# Audio-Vocal Integration Mechanisms and Volitional Control of Vocal Behavior in Marmoset Monkeys (*Callithrix jacchus*)

Dissertation

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät

und

der Medizinischen Fakultät

der Eberhard-Karls-Universität Tübingen

vorgelegt

von

Thomas Pomberger

aus Bad Ischl, Österreich

April - 2019

Tag der mündlichen Prüfung: 17. September 2019

Dekan der Math.-Nat. Fakultät: Prof. Dr. W. Rosenstiel

Dekan der Medizinischen Fakultät: Prof. Dr. I. B. Autenrieth

1. Berichterstatter: PD Dr. Steffen R. Hage

2. Berichterstatter: Prof. Dr. Uwe Ilg

Prüfungskommission: PD Dr. Steffen R. Hage

Prof. Dr. Uwe Ilg

Prof. Dr. Andreas Nieder

Prof. Dr. Ziad Hafed

**Erklärung / Declaration:**

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

„Audio-Vocal Integration Mechanisms and Volitional Control of Vocal Behavior in Marmoset Monkeys (*Callithrix jacchus*)“

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

*I hereby declare that I have produced the work entitled “*Audio-Vocal Integration Mechanisms and Volitional Control of Vocal Behavior in Marmoset Monkeys (*Callithrix jacchus*)*”, submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.*

Tübingen, den ..... .....

Datum / Date Unterschrift /Signature

**Abstract**

As a prerequisite for human speech vocal communication has been intensively investigated in various vertebrate species in the last decades. It enables two or more individuals to rapidly transmit information. Although, many vertebrate taxa possess the ability to vocalize, only a few are able to learn their vocal patterns by imitation or invention. A well-known example for learned vocalizations is human speech. However, up to now there is no evidence that our closest relatives, non-human primates, are able to produce learned vocal patterns. Hence, a major question in science is when and how human speech appeared in the primate lineage. Hereby, studying the neural mechanisms underlying vocal behavior in primates might help elucidating these questions.

A non-human primate model that has gathered increasing interest in neuroscience in the last decades is the common marmoset monkey (Callithrix jacchus), a highly social and vocal New World primate. In the present thesis I worked with this animal species as a model system to study vocal flexibility, audio-vocal integration mechanisms and cognitive control of vocal behavior combining behavioral, neuroethological, psychophysical and electrophysiological recording techniques. Using acoustic perturbation triggered by the monkeys own vocalizations we found high flexibility in their vocal behavior as well as indications that their vocalizations are built out of small distinct units, overturning decade old thoughts about the structure of primate vocalizations.

Furthermore, we showed that marmosets are capable of performing a complex vocal-motor task in a well-controlled environment. Monkeys were trained to vocalize on command in response to a visual cue as well as executing two distinct vocalizations in response to two different visual cues.

Finally, we developed a new electrophysiological method enabling the extracellular electrophysiological recording from many single units at the same time in deep brainstem structures. We found vocal-motor and auditory neurons in the ventrolateral pontine brainstem and could show that this method is suitable to investigate neural circuits underlying vocal behavior.

The results of this thesis demonstrate that vocal behavior of primates is much more flexible than previously though and, thus, making the marmoset monkey a suitable model to study vocal flexibility, a crucial preadaptation for the evolution of human speech in the primate lineage.

# Table of Contents

## 1. Introduction

Communication between individuals is a crucial aspect of evolutionary success and appears in various forms in nature, ranging from olfactory (Eisenberg and Kleiman, 1972; Russell, 1976) to visual signals (Osorio and Vorobyev, 2008) as well as vocalizations (Ackermann et al., 2014). As a preadaptation for the evolution of human speech, vocal communication has been broadly investigated in several vertebrate species in the last years (Arriaga et al., 2012; Egnor and Seagraves, 2016; Engesser et al., 2016; Seyfarth and Cheney, 2017; Smotherman, 2007). In nature, evolutionary success is usually given to callers which produce vocal signals capable of changing the behavior of a listener in a beneficial way as well as to listeners which possess the ability to connect a call to particular events and, therefore, extracting the most information out of it (Seyfarth and Cheney, 2003). Usually, such vocal patterns can be classified in two groups: Innate and learned vocal patterns (Jürgens, 2009). Prominent examples of innate vocalizations are laughing and crying of humans (Scheiner et al., 2004) or vocal utterances of non-human primates (Hammerschmidt et al., 2001). Although, most vertebrate species possess the ability to produce vocal signals, the majority of vocalizations is largely innate and only a few species, including humans, are capable of learning new vocal signals by vocal imitation or even inventing new ones. Beside human speech, a very prominent example for vocal imitation learning is singing in songbirds where juveniles listen to a tutor and, thus, are capable of learning a complete new vocal pattern (Marler and Tamura, 1964).

The vast majority of vertebrate calls, however, is largely innate and highly dependent on their affective state (Brudzynski, 2013). This is especially true for our closest relatives, non-human primates. Vocal utterances of non-human primates predominantly consist of

genetically predetermined vocal patterns and do not have to be heard initially to be produced. This has been shown in studies in which deaf-born squirrel monkeys (*Saimiri sciureus*) and squirrel monkeys raised in isolation developed vocal patterns similar to normally raised adult individuals (Hammerschmidt et al., 2001; Herzog and Hopf, 1984). Basically, learned and innate vocal patterns are underlying different vocal motor networks. It is assumed that genetically predetermined vocalizations are produced by a primary vocal motor network situated in subcortical and brainstem regions. These innate types of vocalizations are highly affective and, therefore, strongly dependent on the motivational state of an individual (Jürgens, 2002). The production of learned vocal patterns, however, requires the cooperation of the before mentioned primary vocal motor network together with a volitional articulatory vocal motor network encompassing several cortical brain regions (Hage, 2018; Hage and Nieder, 2016). While learned vocal motor patterns are generated by a complex cortical network (Hage and Nieder 2016), innate vocalizations seem to be basically generated by genetically-preprogrammed vocal pattern generating mechanisms situated in the brainstem (Jürgens and Hage, 2007). In the last decades, it has been discussed controversially how this vocal pattern generator is composed of and where it might be localized in the mammalian brainstem. Structures playing an important role in the generation of vocal patterns are the periaqueductal gray, the lateral reticular formation, the nucleus retroambiguus and several other structures in the brainstem (Steffen R Hage, 2010; Steffen R Hage and Jürgens, 2006a). Furthermore, many muscles are involved in the production of vocal patterns including laryngeal, respiratory and supralaryngeal components and these muscles are innervated by different motorneurons of the ventral horn and the brainstem (Jürgens, 2002).

Additionally, a vocal pattern generator should be reciprocally connected to other vital pattern generators using the same muscles such as breathing, sucking or mastication to prevent two or more patterns to happen at the same time (Barlow & Estep 2006, Hage 2010b). Here, a previous study found mastication-correlated neurons in the region of the masticatory pattern generator which were inhibited during vocalization underlining this assumption (Steffen R Hage and Jürgens, 2006a). Furthermore, central pattern generators can be modulated by sensory inputs to produce adapted forms of activity (Grillner, 1991; Lund and Kolta, 2006). An electrophysiological study from 2006 in freely behaving and vocalizing squirrel monkeys found audio-vocal and vocal-motor neurons in the pontine reticular formation of the brainstem which seem to possess all the afore mentioned features required for a vocal pattern generator (Steffen R Hage and Jürgens, 2006b). However, it is yet still unclear what the intrinsic properties of this vocal pattern generator are. Morton and Chiel (1994) suggest three basic forms of neural circuits forming pattern generators: A dedicated pattern generator where each vocalization is generated by a separate, dedicated neural circuit. A reorganizing circuit in which the neurons participating change for each vocal output. Or a distributed pattern generator where connections between neurons change between different vocal utterances. It is assumed that pattern generators consist of one or more of these three circuit architectures. A recent study, for example, found a spatially dynamic network generating inspiratory behavior in mice, also located in the reticular formation of the medulla (Baertsch et al., 2019). It is, therefore, important to simultaneously record from an ensemble of many neurons in vocalizing animals to investigate the intrinsic properties of this innate vocal pattern generating network in the brainstem.

