A review of vocal duetting in birds**. *Adv Stud Behav* 2009, **40**:67–121. 40.
17. Fortune ES, et al.: **Neural mechanisms for the coordination of duet singing in wrens**. *Science* 2011, **334**:666–670.
18. Coleman M, Fortune E: **Duet singing in plain-tailed wrens**. *Curr Biol* 2018, **28**:R643–R645.
19. Hoffmann S, et al.: **Duets recorded in the wild reveal that interindividually coordinated motor control enables cooperative behavior**. *Nat Commun* 2019, **10**.

20. Lemazina A, *et al.*: **The multifaceted vocal duets of white-browed sparrow weavers are based on complex duetting rules.** *J Avian Biol* 2021, **52**.
21. Rivera-Caceres KD, *et al.*: **Early development of vocal interaction rules in a duetting songbird.** *R Soc Open Sci* 2018, **5**.
22. Benichov JI, Vallentin D: **Inhibition within a premotor circuit controls the timing of vocal turn-taking in zebra finches.** *Nat Commun* 2020, **11**:221.
23. Carter GG, *et al.*: **Antiphonal calling allows individual discrimination in white-winged vampire bats.** *Anim Behav* 2008, **76**:1343–1355.
24. Knornschild M: **Vocal production learning in bats.** *Curr Opin Neurobiol* 2014, **28**:80–85.
25. Fernandez AA, *et al.*: **Babbling in a vocal learning bat resembles human infant babbling.** *Science* 2021, **373**:923.
26. Knornschild M, Fernandez AA, Nagy M: **Vocal information and the navigation of social decisions in bats: is social complexity linked to vocal complexity?** *Funct Ecol* 2020, **34**:322–331.
27. Nakahara F, Miyazaki N: **Vocal exchanges of signature whistles in bottlenose dolphins (*Tursiops truncatus*).** *J Ethol* 2011, **29**:309–320.
28. Drosopoulos S, Claridge MF. In *Insect sounds and communication*. 1st ed. Boca Raton: CRC Press; 2005:552. Vol.
29. LaRue KM, *et al.*: **Acoustic duetting in *Drosophila virilis* relies on the integration of auditory and tactile signals.** *Elife* 2015, **4**.
30. Schulze L, *et al.*: **Transparent *Danio rerio* as a genetically tractable vertebrate brain model (vol 15, pg 977, 2018).** *Nat Methods* 2018, **15**:1126. 1126.
31. Wilson DR, Ratcliffe LM, Mennill DJ: **Black-capped chickadees, *Parus atricapillus*, avoid song overlapping: evidence for the acoustic interference hypothesis.** *Anim Behav* 2016, **114**:219–229.
32. Hutfluss A, *et al.*: **Does song overlap signal aggressiveness? An experimental study with repeated measures in free-ranging great tits.** *Anim Behav* 2021, **179**:199–211.
33. Pika S, *et al.*: **Taking turns: bridging the gap between human and animal communication.** *Proc Biol Sci* 2018, **285**.
34. Wells KD: **Social-Behavior of Anuran Amphibians.** *Anim Behav* 1977, **25**:666–693.
35. Fischer J, Price T: **Meaning, intention, and inference in primate vocal communication.** *Neurosci Biobehav Rev* 2017, **82**:22–31.
36. Greenfield MD, Tourtellot MK, Snedden WA: **Precedence effects and the evolution of chorusing.** *Proc Biol Sci* 1997, **264**:1355–1361.
37. Allen-Ankins S, Schwarzkopf L: **Spectral overlap and temporal avoidance in a tropical savannah frog community.** *Anim Behav* 2021, **180**:1–11.
38. Evans DH, Currie S. In *The physiology of fishes*. 5th ed. CRC Press; 2020. Vol.
39. Bass AH, Chagnaud BP: **Shared developmental and evolutionary origins for neural basis of vocal-acoustic and pectoral-gestural signaling.** *Proc. Natl. Acad. Sci. U.S.A* 2012, **109**:10677–10684.
40. Noordzij JP, Ossoff RH: **Anatomy and physiology of the larynx.** *Otolaryngol Clin* 2006, **39**:1.
41. Stein RC: **Modulation in bird sounds.** *Auk* 1968, **85**:229–&.
42. Feng NY, Bass AH: **"Singing" fish rely on circadian rhythm and melatonin for the timing of nocturnal courtship vocalization.** *Curr Biol* 2016, **26**:2681–2689.
43. Jarvis ED: **Evolution of vocal learning and spoken language.** *Science* 2019, **366**:50–54.
44. Hage SR: **The role of auditory feedback on vocal pattern generation in marmoset monkeys.** *Curr Opin Neurobiol* 2020, **60**:92–98.
45. Nieder A, Mooney R: **The neurobiology of innate, volitional and learned vocalizations in mammals and birds.** *Philos Trans R Soc Lond B Biol Sci* 2020, **375**:20190054.
