

Distinct vocal strategies to cope with ambient noise in marmoset monkeys

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von

Julia Löschner

aus Stuttgart

Tübingen

2023

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	26.10.2023
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Steffen Hage
2. Berichterstatter/-in:	Prof. Dr. Jan Benda

Erklärung

Ich erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel:

„Distinct vocal strategies to cope with ambient noise in marmoset monkeys“

selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Angabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft wird.

Tübingen, den 16.05.2023

Julia Löschner

Contents

Zusammenfassung	7
Abstract.....	8
I. Synopsis	9
1. Introduction	9
1.1. Different strategies for communicating during noise	9
1.2. The common marmoset	11
1.2.1. Marmosets as a model system	12
1.2.2. Marmoset vocal behavior	14
2. Aim of the studies	16
3. Results	17
3.1. Chapter I: Increase in signal-to-noise ratio	17
3.1.1. Effects of ambient noise on call frequency	17
3.1.2. Effects of ambient noise on call amplitude	18
3.1.3. Summary	19
3.2. Chapter II: Noise avoidance strategies	19
3.2.1. Effects of ambient noise on call duration.....	20
3.2.2. Effects of ambient noise on syllable amount	22
3.2.3. Summary	22
4. Discussion.....	23
4.1. Comparison to previous studies on vocal behavior in noise	23
4.2. Neural principles of communication	27
4.2.1. Audio-vocal integration mechanisms	27
4.2.2. Possible neural networks in marmoset monkeys	28
4.2.3. Vocal pattern generating network	30
4.3. Conclusion and outlook	31

References.....	33
Acknowledgements.....	40
II. Individual studies.....	41
Statement of contributions	41
Publication 1: Compensatory mechanisms affect sensorimotor integration during ongoing vocal motor acts in marmoset monkeys.....	42
Publication 2: Marmoset monkeys use different avoidance strategies to cope with ambient noise during vocal behavior	57

Zusammenfassung

Jede Übertragung von Lauten ist mit Beeinträchtigungen wie starkem Regen, Wind, Tierlauten oder Stadtgeräuschen konfrontiert. Infolgedessen haben sich mehrere Strategien entwickelt, um Störgeräusche während der Lautproduktion zu kompensieren, was zu verschiedenen Veränderungen der temporalen und spektralen Eigenschaften führt. Die bekannteste lärmbedingte Anpassung ist der Lombard-Effekt, eine unwillkürliche Erhöhung der Lautstärke, die oft mit Änderungen der Lautdauer und Tonhöhe einhergeht. Eine andere Strategie besteht darin, die Lautproduktion auf Zeitpunkte zu beschränken, in denen kein Lärm vorhanden ist. Mithilfe von akustischem Rauschen, welches durch das Lautverhalten selbst ausgelöst wurde, konnten wir zeigen, dass Marmosetten in der Lage sind, die Rufamplitude und -frequenz als Reaktion auf Störgeräusche, die nach Beginn des Rufs auftreten, schnell zu modulieren. Der stärkste Anstieg der Tonhöhe wurde bei höherem Lärmpegel festgestellt. Überraschenderweise unterlagen die Phee-Laute nicht dem Lombard-Effekt. Stattdessen verringerten unsere Affen ihre Lautstärke mit zunehmendem Lärmpegel. In einem weiteren Experiment haben wir Phees systematisch zu verschiedenen Zeitpunkten mit Rauschen gestört. Dadurch haben wir Veränderungen im Lautverhalten festgestellt, die sowohl reflexives als auch adaptives Verhalten als Reaktion auf das Rauschen belegen. Die Marmosetten brachen ihre Rufe unmittelbar nach Einsetzen des Rauschens ab. Die Abnahme der Rufdauer begann bereits während des ersten maskierten Lautes. Im Gegensatz dazu blieb die Reduzierung der Silbenanzahl über das Rauschen hinaus bestehen, was auf ein adaptives Verhalten hinweist. Mit Hilfe von maschinellem Lernen, welches auf Rufparametern basierte, konnten wir feststellen, dass ein Teil der einzelsilbigen Laute, die während und nach der Rausch-Phase produziert wurden, ursprünglich als Doppelsilben geplant waren und nach der ersten Silbe aktiv unterbrochen wurden. Insgesamt deuten diese Ergebnisse darauf hin, dass Krallenaffen verschiedene parallele Mechanismen nutzen, um mit Umgebungslärm umzugehen. Sie zeigen vokale Anpassungen als direkte Reaktion auf störende Geräusche, wie z.B. Veränderungen der Lautstärke und Tonhöhe. Darüber hinaus wenden sie Strategien zur Lärmvermeidung an, d. h. sie unterdrücken ihre Laute in Zeiten erhöhter Störgeräusche.

Abstract

Any transmission of vocal signals faces the challenge of acoustic interferences such as heavy rain, wind, and animal or urban sounds. Consequently, multiple strategies have evolved to compensate for masking noise during vocal behavior, leading to several changes in temporal and spectral call features. One prominent noise-related call adjustment is the Lombard effect, an involuntary increase in call amplitude in response to masking noise, which is often accompanied by changes in call duration and frequency. Another strategy involves limiting call production to periods where noise is absent. Using acoustic perturbation triggered by the vocal behavior itself, we showed that marmosets are capable of rapidly modulating call amplitude and frequency in response to perturbing noise bursts presented after call onset. The strongest rise in call frequencies were found for high noise amplitudes. Surprisingly, phee calls did not exhibit the Lombard effect as previously reported. Instead, our monkeys decreased their call intensity with increasing noise intensity. Furthermore, we showed that marmosets are capable of producing calls with durations beyond the natural boundaries of their repertoire by interrupting ongoing vocalizations rapidly after noise onset. This finding suggests a general strategy of avoiding to call in noisy environments. We systematically perturbed ongoing vocalizations with noise presented at different time points and detected changes in vocal behavior that supported both reflexive and adaptive behavior in response to noise perturbation. Marmosets canceled their calls immediately after noise onset. The decrease in call duration started during the first perturbed call. In contrast, the reduction in number of syllables persisted beyond noise perturbation, indicating adaptive behavior in response to perturbing noise. Using machine learning techniques based on call parameters, we found that a fraction of single phees uttered during and after noise perturbation were initially planned as double phees and became actively interrupted after the first syllable. Altogether, these findings indicate that marmosets use different parallel mechanisms to cope with ambient noise. They show vocal adjustments as a direct response to perturbing noise, such as a decrease in call amplitude and changes in call frequency. Additionally, they use noise avoidance strategies, i.e., suppressing vocalizations, during periods of elevated ambient noise levels.

I. Synopsis

1. Introduction

Compared to human language, the way how animals communicate seems to be quite simple. For over 50 years, monkey vocalizations have been thought to be largely innate, highly affective, and stereotyped (Hammerschmidt et al., 2001; Jürgens, 2002). Recently, this perception has dramatically changed. Current studies have demonstrated distinct learning mechanisms during vocal development (Gultekin & Hage, 2017; Takahashi et al., 2017; Takahashi et al., 2015), as well as vocal flexibility, which allows monkeys to cognitively control when (Hage & Nieder, 2013; Roy et al., 2011), where (Choi et al., 2015), and what to vocalize (Hage & Nieder, 2013; Price et al., 2015; Seyfarth et al., 1980). Despite this flexibility, certain call features, such as duration and frequency remain surprisingly robust and consistent. As a result, the monkeys' vocalizations are characterized by stereotyped and discrete call patterns (Agamaite et al., 2015).

1.1. Different strategies for communicating during noise

Communication between animals is a crucial trait for evolutionary success (Balter, 2010). Vocal signals have evolved as one of the dominant forms of direct communication among individuals in several bird and mammal species (Ackermann et al., 2014; Charlton et al., 2019; Hage & Nieder, 2016; Hammerschmidt, 2008; Janik & Knörnschild, 2021; Jürgens, 2002). However, vocal communication often occurs in the presence of competing sound sources. For effective communication, the transmission of a signal produced by a sender must be detected and decoded by one or more receivers (Bradbury & Vehrencamp, 1998). Thus, the sender must be able to adjust the temporal and spectral features of the signal to overcome any potential masking from ambient noise and allow for effective signal transmission (Brumm & Slabbekoorn, 2005).

The mechanisms used to deal with acoustic disturbances can generally be divided into two main types. The first mechanism involves modulating vocalizations as soon as they directly encounter an increase in ambient noise, such as when they are produced during acoustic perturbation. These vocal modifications can occur either involuntarily or under volitional control. One of the most effective mechanisms to increase the signal-to-noise ratio in call production is the so-called Lombard effect. This effect refers to the involuntary increase in call amplitude in response to masking ambient noise (Brumm & Zollinger, 2011; Eliades & Wang, 2012; Hage et al., 2013; Lombard, 1911; Luo et al., 2018; Pomberger et al., 2020) and has been observed in many vertebrate species, from fish to frogs to birds to mammals, including humans (Brown et al., 2021; Brumm & Slabbekoorn, 2005; Brumm & Zollinger, 2011; Luo et al., 2018; Manabe et al., 1998; Stowe & Golob, 2013). This suggests that the Lombard effect is an evolutionarily old behavior that may have emerged about 450 million years ago. This effect is often accompanied by several other changes, such as a shift in call frequency and a change in vocal density, for example, with longer call durations and/or increased repetitions of syllables (Brumm et al., 2004; Courter et al., 2020; Hage et al., 2013; Luo et al., 2015; Osmani & Dooling, 2009; Pomberger et al., 2020; Tressler & Smotherman, 2009).

Besides the mechanism of modulating vocalizations to improve signal transmission in noisy environments, there is another vocal control strategy that prevents animals from producing vocalizations during noisy events and encourages them to vocalize during quieter periods. This approach involves limiting call emission to times when noise is low, absent, or predictable (Brumm, 2006; Roy et al., 2011; Zelick & Narins, 1982). This eliminates the need for call parameter modifications that might still be unable to increase the signal-to-noise ratio sufficiently and reduces the physiological cost of call production at high intensities. Recently, we demonstrated that marmosets are able to interrupt ongoing vocalizations directly after noise onset (Pomberger et al., 2018), suggesting that marmosets can avoid calling in noise on a rapid time scale. This finding challenges long-held concepts regarding vocal pattern generation (Egnor et al., 2006; Miller et al., 2003; Miller et al., 2009a) and highlights the flexibility and adaptability of vocal communication in response to environmental noise.

1.2. The common marmoset

The common marmoset (*Callithrix jacchus*) is a small New World primate endemic to the forests of northeastern Brazil (Abbott et al., 2003; Bezerra & Souto, 2008; Hubrecht, 1985). Adult marmosets typically measure 20–30 cm and weigh 350–400 g. They have distinctive long banded tails, white ear tufts, and a white patch on their forehead. The fur of the common marmoset is multicolored (**Figure 1**).



Figure 1: Image of a common marmoset.

Marmosets primarily feed on plant exudates (20–70%) and insects (24–30%), but also consume fruits, seeds, flowers, fungi, nectar, snails, lizards, bird eggs, nestlings, and infant mammals (Abbott et al., 2003). Unlike many other nonhuman primates, marmosets form stable family groups consisting of around 9 members on average (Tardif et al., 2003). Furthermore, they are also known for their advanced vocal communication abilities (Pistorio et al., 2006), possessing a diverse range of vocalizations with a diverse vocal repertoire (Agamaite et al., 2015; Bezerra & Souto, 2008). that they use to communicate within their family groups.

A marmoset family typically includes one or two breeding females, a breeding male, their offspring and adult relatives. In the event of the breeding male's death, family members merge into new groups (Lazaro-Perea, 2001). Dominance in common marmoset groups is maintained through various behaviors, postures, and vocalizations, such as subordinate members grooming their superiors. However, the social rank within the group is based on age (Digby, 1995).

Most marmosets live monogamously (Mansfield, 2003). Female marmosets usually give birth to two dizygotic twins per delivery. The time between deliveries is around five months, allowing them to give birth twice a year. Because female marmosets need to nurse their infants during the following pregnancy, the breeding pair suppresses the reproduction of the other group members (Baker et al., 1999; Saltzman et al., 1997). Thus, the male partner and other members of the group care for the infants together. The infants reach adult size and sexual maturity at 15 months, but cannot breed until they become dominant (Yamamoto, 1993).

1.2.1. Marmosets as a model system

New World primates (*platyrrhines*) including the common marmoset diverged from Old World primates (*catarrhines*) around 35 million years ago. This separation led to the evolution of adaptations to the neotropical environment among New World species, as well as differences in physiology and susceptibility to disease (Abbott et al., 2003; Mansfield, 2003). Although macaques are more closely related to humans from an evolutionary standpoint, some of the characteristics of marmosets are more akin to those of humans than macaques, likely due to the physical separation of these primate groups (Mansfield, 2003).

Common marmosets are considered a suitable laboratory animal for biomedical research due to their ease of handling, animal welfare, practicality, and scientific suitability. The marmoset has advantages over the macaques in terms of animal welfare and practicality. They are available for laboratory use from well-established captive colonies in national primate research centers, academic institutions, and commercial breeding facilities. Unlike macaques, marmosets do not carry Herpes B virus (Mansfield, 2003), which is a benefit to their handlers. Additionally, their smaller size results in lower costs for caging and feeding, and they require less floor space than macaques. Consequently, using common marmosets is a more cost-effective option which can lead to significant cost savings in equivalent experiments (Smith et al., 2001).

Breeding marmosets in laboratory environments is highly efficient and straightforward. The marmoset's ovarian cycle lasts around 28 days, with ovulation typically occurring around the tenth day, exhibiting hormonal profiles similar to those

of humans (Summers et al., 1985). Female marmosets usually give birth to two or three offspring per delivery and can have two deliveries annually. Consequently, they can have a relatively high number of deliveries (20–30 over their lifespan) and offspring (40–80 over their lifespan), giving them a reproductive advantage compared to other nonhuman primates. For instance, it takes approximately 5 years for macaques to reach sexually maturity, and they only give birth to one offspring per year (Fischer & Austad, 2011). Therefore, while it is possible to obtain three macaque pups over three years, one female marmoset can reasonably provide 14 marmoset pups during the same time frame (Okano et al., 2012). This exceptional reproductive efficiency provides a significant advantage in the development of transgenic animals.

Marmosets are the most common nonhuman primates used for animal research (Abbott et al., 2003) due to their suitability as model organisms in a wide range of areas, such as neurobiology, reproduction, immunology, endocrinology, obesity, and aging, as well as neurodegenerative diseases (Tardif et al., 2011). They display age-related pathologies that are similar to those seen in humans, such as cancer, amyloidosis, diabetes, and chronic renal disease (Tardif et al., 2011). One of the key advantages of using marmosets in research is that their brains share a number of similarities with the human brain, such as an expanded temporal lobe and hierarchically structured sensory cortices (Hackett, 2011; Mitchell & Leopold, 2015), as well as a highly developed prefrontal cortex (Roberts et al., 2007). Although the marmoset brain is much smaller than the brains of macaques and humans (180 times smaller in volume) (Stephan et al., 1981), it is proportional to their weight and neuron number (Azevedo et al., 2009). Moreover, their brain's smooth surface makes it easy to target (Tokuno et al., 2015), which is advantageous for areal mapping, laminar electrode penetration, and two-photon and optical imaging (Mitchell & Leopold, 2015). Marmosets also share various behavioral and cognitive traits with humans, such as pair bonding, cooperative breeding, prosocial behavior, and complex vocal communication (Dell'Mour et al., 2009; Eliades & Wang, 2008; Gordon & Rogers, 2010).

1.2.2. Marmoset vocal behavior

Similar to many New World primate species, marmosets have a wide range of vocalizations that are used in different social contexts, such as social interaction, alarm, mobbing, and food-related calls (Agamaite et al., 2015; Bezerra & Souto, 2008). The most extensively studied vocalization in common marmosets is the phee call, which has been the focus of numerous studies on behavior and neurobiology (Chen et al., 2009; Eliades & Wang, 2008; Jones et al., 1993; Löschner et al., 2023; Miller & Wang, 2006; Miller et al., 2009a; Miller et al., 2009b; Norcross & Newman, 1993; Norcross & Newman, 1997; Norcross et al., 1994; Pistorio et al., 2006; Pomberger et al., 2020; Pomberger et al., 2018). This call serves to attract mates, keep groups together, defend territories, and locate missing group members (Jones, 1997).

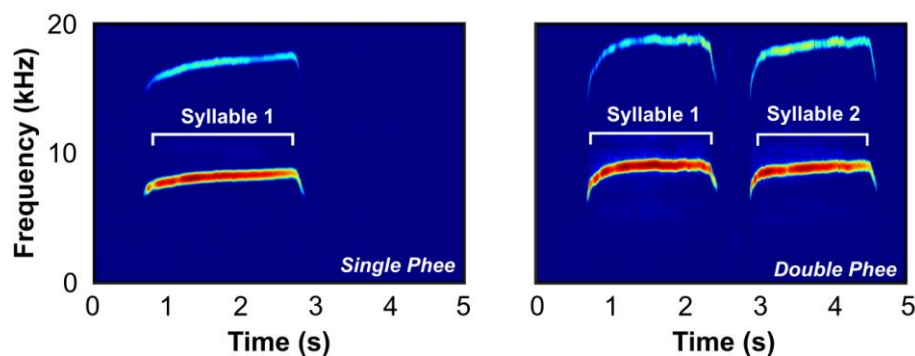


Figure 2: Exemplar sonogram of phee calls. Phees are long-distance contact calls, composed of one (single phees), two (double phees), or more phee syllables, to interact with conspecifics.

However, phee calls are the typical long-distance contact calls produced during antiphonal calling and consist of a series of whistle-like syllables with relatively constant frequency (**Figure 2**). Furthermore, studies have demonstrated that phee calls can be reliably classified based on caller's individual identity (Jones et al., 1993; Miller et al., 2010), their sex (Norcross & Newman, 1993), and their social group (Miller et al., 2010). This indicates that there is some level of vocal control and adaptive learning is involved in the production of these calls, similar to what has been observed in other marmoset species (Elowson & Snowdon, 1994; Snowdon & Elowson, 1999). Interestingly, even marmosets living in captive colonies in close proximity can develop these group-level phee call dialects. Prior to producing a phee

call, marmosets establish a motor plan that includes the entire acoustic structure of the call, suggesting that specific acoustic features are not random but may instead represent deliberate control over the vocal motor output (Miller et al., 2009a).

2. Aim of the studies

The aim with these studies was to gain a better understanding of how marmosets cope with ambient noise on a behavioral level and which strategies have evolved to deal with such perturbing noise. As improvements in signal transmission have been found throughout the animal kingdom, we wanted to study changes in call features in the marmoset in more detail. We focused on analyzing phee calls, which are reliably emitted by monkeys when they are separated from the group.

Previous studies that used continuous noise showed that marmoset vocalizations are rather stereotyped. Therefore, we used noise bursts starting after call onset. In a previous study where we perturbed the calls right after call onset, we found that marmosets are capable of interrupting their phee calls in response to noise bursts, but this happened only in rare cases. Therefore, we decided to play back noise at later timepoints throughout the calls to investigate whether the monkeys would cancel more of their phee calls. Additionally, we aimed to examine how quickly changes in vocal behavior occur and whether they happen in a reflexive or adaptive manner.

We hypothesize that marmosets are able to change their call features in ongoing calls and that the monkeys' vocal behavior is highly dependent on the signal-to-noise ratio. With higher noise amplitudes, and frequency bands overlapping with the fundamental frequency of the phees, the monkeys would shift the call frequency upwards and increase their call amplitude in response to noise exposure. Additionally, we expect that the noise onset timing would play an important role in terms of phee cancelation performance.

3. Results

In the following chapters, I will summarize the main results of the two publications included in this thesis. The full publications are attached as chapters at the end of the thesis (see individual studies).

3.1. Chapter I: Increase in signal-to-noise ratio

The aim of this study was to investigate whether marmosets are capable of exhibiting changes in call structure in response to noise disturbances starting after call onset or whether such effects only occur if noise perturbation starts prior to call onset. We analyzed vocal behavior in marmosets held separate in a soundproofed chamber, with and without acoustic perturbation. We used different noise band conditions (broadband noise and bandpass-filtered noise bands) and different noise amplitudes. The noise perturbation was triggered by the vocal behavior itself to test in a controlled experimental design whether marmosets are capable of rapidly modulating distinct vocal parameters such as call frequency and amplitude in ongoing vocalizations. Overall, four common marmosets were successfully used with a total number of 6,298 phee calls recorded.

3.1.1. Effects of ambient noise on call frequency

We first investigated whether and in what way marmosets modify the fundamental frequency of their ongoing phee vocalizations when exposed to varying noise conditions. For ongoing phee syllables, there was an immediate increase in call frequency when the perturbing noise either overlapped with or was higher than the fundamental frequency of the call. Bandpass-filtered noise bursts, which were above the fundamental frequency of the calls but did not mask it, had no effect on call frequency. This indicates that noise-related frequency modulation is a selective process in marmoset monkeys. Furthermore, higher noise amplitudes resulted in a stronger increase in call frequencies. Second phee syllables, however, did not show significant shifts in fundamental frequencies in response to noise

exposure. Moreover, we found that shifts in fundamental frequency were primarily correlated to the different noise bands rather than the amplitude conditions. As a next step, we tested how fast fundamental frequency shifts occurred within the first phee syllables after noise onset, resulting in a mean latency of around 30 ms suggesting a rapid underlying neural mechanism for frequency modulation.

3.1.2. Effects of ambient noise on call amplitude

Although increasing call frequency improves the signal detectability in noisy environments, the most effective mechanism to improve the signal-to-noise ratio during vocal production is the Lombard effect. Therefore, we investigated how noise perturbation affected call amplitudes.

Noise perturbation starting after phee call onset did not systematically affect the call amplitude of the first syllable. Interestingly, in cases in which significant shifts occurred, call amplitude did not increase as expected, but decreased. This effect was stronger for the second syllables of the phee calls completely produced during the noise perturbation. Furthermore, a stronger decrease in call amplitude could be observed for low-frequency noise conditions. Consequently, call intensity decreased in a stepwise function with increasing noise intensity suggesting a direct impact of noise intensity on call amplitude. To investigate whether changes in frequency were a by-product of an accompanied change in amplitude, we tested the correlation between frequency and amplitude across all noise conditions and noise amplitudes used. We found no significant relationship between the two parameters neither for first nor for second syllables, indicating that changes in call frequency and amplitude occurred independently of each other.

After that, we modified our behavioral experimental design to investigate whether our animals are able to show the Lombard effect in general or whether they suppress it when producing phee calls in a noisy environment. Therefore, we played back all five noise conditions continuously for three minutes each, along with a control condition (silence). We observed that one monkey significantly increased the intensity of both phee syllables in response to predictable continuous noise, demonstrating the Lombard effect. Another monkey significantly decreased the

intensity of the second phee syllable but showed no changes in call intensity for the first syllable. However, two of the monkeys showed no significant change in call amplitude. Overall, these findings suggest that marmosets are capable of both exhibiting and actively suppressing the Lombard effect when producing phee calls in a noisy environment.

3.1.3. Summary

This study reports two main findings: first, marmosets can quickly adjust the frequency of their calls in response to ambient noise. Second, their call amplitudes are also affected by noise, but, interestingly, marmosets did not exhibit the Lombard effect that is commonly found in other animals during the utterance of phee calls. Instead, they reduced their call amplitude. However, when the noise was predictable, one monkey showed the Lombard effect. These results suggest that marmosets have a general strategy to avoid calling in the presence of noise and that they are capable of actively counteracting previously thought involuntary audio-vocal mechanisms. This study opens up avenues for investigating the neural mechanisms underlying these behaviors.