## 2. The Marmoset Monkey as a Model to Study Complex Vocal Behavior

In the last years, the common marmoset (*Callithrix jacchus*) has gathered considerable interest as a model in neuroscience and social communication (Miller et al., 2016). This New World monkey species originates from an ancestor separating from the human lineage about 35 million years ago. In nature, marmoset monkeys live in small groups of up to about 15 individuals with only one breeding female (Stevenson and Poole, 1982) and develop complex social systems where males and non-reproductive individuals show extensive parental care (Goldizen, 1990). Birth is usually given to twins or triplets twice a year and infants reach adulthood between 16-18 months after birth while their lifespan ranges from 12 to 16 years (Schultz-Darken et al., 2016). Their short reproduction cycle compared to other primate species makes them a promising animal model for developing transgenic lines (Sasaki et al., 2009) and, thus, enabling the investigation of neurodegenerative diseases as Parkinson and Alzheimer.

Communication within a marmoset group predominantly happens via visual (Kemp and Kaplan, 2013) and vocal signals (Agamaite et al., 2015; Bezerra and Souto, 2008). Social hierarchies and interactions are also established in captivity (Yamamoto et al., 1996) giving the opportunity to study social communication in primates in a laboratory environment.

The present thesis introduces vocal-motor and audio-vocal integration experiments performed in marmoset monkeys (*Callithrix jacchus*), a highly vocal and social New World monkey species (*Platyrrhini*). We first investigated the vocal behavior of freely moving, spontaneously animals to pin down their vocal pattern generating and audio-vocal

integration mechanisms via neuroethological techniques. Then, we trained marmosets to perform cognitive vocal-motor tasks in a controlled experimental design to decipher their capability to produce calls under volitional control in high numbers. Finally, we performed electrophysiological recordings in deep brainstem structures to reveal vocal pattern generating and audio-vocal integration mechanisms via a self-developed semi-chronic recording device.

## 3. Performed Experiments

### 3.1. Syllable Interruption and Syllable Segmentation in Phee Calls of Marmoset Monkeys

Individual marmoset monkeys which are physically separated from their colony usually tend to produce long distant contact calls, so called phees. This call type is used to get in contact with their group members. Phee calls have a rather simple acoustic structure which is comprised of long duration narrowband fundamental frequency components and usually consist out of one or two, sometimes more syllables (Agamaite et al., 2015). Miller et al. 2009 compared early temporal and spectral features of an ongoing phee call with subsequent features and were able to predict if a marmoset monkey is producing only one or more syllables. These results led them to the hypothesis that the whole vocal-motor plan is already present before vocal onset but can be actively modulated due to external perturbation events. These findings are supported by an earlier study on tamarin monkeys (*Saguinus oedipus*), another New World monkey species, closely related to marmoset monkeys, in which combination long calls, a call type very similar to marmoset phee calls, were perturbed by white noise bursts after call onset (Egnor et al., 2006). The results of this study revealed that tamarin monkeys are able to shorten their calls due to

external noise by reducing the number of syllables within a call. However, the syllable duration itself always remained the same.

We performed a similar experiment in which we placed individual marmoset monkeys in a cage being placed in a sound-proof chamber. Under these conditions marmosets predominantly produced phee calls which were then perturbed by various noise conditions of different amplitude intensities after call onset (Pomberger et al. 2018; see also chapter 3). We could show that marmoset monkeys decrease the amount of double phees under noise perturbation as it has been already shown for tamarin monkeys (Egnor et al., 2006). Furthermore, the vast majority of first syllables did not change their duration when perturbed with noise bursts. A small amount of first syllables, however, was interrupted shortly after noise onset. We found that these interruptions happened most of the time directly after noise perturbation indicating that a fast sensorimotor mechanism is present in these animals. In a next step, we wanted to know, whether marmoset monkeys are able to interrupt their first phee syllables at any time point or if call interruption is restricted to certain distinct time points within a syllable. One finding supporting this hypothesis were the duration distributions of interrupted phees that showed several peaks being multiples of each other. Another observation supporting this hypothesis were so called segmented phees that where produced by a few monkeys. These phee calls had similar syllable durations as normal phees but were segmented by small breaks. Interestingly, we found that unit intervals within segmented phees showed a 7 Hz rhythm, similar to human speech rhythm which is between 3-8 Hz (Macneilage, 1998). Finally, we compared segmented phee unit durations with durations of other marmoset calls, e.g. twitter, tsik, ekk and phee syllables. Except for phee syllables which seemed to be quite

variable in their durations all other syllables seemed to have similar durations and variabilities as segmented phee units. These results indicate that marmoset monkeys are able to rapidly interrupting ongoing calls after acoustic perturbation but call interruption can only occur at specific time points within a syllable. Additionally, our results show that marmoset phee calls are built out of short distinct units similar in duration to short syllables of other marmoset call types.

### 3.2. Audio-Vocal Integration During Vocal Production

Even though that marmoset monkeys are capable to interrupt calls as a response to perturbing noise, we observed that the duration of the majority of perturbed phee calls was not affected by noise perturbation (Pomberger et al., 2018). We, therefore, investigated if and how other call features such as frequency and amplitude were affected by noise perturbation starting after call onset, i.e., if and how marmosets are able to elicit distinct audio-vocal integration mechanisms in ongoing vocalizations to increase call detectability.

One of the most efficient mechanisms to increase signal-to-noise ratio in call production is the so-called Lombard effect, i.e., the involuntary increase in call amplitude in response to masking ambient noise, which has been described for the first time more than 100 years ago (Lombard, 1911). Basically, the Lombard effect is an involuntary rise in call amplitude due to ambient masking noise. It is often accompanied by a shift in call frequency (Hage et al., 2013b) as well as a change in call duration (Luo et al., 2015) and has been shown to be present in many vertebrate species such as in fish (Holt and Johnston, 2014), frogs (Halfwerk et al., 2016), birds (Pytte et al., 2003; Brumm et al.,

2009), cetaceans (Dunlop et al., 2014) and primates (Brumm, 2004; Egnor and Hauser, 2006) including humans (Lombard, 1911). Furthermore, the Lombard effect could be already observed in very young juveniles of different species (Dorado-Correa et al., 2018; Leonard and Horn, 2005; Luo et al., 2017b) and noise playback experiments in bats revealed a rapid increase in call amplitude of about 30 ms after noise onset (Luo et al., 2017a). Comparing the various taxa which exhibit this effect with respect to their phylogeny, it is assumed that the Lombard effect may have emerged ~450 million years ago (Luo et al., 2018). Although, the Lombard effect is extremely robust and stable in its appearance several studies found that songbirds and humans are able to volitionally inhibit it (Kobayasi and Okanoya, 2003; Pick et al., 1989; Therrien et al., 2012). Together, these findings suggest that subcortical regions are sufficient to elicit the Lombard effect but cortical processes may be able to play a modulatory role (Luo et al., 2018).

Another strategy to increase call detectability in a noisy environment is to vocalize in timeslots where noise perturbation is low in amplitude or even absent (Roy et al., 2011). This approach avoids the increased physiological cost of call emission at high intensities that still might be insufficiently increasing signal-to-noise ratio (Roulin, 2001). As mentioned above, the duration of the majority of phee calls was unaffected. We, therefore investigated in this project, if and how fundamental frequencies and amplitudes of these calls were affected by noise perturbation starting after call onset. We found that monkeys significantly increased fundamental frequencies of first syllables with short latency of approximately 30 ms after noise onset. This fast response is in accordance with similar response values for amplitude shifts in bats (Luo et al., 2017a). Amplitudes of first syllables were affected as well, surprisingly, they were slightly decreased and not

increased as expected as a direct influence of the Lombard effect. This effect was even stronger for the second phee syllables which showed a robust decrease in amplitude in a step-wise function under certain noise conditions.

Our results indicate that marmoset monkeys are capable of inhibiting or even counteracting the Lombard effect. These findings suggest that cognitive processes might have a modulatory influence on amplitude modulation as it has been already demonstrated in experiments performed in birds and humans. These studies were able to show that vocal learners are able to volitionally decrease or even block the Lombard effect while simultaneously performing demanding cognitive tasks (Kobayasi and Okanoya, 2003; Pick et al., 1989; Therrien et al., 2012; Vinney et al., 2016).

