Marmoset monkeys are known for their rich vocal repertoire. However, the underlying call production mechanisms remain unclear. By showing that marmoset moneys are capable of interrupting and modulating ongoing vocalizations, recent studies in marmoset monkeys challenged the decades-old concepts of primate vocal pattern generation that suggested that monkey calls consist of one discrete call pattern. The current article will present a revised version of the brainstem vocal pattern-generating network in marmoset monkeys that is capable of responding to perturbing auditory stimuli with rapid modulatory changes of the acoustic call structure during ongoing calls. These audio-vocal integration processes might potentially happen at both the cortical and subcortical brain level.

Addresses
1 Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Str. 25, 72076, Tübingen, Germany
2 Department of Otolaryngology, Head & Neck Surgery, Medical Center, University of Tübingen, 72076 Tübingen, Germany

Corresponding author: Hage, Steffen R (steffen.hage@uni-tuebingen.de)

Introduction
The evolution of the human speech and language system is one of the most difficult and controversial biological questions to answer, mainly because human speech considerably surpasses primate vocal communication systems in scope and flexibility with seemingly no counterpart in the animal kingdom, even among hominids [1,2*,3]. While human speech is above all a learned vocal pattern [4], the vocal repertoire of our closest relatives, nonhuman primates, consists mainly of stereotyped and largely innate calls that are uttered affectively in most contexts [3,5–7]. These assumptions were primarily supported by previous studies showing that monkeys that were deaf-born or deafened [8], raised in social isolation [8], or cross-fostered [9] developed the full adult species-specific call repertoire with only minor differences in call structure from the vocal utterances of their normally raised conspecifics. Furthermore, specific call features, such as duration and frequency, remain surprisingly robust and stable in adult monkeys, resulting in rather stereotyped and discrete call patterns [10]. For example, behavioral experiments have shown that monkeys can stop sequences of calls immediately after acoustic perturbation, but cannot interrupt the acoustic structure of ongoing vocalizations [11,12]. These findings suggested that calls consist of single discrete pulses [11,12] and that the vocal motor plan is already present before vocal onset [13]. In accordance with neurophysiological and anatomical studies, these vocalizations were found to be produced by a vocal pattern-generating network (VPGN) situated in the brainstem [2*,6,14] that is capable of producing rather fixed, inflexible, and discrete utterances.

In the last few years, however, a number of studies have observed several types of vocal flexibility allowing monkeys to cognitively control when [15,16], where [17], and what to vocalize [18–20]. In addition, it has been revealed that marmoset monkeys are able to modulate call patterns during ongoing calls in response to perturbing acoustic feedback [21**,22**,23,24]. These findings challenge previous assumptions that monkey vocalizations are discrete and inflexible patterns and question the idea of a VPGN that produces fixed-action vocal patterns.

In the present article, I will first review recent neuroethological insights suggesting that monkeys exhibit distinct types of vocal flexibility enabling them to modulate the acoustic structure in ongoing calls in response to acoustic perturbation. I will also introduce the VPGN as it has been thought to be until recently. Taking into consideration recent neurophysiological and neuroethological findings, I will then propose necessary alterations to the VPGN that can explain how it might be capable of generating the observed changes in call structure and present a revised version of the VPGN. Finally, I will summarize the potential cortical and subcortical brain levels at which the auditory system and the vocal motor network might interfere to exhibit the observed modulatory effects during ongoing calls.

Rapid vocal flexibility in response to perturbing auditory feedback
Very recently, several labs were able to reveal that monkeys are capable of modulating ongoing vocalizations rapidly after perturbing auditory feedback [21**,22**,23,24]. These
findings challenged the decades-old concepts of primate vocal pattern generation, indicating that vocalizations do not consist of one discrete call pattern.

