



Research Paper

Rapid modulations of the vocal structure in marmoset monkeys

Lingyun Zhao¹, Sabyasachi Roy¹, Xiaoqin Wang*

Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, The Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA



ARTICLE INFO

Article history:

Received 31 May 2019

Received in revised form

5 October 2019

Accepted 7 October 2019

Available online 9 October 2019

Keywords:

Vocal control

Marmoset

Perturbation

ABSTRACT

Humans and some animal species show flexibility in vocal production either voluntarily or in response to environmental cues. Studies have shown rapid spectrotemporal changes in speech or vocalizations during altered auditory feedback in humans, songbirds and bats. Non-human primates, however, have long been considered lacking the ability to modify spectrotemporal structures of their vocalizations. Here we tested the ability of the common marmoset (*Callithrix jacchus*), a highly vocal New World primate species to alter spectral and temporal structures of their species-specific vocalizations in the presence of perturbation signals. By presenting perturbation noises while marmosets were vocalizing phee calls, we showed that they were able to change in real-time the duration or spectral trajectory of an ongoing phee phrase by either terminating it before its completion, making rapid shifts in fundamental frequency or in some cases prolonging the duration beyond the natural range of phee calls. In some animals, we observed fragmented phee calls which were not produced by marmosets in their natural environment. Interestingly, some perturbation-induced changes persisted even in the absence of the perturbation noises. These observations provide further evidence that marmoset monkeys are capable of rapidly modulating their vocal structure and suggested potential voluntary vocal control by this non-human primate species.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Humans have remarkable flexibility in vocal production in speaking or singing. Previous studies have shown the acoustic structure of human speech sounds is under precise voluntary control and is maintained through auditory feedback (Hickok et al., 2011; Houde and Jordan, 1998; Villacorta et al., 2004; Waldstein, 1990). Altering auditory feedback usually leads to real-time changes in speech production. For example, on-going speech may be interrupted or stopped when auditory feedback is delayed (Fairbanks, 1955; Fairbanks and Guttman, 1958; Lee, 1950). The fundamental frequency of speech may be altered when auditory feedback is pitch-shifted (Burnett et al., 1998; Chen et al., 2007; Larson et al., 2000). The alternation of speech production usually occurs soon after the feedback manipulation starts, usually within several hundreds of milliseconds (Burnett et al., 1998), suggesting

that the vocal control circuit utilizes a mechanism that updates the status of the motor output in real-time based on the auditory feedback.

Studies in some animal species such as songbirds and bats generally found similar real-time vocal control behaviors (Brainard and Doupe, 2000; Smotherman, 2007). By manipulating auditory feedback, a number of previous studies have demonstrated modifications of vocal production in songbirds. For example, songs produced by Zebra finches showed stuttering or omitted syllables when auditory feedback was delayed (Cynx and von Rad, 2001; Leonardo and Konishi, 1999). Both songs and calls showed increased amplitude when auditory feedback was masked (Brumm and Todt, 2002; Cynx et al., 1998). Spectral modifications to songs occurred when normal auditory feedback was modified, e.g. pitch shifted (Brainard and Doupe, 2000; Sober and Brainard, 2009). Similar changes in the structure of echo-locating calls in bats have been reported when their ultrasonic vocalizations were jammed by conspecific vocalizations (Chiu et al., 2008; Gillam et al., 2007; Luo and Moss, 2017; Ulanovsky et al., 2004).

Evidence for feedback-dependent vocal production in nonhuman primates is sparse. A few studies have shown that deafened monkeys develop normal vocal structure which suggest

* Corresponding author. Department of Biomedical Engineering, Johns Hopkins University School of Medicine, 720 Rutland Ave., Traylor 410, Baltimore, MD, 21205, USA.

E-mail address: xiaoqin.wang@jhu.edu (X. Wang).

¹ Equal contribution.

that monkey vocalizations followed innate templates, rather than being actively maintained in vocal production (Hammerschmidt et al., 2001; Winter et al., 1973). Some studies have shown that nonhuman primates can alter the amplitude of their vocalizations in the presence of background noises (Lombard effect) or the temporal structure of their vocal utterance when the auditory feedback was modified (Eliades and Miller, 2017; Ruch et al., 2018). The Lombard effect was reported in several species such as macaques (Sinnott et al., 1975), cotton-top tamarins (Egnor and Hauser, 2006) and marmosets (Brumm, 2004; Eliades and Wang, 2012). Reported changes in temporal structure of vocalizations included vocal onset time (Roy et al., 2011), duration of a vocalization (Brumm, 2004; Pomberger et al., 2018), the number of phrases and inter-phrase intervals of multi-phrase calls (Egnor et al., 2006; Miller et al., 2003; Pomberger et al., 2018). However, it has remained largely unclear whether non-human primates can engage in real-time vocal control mediated by auditory feedback as shown in humans, songbirds and bats.

Our laboratory has been investigating these questions using the common marmoset (*Callithrix jacchus*) as a model system. Marmosets are a New World primate species which maintains frequent vocal communications between conspecifics even in captivity (Agamaite et al., 2015; Bezerra and Souto, 2008; Miller and Wang, 2006). In the past decade, we have developed antiphonal calling (Miller and Wang, 2006) and adapted noise perturbation (Brumm, 2004; Egnor et al., 2006; Miller et al., 2003; Roy et al., 2011) techniques to probe marmosets' vocal production behaviors. Using these techniques, we showed that marmosets were able to control the pace and initiation time of their vocalizations while they engaged in vocal exchanges with conspecifics (Roy et al., 2011). The noise perturbation technique has also been used to probe vocal behaviors in bats (Hage et al., 2013; Takahashi et al., 2014; Tressler and Smotherman, 2009). A recent study using perturbation noises found that single phrases of marmoset phee calls can be interrupted (Pomberger et al., 2018).

In our previous experiments, perturbation noises were delivered at constant or predictable intervals or randomly while a marmoset engaged in the antiphonal calling behavior (Roy et al., 2011). Results from this series of experiments showed that marmosets can "navigate" through perturbation noises by adjusting the initiation time of their vocalizations so that most of their calls fell into the silence gaps between perturbation noises. In the present study, we employed a different strategy and delivered perturbation noises during a phee call (shortly after its initiation) which effectively modified auditory feedback signals received by the test subjects. In this case, marmosets cannot avoid being overlapped with the perturbation noises. This approach was meant to probe marmoset's ability (or inability) to make rapid adjustments to the spectrotemporal parameters of their vocalizations. These experiments showed that marmosets were able to alter particular spectral and temporal parameters of their vocalizations in real-time and, furthermore, some of these perturbation-induced changes persisted even in the absence of perturbation noises. The findings from the present study extended observations reported in a recent study that used perturbation noises in probing marmoset vocal production (Pomberger et al., 2018). Preliminary findings from the present study were previously reported in conferences (Zhao et al., 2013).