Based on our neuroethological studies on audio-vocal integration mechanisms, we propose a hypothetical neural model that is able to explain the different strategies to call in a noisy environment (see also Fig. 4 of chapter 3). Animals can either decide not to vocalize at all when noise perturbation is already present before vocal onset or interrupt their vocalizations when noise perturbation occurs within a call and continue to vocalize in time periods where less or no noise is present (Egnor et al., 2006; Pomberger et al., 2018; Roy et al., 2011). This has the advantage of reducing the physiological costs of producing a vocalization that might not be detected by a conspecific. These strategies have to involve cognitive control mechanisms and might, therefore, involve cortical structures such as the prefrontal cortex or the anterior cingulate cortex. Marmosets also exhibit upward shifts in fundamental frequency, which has been previously reported to be present in birds (Wood et al., 2011) and bats (Hage et al., 2013b) as well.

Such basic audio-vocal mechanisms might be directly communicated on lower brainstem level via the cochlear nucleus and the motoneuron pools (Jürgens, 2009). Nevertheless, a recent study in marmoset monkeys revealed that stimulation of auditory cortex causes rapid changes in the fundamental frequency of the monkeys own vocalization (Eliades and Tsunada, 2018). This indicates that such rapid audio-vocal integration mechanisms might also be communicated on cortical level from auditory cortex to premotor and/or motor cortex (Rauschecker and Scott, 2009, Hage 2018). Finally, it seems that marmoset monkeys are capable of volitionally modulating call features such as call amplitude using higher order brain structures (Gavrilov et al., 2017; Hage and Nieder, 2013).

### 3.3. Cognitive Vocal Control of Vocal Behavior in Marmoset Monkeys

Until recently, several brain imaging and electrophysiological methods have been successfully implemented in the marmoset model (Marx, 2016; Miller et al., 2016). Additionally, recent studies have shown that marmoset monkeys can be trained to succeed in basic auditory discrimination tasks (Remington et al., 2012) as well as saccadic eye movements (Mitchell et al., 2014) under constrained and controlled conditions. On the other hand, it is not yet clear if marmoset monkeys are capable of learning complex cognitive tasks as it has been already shown in rhesus macaques (Hage et al., 2013a; Hage and Nieder, 2013). In neuroscience, however, it is important to have animal models showing complex behavior in a well-controlled experimental design to understand certain brain-behavior relationships (Krakauer et al., 2017). Consequently, showing that marmoset monkeys can be trained to perform complex behavioral tasks in combination with all the physiological methods already implemented for this animal

species, would make them a suitable model to study the neural basis of cognitive control mechanisms affecting human behavior in health and disease.

In our study we could show that marmoset monkeys are able to volitionally control a complex vocal behavior, namely their vocal output, on command in response to an arbitrary visual cue in a well-controlled experimental design. Furthermore, we trained one monkey in a vocal discrimination task where it had to produce two different call types in response to two different colored cues.

Together with the recent advent of transgenic marmoset lines (Sasaki et al., 2009), our results demonstrate that marmoset monkeys are a suitable model to study complex motor behavior in a controlled experimental design as well as for the research of neurodegenerative diseases.

### 3.4. Developing an Electrophysiological Setup to Perform Semi-Chronic Laminar Recordings in Vocalizing Marmoset Monkeys

In general, starting in a research group as the first doctoral student is a different situation than joining an already established laboratory. Therefore, I had the chance to contribute in assembling the setup from scratch before starting with the experiments. This work included hardware and software installation for each setup, programming of the TDT system and associated Matlab codes, as well as building up the experimental setup in the sound-proof chambers. In addition, we had to fully develop a behavioral protocol and a new neurophysiological system to record from single neurons in vocalizing monkeys in a controlled experimental design. Such a system was not available off-the-shelf when we started our neurophysiological project.

In our project we wanted to record from deep brainstem structures to decipher the intrinsic properties of the putative vocal pattern generating network. The system that we planned to use for investigating neural vocal pattern generating mechanisms in deep brain structures of awake animals required the possibility to record stable signals over a long period of time as well as the capability of simultaneously recording from a large ensemble of neurons. Several multi-electrode systems have been already developed to record from cortical structures in awake marmoset monkeys (Eliades and Wang, 2008; Roy and Wang, 2012). However, only a few electrophysiological methods exist which allow the recording from brainstem structures and they measure neural activity only at a maximum of two recording sites (Steffen R Hage and Jürgens, 2006b). We, therefore, developed in cooperation with the company Neuronexus (Ann Arbor, MI, U.S.A.) a recording system that allows to get stable recordings from deep brainstem structures by a vertically arranged 32-channel laminar probe (Pomberger and Hage, 2019). Basically, the system consists of a titanium base chamber that is chronically implanted on the skull of the monkey and a synthetic upper chamber which can be semi-chronically implanted on top of the base chamber. The laminar probe can then be flexibly positioned along a vertical line via a microdrive within the chamber enabling the switching of recording positions from session to session. We could show that this system delivers stable recordings and clear signals during a whole recording session of a monkey sitting in a monkey chair. Furthermore, we found cells which show vocal-motor activity as well as audio-playback activity. Together, these results demonstrate that this electrophysiological approach is a suitable method for analyzing neural circuits within deep brain structures involved in vocal-motor control and audio-vocal integration mechanisms.

## 4. Conclusion and Outlook

In the experiments performed for this thesis we combined neuroethological, psychophysical and electrophysiological methods to investigate vocal-motor and audio-vocal integration mechanisms in a highly social and vocal animal model: the marmoset monkey. Our neuroethological experiments revealed high vocal flexibility in these animals. Marmosets are capable of cancelling calls shortly after acoustic perturbation. However, these interruptions only happen at specific time points indicating that their calls are built out of short distinct vocal units of equal duration.

Beside phee interruption, we also investigated if and how fundamental frequency and amplitudes change due to external noise perturbation of an ongoing call. Although, shifts in amplitude caused by the Lombard effect are usually accompanied by shifts in fundamental frequency (Luo et al., 2018), our results suggest that these alterations of acoustic features are independent from each other as it has been already shown in bats (Hage et al., 2013b).

We also trained marmoset monkeys to volitionally control their vocalizations on command while sitting in a monkey chair as it has been already done in a similar approach for rhesus macaques (Hage et al., 2013a). This approach shows that also marmoset monkeys are able to perform complex behavioral tasks in a highly controlled environment and make them a suitable model to study cognitive behavior in human diseases. Furthermore, we also developed and recorded via a new electrophysiological method semi-chronically from up to 32 recording sites at the same time from neurons in deep brainstem structures,

getting first insights into intrinsic properties of vocal pattern generating mechanisms in marmoset monkeys.

Our findings raise several important questions that will have to be tackled in future experiments. For example, it is not yet clear why monkeys interrupted only a small ratio of phee calls since the duration of most perturbed calls was unaffected. The small fraction of interrupted and segmented phees in most animals indicates that marmosets may have stark neuronal and/or anatomical constraints in exhibiting such behavior. This might be the case, since we perturbed most of the calls shortly after call onset. We hypothesize that it might be hard for the animal to interrupt calls at such an early time point. Thus, future studies will have to investigate, whether animals are able to interrupt calls with higher rations when noise perturbation starts more towards the end of a vocalization.

It will be also interesting to further investigate the rhythmicity of vocal units in segmented phee calls. Other questions might be if and how phee units and inter-unit intervals correlate to each other and if segmented phees belong to the phee call type of marmoset monkeys or if these utterances are a completely new call type. Finally, electrophysiological recordings in freely behaving marmoset monkeys will be necessary to investigate the neural mechanisms underlying the production of the discovered phee interruption and segmentation and where the neural mechanisms underlying this behavior are located in the marmoset brain.

In several species frequency and amplitude shifts occur on a very fast timescale suggesting audio-integration processes at low brainstem level (Hage et al., 2013b; Luo et al., 2018, 2017a). However, a recent study in marmoset monkeys revealed that

stimulation of auditory cortex in vocalizing animals causes a frequency shift in the monkey's own vocalization with a latency of about 40 ms (Eliades and Tsunada, 2018). It is, therefore, important to study where and how in the brain auditory-integration mechanisms interact with vocal motor production by electrophysiological methods.