In a recent study, my group was able to show that marmoset monkeys are capable of truncating calls in response to vocalization-triggered noise [21**]. As in previous studies in other primate species [11,12], we used acoustic perturbation triggered by the vocal behavior itself and quantitatively measured the resulting vocal adjustments of marmoset monkeys during phee call production (Figure 1a). During perturbation, the duration of most uttered phee vocalizations was not affected by the interfering noise stimuli and uttered within the normal phee duration range (between 1 and 2 s of duration). However, a small yet distinct number of vocalizations (around 3% of all uttered phees in the noise conditions) were interrupted by the noise stimuli resulting in phee calls with durations well below 0.5 s (Figure 1a and b). These findings suggest that marmoset monkeys are capable of interfering with vocal pattern production after call onset and producing calls being shorter than the natural range of call durations. Calls were interrupted approximately already 80 ms after noise perturbation onset resulting in truncated calls being much shorter than vocalizations uttered under normal conditions, indicating a direct audio-vocal integration mechanism underlies this behavior.

In another study, my group investigated whether the marmoset phee calls that do not show any apparent differences in call duration exhibit changes in call structure when the perturbing noise starts after call onset [23]. We found that marmoset monkeys are capable of rapidly (<100 ms) modulating call frequency in response to these perturbing noise stimuli (Figure 1c). Phee vocalizations swiftly increased in call frequency after noise perturbation in a rapid way, similar to the occasional interruptions of their calls in response to these perturbing stimuli (<100 ms). These findings are supported by another recent study using a similar neuroethological approach where marmoset monkeys were confronted with perturbing band pass filtered noise starting after call onset and with frequencies directly above or below the fundamental frequency of the phee calls [24]. Here, the monkeys shifted the fundamental frequencies of their vocalizations away from the spectrum of the interfering noise bursts. The findings of these studies also indicate a rapid effect of perturbing noise on vocal motor pattern production.

Finally, recent work revealed that marmoset monkeys exhibit feedback-dependent vocal control [22**]. Here, the frequency of the uttered vocalization was shifted in real-time and directly played back to the monkeys during ongoing calls. In response to the altered vocal feedback, the marmoset monkeys showed compensatory changes in vocal production with significantly increased call frequencies during negative frequency shifts and significantly decreased frequencies during positive shifts [22**]. Again, the observed changes in call structure were exhibited with short response latencies of approximately 150 ms after playback alteration, which are comparable to

---

**Figure 1**

Noise-induced interruption of phee call production in marmoset monkeys.
(a) Example spectrogram of a normal, single phee and a phee interrupted by noise perturbation, respectively. (b) Normalized phee duration distribution for calls that were (noise) and were not perturbed (no noise) for five monkeys (adapted from Ref. [21**]). (c) Example spectrogram of a call that was not and was perturbed by noise. The black dotted line in the lower spectrogram indicates the trajectory of the unperturbed phee shown above (adapted from Ref. [23]).
response latencies being observed for human subjects in a similar experiment (228 ms) [25], again indicating a rather direct modulation of the vocal pattern-generating network from the auditory side (see also explanations regarding Figure 2 below).

Overall, these recent studies suggest rapid audio-vocal integration mechanisms can directly modulate ongoing vocal pattern generating processes and challenge the concept of primate vocal pattern generation that vocal utterances are discrete call patterns.

**Vocal pattern generating network revisited: a neuroethological approach**

In non-human primates, the genetically predetermined vocalizations are produced by the primary vocal motor network [2*]. This network consists of two structurally and functionally distinct parts, the VPGN in the brainstem and an upstream limbic vocal-initiating network driving the pattern generator based on affective state [6,7]. The core of the primary vocal motor network is the VPGN in the brainstem, which is highly conserved in all vertebrate taxa [26]. This network consists mainly of the interconnected areas of the periaqueductal grey (PAG), parabrachial nucleus (PB), and ventrolateral pontine reticular formation. The latter contains the vocal pattern generator, which controls all phonatory motoneuron pools (nuclei of cranial nerves V, VII, and XII, and the ambiguous nucleus) that innervate the muscles of the larynx, oral cavity, tongue, and respiratory system to produce the actual acoustic structure of the calls [6,27,28**,29]. The role of these brainstem structures in vocal pattern generation has been deciphered based on a large number of electrophysiological recording, stimulation, and lesion studies in monkeys (for review: Ref. [2*]). The PAG and PB have a more general function in gating the type of vocalization, call duration, and timing of the vocal onset on the basis of the momentary respiratory status [2*,6]. The vocal pattern generator in the ventrolateral pontine brainstem itself is activated by these