2. Material and methods

2.1. Subjects

The subjects in this experiment were six adult marmosets (three males, subject ID: 9000, 9002, 22U; three females, subject ID: 15T, 69V, 6207A; ages ranging from 3 to 7 years old). They were housed

in a captive colony at the Johns Hopkins University School of Medicine (lights on from 7am to 9pm). All subjects were single housed within close proximity of other marmosets within the colony. The subjects were maintained on a diet consisting of monkey chow, fruit and yogurt with ad libitum access to water. No food or water restrictions were applied to any test subjects during the reported experiments. Experiments were conducted over a period of five months between the hours of 0800 and 1800. All experimental procedures were approved by the Johns Hopkins University Animal Care and Use Committee and in compliance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals 8th Edition.

2.2. Behavioral paradigm and apparatus

The experimental setup is illustrated in Fig. 1. The behavioral experiment and acoustic recordings were made within a sound attenuating recording chamber (Roy and Wang, 2012). In each experimental session, one subject was transported to the chamber from the colony in an opaque transport cage and tested in a wire mesh recording cage ($60 \times 30 \times 30$ cm). A loudspeaker (Cambridge Soundworks, M80, North Andover, MA, USA) was placed 2 m in front of the recording cage and was used to present perturbation signals. Two directional microphones (Sennheiser, ME66, Old Lyme, CT, USA) were used to record the vocalizations from the subject and the perturbation signals presented by the speaker. Mic 1 was placed 50 cm in front of the recording cage. Signals from the microphones were amplified by a dual-channel preamplifier (Symetrix, 302, Mountlake Terrace, WA, USA) and then digitized and recorded through a sound card (M-Audio, 1010LT, Cumberland, RI, USA) at a sample rate of 44.1 kHz. The sound level of the perturbation signals was calibrated by a hand-held sound level meter (Brüel & Kjær Type 2250, Nærum, Denmark) with a 1/2 inch prepolarized free field microphone (Brüel & Kjær Type 4189). After a session was finished, the subject was transported back to the colony to its home cage.

Subjects were tested in two types of experimental sessions,

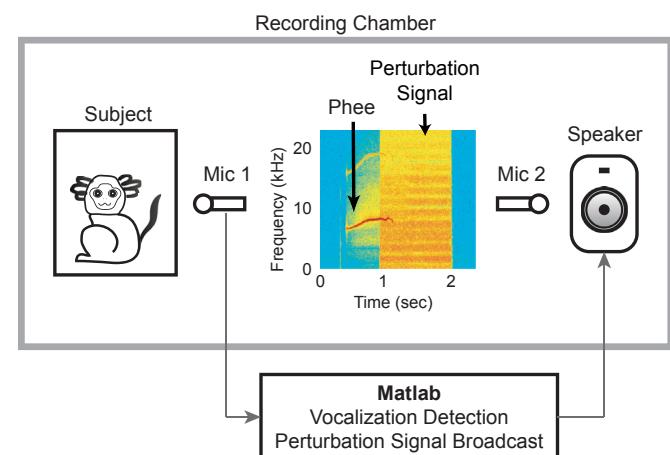


Fig. 1. Schematic of the experimental setup. A test cage (shown on the left) was placed in the sound-attenuating recording chamber. A speaker (shown on the right) was used to broadcast perturbation signals to the subject. A microphone (Mic 1) was placed near the test cage to record the subject's vocalizations. A second microphone (Mic 2) was placed near the speaker to monitor the perturbation signals. Acoustic signals from Mic 1 were monitored by a Matlab program in real time to detect the onset of vocal production. The same program controlled the presentation of perturbation signals after vocal onset. The spectrogram illustrated a phee call detected by the program and a white noise perturbation signal that was broadcasted at a given delay from the onset of the call.

perturbation sessions and baseline sessions. Baseline sessions were recorded two weeks prior to the first exposure of the marmoset to the perturbation signals. Then perturbation sessions with all types of perturbation signals were recorded. For each subject, there were 1–2 baseline sessions and 2–6 perturbation sessions. Repeated sessions have a time interval of 1–5 days in between. Each experimental session lasted 30 min in which one subject was tested. A custom-written Matlab program continuously monitored the subject's vocalizations, synthesized the perturbation signals, and controlled the delivery of perturbation signals. In the perturbation sessions, the program detected the onset of the subject's vocalizations and broadcast the perturbation signal at a given latency for a random subset of the vocalizations. The vocalization detection was processed online using criteria of amplitude threshold and band limited energy within the typical marmoset vocalization frequency range (5–12 kHz). Test subjects generated phee calls spontaneously in the recording chamber. The parameters in the Matlab program used for vocal detection were set according to the sound level of phee calls recorded in the testing chamber and optimized before the perturbation sessions so that all vocalizations were successfully detected. One of four types of perturbation signals, white noise (WN), tone (T), broad-band noise (BBN) or narrow-band noise (NBN) was presented at 75 dB SPL (A-weighting, measured 1 m from the speaker) through the speaker. Perturbation signals were presented with a range of delays with respect to call onset, for a duration of one second in most cases (a small subset had shorter durations of 250 ms; in initial sessions in two subjects). There appeared to be no major differences in vocal behavior in response to short (250 ms) or long (1 s) perturbations. However, we did observe that the short perturbation sometimes seemed to elicit a stress response from the animals (they appeared to be agitated). This was one of the reasons that we did not use the short perturbation in all animals. The tone was played at 500 Hz above or below the fundamental frequency of on-going calls. The narrow band noise was band limited to half of the bandwidth of the fundamental frequency range ($[F_{\max} - F_{\min}]/2$) of phee calls and the broad-band noise was band limited to the entire fundamental frequency range ($F_{\max} - F_{\min}$) (F is the fundamental frequency of a phee call which typically ranges from 5 to 12 kHz). These parameters of perturbation signals were configured in the Matlab program. During each perturbation session, the program generated perturbation signals according to the ongoing vocalizations depending on the perturbation signal type.

Phee calls produced in the baseline sessions were referred to as "baseline" condition. Phee calls produced in the perturbation sessions that received perturbation were referred to as "perturbed" condition. Phee calls that did not receive perturbation but were produced in the perturbation sessions were referred to as "not-perturbed" condition.

2.3. Data analysis

The original acoustic recordings, including vocalizations and perturbation signals, were de-noised using the referenced noise filtering method as described previously (Roy et al., 2011). To quantify the shortening of phrase length (Fig. 3A), we only analyzed calls with phrase length less than the maximum baseline phrase length of each marmoset. This was done to disambiguate the effect of prolonged phrases that some marmosets produced (Fig. 7, see section 3.3). Additionally, the analysis was limited to phrases where the perturbation signal delay was less than half of the mean baseline phrase length. In order to quantify the spectral changes of the phee calls, the rate of fundamental frequency changes was calculated in 100 ms non-overlapping windows. The frequency slope of a call was defined as the mean absolute rate of frequency

changes within that call. When comparing multiple groups of data, one-way ANOVA was applied and *post-hoc* analysis with Bonferroni correction was used to report significant difference for multiple comparisons.