Behavioral and neuroethological approaches are very important and useful in investigating a monkey's natural behavior. However, especially for neurophysiological approaches experiments have to be performed in a highly controlled way. For example, it is hard to control eye movements of a monkey while it is moving in the three dimensional space. Considering the outcome of the work done in rhesus monkeys and our results from an evolutionary perspective, it seems that the ability to volitionally control vocal output might have already existed in the last common ancestor of New World and Old World monkeys (*Catarrhini*) ~35 million years ago. Recent studies in rhesus macaques found neural activity in cortical brain structures underlying the volitional output of vocalizations similar to those which are crucial for speech production in humans (Flinker et al., 2015; Gavrilov et al., 2017). Although, we are just at the beginning of understanding the similarities and differences of neural correlates underlying the vocal production of non-human primates and speech production of humans, marmoset monkeys might be a suitable model to study these evolutionary questions. Further psychophysical experiments in combination with electrophysiological recordings are necessary to study neural activity in cortical as well as subcortical structures in vocalizing monkeys under highly controlled conditions. Additionally, considering the advent of transgenic marmoset lines, it will be probably soon possible to study vocal behavior effected by Parkinson or Alzheimer disease in a non-human primate animal model.

To study all these questions in combination with electrophysiological approaches, especially when recording from deep brain structures, it is inevitable to use a method that allows to perform stable recordings from many neurons simultaneously. Our newly established tool (Pomberger and Hage 2019) gives the unique opportunity to study, for example, vocal pattern generation and audio-vocal integration mechanisms at brainstem level, respectively, in awake and behaving marmoset monkeys. A next step might be to modify the present system to be used in a telemetric approach in freely moving marmoset monkeys and, therefore, to combine neuroethological approaches with electrophysiological recordings. Using this approach for future experiments will help revealing neural mechanisms underlying vocal behavior of marmoset monkeys and non-human primates in general and will have the potential to contribute in understanding the evolution of human speech and language.

## 5. References


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.

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, 2906–2928.

Arriaga, G., Zhou, E.P., Jarvis, E.D., 2012. Of Mice, Birds, and Men: The Mouse Ultrasonic Song System Has Some Features Similar to Humans and Song-Learning Birds. PLoS One 7, e46610.

Baertsch, N.A., Severs, L.J., Anderson, T.M., Ramirez, J., 2019. A spatially dynamic network underlies the generation of inspiratory behaviors. Proc. Natl. Acad. Sci. 116, 7493–7502.

Barlow, S.M., Estep, M., 2006. Central pattern generation and the motor infrastructure for suck, respiration, and speech. J. Commun. Disord. 39, 366–380.

Bezerra, B.M., Souto, A., 2008. Structure and usage of the vocal repertoire of Callithrix jacchus. Int. J. Primatol. 29, 671–701.

Brudzynski, S.M., 2013. Ethotransmission: communication of emotional states through ultrasonic vocalization in rats. Curr. Opin. Neurobiol. 23, 310–317.

Brumm, H., 2004. Acoustic communication in noise: regulation of call characteristics in a New World monkey. J. Exp. Biol. 207, 443–448.

Brumm, H., Schmidt, R., Schrader, L., 2009. Noise-dependent vocal plasticity in domestic fowl. Anim. Behav. 78, 741–746.

Dorado-Correa, A.M., Zollinger, S.A., Brumm, H., 2018. Vocal plasticity in mallards: multiple signal changes in noise and the evolution of the Lombard effect in birds. J. Avian Biol. 49, 1–9.

Dunlop, R.A., Cato, D.H., Noad, M.J., 2014. Evidence of a Lombard response in migrating humpback whales (Megaptera novaeangliae). J. Acoust. Soc. Am. 136, 430–437.

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.

Egnor, S.E.R., Seagraves, K.M., 2016. The contribution of ultrasonic vocalizations to mouse courtship. Curr. Opin. Neurobiol. 38, 1–5.

Egnor, S.E.R., Hauser, M.D., 2006. Noise-induced vocal modulation in cotton-top tamarins (Saguinus oedipus). Am. J. Primatol. 68, 1183–1190.

Eisenberg, J.F., Kleiman, D.G., 1972. Olfactory Communication in Mammals. Annu. Rev. Ecol. Syst. 3, 1–32.

Eliades, S.J., Tsunada, J., 2018. Auditory cortical activity drives feedback-dependent vocal control in marmosets. Nat. Commun. 9, 2540.

Eliades, S.J., Wang, X., 2008. Chronic multi-electrode neural recording in free-roaming monkeys. J. Neurosci. Methods 172, 201–214.

Engesser, S., Ridley, A.R., Townsend, S.W., 2016. Meaningful call combinations and compositional processing in the southern pied babbler. Proc. Natl. Acad. Sci. 113, 5976–5981.

Flinker, A., Korzeniewska, A., Shestyuk, A.Y., Franaszczuk, P.J., Dronkers, N.F., Knight, R.T., Crone, N.E., 2015. Redefining the role of Broca's area in speech. Proc. Natl. Acad. Sci. 112, 2871–2875.

Gavrilov, N., Hage, S.R., Nieder, A., 2017. Functional Specialization of the Primate Frontal Lobe during Cognitive Control of Vocalizations. Cell Rep. 21, 2393–2406.

Goldizen, A.W., 1990. A Comparative Perspective on the Evolution of Tamarin and Marmoset Social Systems. Int. J. Primatol. 11, 63–82.

Grillner, S., 1991. Recombination of motor pattern generators. Curr. Biol. 1, 231–233.

Hage, S.R., 2018. Dual neural network model of speech and language evolution : new insights on flexibility of vocal production systems and involvement of frontal cortex. Curr. Opin. Behav. Sci. 21, 80–87.

Hage, S.R., 2010. Localization of the central pattern generator for vocalization, Handbook of Mammalian Vocalization. Elsevier, Berlin, pp. 329-338.

Hage, S.R., 2010. Neuronal networks involved in the generation of vocalization, Handbook of Mammalian Vocalization. Elsevier, Berlin, pp. 339-350.

Hage, S.R., Gavrilov, N., Nieder, A., 2013a. Cognitive control of distinct vocalizations in rhesus monkeys. J. Cogn. Neurosci. 25, 1692–701.

Hage, S.R., Jiang, T., Berquist, S.W., Feng, J., Metzner, W., 2013b. Ambient noise induces independent shifts in call frequency and amplitude within the Lombard effect in echolocating bats. Proc. Natl. Acad. Sci. 110, 4063–4068.

Hage, S.R., Jürgens, U., 2006a. Localization of a vocal pattern generator in the pontine brainstem of the squirrel monkey. Eur. J. Neurosci. 23, 840–844.

Hage, S.R., Jürgens, U., 2006b. On the Role of the Pontine Brainstem in Vocal Pattern Generation: A Telemetric Single-Unit Recording Study in the Squirrel Monkey. J. Neurosci. 26, 7105–7115.

Hage, S.R., Nieder, A., 2016. Dual Neural Network Model for the Evolution of Speech and Language. Trends Neurosci. 39, 813–829.

Hage, S.R., Nieder, A., 2013. Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. Nat. Commun. 4, 2409.

Halfwerk, W., Lea, A.M., Guerra, M.A., Page, R.A., Ryan, M.J., 2016. Vocal responses to noise reveal the presence of the Lombard effect in a frog. Behav. Ecol. 27, 669–676.

Hammerschmidt, K., Jürgens, U., Freudenstein, T., 2001. Vocal Development in Squirrel Monkeys. Behaviour 138, 1179–1204.

Herzog, M., Hopf, S., 1984. Behavioral responses to species-specific warning calls in infant squirrel monkeys reared in social isolation. Am. J. Primatol. 106, 99–106.

Holt, D.E., Johnston, C.E., 2014. Evidence of the Lombard effect in fishes. Behav. Ecol. 25, 819–826.

Jürgens, U., 2009. The Neural Control of Vocalization in Mammals: A Review. J. Voice 23, 1–10.

Jürgens, U., 2002. Neural pathways underlying vocal control. Neurosci. Biobehav. Rev. 26, 235–258.

Jürgens, U., Hage, S.R., 2007. On the role of the reticular formation in vocal pattern generation. Behav. Brain Res. 182, 308–314.

Kemp, C., Kaplan, G., 2013. Facial expressions in common marmosets (Callithrix jacchus) and their use by conspecifics. Anim. Cogn. 16, 773–788.