![Figure 2](current-opinion-in-neurobiology-2020-60-92-98-1.png)

Phee vocalizations share common principles in call pattern production with other call types. (a) Multi-modal phee distribution of interrupted phees in an individual marmoset monkey. Vertical lines are multiple of each other. (b) and (c) Example spectrograms for phee, twitter, and tsik–ekks. Phee calls consist of phee units of similar duration (b), twitters of twitter syllables, and tsik–ekks of tsik and ekk syllables (c) (adapted from Ref. [2*]).
structures and generates the neural patterns that are driving the motoneuron pools to produce the acoustic structure of the entire calls. Earlier studies have suggested that the vocal pattern generating system consists of a network of dedicated pattern generators [30], each for each call type, since inactivation of distinct areas in the ventrolateral pontine brainstem led to the inhibition of just a few call types and not the entire vocal repertoire [31].

The recent studies discussed above suggest that marmoset vocalizations are not generated as a fixed action pattern, that is, predetermined before call onset, but rather in a more flexible way, allowing immediate modulation of the acoustic structure during ongoing call production. Therefore, these recent findings seem to be in contrast to the predominant concept of vocal pattern production in primates [2*,6,7,14]. However, recent findings might help to sufficiently revise this current concept and suggest an improved general principle of pattern generation in primate vocalization. As discussed above, a recent study revealed that marmoset monkeys were able to interrupt phee call production in response to perturbing noise [21**]. Interestingly, interrupted phee distributions exhibited a multimodal distribution with distinct sharp peaks at multiples of a smaller fraction (Figure 2a). These observations and further investigations suggest that the phee call is not a discrete call pattern by itself but that it is built of many sequentially uttered stereotyped brief units, just like other marmoset call types (Figure 2b and c).

According to these findings, I hypothesize that the vocal pattern generator in the ventrolateral reticular formation is not generating the acoustic structure of phee vocalizations as a whole, but rather generating highly stereotyped and more elementary vocal motor units. The number of vocal motor units that are produced by the vocal pattern generator and, therefore, determine phee call duration, is dependent on the duration of the excitatory input originating from the PAG, which serves a main gating function and is reciprocally connected with the vocal pattern generator within the ventrolateral pontine reticular formation [31,32]. This hypothesis is supported by single-unit recordings in macaques and squirrel monkeys [33,34] where neuronal activity observed in the PAG was closely correlated to call duration, while single-unit recordings in the putative vocal pattern generator of squirrel monkeys revealed syllable-correlated activity during trill vocalizations (comparable to twitter calls in marmosets; Figure 2c) [35].