3. Results

The results reported here were based on a total of 1638 phee calls (2492 phrases) recorded from six marmosets that were collected during 2010–2011 when the reported experiments were carried out (see Table 1 for detailed number of calls for each marmoset in each condition). Perturbation noises that were time-locked to the onset of a phee call resulted in several types of vocal alterations as described below.

3.1. Alterations of temporal parameters of the vocal structure

The phee call is a type of most commonly observed marmoset vocalizations and typically composed of 1–3 phrases (Agamaite et al., 2015). When marmosets engaged in antiphonal calling with a real or simulated conspecific, they almost exclusively produce phee calls (Miller et al., 2009; Miller and Wang, 2006). Fig. 2A–D (top row) shows examples of phee calls produced by four test subjects. When these four subjects were tested with the perturbation noises, we found surprisingly that they all terminated some of phee calls before a call phrase was completed (Fig. 2A–D; bottom row, red arrow). This was observed for the other two subjects as well.

In perturbation sessions, perturbation signals were delivered randomly to only 50% of calls. The average phrase length of the perturbed phee calls was significantly shorter than that of phee calls recorded in baseline sessions (perturbed: 0.9 s, baseline: 1.36 s, $p = 0.002$, One-way ANOVA with *post-hoc* analysis, Fig. 3A). Surprisingly, the mean phrase length of not-perturbed calls in perturbation sessions was also shorter than that of the baseline calls (not-perturbed: 1.1 s, $p = 0.005$, One-way ANOVA with *post-hoc* analysis, Fig. 3A), albeit longer than that of perturbed calls.

We also tested the effect of the delay in delivering the perturbation signals and found that the majority of phee calls terminated within 200 ms after a perturbation signal was delivered between ~50 ms and ~1000 ms after the call onset (Fig. 3B). This means that marmosets can cut short a phee call phrase literally at any time point during a phee phrase. The trend of the data in Fig. 3A is also seen when the number of phrases per phee call was measured. Compared to baseline sessions, the average number of phrases per phee call was significantly reduced for perturbed calls in perturbation sessions (perturbed: 1.4 phrases, baseline: 2.2 phrases, $p = 0.004$, One-way ANOVA with *post-hoc* analysis). Interestingly, the average number of phrases per phee call for not-perturbed calls in perturbation sessions was also shortened, similar to the perturbed calls (not-perturbed: 1.5 phrases, $p = 0.024$, One-way ANOVA with *post-hoc* analysis).

3.2. Alterations of spectral parameters of the vocal structure

In addition to changes in temporal parameters of phee calls such as phrase length (Fig. 3), we also observed changes in spectral parameters during perturbation. For example, some phee calls did not terminate after a perturbation signal was presented but rather displayed abrupt shifts in the contour of the fundamental frequency (Fig. 4A and B). The direction of the shifts can be either upward (Fig. 4A) or downward (Fig. 4B). Such changes were rare if ever observed in the baseline condition.

Phee calls are usually produced such that the fundamental frequency increases over the duration of a call (e.g., Fig. 4A). We

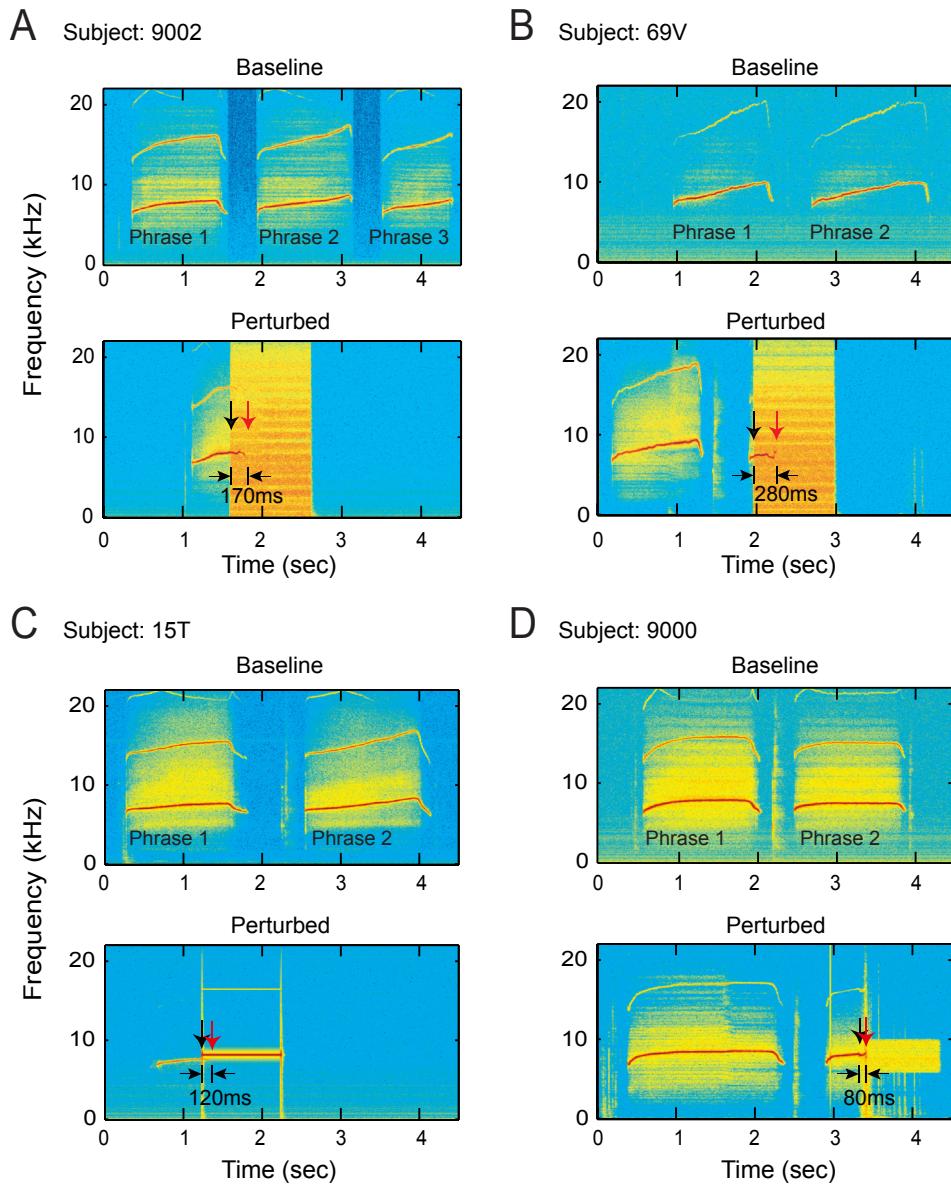


Fig. 2. Examples of regular phee phrases in baseline condition (top) and phee phrases with early termination (red arrows) soon after perturbation signals started (black arrows) in the perturbation condition (bottom). Subject IDs are shown on the top of each panel: (A) 9002; (B) 69V; (C) 15T; (D) 9000. The time intervals indicated by vertical bars and horizontal arrows in the bottom panels are the phrase length after perturbation signal start. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