3.2. Chapter II: Noise avoidance strategies

While control mechanisms underlying vocal adjustments in response to ambient noise have been already well studied, the aim of this study was to improve our understanding of the mechanisms involved in noise avoidance strategies. In a previous study we demonstrated that marmosets are capable of immediately canceling ongoing vocalizations after noise onset (Pomberger et al., 2018). However, this behavior was observed only in rare instances, ranging from 0.3–7.7% depending on the monkey. It remains unclear whether the ability to interrupt ongoing calls in response to ambient noise is based on either reflexive, adaptive, and/or cognitive processes.

To better understand the strategies used by marmosets when vocalizing in noisy environments, we measured the vocal behavior of marmosets while held separate

in a soundproofed chamber, with and without acoustic perturbation. We perturbed ongoing vocalizations by using 80 dB broadband noise at different time points after the call onset once the monkey exceeded a certain call duration. In this experiment, four common marmosets were successfully used and a total of 7945 phee calls, including 5,843 single and 2,102 double phees were recorded.

3.2.1. Effects of ambient noise on call duration

As a first step we investigated the general effect of noise perturbation on call durations. We found that call durations of first syllables were significantly shorter during phases with noise perturbation compared to before. However, after noise perturbation ended, the call durations returned to baseline levels. To determine how rapidly the monkeys modulated phee duration in response to noise perturbation, we measured the duration of the first phee syllables in relation to when they were uttered within the session. We found that the first call perturbed with noise showed a significant decrease in call duration across sessions, indicating an immediate response to noise perturbation. In other words, marmoset calls were affected immediately after noise onset and this behavior stopped directly after the last call was perturbed.

We then examined whether call durations were dependent on the onset of noise perturbation in relation to call offset. Our results showed that the marmosets' call durations deviated significantly from both perfect compensation to noise perturbation and fully unaffected call durations, indicating a more complex vocal behavior in response to different noise onset times. Therefore, we analyzed the call duration distributions in relation to the different noise onset times more closely. We found that for most noise onset conditions, the distribution of call durations was bimodal, implying that calls were either canceled immediately after noise onset or were rather unaffected by the noise.

Our findings also indicate that it was easier for the animals to cancel their calls towards the expected end of the phee calls than right at the beginning of the vocal onset. These results may suggest evidence of an underlying neural mechanism that could inhibit the interruption of a vocalization at the start of the vocal pattern. This

could also explain the rare incidence of interrupted calls in our previous study, in which ongoing phee calls were perturbed right after call onset (Pomberger et al., 2018). This result would also support the hypothesis that call patterns are more stable at vocal onset and can be more readily modified towards the end.

Next, we investigated how the duration of second syllables of double phees, which were fully perturbed by noise, was affected. We found significant differences in the duration of second syllables across experimental phases. Similar to call duration distributions of the first phee syllables, call durations were shorter during noise perturbation than in the pre-phase (0.48 s shorter than in the pre-phase). Interestingly, this effect was much stronger than that for first syllable durations. Moreover, after noise perturbation (post-phase), durations of the second syllables remained significantly shorter than those in the pre-phase. This suggests that in rare cases where a second syllable was produced after noise perturbation, its duration was considerably shorter. However, the effect of noise perturbation on call duration varied among monkeys, and the change in call duration did not occur systematically. While the two female monkeys showed an increase in first syllable duration in the post-phase compared to the pre-phase, one of the males showed a significant shortening of the first syllable when comparing the pre- and post-phases. The other male showed no differences in the duration of the first syllable between the pre- and post-phases.

As a next step, we wanted to test how quickly the monkeys changed the duration of second syllables in response to noise perturbation. Therefore, we measured the duration of these syllables as a function of when they were produced within the session. Like the first syllables, we observed a significant decrease in duration for the first call perturbed with noise compared to the last call of the pre-phase, indicating an immediate response to noise perturbations. To analyze the persistence of this effect, we compared the duration of the last 10 calls in the post-phase with the duration of the first 10 calls in the pre-phase of the following day's session. We found that the duration of the second phee syllable was significantly longer than the duration of the last calls of the previous session, indicating that the effect was not long lasting.

3.2.2. Effects of ambient noise on syllable amount

We first investigated if the ratio of single to double phee was affected by noise perturbation. Our results agreed with previous studies showing that marmosets produce significantly more single phee when exposed to perturbing noise. This behavior persisted beyond the end of noise phases. This finding suggests that the monkeys modified their vocal behavior in response to noise perturbation by producing fewer double phee, consistent with our previous finding (Pomberger et al., 2018).

Since we found that some of the uttered single phee seemed to be more similar to first syllables of double phee, we wanted to determine whether marmosets ended the call sequence directly after noise onset within the first phee syllable, even if a double phee had been planned originally. Therefore, we used a machine learning classification model that was based on call parameters that were not affected by noise. Our results indicate that some of the uttered single phee were initially intended to be double phee, but were canceled after the first syllable during the noise-phase and even after the noise-phase had ended.

3.2.3. Summary

The second study yielded three important results: first, marmosets exhibited reflexive behavior by decreasing call durations or canceling calls in response to noise, but resumed normal behavior immediately after the noise ended. Second, it seemed easier for the monkeys to cancel their calls towards the expected end of them. Third, the monkeys showed adaptive behavior by reducing the number of syllables, which persisted even after the noise had stopped. In other words, the study showed that marmosets can both adapt their vocal behavior to noise and make immediate reflexive changes to their vocalizations. These findings suggest that marmosets have direct control over their vocal output and can make rapid adjustments in response to environmental noise.

4. Discussion

The ability to adjust various vocal features in response to noise is crucial for ensuring successful communication. In my PhD thesis, I demonstrate that marmosets exhibit this behavior by modulating their call features in response to noise perturbations. This finding suggests that the vocalizations of marmosets are more flexible than previously thought.

4.1. Comparison to previous studies on vocal behavior in noise

This chapter aims to provide an overview of the previous understanding of marmoset vocal flexibility and distinct strategies used to deal with ambient noise. Moreover, it will contextualize these findings in relation to the results presented in my thesis.

Vocal changes as direct responses to perturbing noise, such as involuntary increases in call amplitude (Lombard effect) and associated changes, including increases in call frequency and call duration, have been extensively studied in recent decades across vertebrate species such as birds, reptiles, and mammals, including cetaceans, bats, and primates (Brainard & Doupe, 2002; Choi et al., 2015; Cynx, 1990; Gultekin & Hage, 2017; Hage & Nieder, 2013; Hardman et al., 2017; Miller et al., 2009a; Pistorio et al., 2006; Price et al., 2015; Roy et al., 2011; Seyfarth et al., 1980; Takahashi et al., 2017). The observed changes resulted either in an increase in signal-to-noise ratio, such as the Lombard effect, increase in call frequency or increase in signal density, such as the production of longer calls and/or increased repetition of call syllables.

Two previous studies have observed the Lombard effect in marmosets, which seems to conflict with our findings (Brumm, 2004; Eliades & Wang, 2012). However, this discrepancy may be explained by the different call types investigated in each study. While we focused on phee calls, which are produced at high amplitude intensities (Eliades & Wang, 2012), one of the earlier studies investigated the twitter call, a vocalization that is produced at lower amplitude intensities (Brumm, 2004).

Interestingly, the other study investigating the Lombard effect during phee call production found mixed effects with half of the animals not showing an increase in call amplitude (Eliades & Wang, 2012). While playing back continuous noise also one of our animals showed the Lombard effect, whereas the others did not. Surprisingly, our study found that phee calls perturbed after call onset did not exhibit the Lombard effect as previously reported for calls produced in constantly presented ambient noise (Brumm, 2004; Egnor & Hauser, 2006). Our results indicate that marmoset monkeys may either not exhibit the Lombard effect when producing ongoing phee calls or may suppress it and lower their call intensities instead.

We suggest that marmoset monkeys counteract the Lombard effect in a noisy environment to reduce the physiological costs of high intensity phee calls. Phee calls are produced at intensities above 100 dB SPL, resulting in high muscle tensions encompassing almost the entire animal's body during call production. This might have led to the evolution of mechanisms in these animals that ensure the proper transmission of these highly energetic calls, such as calling in silent gaps or decreasing call intensity in situations in which sufficient detectability might be potentially diminished, such as during the presence of ambient noise. Based on our findings, it appears that marmoset monkeys possess an audio-vocal integration mechanism that is capable of counteracting the Lombard effect. Similar mechanisms have already been observed in vocal production learners like birds and humans (Kobayasi & Okanoya, 2003; Pick et al., 1989; Therrien et al., 2012; Vinney et al., 2016) and seems to be mainly driven by higher-order cognitive processes including cortical structures (Luo et al., 2018).

Whether auditory feedback contributes to the Lombard effect remains a topic of debate. Some research suggests that sensory feedback may not be necessary for eliciting the Lombard effect. In one study, greater horseshoe bats exhibited the Lombard effect in the first call after noise onset, indicating that the bats increase their vocalization amplitude without using auditory feedback (Hage et al., 2013). Other studies have shown that the Lombard effect relies on spectral overlap between vocalizations and background noise. Therefore, it might be possible that auditory feedback is used to extract the spectral information of ongoing vocalizations and compare it with background noise.

Although the Lombard effect is already well studied across the whole animal kingdom, noise-dependent shifts in call frequency have been mostly unattended and poorly understood. Only a few studies have reported a rise in call frequencies with increasing ambient noise levels in birds and bats (Hage et al., 2013; Osmanski & Dooling, 2009; Schuster et al., 2012), while only one study investigated the effect of different noise bands on call frequencies. Recent research on marmosets found that they are able to adjust their phee call frequency to shift away from a predictable high- and low-frequency noise, respectively, which did not overlap with the phee call frequencies (Zhao et al., 2019). In bats, the frequencies of echolocation calls increased significantly in response to a variety of noise stimuli, regardless of whether they were directly masking the call's fundamental frequency or presented below the dominant call frequency (Hage et al., 2013). These findings suggest that the observed rises in call frequencies are likely an audio-vocal mechanism elicited to increase call detectability in a noisy environment, as has been found in previous studies including birds (Andalman & Fee, 2009; Bermudez-Cuamatzin et al., 2011; Charlesworth et al., 2011; Nemeth & Brumm, 2010; Pohl et al., 2012). Shifts in song frequencies of around 200 Hz have been predicted to improve call detectability by about 10–20% (Nemeth & Brumm, 2010), which is mainly due to the fact that the spectrum of environmental noise generally shows a decay in amplitude with increasing frequency (Halfwerk & Slabbekoorn, 2009; Nemeth & Brumm, 2010; Pohl et al., 2009; Pohl et al., 2012). Only echolocating bats have been found to exhibit such fast responses to ambient noise, and increase their call amplitude within approximately 30 ms after noise onset (Luo et al., 2017).

While vocal changes in response to noise have been extensively studied, strategies for avoiding noise, such as suppressing vocalizations during periods of elevated ambient noise levels, have received little attention in comparison. Recent studies have shown that monkeys have the ability to avoid calling in noisy environments and time their calls to silent periods (Miller et al., 2003), as well as terminate call sequences immediately after noise onset (Egnor et al., 2006; Pistorio et al., 2006). They are even able to interrupt ongoing vocalizations directly after the onset of perturbing noise (Pomberger et al., 2018). However, the mechanisms underlying these avoidance strategies remain largely unclear.

The findings shown in this thesis differ from those of previous studies conducted by Egnor et al. (2006) and Miller et al. (2003 & 2009a). Recent studies have revealed a high degree of vocal flexibility in marmosets (Ghazanfar et al., 2019), allowing them to control when (Pomberger et al., 2019; Roy et al., 2011), where (Choi et al., 2015), and what to vocalize (Liao et al., 2018). This vocal flexibility enables them to avoid calling in the presence of environmental noise and initiate their vocalizations mainly during silent periods (Roy et al., 2011). In a previous study, we demonstrated that marmosets tend to interrupt their vocalizations shortly after noise onset when perturbation starts after vocal onset, indicating their inclination to avoid calling in ambient noise (Pomberger et al., 2018). However, such call interruptions account for only 2.6% of all calls, indicating stark neuronal and/or anatomical constraints that limit this behavior.

In our second study, marmosets canceled their calls immediately after noise onset. This behavior was evident in the first perturbed call, indicating a reflexive behavioral response to noise perturbation. Here, the animals were more likely to cancel their calls towards the expected end of the vocalization rather than at the beginning. These results suggest that there may be neural mechanisms that inhibit the interruption of the vocalization at the beginning of the pattern, as described in the study by Sober & Brainard (2009). This could also explain the low occurrence of interrupted calls in our earlier study, in which vocalizations were perturbed immediately after call onset. These findings further support the hypothesis that call patterns are more stable at the beginning of a vocalization and can be modulated towards the end.

Our study found that the occurrence of double phee calls significantly decreased during noise perturbation, which is consistent with previous studies (Egnor et al., 2006; Pistorio et al., 2006). This indicated that the marmosets terminated the call sequence directly after perturbation onset within the first phee syllable, regardless of whether they canceled the first syllable or not. Furthermore, our study revealed that this behavior persisted even after noise perturbation ended, suggesting that the monkeys exhibited adaptive changes in their vocal behavior. Interestingly, our machine learning classification model based on call parameters not directly affected by perturbing noise provided first evidence that some single phees were planned to

be doubles phee's, but were canceled after the first syllable during phases with perturbing noise and also in phases where the noise had already ended. These findings further show that marmosets have direct control over their vocal output and are capable of modulating ongoing vocalizations in a rapid and direct manner.

4.2. Neural principles of communication

In this paragraph I would like to hypothesize the possible neural networks underlying the analyzed vocal behavior in response to noise exposure.

The acoustic structure of monkey calls, much like the vocalizations of most other mammals and non-verbal emotional vocalizations of humans, is largely genetically determined (Eibl-Eibesfeldt, 1973; Geissmann, 1984; Winter et al., 1973). Given the innate character of these motor patterns, it is likely that they are generated subcortically rather than a cortically.

4.2.1. Audio-vocal integration mechanisms

Audio-vocal integration mechanisms, refer to the processes by which the nervous system integrates auditory information with vocal motor output. The integration of auditory and vocal motor signals is crucial for effective vocal communication in animals including humans.

To produce and maintain effective vocal communication, the brain needs to finely control the acoustical parameters of self-generated sounds and integrate sensory and motor signals. Such signals also include feedforward motor commands and action generation to produce sounds, in addition to ongoing adjustments of the motor system in cases where errors occur. This involves monitoring and adjusting speech production in humans in response to changes in acoustic feedback. The Lombard reflex is one such representative feedback mechanism, where speech loudness is adapted to environmental loudness (Hanley & Harvey, 1965). The Lombard effect is widespread in animal groups and is often considered a basic audio-vocal integration phenomenon. Another example for such feedback mechanisms has been shown in a study in which human subjects have received

voice pitched feedback (up- or downward) during phonation, which resulted in a change of fundamental frequency in their vocal signals (Burnett et al., 1998). These two examples illustrate the central role of feedback processing in influencing the motor system. While the behavioral effects of feedback changes on speech are well studied, the underlying neural mechanisms remain speculative. However, there is evidence that neural interactions between the auditory system and the vocal motor system exist both at the brainstem and cortical levels.

4.2.2. Possible neural networks in marmoset monkeys

The structures involved in the control of phee call modulation in response to a noisy environment are likely to contain neurons that exhibit vocal motor activity with short pre-vocal latencies that are inhibited by auditory stimulation. Examples of brain structures that contain such neurons include the ventrolateral prefrontal cortex (Hage & Nieder, 2015), as well as the pontine and medullary reticular formation (Hage et al., 2006). Consistent with previous research (Hage, 2019; Hage & Nieder, 2016), our model proposes that a volitional articulatory motor network originating in the prefrontal cortex cognitively controls the vocal output of a phylogenetically conserved primary vocal motor network, which is predominantly composed of a subcortical neuronal network.

The vocal motor network can be modulated by auditory structures on several cortical and subcortical brain levels (Luo et al., 2018). The neural mechanisms controlling the interruption of calls or modulation of call parameters as a response to perturbing noise likely include both cortical and subcortical structures, as well as corticofugal projections. There are various potential anatomical and neurophysiological plausible audio-vocal loops that might be involved in audio-vocal interactions that could underlie the observed call modifications, including call inhibition and interruption, as well as shifts in call frequency and amplitude (**Figure 3**). As potential hubs in audio-vocal integration (Eliades & Wang, 2012; Eliades & Tsunada, 2018; Hage & Nieder, 2015; Hage et al., 2006), a cortical audio-vocal loop from the auditory cortex to the ventrolateral prefrontal cortex to premotor cortex to the pontine reticular formation could be responsible (Hage & Nieder, 2016).

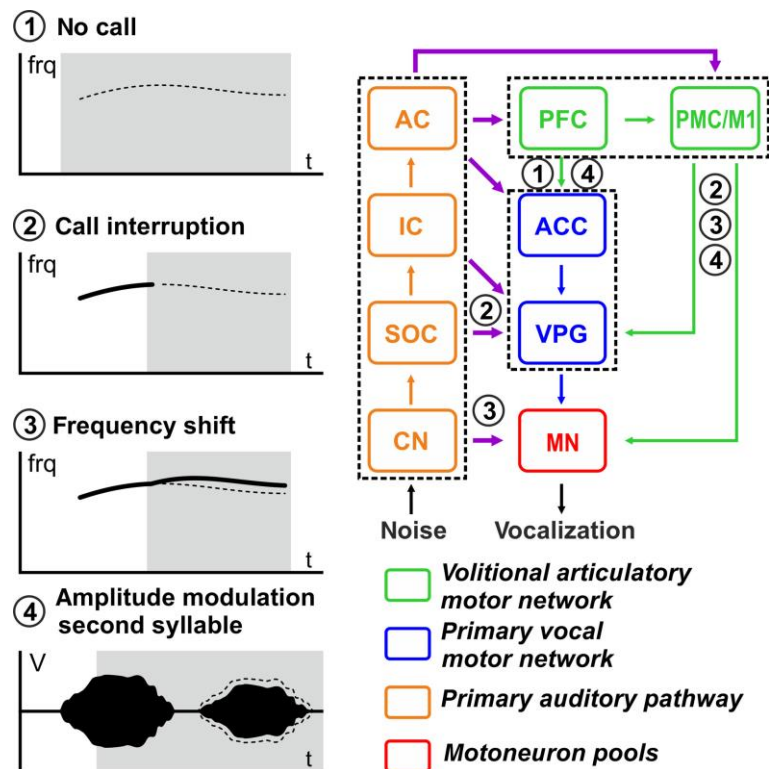


Figure 3: Hypothetical neuronal model for audio-vocal interaction. Audio-vocal integration mechanisms are known to happen between cortical and subcortical structures as well as via corticofugal projections. Call production might be affected by ambient noise at different brain levels. Simplified circuit diagram summarizing the most relevant structures for vocal production and the auditory pathway in monkeys. Arrows indicate anatomically verified and relevant direct connections. Numbers highlight connections that might be involved in noise-related call inhibition (1), call interruption (2), frequency shifts and/or amplitude modulations of the second syllable of call sequences such as phee calls. 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 red) involved in call production. External auditory stimuli are perceived via the auditory pathway (indicated in orange), which is capable of modulating the vocal motor system via direct connections (indicated in purple) at cortical and subcortical levels. See text for further explanation. AC, auditory cortex; ACC, anterior cingulate cortex; CN, cochlear nucleus; IC, inferior colliculus; M1, primary motor cortex; MN, motoneuron pools; PFC, prefrontal cortex; PMC, premotor cortex; SOC, superior olivary complex; VPG, vocal pattern generator in the ventrolateral pontine brainstem.

In addition, it is possible that some call modifications, such as interrupting vocal output or modulating call amplitude, can be achieved through a direct connection from the premotor cortex to single motoneuron pools. In this case, inhibiting specific muscles, such as those involved in expiration, may be sufficient to interrupt vocal output or modulate call amplitude. Another potential subcortical audio-vocal loop involves connections from the cochlear nucleus, superior olivary complex, or inferior

colliculus to the pontine reticular formation, which could mediate some of the observed vocal behaviors such as call interruption or shifts in call frequency (Hage, 2020; Hage & Nieder, 2016). Interestingly, previous studies have identified direct and active connections between the cochlear nucleus and the laryngeal motoneuron pool in mammals, suggesting that such connections may be capable of modulating vocal output directly (Jen & Ostwald, 1977).

4.2.3. Vocal pattern generating network

Vocalizations arise from some of the most complex motor patterns produced by vertebrates. The complexity of the vocalization process makes it challenging to understand its neural underpinnings. Here, I want to hypothesize possible structures involved in determining phee call duration.

In a previous study, we showed that marmoset calls are composed of multiple sequentially uttered units, similar to human speech (Pomberger et al., 2018). In that study, the monkeys were able to interrupt their phees after the first vocal motor unit (less than 100 ms after call onset). Therefore, the duration of a phee call is determined by the number of consecutively produced units rather than a predefined, impartible pulse. This would explain the monkeys' ability to interrupt ongoing phee vocalizations at several moments during vocal production. This only occurs at specific time points, indicating that phee calls can only be interrupted between single units. Similar observations have been made in passeriform birds, where song bouts consist of complex, distinct syllables which are learned during development (Brainard & Doupe, 2002) and can only be interrupted between, not within, syllables (Cynx, 1990; Hardman et al., 2017). Learning processes induced by acoustic perturbation affect acoustic features of the entire song syllable rather than those after the initiation of acoustic perturbation (Sober & Brainard, 2009).

From a neurophysiological perspective, our previous study (Pomberger et al., 2018) suggested the existence of a vocal pattern-generating network that determines phee call duration and may be directly inhibited by perturbing acoustic stimuli. Previous research has indicated that such a vocal pattern generating network is located in the lower brainstem and receives input from higher-order structures (Hage & Nieder,

2016; Jürgens, 2002; Loh et al., 2017). The periaqueductal gray, one of these structures, has shown activity that is correlated with call duration, and may be sufficient for determining phee call duration (Larson, 1991). However, the short latencies of call cancellation after noise onset (<100 ms) and the pre-vocal activity latencies within the PAG (\approx 100 ms) (Düsterhöft et al., 2004; Larson, 1991) make it unlikely that these inputs are sufficient to produce this vocal behavior. These findings suggest that there may be direct interactions between auditory input and a vocal pattern generating network in the brainstem (Hage & Nieder, 2016).

4.3. Conclusion and outlook

In this thesis, we investigated the mechanisms of audio-vocal integration in marmosets, which are highly social and vocal animals. We studied the effects of different noise conditions on the vocalizations of marmosets, specifically the changes in frequency and amplitude of the monkeys' phee vocalizations when perturbed by ambient noise. The results show that noise-related frequency modulation is a selective effect in marmosets, with the strongest rise in call frequencies found for high noise amplitudes. The shifts in fundamental frequency are mainly correlated with different noise bands rather than amplitude conditions. Additionally, we found that noise perturbation starting after phee call onset had no systematic effect on call amplitude of the first syllable, but it decreased call amplitude in a stepwise function with increasing noise intensity, especially for the second syllable. The study also showed that marmosets are capable of exhibiting as well as actively suppressing the Lombard effect in a noisy environment during phee call production.

Furthermore, we investigated the effect of noise perturbation on call durations and syllable amount of marmoset monkeys. We found that call durations of the first syllables were significantly shorter during phases with noise perturbation than before, and that the monkeys immediately canceled their calls after noise onset. We also found that it was easier for the animals to cancel their calls towards the expected end of the phee calls rather than directly at the beginning of the vocal onset. Additionally, we investigated if the ratio of single to double phees was affected by noise perturbation and found that marmoset monkeys emitted

significantly more single phee in phases of perturbing noise and afterwards. Overall, the study adds to previous knowledge about the effects of noise on vocal production in nonhuman primates and provides insights into the mechanisms underlying the Lombard effect.