46. Colquitt BM, *et al.*: **Cellular transcriptomics reveals evolutionary identities of songbird vocal circuits.** *Science* 2021, **371**.
47. Hage SR, Nieder A: **Dual neural network model for the evolution of speech and language.** *Trends Neurosci* 2016, **39**:813–829.
48. Zhang YS, Ghazanfar AA: **A hierarchy of autonomous systems for vocal production.** *Trends Neurosci* 2020, **43**:115–126.
49. Jurgens U: **The role of the periaqueductal grey in vocal behaviour.** *Behav Brain Res* 1994, **62**:107–117.
50. Tschida K, *et al.*: **A specialized neural circuit gates social vocalizations in the mouse.** *Neuron* 2019, **103**:459–472. e4.
51. Jurgens U: **The neural control of vocalization in mammals: a review.** *J Voice* 2009, **23**:1–10.
52. Michael V, *et al.*: **Circuit and synaptic organization of forebrain-to-midbrain pathways that promote and suppress vocalization.** *Elife* 2020, **9**.
53. Chen J, *et al.*: **Flexible scaling and persistence of social vocal communication.** *Nature* 2021, **593**:108–113.
54. Coude G, *et al.*: **Neurons controlling voluntary vocalization in the macaque ventral premotor cortex.** *PLoS One* 2011, **6**:e26822.
55. Hage SR, Gavrilov N, Nieder A: **Cognitive control of distinct vocalizations in rhesus monkeys.** *J Cognit Neurosci* 2013, **25**:1692–1701.
56. Takahashi DY, Fenley AR, Ghazanfar AA: **Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys.** *Philos Trans R Soc Lond B Biol Sci* 2016, **371**.
57. Eliades SJ, Miller CT: **Marmoset vocal communication: behavior and neurobiology.** *Dev Neurobiol* 2017, **77**:286–299.
58. Miller CT, Wang X: **Sensory-motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets.** *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 2006, **192**:27–38.
59. Norcross JL, Newman JD: **Context and gender-specific differences in the acoustic structure of common marmoset (*Callithrix jacchus*) phee calls.** *Am J Primatol* 1993, **30**:37–54.
60. Miller CT, Mandel K, Wang X: **The communicative content of the common marmoset phee call during antiphonal calling.** *Am J Primatol* 2010, **72**:974–980.
61. Miller CT, Wren Thomas A: **Individual recognition during bouts of antiphonal calling in common marmosets.** *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 2012, **198**:337–346.
62. Pomberger T, *et al.*: **Precise motor control enables rapid flexibility in vocal behavior of marmoset monkeys.** *Curr Biol* 2018, **28**:788–794 e3.
63. Wang X, *et al.*: **Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics.** *J Neurophysiol* 1995, **74**:2685–2706.
64. Eliades SJ, Wang X: **Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations.** *J Neurophysiol* 2003, **89**:2194–2207.

65. Wolpert DM, Flanagan JR: **Motor prediction.** *Curr Biol* 2001, **11**: R729–R732.
66. Eliades SJ, Wang X: **Neural substrates of vocalization feedback monitoring in primate auditory cortex.** *Nature* 2008, **453**: 1102–1106.
67. Eliades SJ, Tsunada J: **Auditory cortical activity drives feedback-dependent vocal control in marmosets.** *Nat Commun* 2018, **9**:2540.
68. Miller CT, et al.: **Vocalization induced CFos expression in marmoset cortex.** *Front Integr Neurosci* 2010, **4**:128.
69. Miller CT, et al.: **Responses of primate frontal cortex neurons during natural vocal communication.** *J Neurophysiol* 2015, **114**:1158–1171.
70. Banerjee A, Phelps SM, Long MA: **Singing mice.** *Curr Biol* 2019, **29**:R190–R191.
71. Nottebohm F, Stokes TM, Leonard CM: **Central control of song in the canary, *Serinus canarius*.** *J Comp Neurol* 1976, **165**:457–486.
72. Pfenning AR, et al.: **Convergent transcriptional specializations in the brains of humans and song-learning birds.** *Science* 2014, **346**:1256846.
73. Long MA, Jin DZ, Fee MS: **Support for a synaptic chain model of neuronal sequence generation.** *Nature* 2010, **468**:394–399.
74. Benichov JI, et al.: **The forebrain song system mediates predictive call timing in female and male zebra finches.** *Curr Biol* 2016, **26**:309–318.
75. Coleman MJ, et al.: **Neurophysiological coordination of duet singing.** *Proc. Natl. Acad. Sci. U.S.A* 2021, **118**.
76. Krakauer JW, et al.: **Neuroscience needs behavior: correcting a reductionist bias.** *Neuron* 2017, **93**:480–490.
77. Jourjine N, Hoekstra HE: **Expanding evolutionary neuroscience: insights from comparing variation in behavior.** *Neuron* 2021, **109**:1084–1099.
78. Cisek P: **Resynthesizing behavior through phylogenetic refinement.** *Atten Percept Psychophys* 2019, **81**(7):2265–2287.