characterize this trend by calculating the rate of frequency change per unit time, i.e., the difference in the fundamental frequency across a small time window divided by the length of the window. To quantify the abrupt shifts in the fundamental frequency, we calculated the mean absolute rate of frequency changes (i.e., frequency slope, see Section 2.3, Data Analysis) of phee calls in perturbed, not-perturbed and baseline conditions and found that perturbed calls had a significantly higher frequency slope than calls of the other two conditions ($p < 0.05$, One-way ANOVA with post-hoc analysis, Fig. 5A). There was no significant difference in frequency slope between the not-perturbed and baseline conditions ($p = 0.47$, One-way ANOVA with post-hoc analysis, Fig. 5A). In perturbation sessions, if a phee phrase's frequency slope was higher than the maximum frequency slope in the baseline condition for that subject, we label it as having a significant change in frequency

slope. There were 480 phrases in total that had a significant change in frequency slope. The proportion of phrases having a significant change in frequency slope was different for the four types of perturbation signals ($p = 0.0003$, One-way ANOVA with post-hoc analysis, Fig. 5B). White noise induced the largest proportion (58%). We then examined the direction of abrupt frequency shifts. We compared the rate of frequency change before and after a perturbation signal started (± 200 ms window) for the phrases with a significant change in frequency slope. There was a significantly higher chance ($p = 0.0028$, t -test) for the rate of frequency change to decrease after a perturbation signal started. On average $24.67 \pm 3.2\%$ of perturbed calls showed a decreased rate of frequency change (Fig. 4B) whereas $6.93 \pm 1.1\%$ showed an increased rate of frequency change (Fig. 4A) immediately after perturbation signals started.

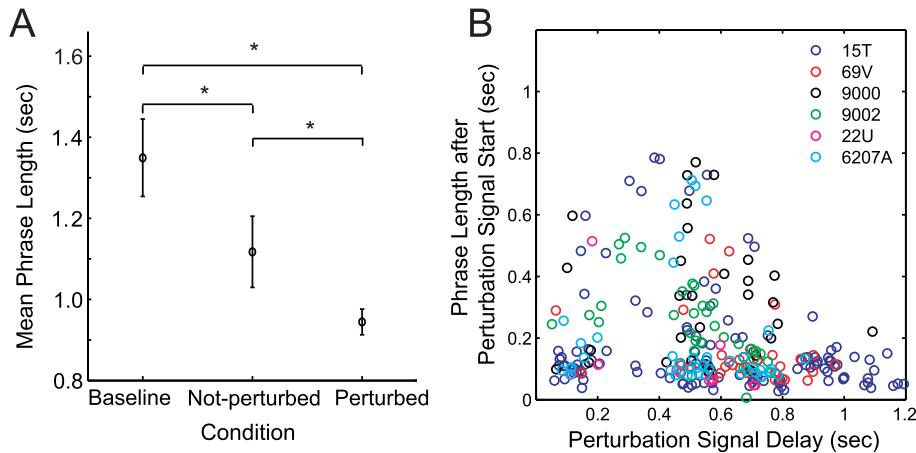


Fig. 3. (A) Statistical summary of the phrase length (mean \pm SEM) of phee calls in the baseline, not-perturbed and perturbed conditions. Asterisks indicate significant difference. (B) Relationship of phrase length after perturbation signal onset and the perturbation signal delay. The color labels individual subjects. The prolonged phrases are excluded (see Fig. 7). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1

Number of Calls [Number of Phrases] {Number of calls used in call length analysis}. Numbers in bold fonts are the total numbers for each subject or each condition.

Animal ID	15T	69V	9000	9002	22U	6207A	Total
Baseline Sessions	23 [38]	21 [35]	20 [35]	20 [55]	23 [45]	53 [89]	160 [297]
Perturbation sessions							
Not-perturbed	171 [197] {115}	88 [95] {70}	86 [121] {35}	167 [274] {50}	174 [280] {88}	79 [102] {75}	765 [1069] {433}
WN	106 [114] {90}	146 [150] {134}	74 [84] {50}	70 [77] {49}	121 [133] {78}	88 [88] {87}	605 [646] {488}
Tone	94 [104] {70}	67 [67] {65}	17 [17] {13}	59 [61] {29}	19 [19] {10}	0 [0] {0}	256 [268] {187}
BBN	5 [5] {4}	23 [23] {23}	22 [22] {15}	23 [23] {11}	16 [16] {12}	0 [0] {0}	89 [89] {65}
NBN	42 [43] {34}	24 [24] {24}	20 [20] {13}	23 [23] {18}	13 [13] {7}	0 [0] {0}	122 [123] {96}
Total calls [phrases] {calls in length analysis} in perturbation sessions	360 [463] {313}	330 [359] {316}	162 [264] {126}	214 [458] {157}	249 [461] {195}	163 [190] {162}	1478 [2195] {1269}

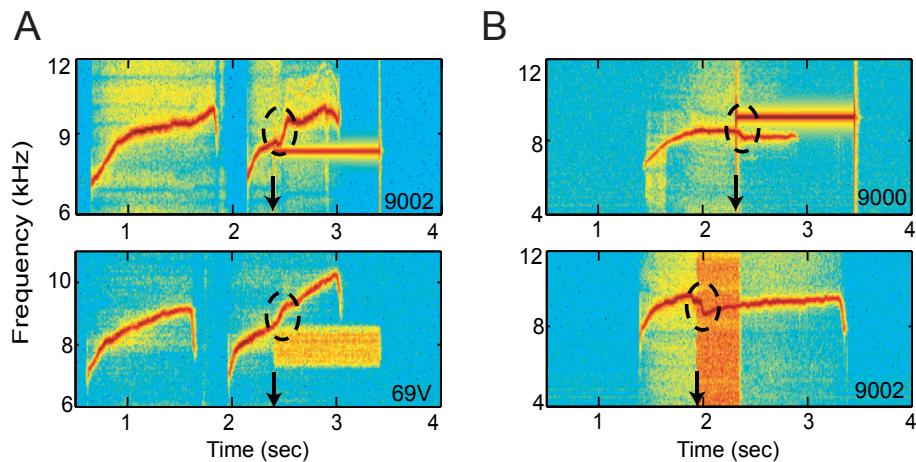


Fig. 4. Examples of abrupt fundamental frequency change, with upward (A) or downward directions (B). The arrows indicate the start time of perturbation signals and the dashed oval indicates the abrupt change in the fundamental frequency. Subject IDs are shown at the lower right corner.