Our findings raise several questions that require further investigation, such as the neural mechanisms underlying the phee call interruptions and the location of these mechanisms in the marmoset brain. To answer these questions, electrophysiological single cell recordings in freely moving marmosets would be useful to investigate the neural mechanisms and how auditory integration mechanisms interact with vocal motor production. In several species, frequency and amplitude shifts occur on a very fast timescale suggesting audio integration processes at lower brainstem levels. However, a recent study in marmosets revealed that stimulation of auditory cortex in vocalizing animals causes a frequency shift with a latency of about 40 ms.

The findings provided in this thesis give evidence that precise vocal motor control mechanisms, can be studied in marmosets. Therefore, marmosets are a suitable model for studying the evolutionary questions surrounding vocal production in nonhuman primates and speech production in humans. To understand the neural activity in cortical and subcortical structures during vocalization, further experiments in combination with electrophysiological recordings are necessary.

References

- Abbott, D. H., Barnett, D. K., Colman, R. J., Yamamoto, M. E., Schultz-Darken, N. J., 2003. Aspects of common marmoset basic biology and life history important for biomedical research. *Comp Med* 53, 339-50.
- Ackermann, H., Hage, S. R., Ziegler, W., 2014. Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behav Brain Sci* 37, 529-546, doi:10.1017/S0140525X13003099.
- 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, 2906-28, doi:10.1121/1.4934268.
- Andalman, A. S., Fee, M. S., 2009. A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc Natl Acad Sci USA* 106, 12518-12523, doi:10.1073/pnas.0903214106.
- Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., Jacob Filho, W., Lent, R., Herculano-Houzel, S., 2009. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol* 513, 532-41, doi:10.1002/cne.21974.
- Baker, J. V., Abbott, D. H., Saltzman, W., 1999. Social determinants of reproductive failure in male common marmosets housed with their natal family. *Anim Behav* 58, 501-513, doi:10.1006/anbe.1999.1200.
- Balter, M., 2010. Animal communication helps reveal roots of language. American Association for the Advancement of Science.
- Bermudez-Cuamatzin, E., Rios-Chelen, A. A., Gil, D., Garcia, C. M., 2011. Experimental evidence for real-time song frequency shift in response to urban noise in a passerine bird. *Biol Lett* 7, 36-8, doi:10.1098/rsbl.2010.0437.
- Bezerra, B. M., Souto, A., 2008. Structure and Usage of the Vocal Repertoire of *Callithrix jacchus*. *Int J Primatol* 29, 671-701, doi:10.1007/s10764-008-9250-0.
- Bradbury, J. W., Vehrencamp, S. L., 1998. Principles of animal communication.
- Brainard, M. S., Doupe, A. J., 2002. What songbirds teach us about learning. *Nature* 417, 351-358, doi:10.1038/417351a.
- Brown, N. A. W., Halliday, W. D., Balshine, S., Juanes, F., 2021. Low-amplitude noise elicits the Lombard effect in plainfin midshipman mating vocalizations in the wild. *Anim Behav* 181, 29-39, doi:10.1016/j.anbehav.2021.08.025.
- Brumm, H., 2004. The impact of environmental noise on song amplitude in a territorial bird. *J Anim Ecol* 73, 434-440, doi:10.1111/j.0021-8790.2004.00814.x.
- Brumm, H., 2006. Signalling through acoustic windows: nightingales avoid interspecific competition by short-term adjustment of song timing. *J Comp Physiol A* 192, 1279-1285, doi:10.1007/s00359-006-0158-x.
- Brumm, H., Slabbekoorn, H., 2005. Acoustic communication in noise. *Adv Stud Behav* 35, 151-209, doi:10.1016/S0065-3454(05)35004-2.
- Brumm, H., Zollinger, S. A., 2011. The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour* 148, 1173-1198, doi:10.1163/000579511x605759.

-
- Brumm, H., Voss, K., Kollmer, I., Todt, D., 2004. Acoustic communication in noise: regulation of call characteristics in a New World monkey. *J Exp Biol* 207, 443-8, doi:10.1242/jeb.00768.
- 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, doi:10.1121/1.423073.
- Charlesworth, J. D., Tumer, E. C., Warren, T. L., Brainard, M. S., 2011. Learning the microstructure of successful behavior. *Nat Neurosci* 14, 373-380, doi:10.1038/nn.2748.
- Charlton, B. D., Owen, M. A., Swaisgood, R. R., 2019. Coevolution of vocal signal characteristics and hearing sensitivity in forest mammals. *Nat Commun* 10, 1-7, doi:10.1038/s41467-019-10768-y.
- Chen, H. C., Kaplan, G., Rogers, L., 2009. Contact calls of common marmosets (*Callithrix jacchus*): influence of age of caller on antiphonal calling and other vocal responses. *Am J Primatol* 71, 165-170.
- Choi, J. Y., Takahashi, D. Y., Ghazanfar, A. A., 2015. Cooperative vocal control in marmoset monkeys via vocal feedback. *J Neurophysiol* 114, 274-283, doi:10.1152/jn.00228.2015.
- Courter, J. R., Perruci, R. J., McGinnis, K. J., Rainieri, J. K., 2020. Black-capped chickadees (*Poecile atricapillus*) alter alarm call duration and peak frequency in response to traffic noise. *PLoS One* 15, e0241035, doi:10.1371/journal.pone.0241035.
- Cynx, J., 1990. Experimental determination of a unit of song production in the zebra finch (*Taeniopygia guttata*). *J Comp Psychol* 104, 3-10, doi:10.1037/0735-7036.104.1.3.
- Dell'Mour, V., Range, F., Huber, L., 2009. Social learning and mother's behavior in manipulative tasks in infant marmosets. *Am J Primatol* 71, 503-9, doi:10.1002/ajp.20682.
- Digby, L. J., 1995. Social organization in a wild population of *Callithrix jacchus*: II. Intragroup social behavior. *Primates* 36, 361-375, doi:10.1007/BF02382859.
- Düsterhöft, F., Häusler, U., Jürgens, U., 2004. Neuronal activity in the periaqueductal gray and bordering structures during vocal communication in the squirrel monkey. *Neuroscience* 123, 53-60, doi:10.1016/j.neuroscience.2003.07.007.
- Egnor, S. E. R., Hauser, M. D., 2006. Noise-induced vocal modulation in cotton-top tamarins (*Saguinus oedipus*). *Am J Primatol* 68, 1183-1190, doi: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, doi:10.1242/jeb.02420.
- Eibl-Eibesfeldt, I., 1973. The expressive behaviour of the deaf-and-blind-born. *Social communication and movement*, 163-194.
- Eliades, S. J., Wang, X., 2008. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* 453, 1102-1106, doi:10.1038/nature06910.
- Eliades, S. J., Wang, X., 2012. Neural correlates of the Lombard effect in primate auditory cortex. *J Neurosci* 32, 10737-10748, doi:10.1523/JNEUROSCI.3448-11.2012.

-
- Eliades, S. J., Tsunada, J., 2018. Auditory cortical activity drives feedback-dependent vocal control in marmosets. *Nat Commun* 9, 2540, doi:10.1038/s41467-018-04961-8.
- Elowson, A. M., Snowdon, C. T., 1994. Pygmy Marmosets, *Cebuella-Pygmaea*, Modify Vocal Structure in Response to Changed Social-Environment. *Anim Behav* 47, 1267-1277, doi:10.1006/anbe.1994.1175.
- Fischer, K. E., Austad, S. N., 2011. The development of small primate models for aging research. *ILAR J* 52, 78-88, doi:10.1093/ilar.52.1.78.
- Geissmann, T., 1984. Inheritance of song parameters in the gibbon song, analysed in 2 hybrid gibbons (*Hylobates pileatus* x *H. lar*). *Folia primatologica* 42, 216-235.
- Ghazanfar, A. A., Liao, D. A., Takahashi, D. Y., 2019. Volition and learning in primate vocal behaviour. *Anim Behav* 151, 239-247, doi:10.1016/j.anbehav.2019.01.021.
- Gordon, D. J., Rogers, L. J., 2010. Differences in Social and Vocal Behavior Between Left- and Right-Handed Common Marmosets (*Callithrix jacchus*). *J Comp Psychol* 124, 402-411, doi:10.1037/a0019736.
- Gultekin, Y. B., Hage, S. R., 2017. Limiting parental feedback disrupts vocal development in marmoset monkeys. *Nat Commun* 8, 14046, doi:10.1038/ncomms14046.
- Hackett, T. A., 2011. Information flow in the auditory cortical network. *Hear Res* 271, 133-146, doi:10.1016/j.heares.2010.01.011.
- Hage, S. R., 2019. Precise vocal timing needs cortical control. *Science* 363, 926-927, doi:10.1126/science.aaw5562.
- Hage, S. R., 2020. The role of auditory feedback on vocal pattern generation in marmoset monkeys. *Curr Opin Neurobiol* 60, 92-98, doi:10.1016/j.conb.2019.10.011.
- Hage, S. R., Nieder, A., 2013. Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat Commun* 4, 2409, doi:10.1038/ncomms3409.
- Hage, S. R., Nieder, A., 2015. Audio-vocal interaction in single neurons of the monkey ventrolateral prefrontal cortex. *J Neurosci* 35, 7030-7040, doi:10.1523/JNEUROSCI.2371-14.2015.
- Hage, S. R., Nieder, A., 2016. Dual Neural Network Model for the Evolution of Speech and Language. *Trends Neurosci* 39, 813-829, doi:10.1016/j.tins.2016.10.006.
- Hage, S. R., Jürgens, U., Ehret, G., 2006. Audio-vocal interaction in the pontine brainstem during self-initiated vocalization in the squirrel monkey. *Eur J Neurosci* 23, 3297-3308, doi:10.1111/j.1460-9568.2006.04835.x.
- 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 USA* 110, 4063-8, doi:10.1073/pnas.1211533110.
- Halfwerk, W., Slabbekoorn, H., 2009. A behavioural mechanism explaining noise-dependent frequency use in urban birdsong. *Anim Behav* 78, 1301-1307, doi:10.1016/j.anbehav.2009.09.015.
- Hammerschmidt, K., 2008. Constraints in primate vocal production. The evolution of communicative creativity: From fixed signals to contextual flexibility, 93-119.

-
- Hammerschmidt, K., Jürgens, U., Freudenstein, T., 2001. Vocal development in squirrel monkeys. *Behaviour* 138, 1179-1204.
- Hanley, C. N., Harvey, D. G., 1965. Quantifying the Lombard effect. *J Speech Hear Disord* 30, 274-277, doi:10.1044/jshd.3003.274.
- Hardman, S. I., Zollinger, S. A., Koselj, K., Leitner, S., Marshall, R. C., Brumm, H., 2017. Lombard effect onset times reveal the speed of vocal plasticity in a songbird. *J Exp Biol* 220, 1065-1071, doi:10.1242/jeb.148734.
- Hubrecht, R. C., 1985. Home-Range Size and Use and Territorial Behavior in the Common Marmoset, *Callithrix-Jacchus-Jacchus*, at the Tapacura Field Station, Recife, Brazil. *Int J Primatol* 6, 533-550, doi:10.1007/Bf02735575.
- Janik, V. M., Knörnschild, M., 2021. Vocal production learning in mammals revisited. *Philos Trans R Soc Lond B Biol Sci* 376, 20200244, doi:10.1098/rstb.2020.0244.
- Jen, P. H., Ostwald, J., 1977. Response of cricothyroid muscles to frequency-modulated sounds in FM bats, *Myotis lucifugus*. *Nature* 265, 77-78, doi:10.1038/265077a0.
- Jones, B. S., Harris, D. H., Catchpole, C. K., 1993. The stability of the vocal signature in phee calls of the common marmoset, *Callithrix jacchus*. *Am J Primatol* 31, 67-75.
- Jones, C., 1997. Quantitative analysis of marmoset vocal communication. *Marmosets and tamarins in biological and biomedical research: proceedings of a workshop*. Salisbury: DSSD Imagery, pp. 145-151.
- Jürgens, U., 2002. Neural pathways underlying vocal control. *Neurosci Biobehav Rev* 26, 235-258, doi:10.1016/s0149-7634(01)00068-9.
- Kobayasi, K. I., Okanoya, K., 2003. Context-dependent song amplitude control in Bengalese finches. *Neuroreport* 14, 521-524.
- Larson, C. R., 1991. On the relation of PAG neurons to laryngeal and respiratory muscles during vocalization in the monkey. *Brain Res* 552, 77-86, doi:10.1016/0006-8993(91)90662-f.
- Lazaro-Perea, C., 2001. Intergroup interactions in wild common marmosets, *Callithrix jacchus*: territorial defence and assessment of neighbours. *Anim Behav* 62, 11-21, doi:10.1006/anbe.2000.1726.
- 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 USA* 115, 3978-3983, doi:10.1073/pnas.1722426115.
- Loh, K. K., Petrides, M., Hopkins, W. D., Procyk, E., Amiez, C., 2017. Cognitive control of vocalizations in the primate ventrolateral-dorsomedial frontal (VLF-DMF) brain network. *Neurosci Biobehav Rev* 82, 32-44, doi:10.1016/j.neubiorev.2016.12.001.
- Lombard, E., 1911. Le signe de l'elevation de la voix. *Ann. Mal. de L'Oreille et du Larynx*, 101-119.
- Löschner, J., Pomberger, T., Hage, S. R., 2023. Marmoset monkeys use different avoidance strategies to cope with ambient noise during vocal behavior. *iScience*, 106219, doi:10.1016/j.isci.2023.106219
- Luo, J., Kothari, N. B., Moss, C. F., 2017. Sensorimotor integration on a rapid time scale. *Proc Natl Acad Sci USA* 114, 6605-6610, doi:10.1073/pnas.1702671114.
- Luo, J., Hage, S. R., Moss, C. F., 2018. The Lombard Effect: From Acoustics to Neural Mechanisms. *Trends Neurosci* 41, 938-949, doi:10.1016/j.tins.2018.07.011.

-
- Luo, J., Goerlitz, H. R., Brumm, H., Wiegrebe, L., 2015. Linking the sender to the receiver: vocal adjustments by bats to maintain signal detection in noise. *Sci Rep* 5, 18556, doi:10.1038/srep18556.
- Manabe, K., Sadr, E. I., Dooling, R. J., 1998. Control of vocal intensity in budgerigars (*Melopsittacus undulatus*): differential reinforcement of vocal intensity and the Lombard effect. *J Acoust Soc Am* 103, 1190-1198, doi:10.1121/1.421227.
- Mansfield, K., 2003. Marmoset models commonly used in biomedical research. *Comp Med* 53, 383-392.
- Miller, C. T., Wang, X., 2006. Sensory-motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. *J comp physiol* 192, 27-38.
- Miller, C. T., Flusberg, S., Hauser, M. D., 2003. Interruptibility of long call production in tamarins: implications for vocal control. *J Exp Biol* 206, 2629-39, doi:10.1242/jeb.00458.
- Miller, C. T., Eliades, S. J., Wang, X., 2009a. Motor planning for vocal production in common marmosets. *Anim Behav* 78, 1195-1203, doi:10.1016/j.anbehav.2009.07.038.
- Miller, C. T., Mandel, K., Wang, X., 2010. The communicative content of the common marmoset phee call during antiphonal calling. *Am J Primatol* 72, 974-80, doi:10.1002/ajp.20854.
- Miller, C. T., Beck, K., Meade, B., Wang, X., 2009b. Antiphonal call timing in marmosets is behaviorally significant: interactive playback experiments. *J Comp Physiol* 195, 783-789.
- Mitchell, J. F., Leopold, D. A., 2015. The marmoset monkey as a model for visual neuroscience. *Neurosci Res* 93, 20-46, doi:10.1016/j.neures.2015.01.008.
- Nemeth, E., Brumm, H., 2010. Birds and anthropogenic noise: are urban songs adaptive? *Am Nat* 176, 465-475, doi:10.1086/656275.
- Norcross, J., Newman, J. D., 1993. Context and gender-specific differences in the acoustic structure of common marmoset (*Callithrix jacchus*) phee calls. *Am J Primatol* 30, 37-54.
- Norcross, J., Newman, J., 1997. Social context affects phee call production by nonreproductive common marmosets (*Callithrix jacchus*). *Am J Primatol* 43, 135-146.
- Norcross, J., Newman, J. D., Fitch, W., 1994. Responses to natural and synthetic phee calls by common marmosets (*Callithrix jacchus*). *Am J Primatol* 33, 15-29.
- Okano, H., Hikishima, K., Iriki, A., Sasaki, E., 2012. The common marmoset as a novel animal model system for biomedical and neuroscience research applications. *Seminars in fetal and neonatal medicine*, Vol. 17. Elsevier, pp. 336-340.
- Osmanski, M. S., Dooling, R. J., 2009. The effect of altered auditory feedback on control of vocal production in budgerigars (*Melopsittacus undulatus*). *J Acoust Soc Am* 126, 911-919, doi:10.1121/1.3158928.
- Pick, H. L., Jr., Siegel, G. M., Fox, P. W., Garber, S. R., Kearney, J. K., 1989. Inhibiting the Lombard effect. *J Acoust Soc Am* 85, 894-900, doi:10.1121/1.397561.
- Pistorio, A. L., Vintch, B., Wang, X., 2006. Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). *J Acoust Soc Am* 120, 1655-1670, doi:10.1121/1.2225899.

-
- Pohl, N. U., Slabbekoorn, H., Klump, G. M., Langemann, U., 2009. Effects of signal features and environmental noise on signal detection in the great tit, *Parus major*. *Anim Behav* 78, 1293-1300, doi:10.1016/j.anbehav.2009.09.005.
- Pohl, N. U., Leadbeater, E., Slabbekoorn, H., Klump, G. M., Langemann, U., 2012. Great tits in urban noise benefit from high frequencies in song detection and discrimination. *Anim Behav* 83, 711-721, doi:10.1016/j.anbehav.2011.12.019.
- Pomberger, T., Löschner, J., Hage, S. R., 2020. Compensatory mechanisms affect sensorimotor integration during ongoing vocal motor acts in marmoset monkeys. *Eur J Neurosci* 52, 3531-3544, doi:10.1111/ejn.14721.
- 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 e3, doi:10.1016/j.cub.2018.01.070.
- Pomberger, T., Risueno-Segovia, C., Gultekin, Y. B., Dohmen, D., Hage, S. R., 2019. Cognitive control of complex motor behavior in marmoset monkeys. *Nat Commun* 10, 3796.
- Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., Fischer, J., 2015. Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Sci Rep* 5, 13220, doi:10.1038/srep13220.
- Roberts, A. C., Tomic, D. L., Parkinson, C. H., Roeling, T. A., Cutter, D. J., Robbins, T. W., Everitt, B. J., 2007. Forebrain connectivity of the prefrontal cortex in the marmoset monkey (*Callithrix jacchus*): an anterograde and retrograde tract-tracing study. *J Comp Neurol* 502, 86-112, doi:10.1002/cne.21300.
- 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, 3619-29, doi:10.1242/jeb.056101.
- Saltzman, W., Severin, J. M., Schultz-Darken, N. J., Abbott, D. H., 1997. Behavioral and social correlates of escape from suppression of ovulation in female common marmosets housed with the natal family. *Am J Primatol* 41, 1-21.
- Schuster, S., Zollinger, S. A., Lesku, J. A., Brumm, H., 2012. On the evolution of noise-dependent vocal plasticity in birds. *Biol Lett* 8, 913-6, doi:10.1098/rsbl.2012.0676.
- Seyfarth, R. M., Cheney, D. L., Marler, P., 1980. Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science* 210, 801-3, doi:10.1126/science.7433999.
- Smith, D., Trennery, P., Farningham, D., Klapwijk, J., 2001. The selection of marmoset monkeys (*Callithrix jacchus*) in pharmaceutical toxicology. *Lab Anim* 35, 117-30, doi:10.1258/0023677011911444.
- Snowdon, C. T., Elowson, A. M., 1999. Pygmy marmosets modify call structure when paired. *Ethology* 105, 893-908, doi:10.1046/j.1439-0310.1999.00483.x.
- Sober, S. J., Brainard, M. S., 2009. Adult birdsong is actively maintained by error correction. *Nat Neurosci* 12, 927-931, doi:10.1038/nn.2336.
- Stephan, H., Frahm, H., Baron, G., 1981. New and Revised Data on Volumes of Brain Structures in Insectivores and Primates. *Folia Primatol* 35, 1-29, doi:10.1159/000155963.
- Stowe, L. M., Golob, E. J., 2013. Evidence that the Lombard effect is frequency-specific in humans. *J Acoust Soc Am* 134, 640-647, doi:10.1121/1.4807645.

-
- Summers, P. M., Wennink, C. J., Hodges, J. K., 1985. Cloprostenol-induced luteolysis in the marmoset monkey (*Callithrix jacchus*). *J Reprod Fertil* 73, 133-138, doi:10.1530/jrf.0.0730133.
- Takahashi, D. Y., Liao, D. A., Ghazanfar, A. A., 2017. Vocal Learning via Social Reinforcement by Infant Marmoset Monkeys. *Curr Biol* 27, 1844-1852 e6, doi:10.1016/j.cub.2017.05.004.
- Takahashi, D. Y., Fenley, A. R., Teramoto, Y., Narayanan, D. Z., Borjon, J. I., Holmes, P., Ghazanfar, A. A., 2015. The developmental dynamics of marmoset monkey vocal production. *Science* 349, 734-738, doi:10.1126/science.aab1058.
- Tardif, S. D., Mansfield, K. G., Ratnam, R., Ross, C. N., Ziegler, T. E., 2011. The marmoset as a model of aging and age-related diseases. *ILAR J* 52, 54-65, doi:10.1093/ilar.52.1.54.
- Tardif, S. D., Smucny, D. A., Abbott, D. H., Mansfield, K., Schultz-Darken, N., Yamamoto, M. E., 2003. Reproduction in captive common marmosets (*Callithrix jacchus*). *Comp Med* 53, 364-368.
- Therrien, A. S., Lyons, J., Balasubramaniam, R., 2012. Sensory attenuation of self-produced feedback: the Lombard effect revisited. *PLoS One* 7, e49370, doi:10.1371/journal.pone.0049370.
- Tokuno, H., Watson, C., Roberts, A., Sasaki, E., Okano, H., 2015. Marmoset neuroscience. *Neurosci Res* 93, 1-2, doi:10.1016/j.neures.2015.03.001.
- Tressler, J., Smotherman, M. S., 2009. Context-dependent effects of noise on echolocation pulse characteristics in free-tailed bats. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 195, 923-934, doi:10.1007/s00359-009-0468-x.
- Vinney, L. A., van Mersbergen, M., Connor, N. P., Turkstra, L. S., 2016. Vocal control: is it susceptible to the negative effects of self-regulatory depletion? *J Voice* 30, 638. e21-638. e31.
- Winter, P., Handley, P., Ploog, D., Schott, D., 1973. Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour* 47, 230-9, doi:10.1163/156853973x00085.
- Yamamoto, M., 1993. From dependence to sexual maturity: the behavioural ontogeny of Callitrichidae. *Marmosets and tamarins: systematics behaviour and ecology*, 235-254.
- Zelick, R. D., Narins, P. M., 1982. Analysis of Acoustically Evoked Call Suppression Behavior in a Neotropical Treefrog. *Anim Behav* 30, 728-733, doi:10.1016/S0003-3472(82)80144-9.
- Zhao, L., Rad, B. B., Wang, X., 2019. Long-lasting vocal plasticity in adult marmoset monkeys. *Proc Biol Sci* 286, 20190817, doi:10.1098/rspb.2019.0817.

Acknowledgements

I would like to express my deepest gratitude to my PI, Steffen Hage, for all the guidance, support, and instruction he provided me throughout my doctoral studies. His feedback and encouragement greatly influenced how I conducted my experiments and interpreted my findings. Furthermore, I want to thank my committee members, Jan Benda, Ziad Hafed, and Lena Veit.

I am also grateful to my former and current lab mates for their help and for all the fun times we had working and socializing together. Special thanks to Thomas and Cristina for the great teamwork on the joint projects, and to Anna and Elena, who in addition, became great friends.