Kobayasi, K.I., Okanoya, K., 2003. Context-dependent song amplitude control in Bengalese finches. Neuroreport 14, 521–524.

Krakauer, J.W., Ghazanfar, A.A., Gomez-Marin, A., MacIver, M.A., Poeppel, D., 2017. Neuroscience Needs Behavior: Correcting a Reductionist Bias. Neuron 93, 480–490.

Leonard, M.L., Horn, A.G., 2005. Ambient noise and the design of begging signals. Proc. R. Soc. B Biol. Sci. 272, 651–656.

Lombard, E., 1911. Le signe de l'elevation de la voix. Ann. Mal. L'Oreille du Larynx XXXVII, 101–119.

Lund, J.P., Kolta, A., 2006. Generation of the central masticatory pattern and its modification by sensory feedback. Dysphagia 21, 167–174.

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.

Luo, J., Hage, S.R., Moss, C.F., 2018. The Lombard Effect: From Acoustics to Neural Mechanisms. Trends Neurosci. 41, 938–949.

Luo, J., Kothari, N.B., Moss, C.F., 2017a. Sensorimotor integration on a rapid time scale. Proc. Natl. Acad. Sci. 114, 6605–6610.

Luo, J., Lingner, A., Firzlaff, U., Wiegrebe, L., 2017b. The Lombard effect emerges early in young bats: implications for the development of audio-vocal integration. J. Exp. Biol. 220, 1032–1037.

Macneilage, P.F., 1998. The frame/content theory of evolution of speech production. Behav. Brain Sci. 21, 499–546.

Marler, P., Tamura, M., 1964. Culturally Transmitted Patterns of Vocal Behavior in Sparrows. Science 146, 1483–1486.

Marx, V., 2016. Neurobiology: learning from marmosets. Nat. Methods 13, 911–916.

Miller, C.T., Eliades, S.J., Wang, X., 2009. Motor planning for vocal production in common marmosets. Anim. Behav. 78, 1195–1203.

Miller, C.T., Freiwald, W.A., Leopold, D.A., Mitchell, J.F., Silva, A.C., Wang, X., 2016. Marmosets: A Neuroscientific Model of Human Social Behavior. Neuron 90, 219–233

Mitchell, J.F., Reynolds, J.H., Miller, C.T., 2014. Active vision in marmosets: a model system for visual neuroscience. J. Neurosci. 34, 1183–1194.

Morton, D.W., Chiel, H.J., 1994. Neural architectures for adaptive behavior. Trends Neurosci. 17, 413–420.

Osorio, D., Vorobyev, M., 2008. A review of the evolution of animal colour vision and visual communication signals. Vision Res. 48, 2042–2051.

Pick, H.L., Siegel, G.M., Fox, P.W., Garber, S.R., Kearney, J.K., 1989. Inhibiting the Lombard effect. J. Acoust. Soc. Am. 85, 894–900.

Pomberger, T., Hage, S.R., 2019. Semi-chronic laminar recordings in the brainstem of behaving marmoset monkeys. J. Neurosci. Methods 311, 186–192.

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.

Pytte, C.L., Rusch, K.M., Ficken, M.S., 2003. Regulation of vocal amplitude by the blue-throated hummingbird, Lampornis clemenciae. Anim. Behav. 66, 703–710.

Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12, 718–724.

Remington, E.D., Osmanski, M.S., Wang, X., 2012. An Operant Conditioning Method for Studying Auditory Behaviors in Marmoset Monkeys. PLoS One 7, e47895.

Roulin, A., 2001. On the cost of begging vocalization: implications of vigilance. Behav. Ecol. 12, 506–515.

Roy, S., Miller, C.T., Gottsch, D., Wang, X., 2011. Vocal control by the common marmoset in the presence of interfering noise. J. Exp. Biol. 214, 3619–3629.

Roy, S., Wang, X., 2012. Wireless multi-channel single unit recording in freely moving and vocalizing primates. J. Neurosci. Methods 203, 28–40.

Russell, M.J., 1976. Human olfactory communication. Nature 260, 520–522.

Sasaki, E., Suemizu, H., Shimada, A., Hanazawa, K., Oiwa, R., Kamioka, M., Tomioka, I., Sotomaru, Y., Hirakawa, R., Eto, T., Shiozawa, S., Maeda, T., Ito, M., Ito, R., Kito, C., Yagihashi, C., Kawai, K., Miyoshi, H., Tanioka, Y., Tamaoki, N., Habu, S., Okano, H., Nomura, T., 2009. Generation of transgenic non-human primates with germline transmission. Nature 459, 523–527.

Scheiner, E., Hammerschmidt, K., Jürgens, U., Zwirner, P., 2004. The influence of hearing impairment on preverbal emotional vocalizations of infants. Folia Phoniatr. Logop. 56, 27–40.

Schultz-Darken, N., Braun, K.M., Emborg, M.E., 2016. Neurobehavioral Development of Common Marmoset Monkeys. Dev. Psychobiol. 58, 141–158.

Seyfarth, R.M., Cheney, D.L., 2017. The origin of meaning in animal signals. Anim. Behav. 124, 339–346.

Seyfarth, R.M., Cheney, D.L., 2003. Signalers and Receivers in Animal Communication. Annu. Rev. Psychol. 54, 145–173.

Smotherman, M.S., 2007. Sensory feedback control of mammalian vocalizations. Behav. Brain Res. 182, 315–326.

Stevenson, M.F., Poole, T.B., 1982. Playful Interactions in Family Groups of the Common Marmoset (Callithris Jacchus Jacchus). Anim. Behav. 30, 886–900.

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? J. Voice 30, 638.e21-638.e31.

Nemeth, E., Pieretti, N., Zollinger, S.A., Geberzahn, N., Partecke, J., Miranda, A.C., Brumm, H., 2011. Bird song and anthropogenic noise: vocal constraints may explain why birds sing higher-frequency songs in cities. Anim. Behav. 23, 201–209.

Yamamoto, M.E., Box, H.O., Albuquerque, F.S., Arruda, M. de F., 1996. Carrying Behaviour in Captive and Wild Marmosets (Callithrix jacchus): A Comparison Between Two Colonies and a Field Site. Primates 37, 297–304.

## 6. List of Papers/Manuscripts Appended

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

Pomberger, T. Löschner, J., Hage, S.R. (under review). Compensatory mechanisms affect sensorimotor integration during ongoing vocal-motor acts in marmoset monkeys

Pomberger, T., Risueno-Segovia, C., Gültekin, Y.B., Dohmen, D., Hage, S.R., (under review). Cognitive control of complex motor behavior in the marmoset monkey

Pomberger, T., Hage, S.R., 2019. Semi-Chronic Laminar Recordings in the Brainstem of Behaving Marmoset Monkeys. J. Neurosci. Methods 311, 186–192

## 7. Statement of Contributions

**Precise Motor Control Enables Rapid Flexibility in Vocal Behavior of Marmoset Monkeys** by Thomas Pomberger, Cristina Risueno-Segovia, Julia Löschner and Steffen R. Hage

In this study we put five individual marmoset monkeys into a cage in a soundproofed chamber and recorded their vocalizations and perturbed them by white noise bursts. We analyzed the durations of first syllables of phee calls of each monkey. Additionally, we recorded the segmented phees of monkeys performing a vocal reinforcement experiment in a monkey chair. T.P. and S.R.H. designed the experiments. T.P. and J.L. conducted the noise playback experiments. T.P. and C.R.-S. conducted the vocal reinforcement experiments. T.P., J.L., and S.R.H. analyzed the noise playback experiment data. C.R.-S., J.L., and S.R.H. analyzed the vocal reinforcement experiment data. All authors interpreted the data and wrote the manuscript. My significant contributions to this project are that I wrote all algorithms for running the setup, programmed the GUI for data analysis in Matlab and analyzed the data of the noise playback experiments. Furthermore, I was strongly involved in writing the manuscript.