3.3. Fragmented phrases and prolonged phrases

Two out of the six subjects produced phee vocalizations that were fragmented. Examples of the fragmented phee calls produced

by subjects 9000 (male, 7 years old) and 15T (female, 5 years old) are shown in Fig. 6 (A, B). A total of 98 fragmented calls were recorded (subject 15T: 68 calls, subject 9000: 30 calls). We define a call as fragmented if there were more than one fragment within a

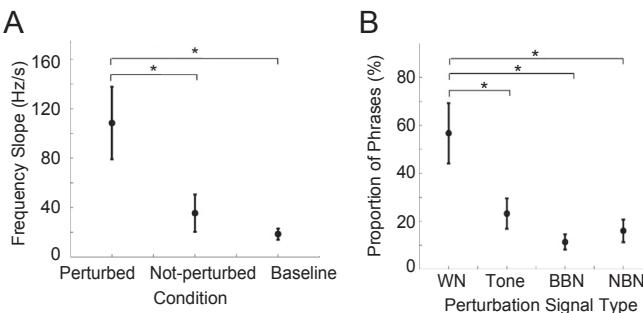


Fig. 5. (A) Statistical summary of the frequency slope, i.e. mean absolute rate of frequency changes (mean \pm SEM) in the baseline, not-perturbed and perturbed conditions. Asterisks indicate significant difference. (B) The proportion of phrases having a significant change in frequency slope (mean \pm SEM) for four different perturbation signal types. WN: white noise; BBN: broad-band noise; NBN: narrow-band noise.

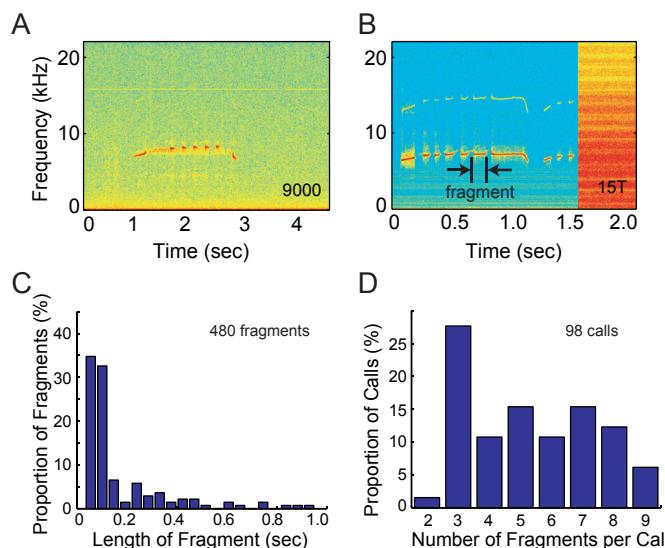


Fig. 6. (A) An example of a fragmented phee call that was not perturbed produced by subject 9000 during a perturbation session. (B) An example of a fragmented phee call that was perturbed by a white noise perturbation signal produced by subject 15T. The perturbation signal started in what appeared to be the second phrase of the call. One of the fragments is illustrated by arrows. (C) Distribution of fragment lengths measured in 480 fragments from 98 phee calls produced by two out of the six marmosets that were tested. (D) Distribution of the number of fragments per call.

single phrase of a phee call. The fragmented calls were observed only during the perturbation sessions but not in the baseline sessions. Both single and multi-phrased fragmented phee calls were observed. The fragmented phee calls were primarily observed when a marmoset was subjected to continuous call perturbations. The temporal structures of the fragmented calls were highly variable within the callers and between them. More than 70% of the fragments measured less than 200 ms in length (Fig. 6C). Fig. 6 (D) shows the distribution of the number of fragments per call. Calls with three fragments were observed most often (~25%) among all possible fragment numbers. The maximum number of observed fragments per call was nine.

While the majority of perturbed phee calls showed shortened phrase length, a small number of phee calls showed unusual long phrase length. If a phee phrase was longer than the maximum phrase length of the same subject in the baseline condition, we defined it as a prolonged phee phrase (see individual examples in Fig. 7A and B). We observed prolonged phee phrases in five out of

six test subjects. Similar to the shortened phrases, marmosets sometimes produced prolonged phee phrases when the calls were not perturbed. Two subjects (15T, 22U) even produced slightly more prolonged phrases in the not-perturbed condition than the perturbed condition (Fig. 7C).

4. Discussion

Using an acoustic perturbation technique, the present study provided three pieces of evidence that shed light on vocal control mechanisms in adult marmoset monkeys. First, we observed rapid modifications of marmoset phee calls in both temporal (Fig. 2) and spectral (Fig. 4) structures. Second, our data provided evidence for potential voluntary vocal control by marmosets, based on changes in vocal structure in not-perturbed condition (Fig. 3A). Third, we showed that marmoset produced highly unusual fragmented phee calls, suggesting a greater extent of flexibility in vocal production by marmosets than previously thought.

4.1. Voluntary vocal control

Previous studies have shown that tamarins and marmosets can voluntarily control the initiation of vocal production (Egnor et al., 2007; Roy et al., 2011). However, little evidence exists on whether marmosets can modify the spectrotemporal structure of their vocalizations (see recent review (Ruch et al., 2018). Although the change of vocal duration could result from non-cognitive modulatory effects, such as arousal level and context (Liao et al., 2018;

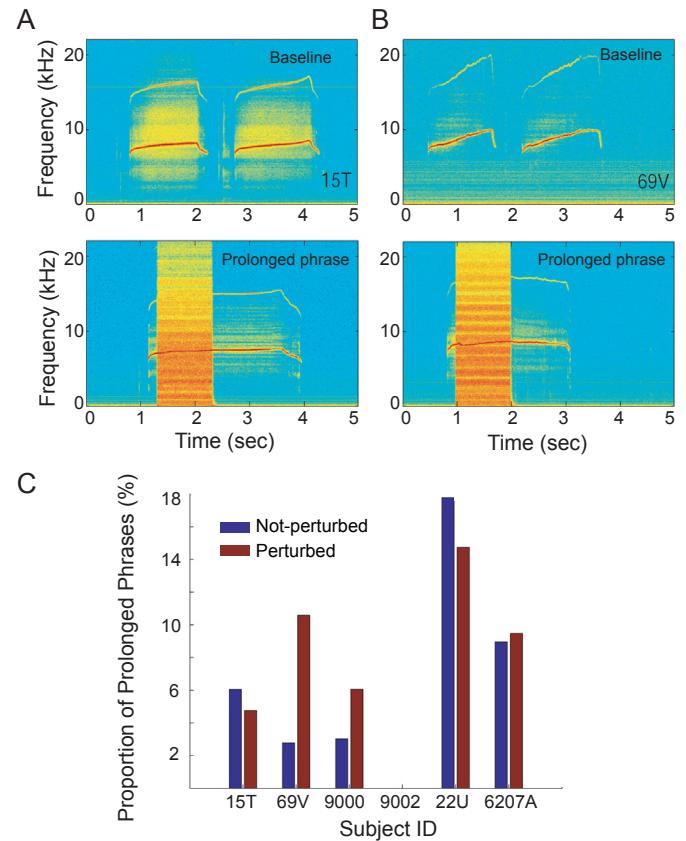


Fig. 7. (A) An example of baseline call phrases (top) and a phrase perturbed by white noise (bottom) from subject 15T that showed prolonged duration (greater than maximum baseline phrase length). (B) Same format with (A), from subject 69V. (C) The proportion of prolonged phrases in the not-perturbed and perturbed conditions for each subject.