Lastly, I would be remiss if I did not mention my family and Dennis. Their belief in me has kept my spirits and motivation high throughout this process. It would have been impossible to finish my studies without their unwavering support over the past few years. Many thanks to Katrin, our great animal caretaker, and to my monkeys (Horst, Sissi, Franz, Willi, Maja, Alice, and Johnny) for all their hard work and entertaining moments. Finally, I would like to thank John Holmes for proofreading.

II. Individual studies

Statement of contributions

This thesis comprises two publications.

1. Pomberger, T.*, **Löschner, J.***, & Hage, S. R. (2020). Compensatory mechanisms affect sensorimotor integration during ongoing vocal motor acts in marmoset monkeys. *European Journal of Neuroscience*, 52(6), 3531-3544.

I conducted the experiments and performed data analyses with T. Pomberger. All authors interpreted the data and wrote the manuscript. The full publication is found in the chapter Publication 1.

2. **Löschner, J.**, Pomberger, T., & Hage, S. R. (2023). Marmoset monkeys use different avoidance strategies to cope with ambient noise during vocal behavior. *Iscience*, 26(3), 106219.

I designed the experiments with T. Pomberger and S.R. Hage. I conducted the experiments and performed data analyses. All authors interpreted the data and wrote the manuscript. The full publication is found in the chapter Publication 2.

Publication 1: Compensatory mechanisms affect sensorimotor integration during ongoing vocal motor acts in marmoset monkeys

Pomberger, T., **Löschner, J.**, & Hage, S. R. (2020). Compensatory mechanisms affect sensorimotor integration during ongoing vocal motor acts in marmoset monkeys. *European Journal of Neuroscience*, 52(6), 3531-3544.

Compensatory mechanisms affect sensorimotor integration during ongoing vocal motor acts in marmoset monkeys

Thomas Pomberger^{1,2} | Julia Löschner¹ | Steffen R. Hage^{1,3} 

¹Neurobiology of Vocal Communication, Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Tübingen, Germany

²Graduate School of Neural & Behavioural Sciences - International Max Planck Research School, University of Tübingen, Tübingen, Germany

³Department of Otolaryngology - Head and Neck Surgery, Hearing Research Center, University of Tübingen, Medical Center, Tübingen, Germany

Correspondence

Steffen R. Hage, Neurobiology of Vocal Communication, Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Str. 25, 72076 Tübingen, Germany.
Email: steffen.hage@uni-tuebingen.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: EXC 307

Abstract

Any transmission of vocal signals faces the challenge of acoustic interferences such as heavy rain, wind, animal or urban sounds. Consequently, several mechanisms and strategies have evolved to optimize signal-to-noise ratio. Examples to increase detectability are the Lombard effect, an involuntary rise in call amplitude in response to masking ambient noise, which is often associated with other vocal changes such as call frequency and duration, as well as the animals' capability of limiting calling to periods where noise perturbation is absent. Previous studies revealed vocal flexibility and various audio-vocal integration mechanisms in marmoset monkeys. Using acoustic perturbation triggered by vocal behaviour, we investigated whether marmosets are capable of exhibiting changes in call structure when perturbing noise starts after call onset or whether such effects only occur if noise perturbation starts prior to call onset. We show that marmosets are capable of rapidly modulating call amplitude and frequency in response to such noise perturbation. Vocalizations swiftly increased call frequency after noise onset indicating a rapid effect of perturbing noise on vocal motor production. Call amplitudes were also affected. Interestingly, however, the marmosets did not exhibit the Lombard effect as previously reported but decreased call intensity in response to noise. Our findings indicate that marmosets possess a general avoidance strategy to call in the presence of ambient noise and suggest that these animals are capable of counteracting a previously thought involuntary audio-vocal mechanism, the Lombard effect. These findings will pave the way to investigate the underlying audio-vocal integration mechanisms explaining these behaviours.

KEYWORDS

acoustic perturbation, audio-vocal integration, *Callithrix jacchus*, vocal communication, vocal motor control, vocalization

Abbreviations: ES, Effect size; FFT, Fast Fourier transformation; GUI, Graphical user interface; PFC, Prefrontal cortex; SNR, Signal-to-noise ratio.

Thomas Pomberger and Julia Löschner contributed equally

Editor: Dr. John Foxxe.

The peer review history for this article is available at <https://publons.com/publon/10.1111/ejn.14721>.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. *European Journal of Neuroscience* published by Federation of European Neuroscience Societies and John Wiley & Sons Ltd

1 | INTRODUCTION

Communication between individuals is a crucial aspect for evolutionary success and appears in various forms in nature ranging from olfactory (Poldrack & Farah, 2015; Stockhorst & Pietrowsky, 2004) to visual (Osorio & Vorobyev, 2008) to vocal signals (Ackermann, Hage, & Ziegler, 2014). For proper communication, the transmission of a signal sent out by a sender has to be detected and decoded by one or more receivers (Bradbury & Vehrencamp, 1998). Therefore, the sender has to be able to modulate the signal in response to potential masking ambient noise to ensure proper signal transmission. For vocal communication in vertebrates, several mechanisms have evolved to compensate for masking acoustic interferences, such as heavy rain, wind, animal or urban sounds, leading to changes in temporal and spectral features of the vocal signals (Brumm & Slabbekoorn, 2005). Such vocal modifications can happen involuntarily as well as under volitional control.

One of the most efficient mechanisms to increase signal-to-noise ratio (SNR) in call production is the so-called Lombard effect, that is the involuntary increase in call amplitude in response to masking ambient noise (Lombard, 1911). It is often accompanied by a shift in call frequency (Hage, Jiang, Berquist, Feng, & Metzner, 2013; Osmanski & Dooling, 2009) as well as a change in call duration (Brumm, 2006; Luo, Goerlitz, Brumm, & Wiegrebe, 2015) and has been shown in many vertebrate species from fish to frogs to birds to mammals including humans (Brumm & Zollinger, 2011; Luo, Hage, & Moss, 2018), suggesting that the Lombard effect is an evolutionary old behaviour that may have emerged about 450 million years ago. Another successful strategy to increase detectability in a noisy environment is the restraint of call emission to timeslots where noise perturbation is low or absent (Brumm, 2006; Roy, Miller, Gottsch, & Wang, 2011; Zelick & Narins, 1982). This approach renders the modification of call parameters unnecessary and avoids the increased physiological cost of call emission at high intensities that might still be insufficiently increasing SNR.

The common marmoset, a small, highly vocal New World monkey indigenous to the dense rainforests of Brazil, has been shown to exhibit vocal flexibility, such as increasing call intensity (Brumm, 2004; Eliades & Wang, 2012) or increasing the duration of specific calls (Brumm, 2004), as well as the attempt to call in silent gaps (Roy et al., 2011), in the presence of perturbing ambient noise. These findings suggest that while these animals generally seem to prefer avoiding calling in a noisy environment, they do exhibit the involuntary audio-vocal effects discussed above when doing so. This idea is supported by a recent study showing that marmosets tend to produce single calls instead of call sequences in response to perturbing noise stimuli (Pomberger, Risueno-Segovia, Löschner, & Hage, 2018). Interestingly, marmoset monkeys are also capable of interrupting ongoing vocalizations rapidly

after noise perturbation onset (Pomberger et al., 2018), overturning decades-old concepts regarding vocal pattern generation (Egnor, Iguina, & Hauser, 2006; Miller, Eliades, & Wang, 2009; Miller, Flusberg, & Hauser, 2003), indicating that vocalizations do not consist of one discrete call pattern but are built of many sequentially uttered units that might be modulated and initiated independently of each other. However, it is yet unclear whether audio-vocal mechanisms, such as the Lombard effect and its accompanied changes in call frequency, can be rapidly elicited in cases where the perturbing noise starts after call onset or whether such effects only occur if noise perturbation starts prior to call onset.

In the present study, we use acoustic perturbation triggered by the vocal behaviour itself to test in a controlled experimental design whether marmosets are capable of rapidly modulating distinct vocal parameters such as call frequency and amplitude in ongoing vocalizations. Performing quantitative measures of resulting adjustments, we show that marmoset monkeys are able to specifically and rapidly modulate call frequency and amplitude as a response to white noise stimuli in ongoing vocal utterances. Hereby, our data indicate that marmosets exhibit a decrease in call amplitude as a result of such noise perturbation, suggesting a mechanism counteracting the rise in amplitude caused by the Lombard effect.

2 | MATERIALS AND METHODS

2.1 | Animal housing and maintenance

Four adult marmoset monkeys (*Callithrix jacchus*) were used in the present study. Monkeys were usually kept in different sex pairs and were all born in captivity. The animals had ad libitum access to water and were fed on a restricted food protocol including a daily basis of commercial pellets, fruits, vegetables, mealworms and locusts. Additional treats, such as marshmallows or grapes, were used as positive reinforcements to transfer the animals from their home cage to the experimental cage. Environmental conditions in the animal husbandry were maintained at a temperature of 26°C, 40%–60% relative humidity and a 12-hr:12-hr day/night cycle. All animal handling procedures were in accordance with the guidelines for animal experimentation and authorized by the national authority, the Regierungspräsidium Tübingen. All vocalizations analysed in this study are a fraction of calls that have been recorded in a previous study (Pomberger et al., 2018).

2.2 | Experimental set-up and procedure

The vocal behaviour of four animals was recorded in a soundproof chamber in response to noise playback that was

initiated after vocal onset as reported earlier (Pomberger et al., 2018). Briefly, the animals were transferred into a recording cage ($0.6 \times 0.6 \times 0.8$ m), which was placed in a soundproof chamber, with ad libitum access to water and food pellets throughout the recording period. In this behavioural set-up, marmoset monkeys predominantly produce phee calls to interact with conspecifics (phee ratio within all uttered calls; monkey S: 99.1%, H: 92.0%, W: 95.6%, F: 96.8%). Other call types such as trill-pees, twitter, trills, tsik-ekks (Agamaite, Chang, Osmanski, & Wang, 2015) or segmented pees (Zürcher & Burkart, 2017) were only rarely uttered (ratios were well below 2.5% for all other call types in all monkeys except trill-pees in monkey H [4.6%]). Monkeys produced a mean of 118 ± 9 (monkey S), 167 ± 31 (H), 117 ± 10 (W) and 87 ± 7 (F) phee calls per session. The vocal behaviour of each individual monkey was recorded once a day in sessions ranging between one and two hours in duration. Data were collected in sessions at various times during the day between 11 a.m. and 5 p.m. Recordings were performed for 10–28 days (mean: 17 ± 3 days) for each individual animal. The monkey's behaviour was constantly monitored and observed with a video camera (ace acA1300-60gc, Basler, with 4.5–12.5 mm CS-Mount Objective H3Z4512CS-IR 1/2, Computar) placed on top of the cage and recorded with standard software (EthoVision XT version 4.2.22, Noldus). The vocal behaviour was collected with eight microphones (MKH 8020 microphone with MZX 8000 preamplifier, Sennheiser), which were positioned in an octagonal design around the cage (Figure 1b), digitized using an A/D interface (Octacapture, Roland; sample rate: 96 kHz) and recorded using standard software (Avisoft-Recorder, Avisoft Bioacoustics). A custom-written program (OpenEX, Tucker-Davis Technologies) running on a workstation (WS-X in combination with an RZ6D multi-I/O processor, Tucker-Davis Technologies) monitored the vocal behaviour in real time via an additional microphone (MKH 8020 microphone with MZX 8000 preamplifier, Sennheiser) placed on top of the cage, which automatically detected vocalizations through online calculation of several acoustic parameters, such as call intensity, minimum duration of call intensity duration, call frequency and several spectral features. The median vocal detection rate was well above 99%, and three out of four vocalizations were detected within the first 146 ms after call onset (Figure 1c).

The eight microphones positioned around the cage were installed to ensure precise calculation of dB SPL values of vocalizations with a corresponding microphone being positioned in front of the monkey (for details, see below).

For two out of three uttered vocalizations, we played back noise bursts of different frequency bands and amplitudes via a loudspeaker (MF1 Multi-Field Magnetic Speakers, Tucker-Davis Technologies) positioned on top of the cage, immediately after vocal detection. Noise bursts had a duration of

4 s (including 10-ms rise times) to ensure noise perturbation throughout the first phee syllable as well as the initiation of the second syllable. Five different noise band conditions (broadband noise and bandpass-filtered noise bands: 0.1–5.1 kHz, 5–10 kHz, 10–15 kHz and 16–21 kHz) were played back at four different amplitudes (50 dB, 60 dB, 70 dB and 80 dB) each (Figure 1d). All 20 noise conditions were played back pseudo-randomly in blocks of 30 uttered vocalizations, resulting in 20 calls being perturbed with noise after call onset and 10 calls without noise playback remaining unaffected (control). After one block ended, a new block was generated. Noise playback generation and presentation were performed with the same custom-written software used for call detection.

2.3 | Data analysis

We programmed a custom-written graphical user interface (GUI; MATLAB, MathWorks) to clock Avisoft, Noldus and Tucker-Davis Technologies recordings offline and to extract the detected calls from the recording channel with the best SNR. Vocal onset to offset was manually flagged as well as noise onset times using the aligned sono- and spectrogram of vocalizations. We used a Hanning window with a 512-window size, 1,024 FFT, overlap of 25% and temporal resolution of 1 ms. We only considered first phee syllables for calculation that were detected/perturbed within 200 ms of call onset and with a minimum duration of 800 ms. Consequently, first phee syllables that were interrupted directly after noise onset as previously reported in an earlier study (Pomberger et al., 2018) were excluded from further analysis. Second phee syllables were only analysed if they had a minimum duration of 500 ms. In rare cases, call termination could not be visually detected due to overlapping noise (mostly during the 80 dB SPL condition). These calls were also excluded from further analysis.

After labelling a call, peak frequencies of the fundamental component were automatically calculated within one-millisecond time bins (8192 FFT, 96 kHz sample rate resulting in a frequency resolution of 11.71 Hz). In cases where the SNR between the call amplitude and playback noise was not high enough for automatic fundamental peak frequency calculation, frequency trajectories were calculated by manually setting call frequencies at several time points and interpolating call frequencies in between the set values. Therefore, call frequencies were analysed without prior noise reduction. The accuracy of manual labelling compared to automatic calculation of peak frequencies was high and median differences between both techniques below the frequency resolution used (Figure S1).

Call amplitudes were calculated for all phee calls during which the animals did not move their heads during call

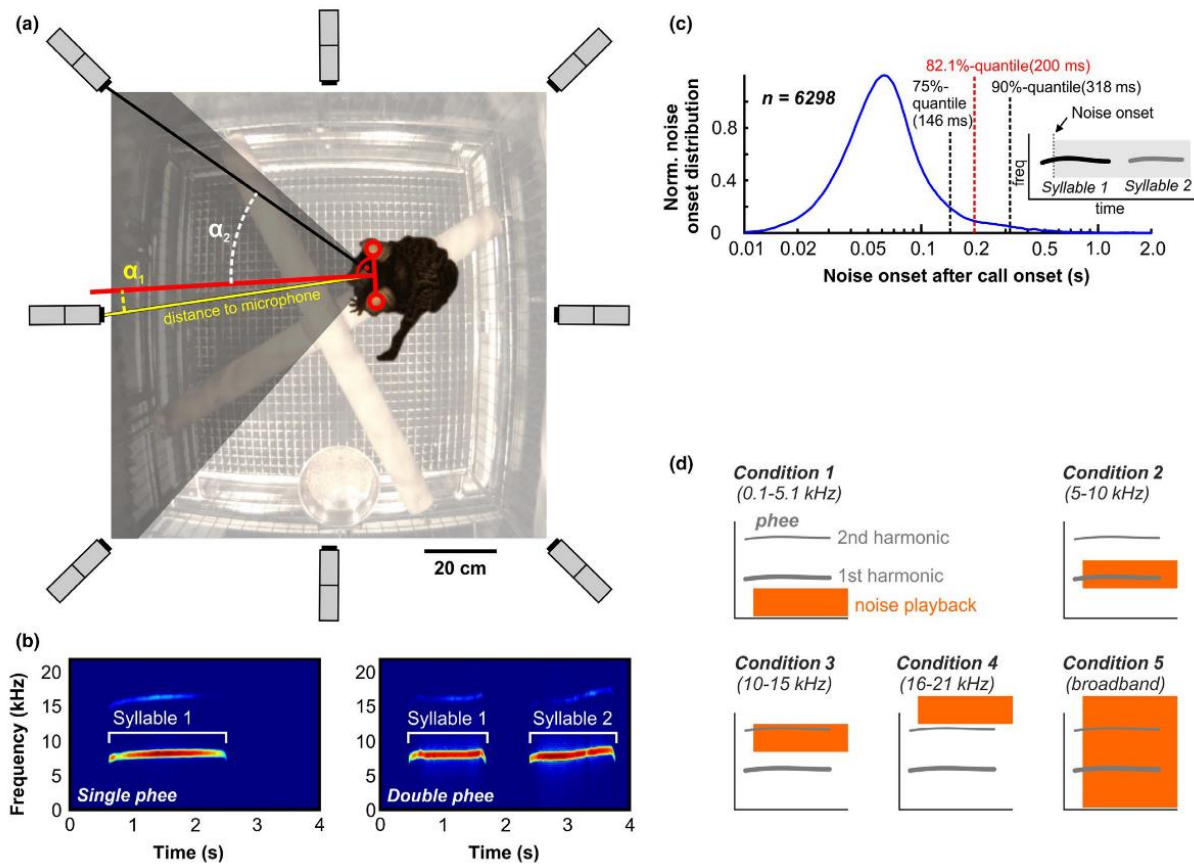


FIGURE 1 Experimental set-up and design. (a) The vocal behaviour of monkeys was recorded in a soundproof chamber. The behaviour was continuously monitored and recorded. The red line shows the monkey's head position in relation to the two closest microphones (yellow and black lines). The acoustic signal recorded with the microphone closest to being directly in front of the monkey's head (i.e. the smallest angle between the monkey's perpendicular and the microphone) was used for amplitude calculation. (b) Exemplary spectrograms of single and double phee calls. (c) Relative vocal detection distribution over time (s). (d) Noise condition overview with masking properties [Colour figure can be viewed at wileyonlinelibrary.com]

production. For these calls, head positions were manually labelled by marking the two white ear tufts in the GUI (Figure 1b). Next, a perpendicular line starting at the centre of the later connection was used to compute angles of the microphones indicating the monkey's relative head position. The microphone with the smallest angle to the perpendicular line was used for further calculation (Figure 1b). Calls that were uttered in the rare cases where the angle between the front of the monkey's head and the microphone was more than 45 degrees were excluded from further analysis. Furthermore, phee calls that were uttered during head movements of the animal were not used for amplitude calculations and only considered for fundamental frequency calculation resulting in a larger data set for frequency analysis.

Call amplitude trajectories (in dB SPL) were calculated using a sliding window approach (window size: 25 ms; step length: 1 ms, time for corresponding amplitude at the centre of the sliding window), from the vocal recordings of the microphone foremost in front of the animal. Sound levels of the

recorded playback noise were determined for all conditions and subtracted from the call amplitude measurements taken, using a modification of the spectral noise subtraction method (Berouti, Schwartz, & Makhoul, 1978). Briefly, we first calculated an estimate for each noise band by calculating the mean of ten recordings of one noise condition for each microphone. Then, we subtracted this noise estimate in the spectral dimension from noise perturbed parts of a call (i.e. from noise onset to the end of the call) and corrected the outcome as shown in formula (1), where $P_S(w)$ is the spectrogram of the signal and the noise; $P_n(w)$ the spectrogram of the noise estimate; and $P'_S(w)$, the modified signal spectrum. Alpha is defined as the subtraction factor and beta as the spectral floor parameter.

$$P'_S = \begin{cases} D(w) = P_S(w) - \alpha P_n(w) & \\ D(w), & \text{if } D(w) > \beta P_n(w), \\ & \alpha \leq 1, \quad 0 < \beta \ll 1 \end{cases} \quad \text{otherwise}$$

Alpha and beta were calculated using the following equation:

$$\alpha = \alpha_0 - \frac{SNR}{s}$$

$$5 \leq SNR \leq 20$$

According to Berouti et al. (1978), we chose $\alpha_0 = 4$ and $s = 20/3$ as a best fit for proper amplitude calculation. A simple empirical test verified the method; a control phee was played and recorded in the recording chamber ten times with broadband noise 70 dB SPL, ten times with a 5–10 kHz noise band and ten times under control conditions (no noise). As reported previously, differences between conditions of <1 dB can be assumed to be negligible (Brumm, Schmidt, & Schrader, 2009). In our case, median differences between control and both noise conditions were below 1 dB (broadband: 0.8 dB, 5–10 kHz noise band: 0.3 dB; Figure S2) indicating successful performance of the used method. The distance of the animal's head to the microphone was considered by adding a distance factor directly after noise subtraction to the measurements resulting in a standardized amplitude trajectory (in dB) of each call as produced 10 cm in front of the animal's head.

2.4 | Frequency/amplitude calculation and normalization

Mean fundamental frequency values were obtained with a sliding window approach (window size: 10 ms, step size: 1 ms; time for corresponding frequency at the centre of the sliding window). As noise onsets showed some variance relative to call onset, we normalized call frequencies after noise onset (for the noise condition) and after call detection (for the control condition), to the call's fundamental frequency just prior to noise onset and call detection, respectively, to decipher the noise-related changes in call frequency. For that, we calculated the mean of the fundamental frequency in a 5- to 20-ms time window prior to noise onset (for the noise conditions) and call detection (for the control condition) for each individual call. This value was then subtracted from each of the frequency values after noise onset. Finally, all values of calls in the noise conditions were normalized by subtracting the mean of the respective frequency value of the control condition. Amplitude values were calculated in a similar way. Here, we also calculated the mean amplitude for each individual call in a 5- to 20-ms time window prior to noise onset and subtracted these values from the mean amplitude values after noise onset. According to the frequency normalization, we then normalized all amplitude values by subtracting the mean of the amplitude values from the corresponding values in the control condition. For the 180-s noise experiment, we

used the calculated amplitude values as described above in data analysis.

2.5 | Phee call discrimination models

Marmoset monkeys tend to interrupt their phee calls after the first syllable in response to noise perturbation (Pomberger et al., 2018). For perturbed phee calls, we consequently assumed that a substantial number of single phees had to be interrupted double phees. Recently, it has been shown that single phees and the first syllables of double phees significantly differ in a number of call parameters, such as call frequency and duration (Miller et al., 2009). We therefore had to find a way to distinguish single phee calls that were interrupted double phees from original single phees prior to data normalization. To address this, we used the findings of Miller et al. (2009) that suggested that early peak frequencies and durations of phee calls are sufficient to predict whether a phee call consists of one or two syllables. Additionally, we found that this is also true for early amplitude values of a call. We applied a quadratic classification model (MATLAB, MathWorks) to discriminate between single and double phees with a two-dimensional classifier for fundamental frequency analysis using 1st syllable durations and peak frequencies at 25 ms after call onset for frequency analyses (Figure S3). As we observed that early amplitude values are also a good predictor, we used a three-dimensional classifier with call amplitude values at 25 ms after call onset as the third measure for amplitude analyses (Figure S4). Basically, in a first step the mean, μ_k , and covariance matrix, Σ_k , of each class are calculated from all control values to obtain the density function of the multivariate normal at a point, x , using the following formula:

$$p(k) = \frac{1}{(2\pi |\Sigma_k|)^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)\right)$$

where $|\Sigma_k|$ is the determinant of Σ_k , and Σ_k^{-1} is the inverse matrix. Using the prior probability $p(k)$ of class k and $p(x)$ as a normalization constant, we obtain the posterior probability $\hat{p}(x)$ that a point x belongs to class k based on:

$$\hat{p}(x) = \frac{p(x)p(k)}{p(x)}$$

These results are then used to classify our phee calls into single and double phees by minimizing the expected classification cost using:

$$\hat{y} = \arg \min_{y=1, \dots, K} \sum_{k=1}^K \hat{p}(x) C(k)$$

where \hat{y} is the predicted classification, K is the number of classes, and $C(k)$ is the cost of classifying an observation as y when its true class is k . In total, the loss for the 2D classification was between 10.8% and 23.2% (mean: 15.1 ± 2.8) and for the 3D classification between 6.8% and 15.7% (mean: 12.3 ± 1.9) for each monkey.