**Cognitive Control of Complex Motor Behavior in the Marmoset Monkey** by Thomas Pomberger, Cristina Risueno-Segovia, Yasemin B. Gültekin, Deniz Dohmen and Steffen R. Hage

In this study we trained four individual marmoset monkeys to sit into a monkey chair in a soundproof chamber and to vocalize on command in response to a visual cue. Furthermore, one monkey (monkey H) was trained to perform a vocal discrimination task

where the animal had to produce two different types of vocalizations in response to two different colored visual cues. S.R.H. conceived the study and designed the experiments. T.P., C.R.-S., Y.G., and D.D. conducted the visual detection experiments. T.P. conducted the visual discrimination experiment; S.R.H., T.P., C.R.-S., and Y.G. performed data analyses. S.R.H. created the visualizations. All authors interpreted the data and wrote the manuscript. My significant contributions to this project are that I programmed the whole setup and trained monkey H performing the operant conditioning task and the vocal discrimination task. Furthermore, I was involved in analyzing the data and writing the manuscript.

**Compensatory Mechanisms Affect Sensorimotor Integration During Ongoing Vocal-Motor Acts in Marmoset Monkeys** by Thomas Pomberger, Julia Löschner and Steffen R. Hage

In this study we put four individual marmoset monkeys into a cage in a soundproofed chamber and recorded their vocalizations and perturbed them by white noise bursts. We analyzed the fundamental frequency and amplitude shifts of vocalizations. S.R.H. conceived the study. T.P. and S.R.H. designed the experiments. TP and JL conducted the experiments and performed data analyses. All authors interpreted the data and wrote the manuscript. My significant contributions to this project are that I programmed the whole setup, programmed the GUI for data analysis in Matlab and analyzed the data of the noise playback experiments. Furthermore, I was strongly involved in writing the manuscript.

Report

Current Biology

# Precise Motor Control Enables Rapid Flexibility in Vocal Behavior of Marmoset Monkeys

## Highlights

- Marmosets interrupt ongoing vocalizations rapidly after acoustic perturbation
- Interruptions happen at periodic time points, indicating precisely tuned call units
- Segmented phee calls have rapid, precise elements matching these time points
- Calls are built from many serially uttered units rather than one discrete pattern

## Authors

Thomas Pomberger, Cristina Risueno-Segovia, Julia Löschner, Steffen R. Hage

## Correspondence

steffen.hage@uni-tuebingen.de

## In Brief

Pomberger et al. show that marmoset monkey calls do not consist of one discrete call pattern but are built out of many sequentially uttered units, like human speech. These findings explain the monkeys' capability to interrupt their calls only at periodic time points within calls and are supported by the occurrence of periodically segmented calls.

Pomberger et al., 2018, Current Biology *28*, 788–794
March 5, 2018 
https://doi.org/10.1016/j.cub.2018.01.070

CellPress

CellPress

Current Biology
Report

# Precise Motor Control Enables Rapid Flexibility in Vocal Behavior of Marmoset Monkeys

Thomas Pomberger,[1,2,3] Cristina Risueno-Segovia,[1,2,3] Julia Löschner,[1] and Steffen R. Hage[1,4,*]
[1]Neurobiology of Vocal Communication, Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Str. 25, 72076 Tübingen, Germany
[2]Graduate School of Neural & Behavioural Sciences - International Max Planck Research School, University of Tübingen, Österberg-Str. 3, 72074 Tübingen, Germany
[3]These authors contributed equally
[4]Lead Contact
*Correspondence: steffen.hage@uni-tuebingen.de
https://doi.org/10.1016/j.cub.2018.01.070

## SUMMARY

**Investigating the evolution of human speech is difficult and controversial because human speech surpasses nonhuman primate vocal communication in scope and flexibility [1–3]. Monkey vocalizations have been assumed to be largely innate, highly affective, and stereotyped for over 50 years [4, 5]. Recently, this perception has dramatically changed. Current studies have revealed distinct learning mechanisms during vocal development [6–8] and vocal flexibility, allowing monkeys to cognitively control when [9, 10], where [11], and what to vocalize [10, 12, 13]. However, specific call features (e.g., duration, frequency) remain surprisingly robust and stable in adult monkeys, resulting in rather stereotyped and discrete call patterns [14]. Additionally, monkeys seem to be unable to modulate their acoustic call structure under reinforced conditions beyond natural constraints [15, 16]. Behavioral experiments have shown that monkeys can stop sequences of calls immediately after acoustic perturbation but cannot interrupt ongoing vocalizations, suggesting that calls consist of single impartible pulses [17, 18]. Using acoustic perturbation triggered by the vocal behavior itself and quantitative measures of resulting vocal adjustments, we show that marmoset monkeys are capable of producing calls with durations beyond the natural boundaries of their repertoire by interrupting ongoing vocalizations rapidly after perturbation onset. Our results indicate that marmosets are capable of interrupting vocalizations only at periodic time points throughout calls, further supported by the occurrence of periodically segmented phees. These ideas overturn decades-old concepts on primate vocal pattern generation, indicating that vocalizations do not consist of one discrete call pattern but are built of many sequentially uttered units, like human speech.**

## RESULTS AND DISCUSSION

We measured vocal behavior in marmoset monkeys (*Callithrix jacchus*, n = 5), a highly vocal New World monkey species, while separated in a soundproofed chamber, with and without acoustic perturbation. In this setting, marmoset monkeys predominantly produce phee calls (monkey S: 99.1%, H: 92.0%, W: 95.6%, L: 90.8%, F:96.8%, Figures 1A and S1A), long-distance contact calls, composed of one (so-called single phees), two (double phees), or more phee syllables, to interact with conspecifics [14] (Figure 1A). Other call types such as trill-phees, twitters, trills, tsik-ekks [14, 19] or segmented phees [20] were rarely uttered (all other call types were well below 2.5% in all monkeys except segmented phees in monkey L [9.1%] and 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). To investigate whether perturbation of different frequency bands within the hearing range of the monkeys has different effects on their vocal behavior, 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 call frequency [noise bands of 10–15 kHz and 16–21 kHz] at four different amplitudes [50 dB, 60 dB, 70 dB, 80 dB] each). All 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 not being perturbed with noise (control). Our monkeys produced 7,485 phees (monkey F = 1,553 calls, H = 1,749, L = 981, S = 1,553, W = 1,649). Monkeys uttered mostly single and double phees (multi-syllabic phees with more than two syllables were rare or absent: monkey F = 1.5%, H = 0.3%, L = 2.5%, S = 0.8%, absent in W), with double phee rates between 8.0% and 75.3% (mean: 38.0% ± 12.1%, n = 5 monkeys) in the control condition.

Similar to results from cotton-top tamarins [17, 18], double phee rates dropped with increasing noise amplitude (Figure 1C; p = 0.025, n = 5 monkeys, Kruskal-Wallis test with post hoc multiple-comparison test) indicating that monkeys stopped calling after acoustic perturbation of the first phee syllable. Next, we evaluated whether call duration of the first phee syllable (hereafter referred to as phee) was affected by noise playback. Median phee duration varied from 1.2–1.9 s between individuals (mean: 1.6 ± 0.1 s) (Figures 1D, 1E, and S1B). Consistent with an

 