Norcross et al., 1999) or audio-vocal integration mechanisms in subcortical brain regions (Luo et al., 2018), we provided in this study one important piece of intriguing evidence for potential voluntary vocal control in marmoset monkeys, based on the analyses of not-perturbed calls in the perturbation sessions. For the not-perturbed condition, since no perturbation signal was presented when a marmoset vocalized, one would not necessarily expect a change in vocal parameters as in the perturbed condition. To our surprise, both the number of phrases per phee call and the phrase length dropped in the not-perturbed condition (1.5 phrases, 1.1 s) compared to that in the baseline condition (~2.2 phrases, 1.36 s), showing the same trend as in the perturbed condition (~1.4 phrases, 0.9 s) (Fig. 3A). This suggests that marmosets may have learned the context of the perturbation sessions and adjust their vocalizations in anticipation of the perturbation signals even when their vocalizations were not perturbed. It also argues against the possibility that the shortening of perturbed phee calls result simply from a reflex of the vocal production system when a perturbation signal is heard by the subject. An alternative interpretation is that vocal modifications in perturbation sessions (in the presence or absence of perturbation signals) might have resulted from other factors such as states of arousal. One way to test whether the state of arousal could explain the perturbation-induced vocal modifications is to show if such modifications persist outside the test session. A recent study from our laboratory with a different perturbation paradigm has provided the evidence to support the notion that marmosets have the ability to voluntarily modify their vocal production (Zhao et al., 2019). In this study, we showed that marmosets exhibited directional long-lasting spectral shifts in their vocalizations long after they had experienced a particular type of interfering sounds.

4.2. Comparison with previous studies

By using perturbation signals, we observed that the number of phrases of phee calls was reduced during perturbation sessions compared to that in the baseline sessions, consistent with what has been shown in tamarins (Egnor et al., 2006; Miller et al., 2003). It has been suggested in previous studies that a phrase of a monkey vocalization is the basic "unit" in its vocal organization and was considered unbreakable (Miller et al., 2003). Surprisingly, most of the perturbed calls in our experiments terminated soon after the perturbation signal was delivered and some terminated even within 50 ms (Fig. 3B). These incomplete calls were outside of the distribution observed in naturally vocalizing marmosets in our colony (Agamaite et al., 2015). We have never observed such partial phee calls before. The fact that marmosets can cut short a phee call phrase at almost arbitrary time points indicates that a phee phrase is not the smallest unit of vocal production.

A recent study using a noise perturbation technique has found shortened and fragmented phee calls as well (Pomberger et al., 2018). The results of Pomberger et al. (2018) study are similar to some of the results of the present study. There are, however, important differences between these two studies due to differences in experimental designs and analysis methodology. The present study measured and compared call parameters between three conditions: baseline, perturbed and not-perturbed. Pomberger et al. (2018) study did not measure parameters of the calls produced prior to the perturbation sessions (i.e., the baseline condition in the present study). By comparing the not-perturbed condition to the baseline condition, we were able to reveal the evidence of anticipation in vocal production by marmosets in the perturbation sessions (i.e., modifications of not-perturbed calls), which suggested the ability of voluntary vocal control. This is an important observation that has not been reported by previous studies. In

Pomberger et al. (2018) study, perturbation signals were presented to 2/3 of phee calls produced by a marmoset and the acoustic parameters of these calls were compared to the rest 1/3 of calls (referred to as the "control condition", equivalent to the not-perturbed condition in the present study).

Moreover, the overall phrase length (population average) of the perturbed calls did not seem to be shortened in Pomberger et al. (2018) study when comparing to the not-perturbed calls, whereas in the present study the phrase length of perturbed calls was shorter than that of not-perturbed calls (Fig. 3A). We postulate that this may be due to how perturbation signals were delivered. In Pomberger et al. (2018) study, perturbation signals always started right after the onset of a call which may cause marmosets to decide either to immediately terminate a phrase or complete the originally planned phrases. The latter turned out to be the majority of the cases (see Fig. 1E in Pomberger et al. (2018)). In our experiments, the starting time of perturbation signals varied with a range of delays after the onset of a phee call (Fig. 3B). This uncertainty may cause marmosets to produce shorter calls in anticipation even if when no perturbation signals were presented. Finally, we observed abrupt changes to the spectral structure of phee calls in the forms of sudden shifts in the fundamental frequency and prolonged call phrases, which were not reported in Pomberger et al. (2018) study. A recent study using pitch-shifted auditory feedback has found compensatory changes in the fundamental frequency of trill and trillphee calls (Eliades and Tsunada, 2018). In that experiment, marmosets heard an auditory feedback of its own production with slight frequency difference and presumably adjusted their production because they sensed an error between the motor goal and the sensory input. The result of the present study on noise perturbed phee calls provides complementary evidence for vocal control, as it demonstrates that marmosets have the ability to produce rapid and sharp frequency changes during ongoing vocal production.

4.3. Implications for vocal behaviors

The perturbation paradigm in our study revealed alterations of marmoset vocal production on two different timescales. On the longer timescale, we showed changes in the number and length of phee call phrases in the perturbation sessions compared with the baseline sessions. On the shorter timescale, we showed rapid termination of phrases and abrupt shifts in the fundamental frequency of phee calls. These observations suggest that marmoset vocal control may have two components on different timescales with one reflecting a cognitive level processing for planning and the other reflecting a rapid sensorimotor processing.

Studies in songbird using altered feedback reported that the earliest temporal shortening and spectral alterations occurred less than 100 ms from the onset of feedback perturbations (Sakata and Brainard, 2006). This suggested that auditory feedback in songbirds has immediate access to the motor production of song syllables. Here we reported that marmosets were also able to dramatically shorten the length of the individual phrases during the perturbation sessions. Overall, a majority of calls were terminated within 200 ms of the onset of the perturbation signal and the earliest possible termination occurred at a latency of approximately 50 ms (Fig. 3B). Such a short latency between perturbation onset and call termination is comparable to the observations in songbirds.

We observed fragmented calls in two out of six marmosets (one male and one female; both at similar age with the other four marmosets). Fragmented phee calls have been reported in previous studies in marmosets in different behavioral situations. Most of the fragmented phee calls in Pomberger et al. (2018) study were observed when marmosets were motivated to vocalize by food

rewards. Fragmented calls in Zürcher and Burkart (2017) study were observed in marmoset colony environment. In both cases, it was unclear whether the fragmented calls were produced in a specific social/training context or can be induced by particular acoustic perturbation.