2.6 | Statistical analysis

Statistical analyses were performed with MATLAB (2016b, MathWorks). We performed a three-way ANOVA with noise amplitude, noise frequency and individual monkey as independent factors to test for significant differences in shifts of fundamental call frequency and amplitude within all noise band conditions ($\alpha = 0.05$, Bonferroni-corrected). We used Pearson's linear correlation coefficient to test for correlations between frequency and amplitude shifts. Effect sizes (ESs) were calculated using the following formula:

$$f_p^2 = \frac{\eta_p^2}{1 - \eta_p^2}$$

where f_p^2 represents the effect size of factor p and η_p^2 is calculated as:

$$\frac{\text{explained sum of squares of } p}{\text{(explained sum of squares of } p + \text{residual sum of squares)}}$$

3 | RESULTS AND STATISTICAL ANALYSIS

We measured vocal behaviour in marmoset monkeys (*Callithrix jacchus*, $n = 4$), a highly vocal New World monkey species, while separated in a soundproofed chamber, with and without acoustic perturbation (Figure 1a). In this setting, marmoset monkeys predominantly produced phee calls (monkey H: 92.0%, S: 99.1%, F: 96.8%, W: 95.6%), long-distance contact calls, composed of one (so-called single phees), two (double phees) or more phee syllables, to interact with conspecifics (Agamaite et al., 2015) (Figure 1b). Other call types such as trill-phees, twitters, trills, tsik-ekks (Agamaite et al., 2015; Pistorio, Vintch, & Wang, 2006) and segmented phees (Zürcher & Burkart, 2017) were rarely uttered (all other call types were well below 2.5% in all monkeys except trill-phees in monkey H [4.6%]).

We perturbed 2/3 of calls with noise playback after vocal onset to ensure perturbation starting after call initiation (Figure 1b,c). To investigate whether perturbation of different frequency bands within the hearing range of the monkeys has different effects on their vocal behaviour, we played back five different noise band conditions (broadband

noise and bandpass-filtered noise bands below [0.1–5 kHz], around [5–10 kHz] or above the fundamental frequency of phee calls [noise bands of 10–15 kHz and 16–21 kHz] at four different amplitudes [50 dB, 60 dB, 70 dB and 80 dB] each) (Figure 1d,e). All noise conditions were played back pseudo-randomly in blocks of 30 uttered vocalizations, resulting in 20 calls being perturbed by noise after call onset and 10 calls not being perturbed by noise (control). In total, our monkeys produced 6,298 phees (monkey F = 1,544 phees, H = 1,471, S = 1,631, W = 1,652). Monkeys uttered mostly single and double phees (multi-syllabic phees with more than two syllables were rare or absent: monkey F = 6.5%, H = 0.4%, S = 1.3%, absent in W), with double phee rates between 8.4% and 55.5% (mean: $29.5\% \pm 9.8\%$, $n = 4$ monkeys) in the control condition.

We first investigated whether and how marmosets changed the fundamental frequency of their ongoing phee syllables when perturbed by different noise conditions. We found an increase in first syllable frequencies ($F(3, 4,901)=6.98$, $p = 1.11e-04$ for amplitude, $F(4, 4,901)=22.32$, $p < .0001$ for frequency, $F(3, 4,901)=12.49$, $p < .0001$ for monkeys, $n = 3,180$, three-way ANOVA). Those frequency shifts were significant in the 0.1–5.1 kHz at 80 dB noise condition (38.5 ± 13.8 Hz, $p = 1.02e-02$, $n = 168$), in the two loudest conditions of the 5–10 kHz noise band (70 dB: 56.9 ± 14.5 Hz, $p = 3.20e-03$, $n = 165$; 80 dB: 76.7 ± 16.4 Hz, $p = 4.12e-08$, $n = 134$) and in all four amplitude conditions of broadband noise (50 dB: 39.2 ± 13.7 Hz, $p = 5.03e-10$, $n = 159$; 60 dB: 68.6 ± 13.4 Hz, $p = 1.21e-08$, $n = 143$; 70 dB: 104.1 ± 12.5 Hz, $p = 3.99e-13$, $n = 135$; 80 dB: 101.7 ± 16.1 Hz, $p = 1.66e-20$, $n = 118$; control: $n = 1,733$; Figure 2a,b). The largest frequency shift could be observed for 70 dB broadband noise, while in the next higher intensity condition (80 dB), there was no further increase in frequency ($p = 1$, $n = 253$), indicating that marmosets are only capable of altering their fundamental frequency within a certain range. Frequency shifts were not observed in calls that were produced during 10–15 kHz and 16–21 kHz noise band perturbations ($p = 1$, $n = 669$ for the 10–15 kHz noise band, $n = 652$ for the 16–21 kHz noise band, Figure 2b). Second phee syllables showed no significant shift in fundamental frequencies when perturbed by noise ($F(3, 1,340)=1.44$, $p = .23$ for amplitude, $F(4, 1,340)=1.23$, $p = .30$ for frequency, $F(3, 1,340)=1.24$, $p = .29$ for monkeys, $n = 761$, three-way ANOVA, Figure 2c).

Next, we tested whether the observed changes in call frequency were more influenced by the different noise conditions or by the noise amplitude. Therefore, we quantified the magnitude of the observed frequency shifts by calculating population effect sizes (ES_p) of the factors frequency (ES_{freq}), amplitude (ES_{ampl}) and the combination of both conditions ($ES_{\text{freq} \times \text{ampl}}$) according to Cohen (1992) (see Material and Methods). An effect would be given if the corresponding ES value of a factor was above the threshold of 0.02 as suggested

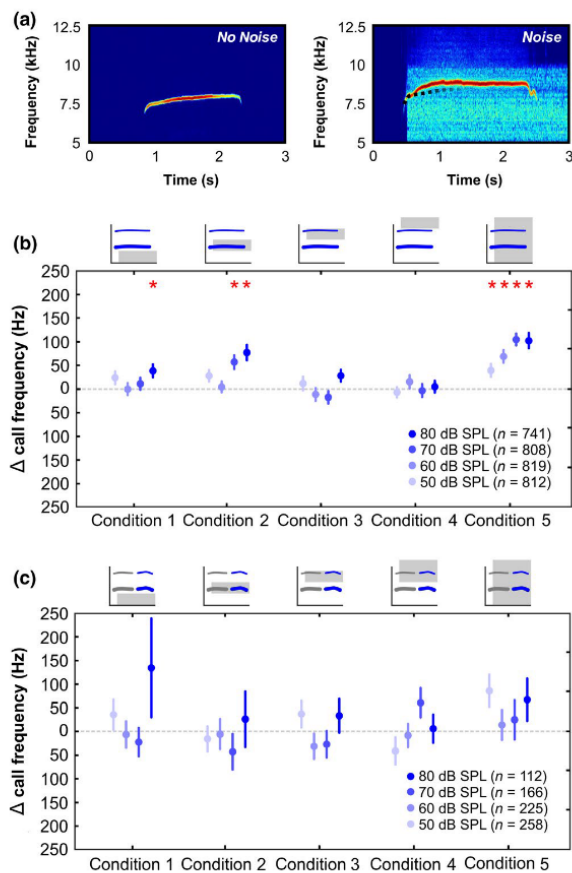


FIGURE 2 Increasing frequency shifts in response to noise bursts. (a) Exemplary spectrograms of a single phee with and without interfering noise presentation. Dashed line in the right spectrogram indicates the frequency trajectory from the example on the left to indicate the rapid frequency shift after noise onset (b and c) Δ call frequency (Hz) per corresponding noise condition normalized to control data (dashed lines). Mean of median frequencies after noise onset of each call pooled over all monkeys \pm SEM (b) for first phee syllables 0–800 ms after noise onset and (c) for second phee syllables 100–400 ms after second syllable onset [Colour figure can be viewed at wileyonlinelibrary.com]

by Cohen (1992). We found $ES_{\text{freq} \times \text{ampl}}$ values of 0.035 for first syllables and 0.019 for second syllables, indicating an effect for first syllables (Figure 2b,c). ES_{freq} for the first syllable was above the threshold ($ES_{\text{freq}} = 0.023$), while ES_{ampl} was below ($ES_{\text{ampl}} = 0.01$), indicating that the shifts in fundamental frequency were mainly correlated with the different noise rather than amplitude conditions.

We then tested how fast fundamental frequency shifts occurred within the first phee syllables after noise onset. Therefore, we plotted the mean fundamental frequency courses starting at noise onset times (Figure 3). The shortest latency of fundamental frequency shifts within a noise condition was defined as the moment where fundamental frequency

shifts were significant for a minimum of five consecutive milliseconds after noise onset. Shortest latencies were found for the 0.1–5.1 kHz noise condition at 80 dB (33 ms, $n = 168$) and all broadband conditions (50 dB: 29 ms, $n = 159$; 60 dB: 34 ms, $n = 143$; 70 dB: 25 ms, $n = 135$; 80 dB: 25 ms, $n = 118$), resulting in a mean latency of 29.2 ± 1.9 ms.

Subsequently, we investigated how call amplitudes changed in response to noise perturbation. We calculated mean amplitude shifts after noise onset for first and second phee syllables (Figure 4a and b). We found a significant decrease in call amplitude for first phee syllables ($F(3, 3,081) = 0.99$, $p = .40$ for amplitude, $F(4, 3,081) = 5.24$, $p = .0003$ for frequency, $F(3, 3,081) = 0.56$, $p = .6397$ for monkeys, $n = 2,019$, three-way ANOVA). These shifts were significant for the two middle intensity levels of the 0.1–5.1 kHz noise (60 dB: -1.7 ± 0.5 dB, $p = 3.28e-03$, $n = 103$; 70 dB: -2.7 ± 0.5 dB, $p = 8.17e-04$, $n = 119$) as well as for the two middle intensity levels of the broadband noise (60 dB: -2.3 ± 0.6 dB, $p = 5.15e-03$, $n = 93$; 70 dB: -2.0 ± 0.6 dB, $p = 8.59e-04$, $n = 85$). However, we could not find any systematic increase in amplitude shifts or significant amplitude shifts in any of the five noise conditions ($n = 3,093$; Figure 4a). Furthermore, the combined effect size ($ES_{\text{freq} \times \text{ampl}} = 0.024$) was above 0.02, while the effect size for the frequency ($ES_{\text{freq}} = 0.014$) and amplitude ($ES_{\text{ampl}} = 0.007$) factors was below 0.02, indicating that noise perturbation of ongoing first syllables has only a small or no effect on amplitude shifts.

However, there was also an amplitude decrease in second phee syllables ($F(3, 947) = 3.75$, $p = .01$ for amplitude, $F(4, 947) = 1.38$, $p = .24$ for frequency, $F(3, 947) = 3.8$, $p = .01$ for monkeys, $n = 554$, three-way ANOVA). The amplitude shifts in the 0.1–5.1 kHz and 5–10 kHz noise conditions were significant at the highest intensity levels (-7.2 ± 1.3 dB, $p = 3.90e-02$, $n = 19$ and -7.9 ± 3.1 dB, $p = 2.68e-03$, $n = 16$, respectively; Figure 4b). Monkeys decreased their call amplitudes in these two conditions with increasing noise intensity levels, while no significant call amplitude changes were observed in the other three conditions. All three ES values were above 0.02 ($ES_{\text{freq} \times \text{ampl}} = 0.064$, $ES_{\text{freq}} = 0.030$, $ES_{\text{ampl}} = 0.024$) suggesting an effect of specific noise perturbation on amplitude shifts of second phee syllables in marmoset monkeys. Although it has been already shown that marmoset monkeys show the Lombard effect while producing twitter calls (Brumm, 2004), our results might indicate that marmoset monkeys do not exhibit this reflex when producing phee calls or suppress it and lower their call intensities instead.

To test whether the change in frequency is just a by-product of an accompanied change in amplitude, we investigated the correlation between these two call parameters for all noise conditions and all noise amplitudes used. We observed no significant relationship between call amplitude and call frequency neither for first phee syllables ($R = -0.30$, $p = .19$,

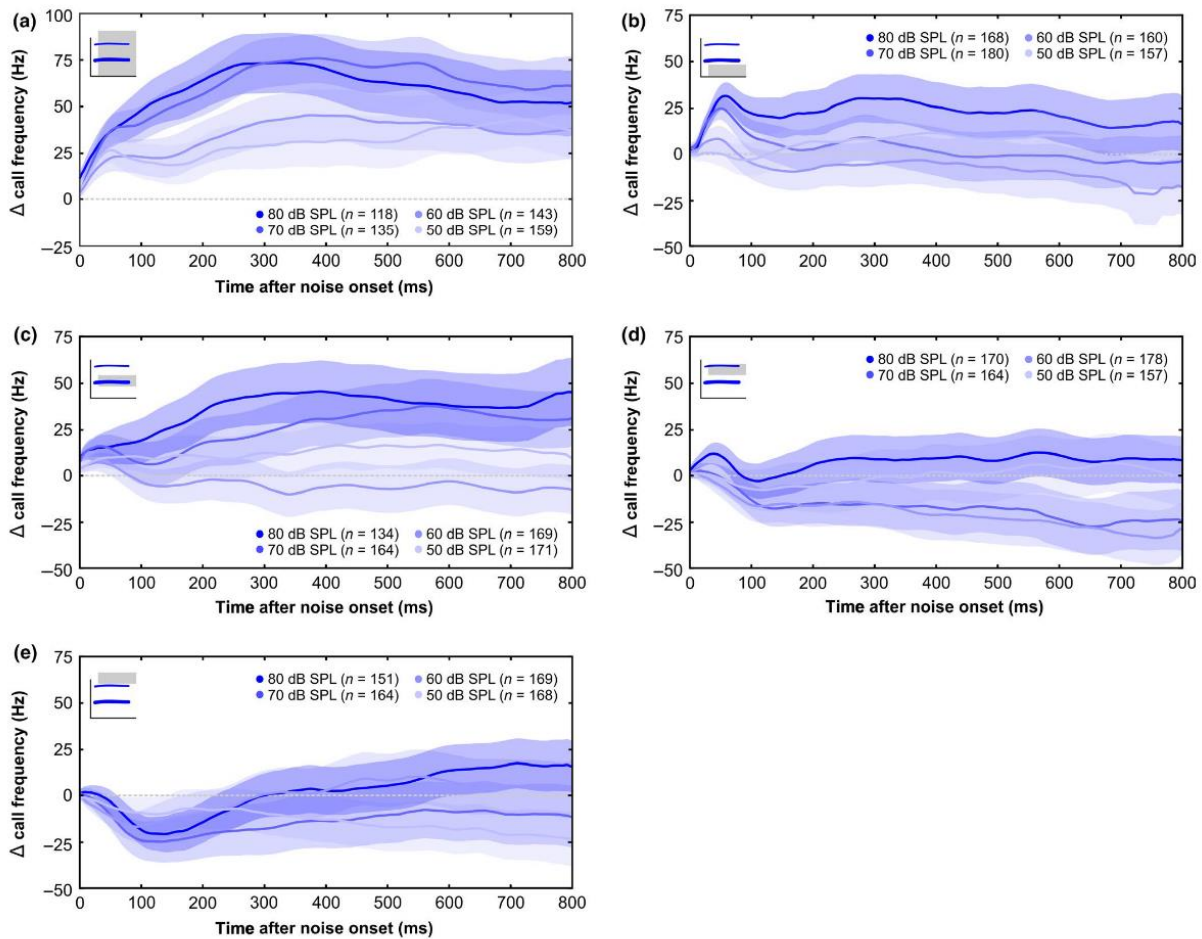


FIGURE 3 Frequency courses in response to noise bursts. First syllable mean Δ frequency courses (Hz) \pm SEM of all amplitude conditions normalized to control data (dashed lines) during (a) broadband, (b) 0.1–5.1 kHz, (c) 5–10 kHz, (d) 10–15 kHz and (e) 16–21 kHz noise conditions over time after noise onset (ms) [Colour figure can be viewed at wileyonlinelibrary.com]

Figure 4c) nor for second phee syllable ($R = -0.33$, $p = .16$, Figure 4d). This indicates that changes in call frequency and amplitude occurred independently from each other.

Finally, we modified our behavioural experiment scheme to test whether our animals are able to show a Lombard effect or suppress it in a noisy environment in general when producing phee calls. We played back all five noise conditions [0.1–5 kHz, 5–10 kHz, 10–15 kHz, 16–21 kHz and broadband] at 70 dB SPL amplitude intensity plus two control conditions with a duration of 180 s each, resulting in a block of seven pseudo-randomized playback conditions with a total duration of 1,260 s. In this new experiment, our monkeys produced a total of 803 phee calls (monkey $F = 222$ phees, $H = 270$, $S = 158$, $W = 153$), which were more commonly uttered ($F = 82.5\%$, $H = 80.4\%$, $S = 84.0\%$, $W = 100\%$) than other produced call types. The relative amounts of single phees ranged between 34.8% and 56.3%, and the relative amounts of double phees ranged between 43.71% and

59.49%. Multi-syllabic phees ($F = 0.5\%$, $H = 1.9\%$, $S = 5.7\%$, $W = 0\%$) and segmented phees ($F = 0.4\%$, $H = 2.4\%$, $S = 0\%$, $W = 0\%$) were nearly absent. Monkey H produced 14.3% trill-phees and monkeys F and S produced 15.2% and 13.8% tsik-ekks, respectively. All other call types were below 2.5% for all monkeys. Under these experimental conditions, we found that monkey W significantly increased its call intensity for both phee syllables when perturbed by noise (first syllable: 6.4 ± 0.8 dB, $p = 1.57e-03$, $n = 107$; second syllable: 8.4 ± 0.9 dB, $p = 6.52e-03$, $n = 46$; Figure 5), thus exhibiting the Lombard effect. Furthermore, monkey S significantly decreased the intensity of the second phee syllable and exhibited no changes in call intensity of the first syllable (second syllable: -4.2 ± 1.0 dB, $p = 2.15e-03$, $n = 52$; first syllable: $p = .10$, $n = 68$), while monkeys F and H showed no significant amplitude change under noise perturbation (first syllable (H): $p = .89$, second syllable (H): $p = .15$, $n = 234$; first syllable (F): $p = .91$, second syllable (F): $p = .06$, $n = 184$). Taken

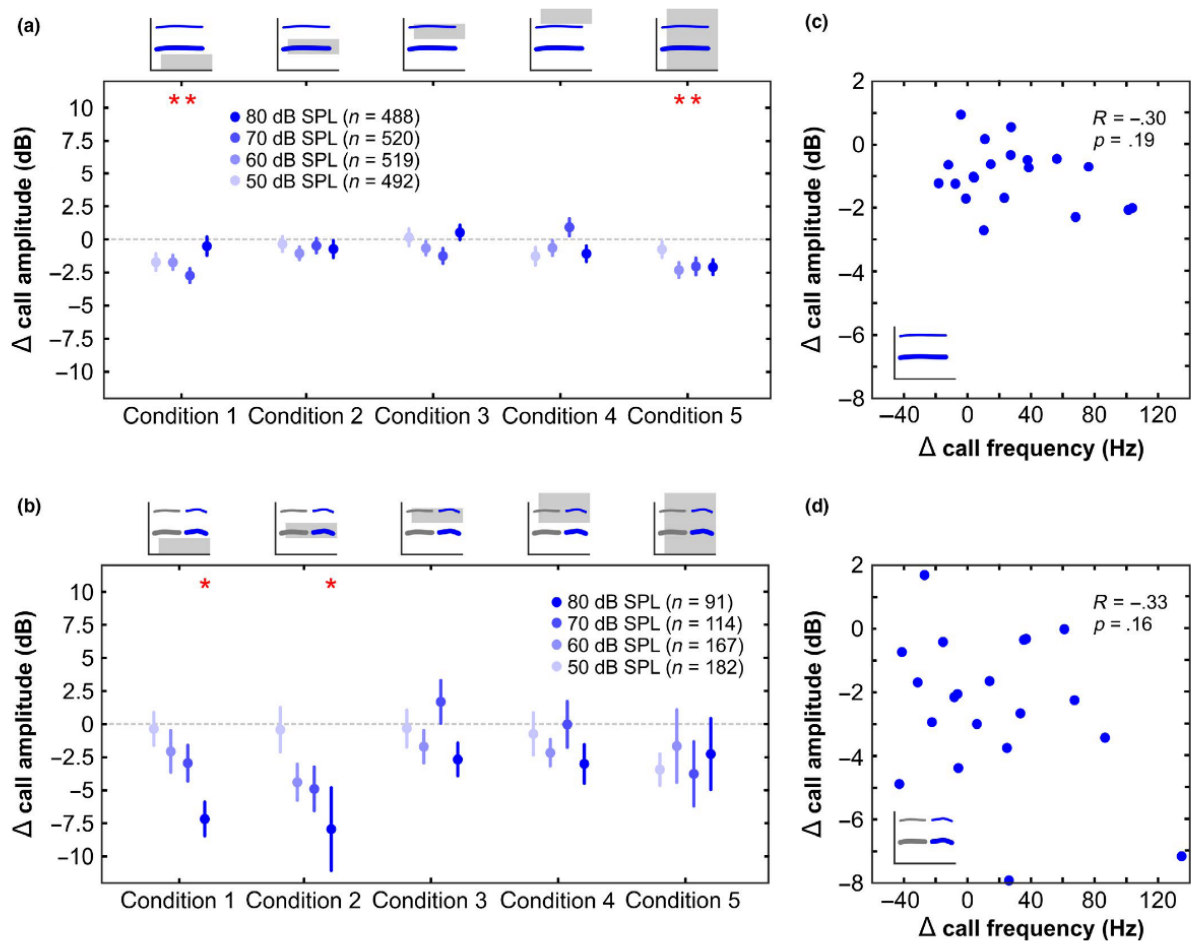


FIGURE 4 Decreasing amplitude shifts in response to noise bursts. Δ call amplitude (dB) per corresponding noise condition normalized to control data (dashed lines). Mean of median amplitudes after noise onset of each call pooled over all monkeys \pm SEM (a) for first phee syllables and (b) for second phee syllables. (c and d) Correlation of mean Δ call amplitude and corresponding Δ call frequency for each noise and amplitude condition, respectively, for first phee syllables (c) and for second phee syllables (d) [Colour figure can be viewed at wileyonlinelibrary.com]

together, the present results suggest that marmosets are capable of exhibiting as well as actively suppressing the Lombard effect in a noisy environment during phee call production.

4 | DISCUSSION

Our results demonstrate that marmoset monkeys show rapid modulation of call parameters in response to perturbing noise bursts presented after call onset. Ongoing phee vocalizations perturbed by ambient noise rapidly increased call frequency in cases where the fundamental frequency was above or directly masked by the perturbing noise. Bandpass-filtered noise bursts, which did not mask but were above the fundamental frequencies of the calls, had no effect on call frequency. Therefore, noise-related frequency modulation is a selective effect in marmoset monkeys. Additionally, call

amplitudes of phee calls were affected by low-frequency noise bands and broadband noise. Surprisingly, phee calls perturbed after call onset did not exhibit a Lombard effect as previously reported for calls that were produced in constantly presented ambient noise (Brumm, 2004; Egnor & Hauser, 2006). Instead, our monkeys decreased their call intensity in a stepwise function with increasing noise intensity. Our findings suggest a general strategy of avoiding calling in a noisy environment in marmoset monkeys.