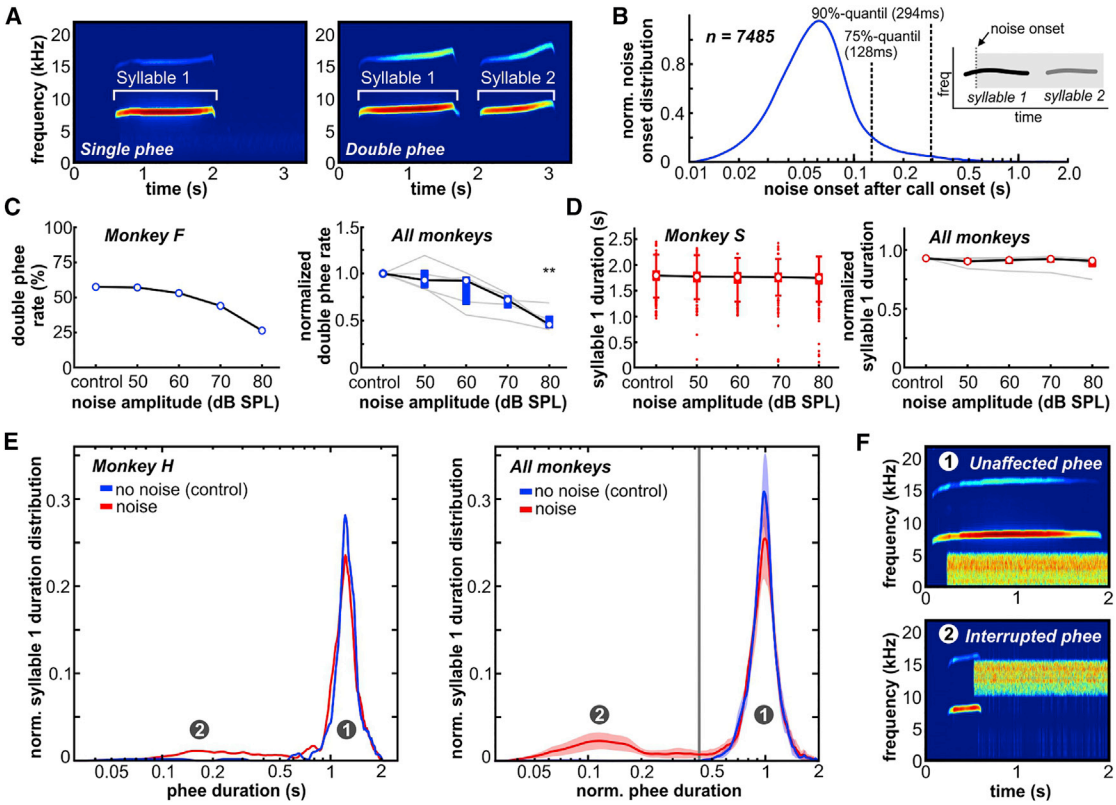


**Figure 1. Marmoset Monkeys Interrupt Their Calls during Vocal Production as a Response to Perturbing Noise Playback**

(A) Example spectrograms of a single and a double phee uttered by monkey H.

(B) Distribution of call-detection and noise-onset times after call onset; 75% of calls were detected within 128 ms. Inset: schematic noise perturbation of phees. Noise playback started after call onset.

(C) Double phee rate as a function of noise amplitude for an individual monkey (left) and normalized for all five monkeys (right). **$p < 0.01$, Kruskal-Wallis test with post hoc multiple-comparison test.

(D) Distribution of phee durations as a function of noise amplitude for an individual monkey (left) and normalized for all five monkeys (right). Medians: white circle inside boxes; first and third quartiles: upper and lower margins of boxes, respectively; 0.4% and 99.6% quantile: end of whiskers above and below boxes, respectively. Outliers: red circles above and below whiskers.

(E) Phee duration distributions for calls that were (noise) and were not perturbed (control) for individual monkey H (left) and normalized for all five monkeys (right). Vertical line marks the lower edge of phee duration distribution and defines calls affected by noise playback.

(F) Example spectrograms of phees unaffected (1) and interrupted (2) by noise perturbation.

See also Figures S1A and S1B.

earlier study, we did not find any population-level effect of noise playback on phee duration (Figure 1D; p = 0.8447, n = 5, Kruskal-Wallis). A significant decrease of call duration only occurred in one monkey (monkey W) when perturbed by the highest noise amplitude tested, with a change of approximately 20% (p = 5.9e-39, n = 1,649 calls, Kruskal-Wallis test with post hoc multiple-comparison test).

Next, we plotted phee duration distributions in noise perturbation and control conditions (Figures 1E and S1B). Phee duration distributions were similar in both cases (except for monkey W). However, we observed that all animals produced a small number of phee vocalizations during the noise condition shorter than 43.5% of their median normalized phee duration in the control condition (range: 0.3%–7.7% between monkeys, mean: 2.6% ± 1.3%), which were defined as interrupted phees (Figure 1F). Although the fraction of interrupted phees (Figure 1E) was small within individual monkeys, these phees were almost exclusively produced in the noise condition (250 in noise condition versus 3 in control, p = 6.2e-36, df = 1, Fisher's exact test). Different noise conditions and amplitudes were differentially effective in interrupting phee calls. Significantly more phee calls were interrupted during broadband noise (p = 2.08e-31, one-sample chi-square test, chi-square = 150.0, df = 4, n = 250) and high noise amplitude (p = 2.24e-7, one-sample chi-square test, chi-square = 33.7, df = 4, n = 250; Figure 2A). Interrupted phees were exhibited throughout recording sessions in most monkeys (except monkey F, which stopped producing interrupted phees after a few sessions). We did not find any significant differences between interrupted phee ratios exhibited within the first three, following three, and last three recording days (p = 0.368, n = 5 monkeys, Friedman test), nor between the first and last three recording days (p = 0.313, n = 5, Wilcoxon signed rank test).

To test whether interrupted phee occurrence was correlated with noise playback onset, we analyzed the phee duration distribution prior to noise onset as a function of syllable duration after noise onset of all interrupted phees (Figure 2B). First, we divided

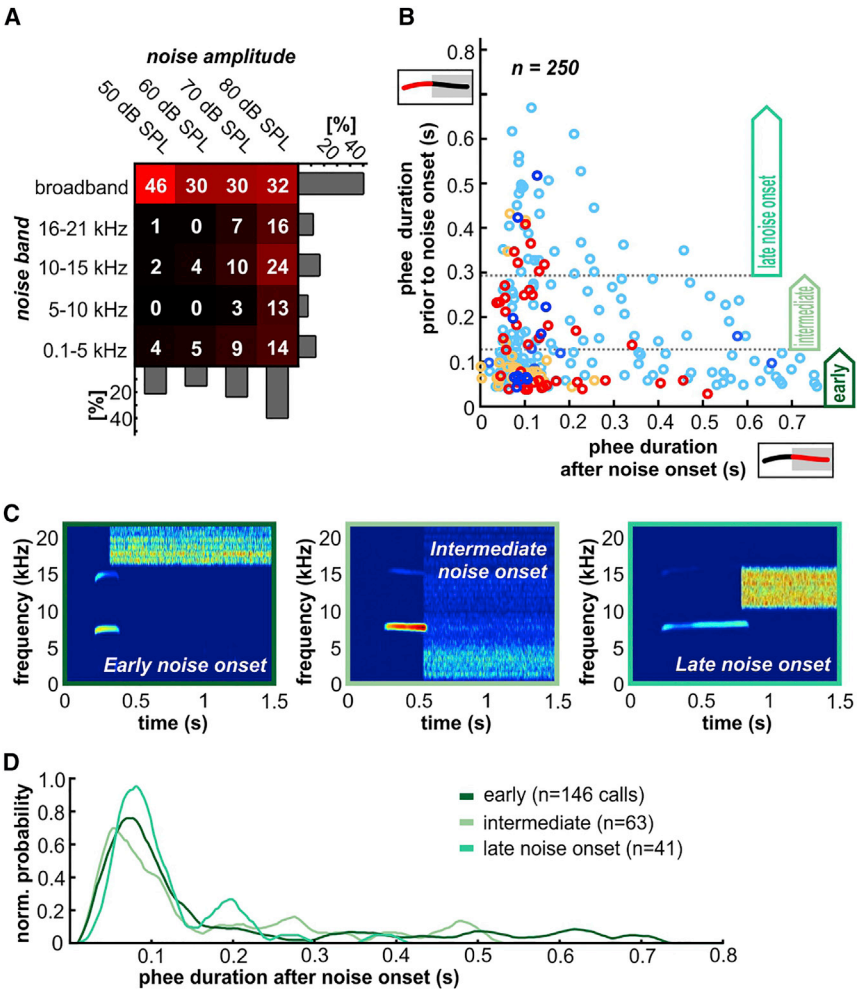


**Figure 2. Occurrence of Phee Call Interruption Is Dependent on Noise Conditions and Directly Related to the Onset of Noise Perturbation**

(A) Occurrence of interrupted phees in response to the different combinations of noise band and amplitude presented after vocal onset. Phee calls were predominantly interrupted in response to broadband noise and high noise amplitude. Color intensity is directly correlated to the number of phee call interruptions within different noise band/amplitude combinations.

(B) Correlation between noise onset and interruption of phee calls. Circles represent the relation between phee duration prior and after noise onset for each call. Different colors represent different subjects (n = 5). Horizontal lines group calls of early (0–128 ms after call onset, n = 146 calls), intermediate (128–294 ms, n = 63), and late noise onset (> 294 ms, n = 41) relative to call onset.

(C) Example spectrograms for each noise onset time group.