In the present study, fragmented calls were only found within perturbation sessions and were typically observed when a marmoset's calls were repeatedly perturbed (>5 times in a sequence). Therefore, the appearance of fragmented phee calls in the present study seemed to be associated with the noise perturbation context. It is not clear whether the mechanism that generates fragmenting phee calls is the same as that responsible for generating shortened phee calls. Because the fragmented phee calls could be observed after the termination of perturbation signals, marmosets appeared to be able to produce such unusual calls in anticipation of perturbation signals. The observation of the fragmented phee calls support the notion that the basic unit of marmoset vocal production is smaller than the length of a phee phrase (~1 s). Collectively, the observations from the present and previous studies suggest that fragmented phee calls in marmosets may result from a variety of behavioral conditions.

Together, our results indicated that marmoset monkeys have a greater flexibility in their vocal production than previously thought. The ability for the voluntary control of their vocal production and the rapid modulation in the spectrotemporal parameters may enable richer vocal behaviors during social communication. For example, marmosets may be able to make fast modifications to the duration of calls or the contour of the fundamental frequency in particular behavioral contexts. Observations from the present study and previous studies further demonstrate that the marmoset is a promising non-human primate model for studying vocal communication behaviors and underlying neural mechanisms.

Acknowledgment

S.R. and X.W. designed the study; S.R. collected and analyzed data; L.Z. performed further analyses and prepared figures; L.Z. and X.W. wrote the manuscript. We thank Jenny Estes and Nate Sotuyo for assistance with animal care and Sarah Smith for help with experiments and data analysis. This work was supported by NIH grants DC005808 and DC014503.

References

- Agamaite, J.A., Chang, C.-J., Osmanski, M.S., Wang, X., 2015. A quantitative acoustic analysis of the vocal repertoire of the common marmoset (*Callithrix jacchus*). *J. Acoust. Soc. Am.* 138, 2906. <https://doi.org/10.1121/1.4934268>.
- Bezerra, B.M., Souto, A., 2008. Structure and usage of the vocal repertoire of *Callithrix jacchus*. *Int. J. Primatol.* 29, 671–701. <https://doi.org/10.1007/s10764-008-9250-0>.
- Brainard, M.S., Doupe, A.J., 2000. Auditory feedback in learning and maintenance of vocal behaviour. *Nat. Rev. Neurosci.* 1, 31–40. <https://doi.org/10.1038/35036205>.
- Brumm, H., 2004. Acoustic communication in noise: regulation of call characteristics in a New World monkey. *J. Exp. Biol.* 207, 443–448. <https://doi.org/10.1242/jeb.00768>.
- Brumm, H., Todt, D., 2002. Noise-dependent song amplitude regulation in a territorial songbird. *Anim. Behav.* 63, 891–897. <https://doi.org/10.1006/anbe.2001.1968>.
- Burnett, T.A., Freedland, M.B., Larson, C.R., Hain, T.C., 1998. Voice F0 responses to manipulations in pitch feedback. *J. Acoust. Soc. Am.* 103, 3153–3161. <https://doi.org/10.1121/1.423073>.
- Chen, S.H., Liu, H., Xu, Y., Larson, C.R., 2007. Voice F0 responses to pitch-shifted voice feedback during English speech. *J. Acoust. Soc. Am.* 121, 1157. <https://doi.org/10.1121/1.2404624>.
- Chiou, C., Xian, W., Moss, C.F., 2008. Flying in silence: echolocating bats cease vocalizing to avoid sonar jamming. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13116–13121. <https://doi.org/10.1073/pnas.0804408105>.
- Cynx, J., Lewis, R., Tavel, B., Tse, H., 1998. Amplitude regulation of vocalizations in noise by a songbird, *Taeniopygia guttata*. *Anim. Behav.* 56, 107–113. <https://doi.org/10.1006/anbe.1998.0746>.
- Cynx, J., von Rad, U., 2001. Immediate and transitory effects of delayed auditory feedback on bird song production. *Anim. Behav.* 62, 305–312. <https://doi.org/10.1006/anbe.2001.1744>.
- Egnor, S.E.R., Hauser, M.D., 2006. Noise-induced vocal modulation in cotton-top tamarins (*Saguinus oedipus*). *Am. J. Primatol.* 68, 1183–1190. <https://doi.org/10.1002/ajp.20317>.
- Egnor, S.E.R., Iguina, C.G., Hauser, M.D., 2006. Perturbation of auditory feedback causes systematic perturbation in vocal structure in adult cotton-top tamarins. *J. Exp. Biol.* 209, 3652–3663. <https://doi.org/10.1242/jeb.02420>.
- Egnor, S.E.R., Wickelgren, J.G., Hauser, M.D., 2007. Tracking silence: adjusting vocal production to avoid acoustic interference. *J. Comp. Physiol. A* 193, 477–483. <https://doi.org/10.1007/s00359-006-0205-7>.
- Eliades, S.J., Miller, C.T., 2017. Marmoset vocal communication: behavior and neurobiology. *Dev. Neurobiol.* 77, 286–299. <https://doi.org/10.1002/dneu.22464>.
- Eliades, S.J., Tsunada, J., 2018. Auditory cortical activity drives feedback-dependent vocal control in marmosets. *Nat. Commun.* 9, 2540. <https://doi.org/10.1038/s41467-018-04961-8>.
- Eliades, S.J., Wang, X., 2012. Neural correlates of the lombard effect in primate auditory cortex. *J. Neurosci.* 32, 10737–10748. <https://doi.org/10.1523/JNEUROSCI.3448-11.2012>.
- Fairbanks, G., 1955. Selective vocal effects of delayed auditory feedback. *J. Speech Hear. Disord.* 20, 333. <https://doi.org/10.1044/jshd.2004.333>.
- Fairbanks, G., Guttman, N., 1958. Effects of delayed auditory feedback upon articulation. *J. Speech Hear. Res.* 1, 12–22. <https://doi.org/10.1044/jshr.0101.12>.
- Gillam, E.H., Ulanovsky, N., McCracken, G.F., 2007. Rapid jamming avoidance in biosonar. *Proc. Biol. Sci.* 274, 651–660. <https://doi.org/10.1098/rspb.2006.0047>.
- Hage, S.R., Jiang, T., Berquist, S.W., Feng, J., Metzner, W., 2013. Ambient noise induces independent shifts in call frequency and amplitude within the Lombard effect in echolocating bats. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4063–4068. <https://doi.org/10.1073/pnas.1211533110>.
- Hammerschmidt, K., Jürgens, U., Freudentstein, T., 2001. Vocal development in squirrel monkeys. *Behaviour* 138 (9), 1179–1204. <https://doi.org/10.1163/156853901753287190>.
- Hickok, G., Houde, J., Rong, F., 2011. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69, 407–422. <https://doi.org/10.1016/j.neuron.2011.01.019>.
- Houde, J.F., Jordan, M.I., 1998. Sensorimotor adaptation in speech production. *Science* 279, 1213–1216.
- Larson, C.R., Burnett, T.A., Kiran, S., Hain, T.C., 2000. Effects of pitch-shift velocity on voice F0 responses. *J. Acoust. Soc. Am.* 107, 559–564. <https://doi.org/10.1121/1.428323>.
- Lee, B.S., 1950. Effects of delayed speech feedback. *J. Acoust. Soc. Am.* 22, 824–826. <https://doi.org/10.1121/1.1906696>.
- Leonardo, A., Konishi, M., 1999. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature* 399, 466–470. <https://doi.org/10.1038/20933>.
- Liao, D.A., Zhang, Y.S., Cai, L.X., Ghazanfar, A.A., 2018. Internal states and extrinsic factors both determine monkey vocal production. *Proc. Natl. Acad. Sci. U.S.A.* 115, 3978–3983. <https://doi.org/10.1073/pnas.1722426115>.
- Luo, J., Hage, S.R., Moss, C.F., 2018. The lombard effect: from acoustics to neural mechanisms. *Trends Neurosci.* 41, 938–949. <https://doi.org/10.1016/j.tins.2018.07.011>.
- Luo, J., Moss, C.F., 2017. Echolocating bats rely on audiovocal feedback to adapt sonar signal design. *Proc. Natl. Acad. Sci. U.S.A.* 114, 10978–10983. <https://doi.org/10.1073/pnas.1711892114>.
- Miller, C.T., Beck, K., Meade, B., Wang, X., 2009. Antiphonal call timing in marmosets is behaviorally significant: interactive playback experiments. *J. Comp. Physiol. A. Neuroethol. Sens. Neural. Behav. Physiol.* 195, 783–789. <https://doi.org/10.1007/s00359-009-0456-1>.
- Miller, C.T., Flusberg, S., Hauser, M.D., 2003. **Interruptibility of long call production in tamarins: implications for vocal control.** *J. Exp. Biol.* 206, 2629–2639.
- Miller, C.T., Wang, X., 2006. Sensory-motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. *J. Comp. Physiol. A. Neuroethol. Sens. Neural. Behav. Physiol.* 192, 27–38. <https://doi.org/10.1007/s00359-005-0043-z>.
- Norcross, J.L., Newman, J.D., Cofrancesco, L.M., 1999. Context and sex differences exist in the acoustic structure of phee calls by newly-paired common marmosets (*Callithrix jacchus*). *Am. J. Primatol.* 49, 165–181. [https://doi.org/10.1002/\(SICI\)1098-2345\(199910\)49:2<165::AID-AJP7>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1098-2345(199910)49:2<165::AID-AJP7>3.0.CO;2-S).
- Pomberger, T., Risueno-Segovia, C., Löschner, J., Hage, S.R., 2018. Precise motor control enables rapid flexibility in vocal behavior of marmoset monkeys. *Curr. Biol.* 28, 788–794. <https://doi.org/10.1016/j.cub.2018.01.070> e3.
- Roy, S., Miller, C.T., Gottsch, D., Wang, X., 2011. Vocal control by the common marmoset in the presence of interfering noise. *J. Exp. Biol.* 214, 3619–3629. <https://doi.org/10.1242/jeb.056101>.
- Roy, S., Wang, X., 2012. Wireless multi-channel single unit recording in freely moving and vocalizing primates. *J. Neurosci. Methods* 203, 28–40. <https://doi.org/10.1016/j.jneumeth.2011.09.004>.
- Ruch, H., Zürcher, Y., Burkart, J.M., 2018. The function and mechanism of vocal accommodation in humans and other primates. *Biol. Rev.* 93, 996–1013. <https://doi.org/10.1111/brv.12382>.
- Sakata, J.T., Brainard, M.S., 2006. Real-time contributions of auditory feedback to avian vocal motor control. *J. Neurosci.* 26, 9619–9628. <https://doi.org/10.1523/JNEUROSCI.2027-06.2006>.
- Sinnott, J.M., Stebbins, W.C., Moody, D.B., 1975. Regulation of voice amplitude by the