4.1 | Effects of ambient noise on call frequency

Noise-dependent shifts in call frequency are not well-studied and relatively poorly understood. Only a few studies have reported a rise in call frequencies with increasing amplitudes of

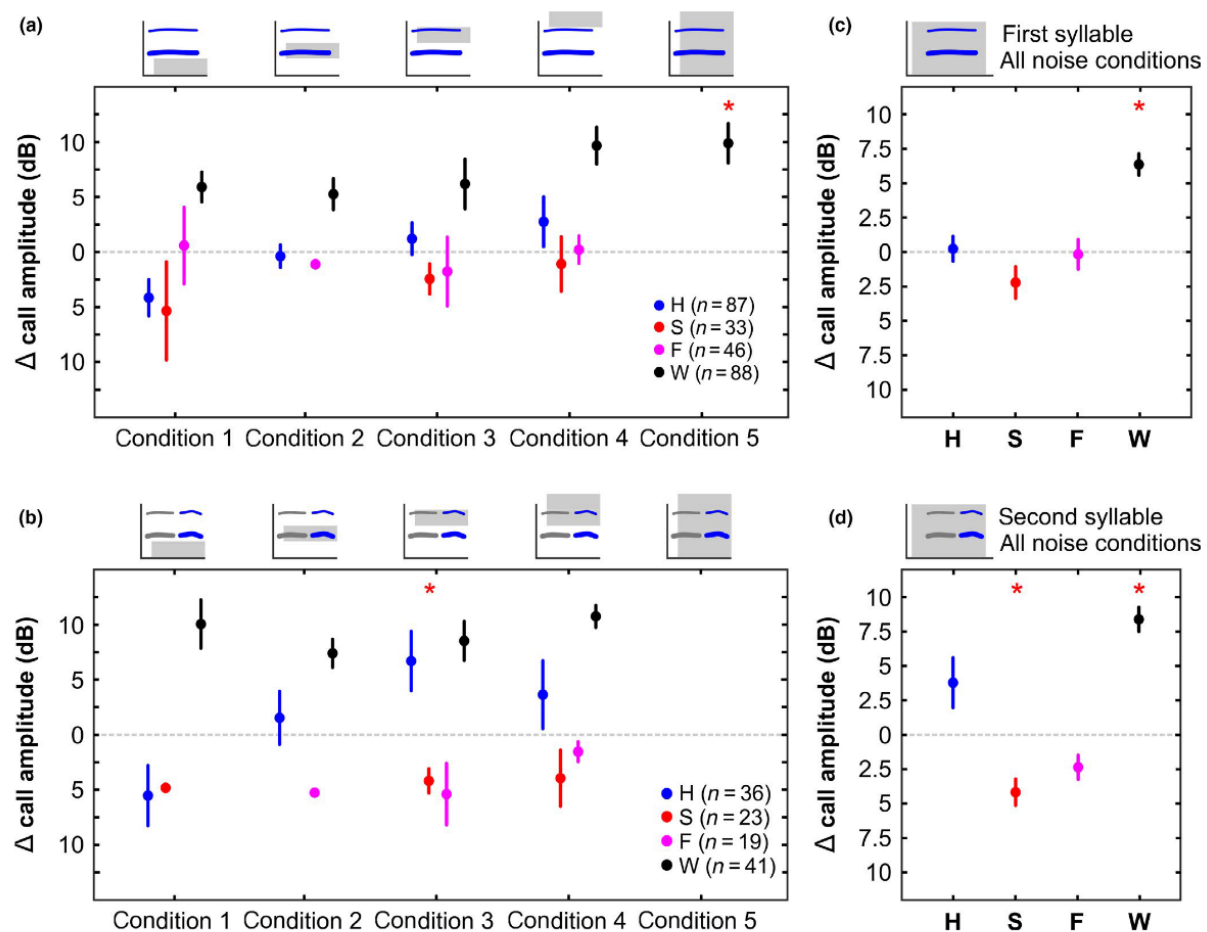


FIGURE 5 Amplitude shifts in response to constant noise perturbation. Pooled maximum Δ amplitudes (dB) \pm SEM during 180-s noise per monkey per noise condition normalized to control data (dashed lines) (a) for first syllables and (b) for second syllables. Over all noise conditions pooled max Δ amplitudes (dB) \pm SEM during 180-s noise per monkey (c) for first syllables and (d) for second syllables. Asterisks denote significant differences [Colour figure can be viewed at wileyonlinelibrary.com]

ambient noise in birds and bats (Hage et al., 2013; Osmanski & Dooling, 2009; Schuster, Zollinger, Lesku, & Brumm, 2012), and only one study investigated the effect of different noise bands on call frequencies. In a recent study, marmosets were able to adjust their phee call frequency to shift away from a predictable high- and low-frequency noise, respectively, which were not overlapping with the phee call frequencies (Zhao, Rad, & Wang, 2019). In bats, the frequencies of echolocation calls increased significantly for a variety of noise stimuli no matter whether they were directly masking the call's fundamental frequency or presented below the dominant call frequency (Hage et al., 2013). In contrast, the present study shows that in marmosets, call frequency was predominantly only affected when we directly masked the call's fundamental frequency. As a result, the strongest rise in call frequencies was found for high noise amplitudes. These findings suggest that the observed rises in call frequencies are an audio-vocal mechanism elicited to increase

call detectability in a noisy environment, as has been found in previous studies involving birds (Bermúdez-Cuamatzin, Ríos-Chelén, Gil, & García, 2011; Nemeth & Brumm, 2010; Pohl, Leadbeater, Slabbekoorn, Klump, & Langemann, 2012). Here, it has been predicted that shifts in song frequencies of about 200 Hz increase call detectability by about 10 to 20% (Nemeth & Brumm, 2010), which is mainly due to the fact that the spectrum of environmental noise generally shows a decay in amplitude with increasing frequency (Halfwerk & Slabbekoorn, 2009; Nemeth & Brumm, 2010; Pohl et al., 2012; Pohl, Slabbekoorn, Klump, & Langemann, 2009). In the present study, shifts in call frequency occurred with a mean latency of about 30 ms after noise onset suggesting a rapid underlying neural mechanism for frequency modulation. Such fast responses to ambient noise have yet only been found in echolocating bats, which exhibit an increase in call amplitude in about 30 ms after noise onset as well (Luo, Kothari, & Moss, 2017).

4.2 | Effects of ambient noise on call amplitude

Despite the positive effect of rises in call frequency on signal detectability, the most effective mechanism to improve SNR ratio in a noisy environment during vocal production is the Lombard effect, that is the involuntary rise in call amplitude as a response to masking noise (Brumm & Zollinger, 2011; Luo et al., 2018). In the present study, noise perturbation starting after phee call onset had no systematic effect on call amplitude of the first syllable, that is the syllable during which noise perturbation started. In cases in which significant shifts occurred, call amplitude did not increase, as expected, but decreased with small effect sizes. This effect was stronger for the second syllables of the phee calls, in which a strong decrease in call amplitude could be observed for low-frequency noise conditions. Consequently, call intensity decreased in a stepwise function with increasing noise intensity suggesting a direct effect of noise intensity on call amplitude. In contrast to our study, the Lombard effect has been observed in marmoset monkeys in two previous studies (Brumm, 2004; Eliades & Wang, 2012). For one study, the different call types that were investigated might explain the apparent discrepancy. While we focused on phee calls, a high amplitude call that is produced at the upper end of the amplitude scale (Eliades & Wang, 2012), one of the earlier studies investigated the twitter call, a vocalization that is produced at lower amplitude intensities (Brumm, 2004). Interestingly, the other study investigating the Lombard effect during phee call production found mixed effects with half of the animals not showing an increase in call amplitude (Eliades & Wang, 2012).

Our results suggest an audio-vocal integration mechanism in marmoset monkeys that is capable of counteracting the Lombard effect. Such a mechanism has been already shown to exist in vocal production learners such as birds and humans (Kobayasi & Okanoya, 2003; Pick, Siegel, Fox, Garber, & Kearney, 1989; Therrien, Lyons, & Balasubramaniam, 2012; Vinney, Mersbergen, Connor, & Turkstra, 2016) and seems to be mainly driven by higher-order cognitive processes including cortical structures (Luo et al., 2018).

4.3 | Vocal flexibility during perturbing noise in marmoset monkeys

Current studies have revealed a high degree of vocal flexibility in marmoset monkeys (Ghazanfar, Liao, & Takahashi, 2019), allowing them to control when (Roy et al., 2011), where (Choi, Takahashi, & Ghazanfar, 2015) and what to vocalize (Liao, Zhang, Cai, & Ghazanfar, 2018). In addition, recent studies revealed that marmosets are able to modulate distinct call parameters in response to acoustic feedback

(Eliades & Tsunada, 2018; Pomberger et al., 2018). This vocal flexibility allows marmosets to avoid calling in the presence of environmental noise and predominantly initiate their vocalizations in silent periods (Roy et al., 2011). In a previous study, we demonstrated that marmosets interrupt their vocalizations shortly after noise onset when perturbation starts after vocal onset (Pomberger et al., 2018), supporting the idea that these animals tend to avoid calling in ambient noise. Such call interruptions, however, were rare (2.6% of all calls), indicating stark neuronal and/or anatomical constraints in exhibiting such behaviour (Pomberger et al., 2018) and resulting in a large fraction of phee calls being perturbed by noise bursts. In the present study, we show that the call amplitude of such vocalizations is lower.

We suggest that marmoset monkeys exhibit this vocal behaviour in a noisy environment to reduce the physiological costs of high-intensity phee calls. Marmoset phee calls are elicited at high intensities above 100 dB SPL, resulting in high muscle tensions encompassing almost the entire animal's body during call production (own observation). Therefore, mechanisms might have evolved in these animals that ensure the proper transmission of these high energetic calls resulting in calling in silent gaps and decreasing call intensity in situations in which sufficient detectability might be potentially diminished, such as during the presence of ambient noise.

4.4 | Mechanisms counteracting involuntary audio-vocal effects

Based on the current work and earlier studies (Pomberger et al., 2018; Roy et al., 2011), we propose a hypothetical neuronal model suggesting various audio-vocal control mechanism involving cortical, subcortical and corticofugal connections capable of modulating vocal behaviour in a noisy environment (Figure 6). In accordance with earlier work (Hage, 2019; Hage & Nieder, 2016), our model consists of a volitional articulatory motor network originating in the prefrontal cortex (PFC) cognitively controlling vocal output of a phylogenetically conserved primary vocal motor network predominantly consisting of a subcortical neuronal network. The vocal motor network can be modulated by auditory structures on several cortical and subcortical brain levels (Luo et al., 2018). The ability to interrupt calls or modulate call parameters as a response to perturbing noise might be controlled by both subcortical mechanisms and corticofugal projections. Our model basically suggests two potential anatomically and neurophysiologically plausible audio-vocal loops, which might be involved in the observed call modifications such as call inhibition and call interruptions as well as shifts in call frequency and amplitude. First, a cortical audio-vocal loop from the auditory cortex to ventrolateral prefrontal cortex to premotor cortex

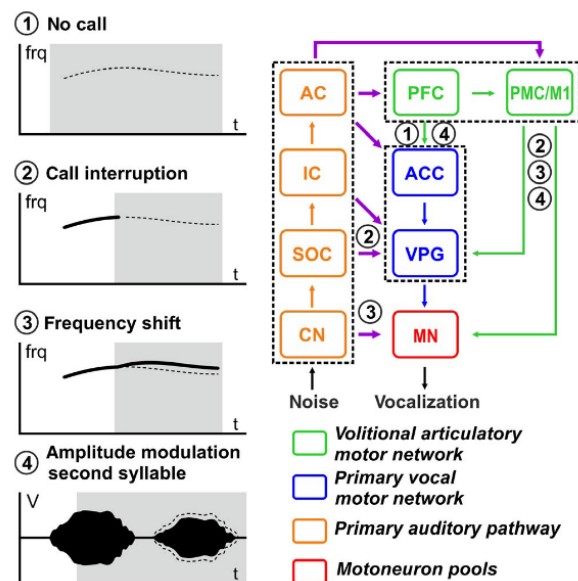


FIGURE 6 Hypothetical neuronal model for audio-vocal interaction. Audio-vocal integration mechanisms are known to happen between cortical and subcortical structures as well as via corticofugal projections. Call production might be affected by ambient noise at different brain levels. Simplified circuit diagram summarizing the most relevant structures for vocal production and the auditory pathway in monkeys. Arrows indicate anatomically verified and relevant direct connections. Numbers highlight connections that might be involved in noise-related call inhibition (1), call interruption (2), frequency shifts and/or amplitude modulations of the second syllable of call sequences such as phee calls. 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 red) involved in call production. External auditory stimuli are perceived via the auditory pathway (indicated in orange), which is capable of modulating the vocal motor system via direct connections (indicated in purple) at cortical and subcortical levels. See text for further explanation. AC, auditory cortex; ACC, anterior cingulate cortex; CN, cochlear nucleus; IC, inferior colliculus; M1, primary motor cortex; MN, motoneuron pools; PFC, prefrontal cortex; PMC, premotor cortex; SOC, superior olivary complex; VPG, vocal pattern generator in the ventrolateral pontine brainstem [Colour figure can be viewed at wileyonlinelibrary.com]

to pontine reticular formation (Hage & Nieder, 2016), all of which may serve as potential hubs in audio-vocal interaction (Eliades & Tsunada, 2018; Eliades & Wang, 2012; Hage, 2018; Hage, Jürgens, & Ehret, 2006; Hage & Nieder, 2015). Furthermore, a direct connection from the premotor cortex to single motoneuron pools might be sufficient for some call modifications, as the inhibition of single muscles, for example muscles involved in expiration, might be sufficient to interrupt vocal output or modulate call amplitude. Second, several anatomically plausible subcortical audio-vocal loop from the cochlear nucleus, superior olivary complex or inferior colliculus to the pontine reticular formation

might be sufficient to mediate some of the observed behaviours such as call interruption or shifts in call frequency (Hage, 2020; Hage & Nieder, 2016). Earlier studies even found direct and active connections between cochlear nucleus and the laryngeal motoneuron pool in mammals, which might be able to modulate vocal output (Jen & Ostwald, 1977). Neurophysiological studies will now have to elucidate at which brain levels audio-vocal integration mechanisms exist that explain the observed capabilities of marmoset monkeys to counteract a previously thought involuntary audio-vocal mechanism, the Lombard effect.

ACKNOWLEDGEMENTS

We 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).

CONFLICT OF INTEREST

We disclose that there are no potential competing interests.

AUTHOR CONTRIBUTIONS

S.R.H. conceived the study; T.P. and S.R.H. designed the experiments; T.P. and J.L. conducted the experiments and performed data analyses; all authors interpreted the data and wrote the manuscript; and S.R.H. provided the animals and supervised the project.

DATA AVAILABILITY STATEMENT

All data needed to evaluate the conclusions in the paper are present in the paper. Additional data related to this paper may be requested from the corresponding author.

ORCID

Steffen R. Hage  <https://orcid.org/0000-0002-7018-543X>

REFERENCES

- Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *The Behavioral and Brain Sciences*, 37, 529–546.
- 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*). *Journal of the Acoustical Society of America*, 138, 2906–2928.
- Bermúdez-Cuamatzin, E., Ríos-Chelén, A. A., Gil, D., & Garcia, C. M. (2011). Experimental evidence for real-time song frequency shift in response to urban noise in a passerine bird. *Biology Letters*, 7, 36–38.
- Berouti, M., Schwartz, R., & Makhoul, J. (1978). *Enhancement of speech corrupted by acoustic noise*. In ICASSP '79. IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 208–211.

- Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication*. Sunderland, MA: Sinauer Associates Inc.
- Brumm, H. (2004). Acoustic communication in noise: Regulation of call characteristics in a New World monkey. *Journal of Experimental Biology*, *207*, 443–448.
- Brumm, H. (2006). Signalling through acoustic windows: Nightingales avoid interspecific competition by short-term adjustment of song timing. *Journal of Comparative Physiology A*, *192*, 1279–1285.
- Brumm, H., Schmidt, R., & Schrader, L. (2009). Noise-dependent vocal plasticity in domestic fowl. *Animal Behavior*, *78*, 741–746.
- Brumm, H., & Slabbekoorn, H. (2005). Acoustic communication in noise. *Advances in the Study of Behaviour*, *35*, 151–209.
- Brumm, H., & Zollinger, A. (2011). The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour*, *148*, 1173–1198.
- Choi, J. Y., Takahashi, D. Y., & Ghazanfar, A. A. (2015). Cooperative vocal control in marmoset monkeys via vocal feedback. *Journal of Neurophysiology*, *114*, 274–283.
- Egnor, S. R., & Hauser, M. (2006). Noise-induced vocal modulation in cotton-top tamarins (*Saguinus oedipus*). *American Journal of Primatology*, *68*, 1183–1190.
- 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. *Journal of Experimental Biology*, *209*, 3652–3663.
- Eliades, S. J., & Tsunada, J. (2018). Auditory cortical activity drives feedback-dependent vocal control in marmosets. *Nature Communications*, *9*, 2540.
- Eliades, S. J., & Wang, X. (2012). Neural correlates of the lombard effect in primate auditory cortex. *Journal of Neuroscience*, *32*, 10737–10748.
- Ghazanfar, A. A., Liao, D. A., & Takahashi, D. Y. (2019). Volition and learning in primate vocal behaviour. *Animal Behavior*, *151*, 239–247.
- Hage, S. R. (2018). Auditory and audio-vocal responses of single neurons in the monkey ventral premotor cortex. *Hearing Research*, *366*, 82–89.
- Hage, S. R. (2019). Precise vocal timing needs cortical control. *Science*, *363*, 926–928.
- Hage, S. R. (2020). The role of auditory feedback on vocal pattern generation in marmoset monkeys. *Current Opinion in Neurobiology*, *60*, 92–98.
- 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. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 4063–4068.
- Hage, S. R., Jürgens, U., & Ehret, G. (2006). Audio-vocal interaction in the pontine brainstem during self-initiated vocalization in the squirrel monkey. *European Journal of Neuroscience*, *23*, 3297–3308.
- Hage, S. R., & Nieder, A. (2015). Audio-vocal interaction in single neurons of the monkey ventrolateral prefrontal cortex. *Journal of Neuroscience*, *35*, 7030–7040.
- Hage, S. R., & Nieder, A. (2016). Dual neural network model for the evolution of speech and language. *Trends in Neurosciences*, *39*, 813–829.
- Halfwerk, W., & Slabbekoorn, H. (2009). A behavioural mechanism explaining noise-dependent frequency use in urban birdsong. *Animal Behavior*, *78*, 1301–1307.
- Jen, P. H. S., & Ostwald, J. (1977). Response of cricothyroid muscles to frequency-modulated sounds in FM bats, *Myotis lucifugus*. *Nature*, *265*, 77–78.
- Kobayasi, K. I., & Okanoya, K. (2003). Context-dependent song amplitude control in Bengalese finches. *NeuroReport*, *14*, 521–524.
- Liao, D. A., Zhang, Y. S., Cai, L. X., & Ghazanfar, A. A. (2018). Internal states and extrinsic factors both determine monkey vocal production. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 201722426.
- Lombard, E. (1911). Le signe de l'élevation de la voix. *Annales Des Maladies De L'oreille Et Du Larynx*, *37*, 101–119.
- Luo, J., Goerlitz, H. R., Brumm, H., & Wiegrebe, L. (2015). Linking the sender to the receiver: Vocal adjustments by bats to maintain signal detection in noise. *Scientific Reports*, *5*, 1–11.
- Luo, J., Hage, S. R., & Moss, C. F. (2018). The Lombard Effect: from acoustics to neural mechanisms. *Trends in Neurosciences*, *41*, 938–949.
- Luo, J., Kothari, N. B., & Moss, C. F. (2017). Sensorimotor integration on a rapid time scale. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 6605–6610.
- Miller, C. T., Eliades, S. J., & Wang, X. (2009). Motor planning for vocal production in common marmosets. *Animal Behavior*, *78*, 1195–1203.
- Miller, C. T., Flusberg, S., & Hauser, M. D. (2003). Interruptibility of long call production in tamarins: Implications for vocal control. *Journal of Experimental Biology*, *206*, 2629–2639.
- Nemeth, E., & Brumm, H. (2010). Birds and anthropogenic noise: Are urban songs adaptive? *American Naturalist*, *176*, 465–475.
- Osmanski, M. S., & Dooling, R. J. (2009). The effect of altered auditory feedback on control of vocal production in budgerigars (*Melopsittacus undulatus*). *Journal of the Acoustical Society of America*, *126*, 911–919.
- Osoorio, D., & Vorobyev, M. (2008). A review of the evolution of animal colour vision and visual communication signals. *Vision Research*, *48*, 2042–2051.
- Pick, H. L., Siegel, G. M., Fox, P. W., Garber, S. R., & Kearney, J. K. (1989). Inhibiting the Lombard effect. *Journal of the Acoustical Society of America*, *85*, 894–900.
- Pistorio, A. L., Vintch, B., & Wang, X. (2006). Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). *Journal of the Acoustical Society of America*, *120*, 1655–1670.
- Pohl, N. U., Leadbeater, E., Slabbekoorn, H., Klump, G. M., & Langemann, U. (2012). Great tits in urban noise benefit from high frequencies in song detection and discrimination. *Animal Behavior*, *83*, 711–721.
- Pohl, N. U., Slabbekoorn, H., Klump, G. M., & Langemann, U. (2009). Effects of signal features and environmental noise on signal detection in the great tit, *Parus major*. *Animal Behavior*, *78*, 1293–1300.
- Poldrack, R. A., & Farah, M. J. (2015). Progress and challenges in probing the human brain. *Nature*, *526*, 371–379.
- Pomberger, T., Risueno-Segovia, C., Löschner, J., & Hage, S. R. (2018). Precise motor control enables rapid flexibility in vocal behavior of marmoset monkeys. *Current Biology*, *28*, 788–794.e3.
- Roy, S., Miller, C. T., Gottsch, D., & Wang, X. (2011). Vocal control by the common marmoset in the presence of interfering noise. *Journal of Experimental Biology*, *214*, 3619–3629.
- Schuster, S., Zollinger, S. A., Lesku, J. A., & Brumm, H. (2012). On the evolution of noise-dependent vocal plasticity in birds. *Biology Letters*, *8*, 913–916.
- Stockhorst, U., & Pietrowsky, R. (2004). Olfactory perception, communication, and the nose-to-brain pathway. *Physiology & Behavior*, *83*, 3–11.