Figure 3

Audio-vocal integration mechanisms exist at cortical and subcortical levels. Call production might be affected by ambient noise at different brain levels. Audio-vocal integration mechanisms are known to exist between cortical and subcortical structures, as well as via corticofugal projections. See text for further explanation. Simplified circuit diagram summarizing the most relevant structures for vocal production and the auditory pathway in monkeys. Arrows indicate anatomically verified and relevant direct connections. The volitional articulatory motor network (indicated in green) is capable of modulating the primary vocal motor network (indicated in blue) and the motoneuron pools (indicated in orange) involved in call production. External auditory stimuli are perceived via the auditory pathway (indicated in purple), which is capable of modulating the vocal motor system via direct connections (indicated in red) at cortical and subcortical levels (adapted from Refs. [2*,23,36,50]). Abbreviations: AC, primary auditory cortex; ACC, anterior cingulate cortex; CN, cochlear nucleus; IC, inferior colliculus; M1, ventral primary motor cortex; MN, motoneuron pools involved in call production; PFC, prefrontal cortex; PMC, ventral premotor cortex; SOC, superior olivary complex; VPG, vocal pattern generating network in the brainstem encompassing the periaqueductal gray, retroambigual nucleus, and parabrachial nucleus.
In contrast to the previous version of the vocal motor network, the proposed revision of the vocal motor network now allows rapid modulations of the acoustic structure during ongoing calls and abruptions of vocalizations in response to perturbing auditory stimuli. Interestingly, perturbing acoustic stimuli are capable of interrupting vocalizations or modulating the acoustic call structure with latencies within a range of the duration of one to two vocal motor units, that is, 80–160 ms, suggesting a rather direct interaction between the auditory system and vocal motor network [21**,22**,23]. Such connections from the auditory to vocal motor side are present at several cortical and subcortical levels, any or all of which might be capable of eliciting the observed audio-vocal interactions discussed above.

Considering the current literature, there are several anatomically and physiologically verified circuits that are capable of mediating audio-vocal integration processes between the auditory system and vocal motor network within the primate brainstem (Luo et al. [36*]). The first connection between the auditory and vocal motor side is the direct projections from the cochlear nucleus to the laryngeal and respiratory motor neuron pools involved in vocal production [36*]. This connection seems to be solely responsible for the laryngeal reflex, an extremely rapid reflexive activation of the recurrent laryngeal nerve in response to external acoustical stimulation in humans and bats [37,38]. Its potential role in audio-vocal integration processes in non-human primates is yet unknown. Two more likely connections between the primary auditory pathway and vocal motor network can be found at the brainstem level (Figure 3). One is between the superior olivary complex and the putative vocal pattern generator within the ventrolateral pontine reticular formation, and one is between the inferior colliculus and the putative vocal pattern generator in the brainstem [36*], which contains single neurons that respond to external acoustical stimuli and also show changes in neural activity before and during vocal production [39].

At the cortical level, the auditory cortex (AC) is reciprocally interconnected with all cortical structures of the vocal motor network, such as the prefrontal cortex (PFC), premotor and motor cortices (PMC and M1) of the volitional articulatory motor network, and anterior cingulate cortex of the primary vocal motor network [2,40–42]. While several studies have already highlighted the potential role of the AC in feedback-dependent vocal control [43,44], a recent study revealed the first evidence of a causal role of the AC in the modulation of vocal output [22**]. As discussed above, marmoset monkeys showed compensatory changes in vocal production in response to real-time shifted playback. Interestingly, electrical microstimulation of the AC rapidly evoked similar changes during vocal output and the activity of single neurons in the AC predicted the observed vocal changes in response to altered feedback [22**].

**Conclusion**

Marmoset monkeys have been shown to be capable of modulating ongoing vocalizations rapidly after perturbing auditory feedback, which challenges the decades-old concepts of primate vocal pattern generation, indicating that vocalizations do not consist of one discrete call pattern. However, recent studies have also revealed that marmoset vocalizations are not a discrete call pattern in and of themselves but built of many sequentially uttered stereotyped brief units enabling audio-vocal integration processes to interfere within ongoing calls produced by the vocal pattern generating network. In the future, neurophysiological studies will have to elucidate at which level the auditory system interferes with the vocal motor network to modulate call production. The marmoset monkey will be a particularly suitable model system due to its rich vocal repertoire, produced either spontaneously [45] or under behavioral control [46], as well due to several neurophysiological techniques that have been established in recent years [47–49].

**Conflict of interest statement**

Nothing declared.

**Acknowledgements**

I thank John Holmes for proofreading. This work was supported by the Werner Reichardt Centre for Integrative Neuroscience (CIN) at the Eberhard Karls University of Tübingen (CIN is an Excellence Cluster funded by the Deutsche Forschungsgemeinschaft within the framework of the Excellence Initiative EXC 307).

**References and recommended reading**

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