- monkey. *J. Acoust. Soc. Am.* 58, 412. <https://doi.org/10.1121/1.380685>.
- Smotherman, M.S., 2007. Sensory feedback control of mammalian vocalizations. *Behav. Brain Res.* 182, 315–326. <https://doi.org/10.1016/j.bbr.2007.03.008>.
- Sober, S.J.S., Brainard, M.M.S., 2009. Adult birdsong is actively maintained by error correction. *Nat. Neurosci.* 12, 927–931. <https://doi.org/10.1038/nn.2336>.
- Takahashi, E., Hyomoto, K., Riquimaroux, H., Watanabe, Y., Ohta, T., Hiryu, S., 2014. Adaptive changes in echolocation sounds by *Pipistrellus abramus* in response to artificial jamming sounds. *J. Exp. Biol.* 217.
- Tressler, J., Smotherman, M.S., 2009. Context-dependent effects of noise on echolocation pulse characteristics in free-tailed bats. *J. Comp. Physiol. A* 195, 923–934. <https://doi.org/10.1007/s00359-009-0468-x>.
- Ulanovsky, N., Fenton, M.B., Tsoar, A., Korine, C., 2004. Dynamics of jamming avoidance in echolocating bats. *Proc. Biol. Sci.* 271, 1467–1475. <https://doi.org/10.1098/rspb.2004.2750>.
- Villacorta, V., Perkell, J., Guenther, F., 2004. Sensorimotor adaptation to acoustic perturbations in vowel formants. *J. Acoust. Soc. Am.* 115, 2430. <https://doi.org/10.1121/1.4781504>.
- Waldstein, R.S., 1990. Effects of postlingual deafness on speech production: implications for the role of auditory feedback. *J. Acoust. Soc. Am.* 88, 2099. <https://doi.org/10.1121/1.400107>.
- Winter, P., Handley, P., Ploog, D., Schott, D., Winter, P., Handley, P., 1973. Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour* 47, 230–239. <https://doi.org/10.1163/156853973X00085>.
- Zhao, L., Rad, B.B., Wang, X., 2019. Long-lasting vocal plasticity in adult marmoset monkeys. *Proc. R. Soc. B Biol. Sci.* 286, 20190817. <https://doi.org/10.1098/rspb.2019.0817>.
- Zhao, L., Roy, S., Wang, X., 2013. Evidence of voluntary vocal control by the common marmosets (*Callithrix jacchus*). *Soc. Neurosci. Abstr.* 44, 584.16.
- Zürcher, Y., Burkart, J.M., 2017. Evidence for dialects in three captive populations of common marmosets (*Callithrix jacchus*). *Int. J. Primatol.* 38, 780–793. <https://doi.org/10.1007/s10764-017-9979-4>.