- Therrien, A. S., Lyons, J., & Balasubramaniam, R. (2012). Sensory attenuation of self-produced feedback: The Lombard effect revisited. *PLoS ONE*, *7*, e49370.
- Vinney, L. A., van Mersbergen, M., Connor, N. P., & Turkstra, L. S. (2016). Vocal control: Is it susceptible to the negative effects of self-regulatory depletion? *Journal of Voice*, *30*, 638.e21–638.e31.
- Zelick, R. D., & Narins, P. M. (1982). Analysis of acoustically evoked call suppression behaviour in a neotropical treefrog. *Animal Behavior*, *30*, 728–733.
- Zhao, L., Rad, B. B., & Wang, X. (2019). Long-lasting vocal plasticity in adult marmoset monkeys. *Proceedings of the Royal Society B-Biological Sciences*, *286*, 1–10.
- Zürcher, Y., & Burkart, J. M. (2017). Evidence for dialects in three captive populations of common marmosets (*Callithrix jacchus*). *International Journal of Primatology*, *38*, 780–793.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Pomberger T, Löschner J, Hage SR. Compensatory mechanisms affect sensorimotor integration during ongoing vocal motor acts in marmoset monkeys. *Eur J Neurosci*. 2020;52:3531–3544. <https://doi.org/10.1111/ejn.14721>

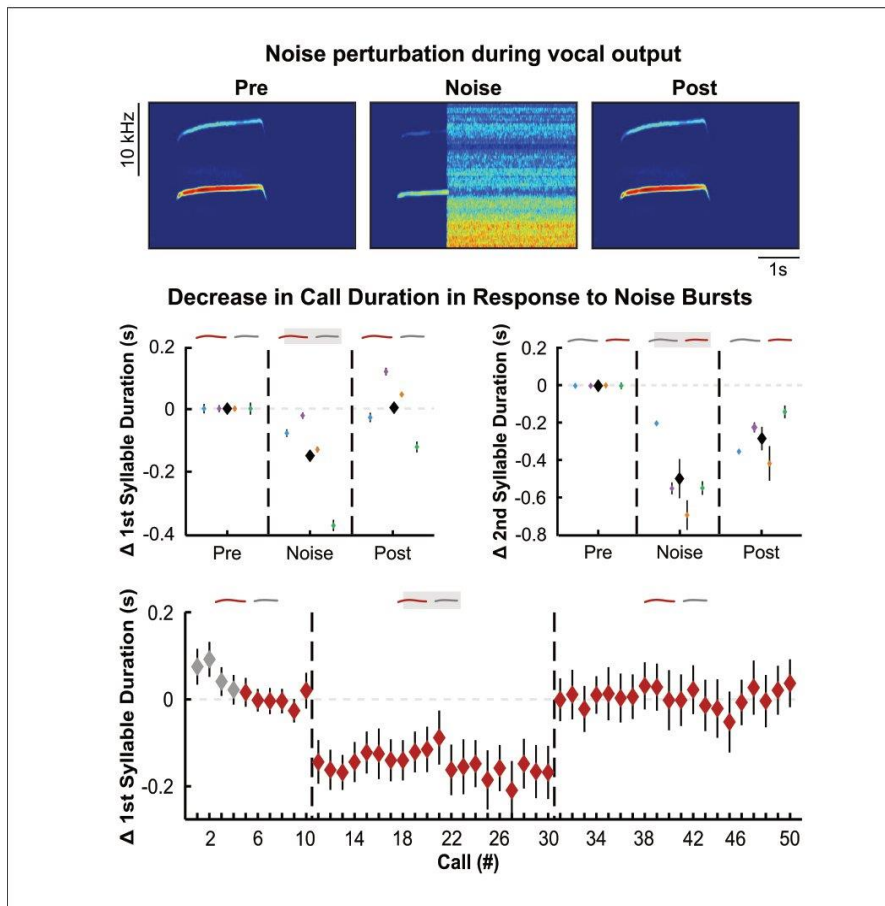
Publication 2: Marmoset monkeys use different avoidance strategies to cope with ambient noise during vocal behavior

Löschner, J., Pomberger, T., & Hage, S. R. (2023). Marmoset monkeys use different avoidance strategies to cope with ambient noise during vocal behavior. *Iscience*, 26(3), 106219.



Article

Marmoset monkeys use different avoidance strategies to cope with ambient noise during vocal behavior



Julia Löschner,
Thomas
Pomberger,
Steffen R. Hage

steffen.hage@uni-tuebingen.de

Highlights

Monkeys are able to cancel their calls to avoid interfering ambient noise

Animals seem to find it easier to cancel calls toward their expected end

Marmosets change their vocal behavior in reflexive and adaptive ways

We found indications of active noise-dependent interruption of call sequences

Löschner et al., iScience 26, 106219
March 17, 2023 © 2023 The Author(s).
<https://doi.org/10.1016/j.isci.2023.106219>





Article

Marmoset monkeys use different avoidance strategies to cope with ambient noise during vocal behavior

Julia Löschner,^{1,2} Thomas Pomberger,^{1,3,4} and Steffen R. Hage^{1,2,5,*}

SUMMARY

Multiple strategies have evolved to compensate for masking noise, leading to changes in call features. One call adjustment is the Lombard effect, an increase in call amplitude in response to noise. Another strategy involves call production in periods where noise is absent. While mechanisms underlying vocal adjustments have been well studied, mechanisms underlying noise avoidance strategies remain largely unclear. We systematically perturbed ongoing phee calls of marmosets to investigate noise avoidance strategies. Marmosets canceled their calls after noise onset and produced longer calls after noise-phases ended. Additionally, the number of uttered syllables decreased during noise perturbation. This behavior persisted beyond the noise-phase. Using machine learning techniques, we found that a fraction of single phees were initially planned as double phees and became interrupted after the first syllable. Our findings indicate that marmosets use different noise avoidance strategies and suggest vocal flexibility at different complexity levels in the marmoset brain.

INTRODUCTION

Communication between animals is a crucial trait for evolutionary success.¹ In many bird and mammalian species, vocal signals have evolved as one of the dominant forms of direct communication between individuals.^{2–4} These vocal signals are usually produced within noisy environments; therefore, several mechanisms have evolved to compensate for masking acoustic interference, such as heavy rain, wind, and animal or urban sounds.⁵

The mechanisms used to deal with acoustic disturbances can generally be divided into two main types. The first type is modulation of vocalizations that are directly confronted with an increase in ambient noise, i.e., when they are produced during acoustic perturbation. Here, the Lombard effect, an involuntary increase in call amplitude in response to masking ambient noise, represents one of the most efficient mechanisms for optimizing the signal-to-noise ratio.^{6–10} This effect is often accompanied by several other vocal changes, such as an increase in call frequency or an increase in signal density, i.e., the production of longer calls and/or increased repetition of call syllables.^{10–13} The latter increases the probability that a transmitted message will be received by another individual. These mechanisms elicited during the presence of masking noise are well known and occur across vertebrates from fish to birds to mammals, including humans.^{5,9,14–16}

In addition to enhancing signal transmission in noisy environments, there is another vocal control mechanism that prevents vocal output during noisy events and induces animals to vocalize during periods when there is little or no acoustic disturbance. Previous studies showed that monkeys are able to time their vocalizations to silent epochs when time windows of ambient noise were predictable to avoid call transmissions with a low signal-to-noise ratio¹⁷ and to stop sequences of calls immediately after acoustic perturbation.^{18,19} Recently, we showed that marmosets are able to interrupt ongoing vocalizations directly after noise onset indicating that marmosets can avoid calling in noise on a rapid timescale.²⁰ However, this behavior was very rare and happened only in 0.3%–7.7% of all cases (depending on the monkey). While the control mechanisms underlying vocal adjustments in response to environmental noise are already well studied, the mechanisms underlying noise avoidance strategies are still largely unclear. For example, it is not yet clear whether the ability to time and interrupt vocalizations in response to environmental noise is based on reflexive and/or adaptive processes.

¹Neurobiology of Social Communication, Department of Otolaryngology – Head and Neck Surgery, Hearing Research Center, University of Tübingen, Medical Center, Elfriede-Aulhorn-Strasse 5, 72076 Tübingen, Germany

²Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Str. 25, 72076 Tübingen, Germany

³Graduate School of Neural & Behavioural Sciences - International Max Planck Research School, University of Tübingen, Osterberg-Str. 3, 72074 Tübingen, Germany

⁴Present address: Neurobiology of Communication, Department of Neurobiology, Duke University School of Medicine, 311 Research Drive, Durham, NC 27710-4432

⁵Lead contact

*Correspondence: steffen.hage@uni-tuebingen.de

<https://doi.org/10.1016/j.isci.2023.106219>



Using a neuroethological approach, we systematically perturbed ongoing vocalizations of marmoset monkeys to investigate the nature of the strategies underlying the exhibited changes in vocal behavior and to decipher, why call interruption occurred only very rarely in our previous study.²⁰ To do so, we perturbed ongoing vocalizations with noise presented at different time points and detected changes in vocal behavior that supported both reflexive and adaptive behavior in response to noise perturbation. Marmosets canceled their calls immediately after noise onset. Hereby, monkeys are more likely to cancel calling toward the expected end of calls rather than at the beginning of the vocalization suggesting that the rare occurrence of this behavior in an earlier study might be explained by the short noise onset latencies used.²⁰ This behavior started during the first perturbed call, indicating a reflexive behavior in response to noise perturbation. In contrast, the reduction in number of syllables persisted beyond noise perturbation, indicating adaptive behavior in response to perturbing noise. Using machine learning techniques based on call parameters, we were able to show that even after the end of noise perturbation, some of the calls produced were initially planned by the monkeys as double phees but ended up being uttered with only one syllable. These findings suggest that marmoset monkeys use two different strategies to avoid noise perturbation during vocal behavior: a rapid one driven by reflexive responses and a more flexible, later occurring, and longer lasting adaptive one.

RESULTS

Marmoset monkeys are able to change their vocal behavior in response to perturbing noise playback

We measured the vocal behavior of marmoset monkeys (*Callithrix jacchus*, $n = 4$), while separated in a soundproofed chamber, with and without acoustic perturbation. In this setting, marmoset monkeys predominantly produced phee vocalizations.^{10,20} In total, we recorded 7945 phees (5843 single and 2102 double phees, Figure 1A). The first 10 calls of each session were not perturbed (pre-phase) and served as a control for each recording day (Figure 1D). The next 20 calls were perturbed in a precise manner if they exceeded a certain duration (noise-phase, Figures 1B and 1C). The rest of the calls produced after the noise-phase were again not perturbed (post-phase). With this experimental design, we were able to investigate changes of vocal behavior in response to perturbing noise as well as long-lasting, adaptive vocal changes after noise perturbation. We first investigated if the ratio of single to double phees was affected by noise perturbation. We found that marmoset monkeys emitted significantly more single phees in the noise- and post-phase than in the pre-phase (likelihood ratio [LR] test, $\chi^2(2) = 687.781$, $p < 0.0001$, Figure 1E). This indicated that the monkeys changed their vocal behavior in response to noise perturbation by producing fewer double phees, consistent with our previous findings.²⁰

The extent of phee duration changes is dependent on noise onset

As a next step, we investigated the general effect of noise perturbation on call durations of first phee syllables (i.e., single phees and the first syllables of double phees) independently of the exact noise onset latency after call onset. We found a significant difference in call duration between the three experimental phases (linear mixed model [LMM], $F(24,322) = 185.006$, $p < 0.0001$, n : pre = 681, noise = 2023, post = 1624). Overall, call durations were significantly shorter during noise perturbation than during pre-phase (approximately 0.15 s); vocalizations produced in the post-phase did not show any differences to calls in the pre-phase (post-hoc Tukey: Pre vs. noise & noise vs. post $p < 0.0001$, pre vs. post $p = 0.910$, Figure 2A). However, the change in call duration across phases did not occur systematically across monkeys. The two female monkeys (orange and purple) showed an increase in first syllable duration in the post-phase compared to the pre-phase ($p < 0.0001$). In contrast, one of the males (green) showed a significant shortening of the first syllable when comparing the pre- and post-phases ($p < 0.0001$); the other male (blue) showed no differences in the length of the first syllable between the post- and pre-phases ($p = 0.505$).

To test how fast the monkeys modulated phee duration in response to noise perturbation, we measured the duration of the first phee syllables with respect to when they were uttered within the session. We found that across sessions, the first call perturbed with noise showed a significant decrease in call duration compared to the last call of the pre-phase, indicating an immediate response to noise perturbation (session calls 10 and 11; paired t-test, $t(113) = 4.979$, $p < 0.0001$, $n = 117$ pre-noise pairs). Similarly, the first unperturbed call after the noise-phase was longer than the last perturbed call (session calls 30 and 31; paired t-test, $t(91) = 4.240$, $p < 0.0001$, $n = 98$ noise-post pairs, Figure 2B). Furthermore, we wanted to know how persistent and long-lasting the change in call duration was. Therefore, we compared the duration of the last 10 calls within the post-phase with the duration of the first 10 calls of the pre-phase of the next day's session.

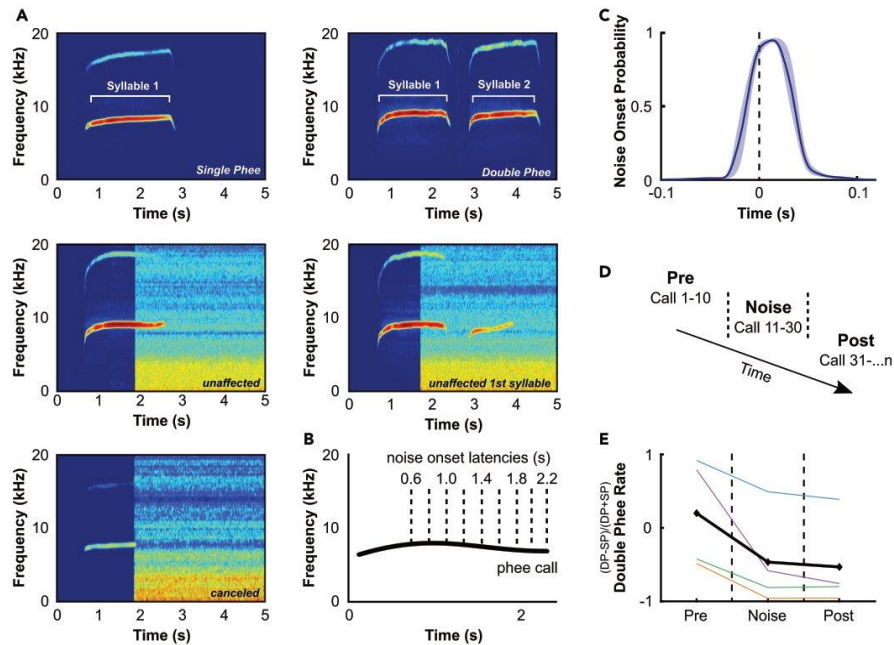


Figure 1. Marmoset monkeys are able to change their vocal behavior in response to perturbing noise playback
 (A) Exemplar spectrograms of single and double phees without and with noise perturbation. Unaffected: Calls with durations being unaffected by noise perturbation; Canceled: calls being canceled after noise onset.
 (B) Schematic of a phee call and the different used noise onset latencies indicated by the dashed lines.
 (C) Pooled noise onset probability normalized to set latencies (dashed line) \pm SEM.
 (D) Schematic of the experimental procedure: pre-phase consisted of the first 10 calls without noise perturbation. Noise-phase consisted of the following 20 calls which were perturbed by noise. All subsequent calls (post) were again unperturbed. Arrow indicates the temporal trajectory of each session.
 (E) Double phee rate per experimental phase. 1: only double phees (DP) were emitted, -1: only single phees (SP) were produced. Different colors indicate different monkeys. The average is displayed in black. Noise-phase is indicated by dashed lines.

We found that the duration of first phee syllables was significantly greater than the duration of the last calls of the previous session (paired t-test, $t(89) = 6.534$, $p < 0.0001$, $n = 90$ post-pre pairs, Figure 2C). Next, we investigated whether call duration was dependent on the onset of noise perturbation related to call offset. During noise-phases, we perturbed calls with noise starting 0.6–2.2 s after call onset (see Methods for details). Because the monkeys exhibited different median call durations, we normalized noise onset latencies (which were initially set by call onset) to the time between noise onset and median call offset. For example, monkey H had a median first phee syllable duration of 1.6 s and monkey W of 2.2 s. Therefore, a noise onset time of 1.6 s resulted in 0 ms time between noise onset and median call offset for monkey H and 600 ms for monkey W. We assumed that if the monkeys perfectly responded to noise perturbation, they would routinely stop their vocalizations directly after perturbation onset (<200 ms after noise onset, horizontal dashed lines, Figure 2D). Alternatively, if the call durations were unaffected by the perturbing noise, they would still be similar to their regular median call duration as indicated by the angular line. However, we found that the monkeys did not purely show one or the other of these assumed behaviors. Their call durations showed both a significant difference from perfect compensation to noise perturbation (z-test, $z(2596) = 34.171$, $p < 0.0001$), as well as from fully unaffected call durations (z-test, $z(2596) = 52.2$, $p < 0.0001$), indicating a more complex vocal behavior in response to different noise onset times.

Consequently, we took a closer look at call duration distributions related to the different noise onset times. Particularly for the short noise onset latencies (time before noise onset), we found two distinct clusters for call duration after noise onset (Figure 2E). These two clusters were also evident when considering call durations in general (Figure 2E). In addition, call durations during noise perturbations were shorter than

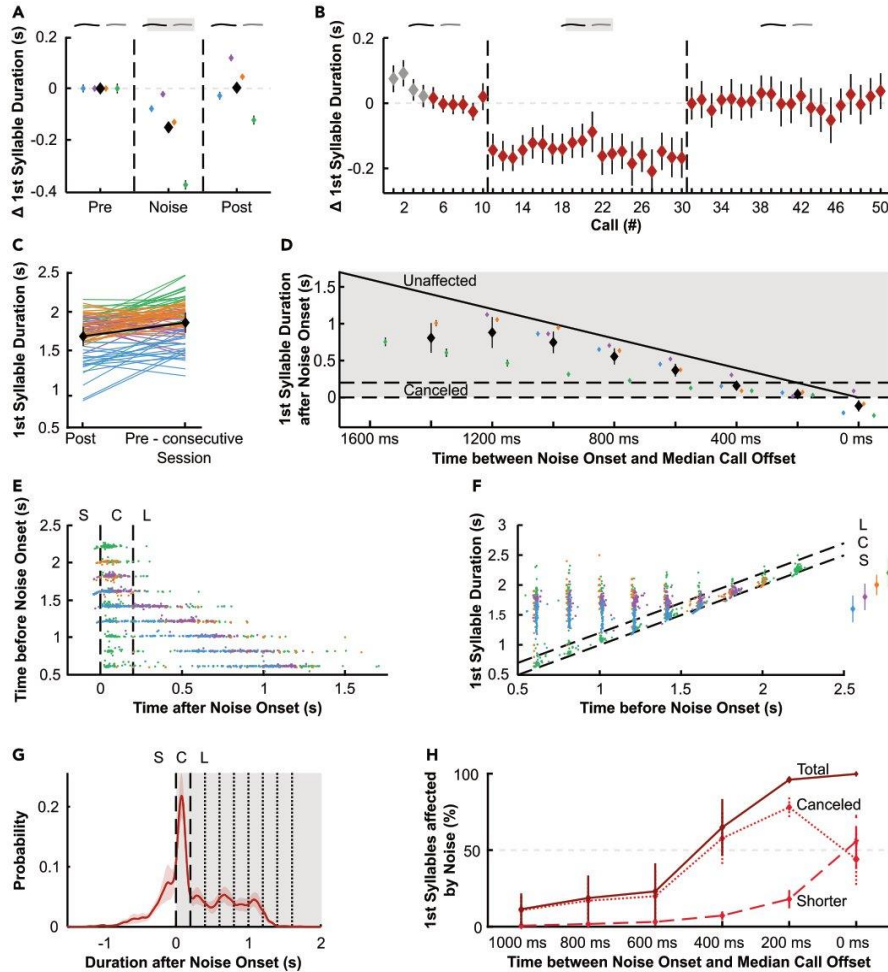


Figure 2. The extent of phee duration changes is dependent on noise onset

Pooled Δ first syllable durations \pm SEM in (A) noise- and post-phases per monkey (monkeys' average displayed in black) normalized to pre data (gray dashed line) in (B) pooled call by call. The first 4 calls (indicated in gray) were discarded for a more stable baseline (see Methods for detail).

(C) Average first syllable duration \pm SEM of every 10 last calls per session (post) and first 10 calls of the consecutive session (pre-consecutive session), respectively.

(D) Pooled Δ first syllable duration \pm SEM after noise onset over time between noise onset and median call offset. Dashed lines show canceled calls (directly affected by noise). Angular line shows the values for unaffected calls. Noise perturbation is shaded in gray.

(E) Call duration before noise onset as a function of call duration after noise onset. Dashed lines indicate canceled calls. S: Calls which were shorter than the given noise onset latency (calls ended before noise perturbation). C: Canceled calls (calls ended <200 ms after noise onset). L: Calls which ended >200 ms after noise onset.

(F) Call duration as a function of call duration before noise onset. Inset on the right: median call duration for each monkey \pm SE. Dashed lines indicate canceled calls.

(G) Average first syllable call duration after noise onset \pm SEM. Dashed lines indicate canceled calls. Dotted lines indicate noise onset latencies. Noise perturbation is shaded in gray.

(H) Proportion of first syllables affected by noise perturbation against time between noise onset and median call offset \pm SEM. Different colors indicate individual monkeys in (A), (C), (D), (E), and (F). See also [Figure S1](#).

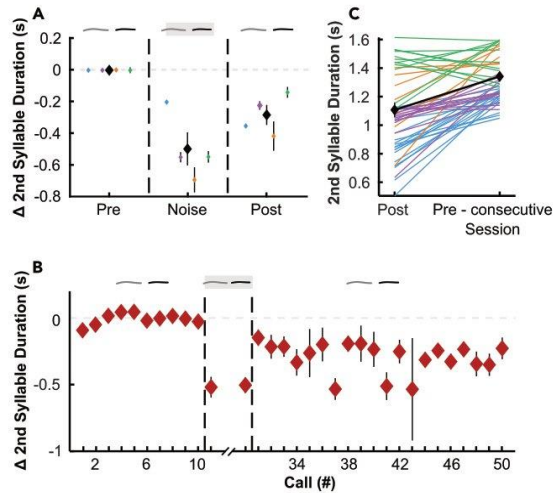


Figure 3. Second syllable duration changes due to noise perturbation
Pooled Δ second syllable durations \pm SEM in (A) noise- and post-phases per monkey normalized to pre data (gray dashed line) in (B) pooled call by call.
(C) Average first syllable duration \pm SEM of every 10 last calls per session (post) and first 10 calls of the consecutive session (pre-consecutive session), respectively. Different colors indicate different monkeys. The average is displayed in black. Noise-phase is indicated by dashed vertical lines. See also [Figure S2](#).

median call durations ([Figure 2F](#)). When analyzing call duration probabilities, we found a multimodal distribution for the pooled data including all used noise onset latencies ([Figure 2G](#)) as well as for the individual animals for most noise onset conditions ([Figure S1](#)), indicating that calls were either canceled directly after noise onset (call offset <200 ms after noise onset) or were unaffected by noise onset. Therefore, we divided call durations into three groups with respect to the relationship between call offset and noise onset: “shorter” for calls ending prior to noise onset, “canceled” for calls ending within 200 ms after noise onset, and “longer” for vocalizations that did not appear to be directly affected by noise onset, i.e., calls that did not end within the first 200 ms after noise onset. We found that the occurrence of these three durations showed significant differences within noise onset conditions (LR test, $\chi^2(10) = 1984.646$, $p < 0.0001$, [Figure 2H](#)). Canceled and shorter calls occurred predominantly in noise onset conditions that were close to the median call offset (400, 200, and 0 ms before median call offset). The presence of these call duration groups decreased rapidly with increasing time between noise onset and median call offset indicating that the monkeys have more control over call duration at the end of their vocalizations. Our results suggest that marmoset monkeys are capable of adaptively modifying the duration of ongoing phoe vocalizations when facing perturbing noise.

Second syllable duration changes due to noise perturbation

Next, we investigated how the duration of second syllables of double phees, which were fully perturbed by noise, was affected. We found significant differences in second syllable durations between experimental phases (LMM, $F(32,095) = 120.719$, $p < 0.0001$, n : pre = 675, noise = 417, post = 764). Similar to call duration distributions of the first phoe syllables, call durations were shorter during noise perturbation than in the pre-phase (0.48 s shorter than in the pre-phase). Interestingly, this effect was much stronger than for first syllable durations. Moreover, after noise perturbation (post-phase), durations of the second syllables were significantly shorter than in the pre-phase (0.28 s; post-hoc Tukey all comparisons $p < 0.0001$, [Figure 3A](#)). This means that in the rare cases where a second syllable was produced after noise perturbation, the duration of these calls was significantly shorter. As a next step, we wanted to test how fast the monkeys modulated the duration of the second syllables in response to noise perturbation. Therefore, we measured the duration of the second phoe syllables as a function of when they were uttered within the session. As with the first syllables, we found that across all sessions, the first call perturbed with noise showed a significant decrease in call duration compared to the last call of the pre-phase, indicating an immediate response to noise perturbation (session calls 10 and 11; paired t-test, $t(52) = 7.562$, $p < 0.0001$, $n = 53$ pre-noise pairs).

This effect could be found in all four monkeys (Figure S2). Similarly, the first unperturbed call after the noise-phase was again longer than the last perturbed call (session calls 30 and 31; paired t-test, $t(35) = 3.708$, $p < 0.0001$, $n = 36$ noise-post pairs, Figure 3B). Additionally, we also wanted to know how persistent and permanent the change in call duration was for the second phee syllables. Therefore, we also compared the duration of the last 10 calls in the post-phase with the duration of the first 10 calls in the pre-phase of the next day's session. We found that the duration of the second phee syllables was significantly longer than the duration of the last calls of the previous session (paired t-test, $t(47) = 6.856$, $p < 0.0001$, $n = 48$ post-pre pairs, Figure 3C).

Accuracy of single phee prediction for single phees produced during different experimental phases

Nevertheless, all monkeys barely produced double phees during the noise-phase. Does this indicate that monkeys switch their vocal behavior from double phees to single phees or that they cancel double phees after the first syllable? This is an important point because it allows us to determine the level of marmoset vocal plasticity with regard to vocal motor control. In the first case, marmosets would be able to adaptively change call initiation in response to perturbing noise; in the second case, they would be able to adaptively change a pre-determined behavior. This question was driven by our observation that while single phees were relatively stable in duration, single phees in the noise- and post-phase often had the same temporal structure and duration as the first syllable of a double phee (Figure 4A). Furthermore, the duration distribution of single phees in the post-phase encompassed the duration distribution of single phees and the first syllables of double phees (Figure 4B). Therefore, we hypothesized that some of the single phees in the noise- and post-phase are first syllables of double phees whose second syllable have been canceled after the production of the first syllable. To test whether single phees during noise- and post-phases were true single phees or the first syllable of double phees, we trained a machine learning classification model (medium support vector machine with 25% holdout) with several call features (for details see Table S1) of single and double phees produced in the pre-phase. Two monkeys produced only a few single phees and were, therefore, excluded from the analysis. For the other two monkeys, we obtained a predictive power of 95% and 96%, respectively, for the correct classification of single phees during the pre-phase, indicating that the classification model worked as expected. We then tested the model's predictive power for single phees in the noise- and post-phases. We argue that if a fraction of produced single phees are first syllables of double phees that have been canceled, the predictive power of our classification model should drop significantly. Indeed, the predictive power was significantly decreased for monkey W in the noise- and post-phases (LR test, $\chi^2(4) = 26.158$, $p < 0.0001$; Figure 4C). Furthermore, for monkey M, we found a similar effect, with a decrease in predictive power in the noise- and first post-phase (LR test, $\chi^2(4) = 9.474$, $p = 0.0503$; Figure 4C). While the predictive power was also significantly lower in the post-phase for monkey W, it recovered for monkey M in the post-phase for single phees and increased during the post-phase in monkey W. These results provide the first evidence that marmoset monkeys can change the number of phee syllables after call onset.

DISCUSSION

In the current study, we showed that marmoset monkeys exhibit adaptive and reflexive noise-dependent changes in acoustic call structure. They produced fewer phee syllables while vocalizing in noise. This behavior remained after noise presentation, indicating an adaptive behavior. Furthermore, marmosets decreased their call durations as a direct response to noise perturbation during ongoing vocalizations in a reflexive manner. After periods of noise perturbation, calls were uttered immediately with durations similar to those being exhibited prior to noise perturbation. In double phees, this effect could be observed for the first syllable, which was perturbed by noise after syllable onset, as well as for the second syllable, which was perturbed by noise from beginning to end.

Vocal changes as direct responses to perturbing noise, such as involuntary increases in call amplitude (Lombard effect) and associated changes, including increases in call frequency and call duration, have been extensively studied in recent decades across vertebrate species such as birds, reptiles, and mammals, including cetaceans, bats, and primates.^{13–17,19,21–24} The observed changes resulted either in an increase in signal-to-noise ratio, such as the Lombard effect, and an increase in call frequency or an increase in signal density, such as in the production of longer calls and/or increased repetition of call syllables, both of which resulted in a potential increase in signal transmission probability.

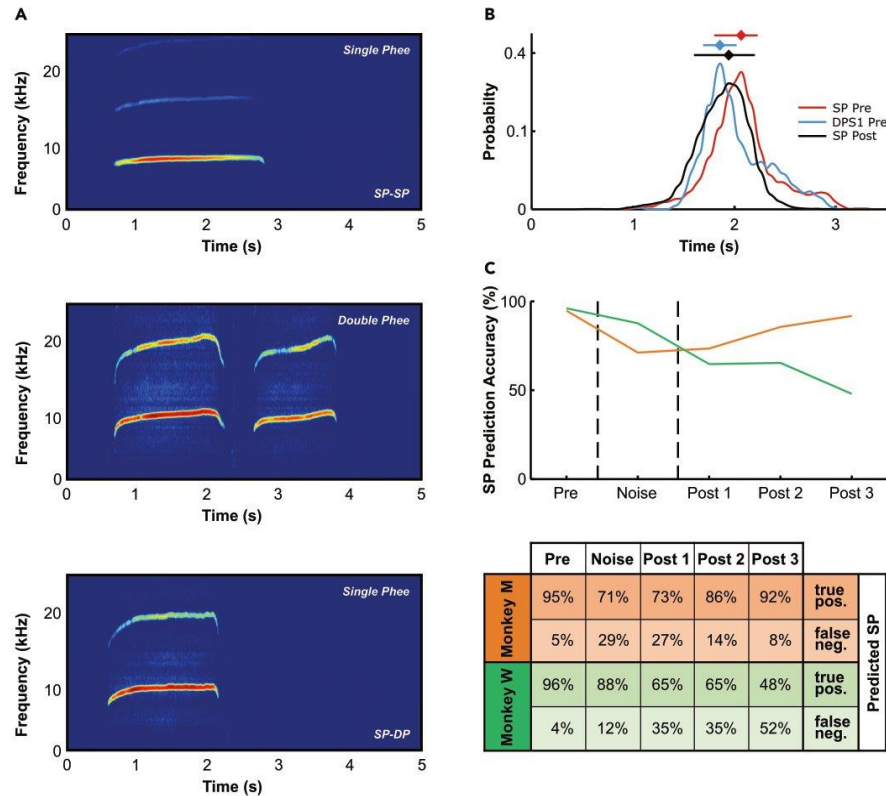


Figure 4. Single phree prediction accuracies for single phrees produced during different experimental phases
 (A) Exemplar spectrograms of a single phree that was classified as a single phree, a double phree, and a single phree that was classified as the first syllable of a double phree.
 (B) Duration distributions of single phrees (SP Pre) and first syllables of double phrees (DPS1 Pre) in the pre-phase as well as single phree durations (SP Post) in the post-phase (monkeys M & W). Horizontal lines indicate full widths at half maximum and diamonds the positions of the maxima of the respective duration distributions.
 (C) Pre = first 10 calls, Noise = next 20 calls, Post = subsequent calls in groups of 20. Different colors indicate different monkeys. Noise-phase is indicated by dashed lines. See also Table S1.

In contrast, noise avoidance strategies, i.e., suppressing vocalizations during periods of elevated ambient noise levels, have received little attention so far. Monkeys have been shown to avoid calling in noisy environments and time their calls to silent epochs¹⁷ and stop call sequences immediately after noise onset.^{18,19} Marmosets are even able to interrupt ongoing vocalizations directly after the onset of perturbing noise.²⁰ However, the underlying mechanisms remain largely unclear.

In our study, marmosets were able to cancel their calls immediately after noise onset. This behavior started during the first perturbed call, indicating a reflexive behavioral reaction to noise perturbation. Here, it was easier for the animals to cancel calls toward the expected end of the call rather than at the beginning of the vocalization. These results suggest underlying neural mechanisms that might inhibit the interruption of vocalizations at the beginning of the pattern.²⁵ This would explain the low occurrence of interrupted calls (0.3%–7.7%) in our earlier study, in which we perturbed vocalizations immediately after call onset.²⁰ These findings further support the hypothesis that call patterns are more stable at the beginning of a vocalization and can be modulated toward the end. Neurophysiological studies are now needed to reveal the underlying neural mechanisms that explain more stable calling behavior in the initial part of a vocalization and a release of the above-mentioned inhibition toward the end of calls.

Interestingly, the monkeys seem to have different strategies to deal with noise when uttering first and second phee syllables. For the second phee syllables, all monkeys showed the same strategy and continued to utter these syllables with shorter duration even after noise perturbation ended. For first phee syllables, however, there was not such a systematic change in vocal behavior. While one monkey remained shorter even after noise perturbing ended, as with the second syllables, the two female monkeys showed a significant increase in call duration even after noise perturbation ended.

Regarding call sequence structure, our study shows, similar to previous studies,^{18,19} that the occurrence of double phee calls significantly decreased during noise perturbation, indicating that marmosets terminated the sequence directly after perturbation onset within the first phee syllable, irrespective of whether they canceled the first syllable. Moreover, we revealed that this behavior persisted after noise perturbation ended, indicating that the monkeys exhibited adaptive changes in their vocal behavior. Interestingly, our machine learning classification model based on call parameters that were not directly affected by the perturbing noise gave the first evidence that some uttered single phees were planned to be doubles phees that were canceled after the first syllable during noise- and post-phases. These findings further show that marmosets have direct control over their vocal output and are capable of modulating ongoing vocalizations in a rapid and direct way.

Limitations of the study

We observed sex-related differences in one aspect of the calling behavior studied in response to noise perturbation. However, to verify whether these are truly sex-specific differences, the sample size used in the present study is too small. Future work with larger number of animals will have to show whether sex-specific differences in noise avoidance strategies exist in marmoset vocal behavior.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availabilities
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Animals
- METHOD DETAILS
 - Experimental setup and procedure
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.106219>.

ACKNOWLEDGMENTS

We thank John Holmes for proofreading and his help in data collection. We acknowledge support by Open Access Publishing Fund of University of Tübingen. 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).

AUTHOR CONTRIBUTIONS

S.R.H. conceived the study; J.L., T.P., and S.R.H. designed the experiments; J.L. conducted the experiments and performed data analyses; all authors interpreted the data and wrote the manuscript. S.R.H. provided the animals and supervised the project.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community. We support inclusive, diverse, and equitable conduct of research.

Received: August 26, 2022

Revised: December 23, 2022

Accepted: February 10, 2023

Published: February 16, 2023

REFERENCES

- Bradbury, J.W., and Vehrencamp, S.L. (1998). *Principles of Animal Communication*.
- Ackermann, H., Hage, S.R., and Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. *Behav. Brain Sci.* 37, 529–546. <https://doi.org/10.1017/S0140525X13003099>.
- Charlton, B.D., Owen, M.A., and Swaisgood, R.R. (2019). Coevolution of vocal signal characteristics and hearing sensitivity in forest mammals. *Nat. Commun.* 10, 2778. <https://doi.org/10.1038/s41467-019-10768-y>.
- Janik, V.M., and Knörnschild, M. (2021). Vocal production learning in mammals revisited. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 376, 20200244. <https://doi.org/10.1098/rstb.2020.0244>.
- Brumm, H., and Slabbekoorn, H. (2005). Acoustic communication in noise. *Adv. Study Behav.* 35, 151–209. [https://doi.org/10.1016/S0065-3454\(05\)35004-2](https://doi.org/10.1016/S0065-3454(05)35004-2).
- Brumm, H., and Zollinger, S.A. (2011). The evolution of the Lombard effect: 100 years of psychoacoustic research. *Beyond Behav.* 148, 1173–1198. <https://doi.org/10.1163/000579511x605759>.
- Eliades, S.J., and 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>.
- Hage, S.R., Jiang, T., Berquist, S.W., Feng, J., and Metzner, W. (2013). Ambient noise induces independent shifts in call frequency and amplitude within the Lombard effect in echolocating bats. *Proc. Natl. Acad. Sci. USA* 110, 4063–4068. <https://doi.org/10.1073/pnas.1211533110>.
- Luo, J., Hage, S.R., and 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>.
- Pomberger, T., Löschner, J., and Hage, S.R. (2020). Compensatory mechanisms affect sensorimotor integration during ongoing vocal motor acts in marmoset monkeys. *Eur. J. Neurosci.* 52, 3531–3544. <https://doi.org/10.1111/ejnn.14721>.
- Brumm, H., Voss, K., Köllmer, I., and Todt, D. (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>.
- Courter, J.R., Perruci, R.J., McGinnis, K.J., and Rainieri, J.K. (2020). Black-capped chickadees (*Parus atricapillus*) alter alarm call duration and peak frequency in response to traffic noise. *PLoS One* 15, e0241035. <https://doi.org/10.1371/journal.pone.0241035>.
- Tressler, J., and Smotherman, M.S. (2009). Context-dependent effects of noise on echolocation pulse characteristics in free-tailed bats. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 195, 923–934. <https://doi.org/10.1007/s00359-009-0468-x>.
- Brown, N.A., Halliday, W.D., Balshine, S., and Juanes, F. (2021). Low-amplitude noise elicits the Lombard effect in plainfin midshipman mating vocalizations in the wild. *Anim. Behav.* 181, 29–39. <https://doi.org/10.1016/j.anbehav.2021.08.025>.
- Manabe, K., Sadr, E.I., and Dooling, R.J. (1998). Control of vocal intensity in budgerigars (*Melopsittacus undulatus*): differential reinforcement of vocal intensity and the Lombard effect. *J. Acoust. Soc. Am.* 103, 1190–1198. <https://doi.org/10.1121/1.421227>.
- Stowe, L.M., and Golob, E.J. (2013). Evidence that the Lombard effect is frequency-specific in humans. *J. Acoust. Soc. Am.* 134, 640–647. <https://doi.org/10.1121/1.4807645>.
- Roy, S., Miller, C.T., Gottsch, D., and 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>.
- Miller, C.T., Flusberg, S., and Hauser, M.D. (2003). Interruptibility of long call production in tamarins: implications for vocal control. *J. Exp. Biol.* 206, 2629–2639. <https://doi.org/10.1242/jeb.00458>.
- Egnor, S.E.R., and 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>.
- Pomberger, T., Risueno-Segovia, C., Löschner, J., and Hage, S.R. (2018). Precise motor control enables rapid flexibility in vocal behavior of marmoset monkeys. *Curr. Biol.* 28, 788–794.e3. <https://doi.org/10.1016/j.cub.2018.01.070>.
- Brumm, H., and Zollinger, S.A. (2017). Vocal plasticity in a reptile. *Proc. Biol. Sci.* 284, 20170451. <https://doi.org/10.1098/rspb.2017.0451>.
- Hardman, S.I., Zollinger, S.A., Koselj, K., Leitner, S., Marshall, R.C., and Brumm, H. (2017). Lombard effect onset times reveal the speed of vocal plasticity in a songbird. *J. Exp. Biol.* 220, 1065–1071. <https://doi.org/10.1242/jeb.148734>.
- Luo, J., Kothari, N.B., and Moss, C.F. (2017). Sensorimotor integration on a rapid time scale. *Proc. Natl. Acad. Sci. USA* 114, 6605–6610. <https://doi.org/10.1073/pnas.1702671114>.
- Scheifele, P.M., Andrew, S., Cooper, R.A., Darre, M., Musiek, F.E., and Max, L. (2005). Indication of a lombard vocal response in the St. Lawrence river beluga. *J. Acoust. Soc. Am.* 117, 1486–1492. <https://doi.org/10.1121/1.1835508>.
- Risueno-Segovia, C., and Hage, S.R. (2020). Theta synchronization of phonatory and articulatory systems in marmoset monkey vocal production. *Curr. Biol.* 30, 4276–4283.e3. <https://doi.org/10.1016/j.cub.2020.08.019>.
- Agamaite, J.A., Chang, C.J., Osmanski, M.S., and Wang, X. (2015). A quantitative acoustic analysis of the vocal repertoire of the common marmoset (*Callithrix jacchus*). *J. Acoust. Soc. Am.* 138, 2906–2928. <https://doi.org/10.1121/1.4934268>.
- Bezerra, B.M., and 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>.
- Risueno-Segovia, C., Koç, O., Champéroux, P., and Hage, S.R. (2022). Cardiovascular mechanisms underlying vocal behavior in freely moving macaque monkeys. *iScience* 25, 103688. <https://doi.org/10.1016/j.isci.2021.103688>.



STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/Strains		
<i>Callithrix jacchus</i>	German Primate Center, Göttingen, Germany, and Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Germany	N/A
Software and algorithms		
MATLAB	MathWorks	R2021b
OpenEx	Tucker-Davis Technologies	N/A
JMP	SAS Institute	version 16
SASLab Pro	Avisoft Bioacoustics	version 5.2.09

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Steffen R. Hage (steffen.hage@uni-tuebingen.de).

Materials availabilities

This study did not generate new unique reagents.

Data and code availability

- Data

All data needed to evaluate the conclusions in the paper are present in the paper. Additional data related to this paper is available from the [lead contact](#) upon request.

- Code

This paper does not report original code.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

Four adult marmoset monkeys (*Callithrix jacchus*, two females aged 6 and 7 years and two males aged 3 and 6 years, respectively, at the beginning of the experiments) were used in this study. The monkeys were usually kept in different sex pairs and were all born in captivity. The animals had *ad libitum* access to water and were fed on a restricted food protocol including a daily basis of commercial pellets, fruits, vegetables, mealworms, and locusts. Additional treats, such as marshmallows or grapes, were used as positive reinforcements to transfer animals from their home cage to the experimental cage. Environmental conditions in the animal husbandry were maintained at a temperature of 26°C, 40–60% relative humidity, and a 12 h:12 h day/night cycle, including periods of twilight in the morning and evening. All animal handling procedures were in accordance with the guidelines for animal experimentation and authorized by the national authority, the Regierungspräsidium Tübingen.

METHOD DETAILS

Experimental setup and procedure

The vocal behavior of four animals in response to noise playback, which was initiated with different latencies relative to the vocal onset, was recorded in a soundproof chamber. The tested animal was transferred into a recording cage (0.6 × 0.6 × 0.8 m) with *ad libitum* access to water and food. In this behavioral setup, marmoset monkeys predominantly produce phee calls (long distance contact calls)



spontaneously.^{10,20,26,27} The vocal behavior of each individual monkey was recorded in daily sessions ranging between 30 and 150 min in duration. Data was collected in sessions at various times during the day between 11 am and 5 pm. Recordings were performed for 21–39 days for each individual animal (mean: 29 ± 7 days). The monkey's behavior was constantly monitored and observed using a webcam (Logitech, Switzerland) placed on top of the cage and recorded with standard software (Synapse Tucker-Davis Technologies, U.S.A.). The vocal behavior was recorded using a microphone (MKH 8020 microphone with MZX 8000 preamplifier, Sennheiser, Germany) and digitized using an A/D interface (RX8, Tucker-Davis Technologies, U.S.A.). A custom-written program (OpenEX and Synapse, Tucker-Davis Technologies, U.S.A.) running on a workstation (WS-X in combination with an RZ6D multi I/O processor, Tucker-Davis Technologies, U.S.A.) recorded the emitted vocalizations and monitored them in real-time. Our vocal detector automatically detected calls in real-time through online calculation of several acoustic parameters, such as call intensity, minimum call intensity duration, call frequency, and several spectral features.

Experimental design

To investigate the effect of noise perturbation during ongoing vocalizations, we partitioned each session into three phases. The animal's first 10 phee calls uttered during a session were kept unperturbed and used as a control for the subsequent performance (pre-phase). For the next 20 phee calls that were longer than the set noise onset latency (between 0.6 s and 2.2 s; see below), we played broadband noise bursts (0.1–60 kHz) with an intensity of 80 dB SPL via a loudspeaker (MF1 Multi-Field Magnetic Speakers, Tucker-Davis Technologies, U.S.A.) positioned on top of the cage (noise-phase). All phee calls following the noise-phase were again unperturbed (post-phase) (Figures 1A, 1B and 1D).

Noise bursts had a duration of 4 s (including 10 ms rise times) to ensure noise perturbation throughout the first syllable after noise onset as well as the entire potential second phee syllable. In total, we used up to nine different noise latencies depending on the monkeys' median call duration. The shortest latency for noise onset after call onset was 0.6 s. The longest latency was defined as the individual monkey's median call duration, which was 2.2 ± 0.234 s (SE) for monkey W, 2.0 ± 0.170 s for monkey M, 1.8 ± 0.219 s for monkey S, and 1.6 ± 0.223 s for monkey H. Additionally, we used noise latencies in steps of 200 ms between the shortest latency and the monkeys' median call duration resulting in nine noise latencies for monkey W (0.6–2.2 s in 200 ms steps), eight latencies for monkey M (0.6–2.0 s), seven latencies for monkey S (0.6–1.8 s), and six latencies for monkey H (0.6–1.6 s). All noise latencies were tested for at least three sessions in a block design per individual monkey. For noise onset determination, we used our call detector and added the corresponding noise latency used in the appropriate session. Figure 1C shows the distribution of noise onsets normalized to the set latencies with a mean shift of 13 ± 0.06 ms (SEM), indicating precise signal detection.

Data analysis

Vocal onsets and offsets were manually flagged, as well as noise onset times, using standard software (Avisoft-SASLab Pro, Germany). Call duration was calculated as the difference between the beginning and end of the vocalization. The spectrograms were calculated using a 512-point FFT Hanning window (256 samples), and 128-sample overlap resulting in a frequency resolution of 191 Hz and temporal resolution of 2.6 ms.

Data normalization

All used call values were normalized by subtracting the mean of the last six calls produced in the pre-phase per respective session. We used the last six calls instead of all calls emitted during pre-phase to get a stable baseline since the monkeys needed some time to settle down (comparison between first 3 and last 3 calls of the 6 calls used for the new baseline, $t(20) = 1.527$, $p = 0.141$).

Single/double phee classifier

We used a medium gaussian support vector machine (SVM) classifier with 25% holdout validation (standard Matlab classification learner app; R2020a MathWorks, U.S.A.) to evaluate the predictive power of certain call features of the first phee syllable for whether they are single phees or the first syllable of a double phee.²⁸ Twenty-four custom call features (Table S1) were defined and computed in order to predict the phee syllable including entropy, bandwidth, maximum and peak frequency at specific time points (20-, 250- and 500 ms after call onset), and the slope between peak and maximum frequencies between 20 ms and 500 ms. We only used call features that could be calculated from the first 500 ms of the calls, since



noise perturbation, which was capable of modifying call structure, started in some sessions at 600 ms after call onset.

QUANTIFICATION AND STATISTICAL ANALYSIS

To evaluate a balanced single to double phee ratio among experimental phases, we performed a two-dimensional likelihood ratio test to examine whether observed frequencies differ between pre, noise, and post-phases. To understand variations in phee call first and second syllable durations, we constructed an LMM with experimental phase (pre, noise, post) as a nominal predictor variable. We further added monkey ID as a random factor to correct for variances in inter-individual differences. To evaluate if there were any differences in first and second syllable call duration between the last call in the pre-phase and first call in the noise-phase, as well as the last call in the noise-phase and first call in the post-phase, we performed paired t-tests. To test if there were any differences in first and second syllable call durations between the sessions' last 10 calls (post) and the first 10 calls (pre) of the following session, we performed a paired t-test. To understand variations in phee call first syllable duration after noise onset, we used z-tests with subsequent Bonferroni correction to compare the measured distributions to either a perfect noise compensated distribution or a fully unaffected call length distribution. To evaluate balanced occurrences of phee durations among latency conditions, we performed a two-dimensional likelihood ratio test to examine whether observed frequencies of shorter, canceled, and longer calls differed between noise latencies. We also performed a two-dimensional likelihood ratio test to examine whether the number of manually labeled single phee calls differed from the count labeled by the classifier over the experimental phases (pre, noise, and post in groups of 20 calls). All statistical analyses were performed using JMP16 (SAS Institute, U.S.A.). In all performed tests, significance was tested at an alpha level of 0.05.