(D) Normalized distributions of phee durations within the three noise onset time groups indicate a direct effect of noise onset on call offset.

calls into three groups according to noise onset latency in relation to call onset—early, intermediate, and late (Figure 2C)—to control for the uneven distribution of noise onset times after call onset. Phee interruptions were significantly more frequent within the first 150 ms after noise onset than at later time points (Figure 2D, p = 2.0e-4, df = 1, Fisher's exact test; n = 250 calls). This behavior was consistent and did not differ between onset time groups (p = 0.8175, Kruskal-Wallis test; n = 250), showing that the median time point of call interruption was independent of whether perturbing noise was early, intermediate, or late after call onset. These results show that, in contrast to earlier findings [17, 18, 21], monkeys are capable of interrupting ongoing vocalizations in response to a perturbing acoustic signal.

What does this mean for vocal pattern generation? Is monkey vocalization not determined prior to vocal onset? Are marmosets able to interrupt phees at any time point? When examining the distribution of interrupted phee durations, we found a multimodal pattern with modes located at multiples of smaller fractions or distinct sharp peaks at multiples of smaller fractions (Figure 3A and Figure S1C). This indicates that calls cannot be interrupted at any point but that phee vocalizations consist of impartible small vocal motor units with potential subsequent abruptions at unit offset. In rare cases (less than 10%), we were able to detect phee calls, i.e., initiate noise onset, within 50 ms of call onset (Figure 2B). Among these, some monkeys were able to interrupt their phees as early as after the first vocal motor unit, i.e., less than 100 ms after call onset (monkey H in Figure 3A and monkey L in Figure S1C).

Next, we investigated the rare yet consistent occurrence of segmented phees [20], which were uttered occasionally and non-systematically (monkey L: 88 segmented phees, W:32, H:28; Figures 3B and 3C; see also STAR Methods; for audio-files of exemplar segmented phees shown in Figures 3E and 4G, see Audio S1–S3), further supporting the idea of impartible small vocal motor units. Phee segments showed variable durations with most segments < 500 ms (Figure 3D). Inter-segment intervals were sharply tuned with most durations < 100 ms and were significantly shorter than inter-syllable intervals, typically > 300 ms (p = 4.3496e-38, n = 231 for monkey L, p = 4.7378e-38, n = 252 for H, p = 2.0495e-13, n = 83 for W, Wilcoxon rank sum test). These findings indicate that the observed segmentation of phee calls is based on the introduction of gaps in a proper phee pattern rather than generating a multi-syllabic phee call consisting of short phee syllables.

To further investigate the acoustic structure of segmented phees, we reinforced three marmosets to vocalize. Monkeys were sitting in a primate chair and received a reward whenever they uttered a vocalization. With this approach, we were able to obtain a high number of vocalizations resulting in a corresponding high number of segmented phees under controlled experimental conditions (monkey L: 2,064 vocalizations, including 15.8% phee calls and 15.7% segmented phees; monkey P: 1,018 vocs, including 28.8% phees and 21.2% segmented phees; monkey H: 201 vocs, including 27.4% phee calls and

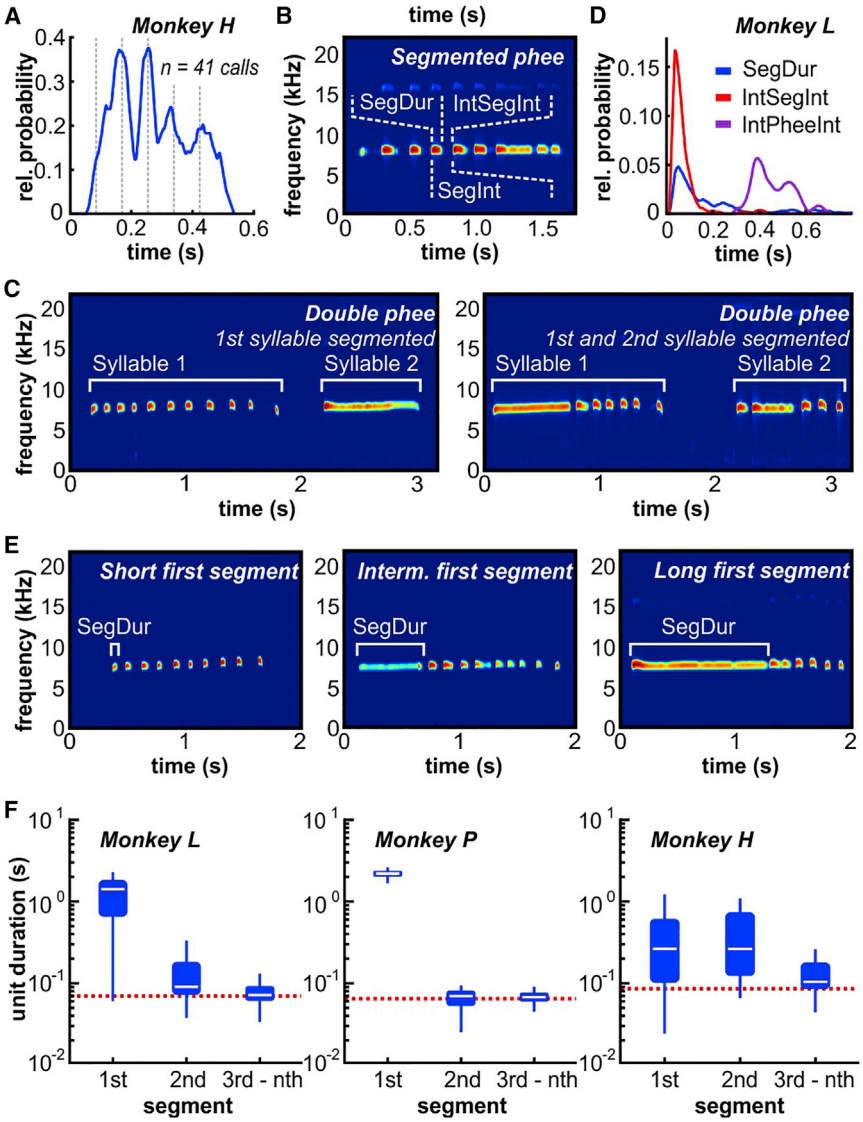


**Figure 3. Segmented Phees of Marmoset Monkeys**

(A) Multi-modal phee duration distribution of interrupted phees in an individual monkey. Vertical lines are multiples of each other.

(B) Example spectrogram of a segmented phee. SegDur, segment duration; IntSegInt, inter-segment interval (time between segment offset and onset of consecutive segment); SegInt, segment interval (time between two consecutive segment onsets).

(C) Example spectrograms of double phees indicate that either or both of the syllables can be segmented.

(D) Distributions of unit durations, inter-segment intervals, and inter-phee intervals in an individual monkey.

(E) Example spectrograms of segmented single phees show high variability in the duration of the first segment.

(F) Distribution of segment duration as a function of the position within a single phee for three individual monkeys. Phee segments were subdivided into first (monkey L: n = 308, monkey P: n = 209, monkey H: n = 41) and second segments (monkey L: n = 308, monkey P: n = 209, monkey H: n = 41). All following segments were grouped as 3rd-nth segments (monkey L: n = 750, monkey P: n = 120, monkey H: n = 40). First- and second-segment durations were typically longer and/or showed higher variability compared with the subsequent segments. Red horizontal line indicates the mode of the animal's corresponding unit duration (see Figure 4E). Medians: horizontal lines inside boxes; first and third quartiles: upper and lower margins of boxes, respectively; 0.4% and 99.6% quantile: end of whiskers above and below boxes, respectively.

See also Figure S1C.

8.0% segmented phees). Segmentation could be seen in both first and second phee syllables. However, since segmented second syllables were scarce due to low double-phee rates in monkey L and P, we focused on the first syllable for in-depth analyses of phee segmentation. We observed that the first and second phee segments showed great variability in duration within individual segmented phees (Figures 3E and 3F). This was due to the fact that phee call segmentation was predominantly initiated toward the end of calls (as indicated in Figure 4A) at different time points. However, as soon as segmentation started (after the first or second multi-unit segment), calls remained fully segmented from this time point to the end of the call in most cases (see also Figures 3E and 3F). Therefore, we aligned the end of the second unit of all segmented phees with more than two segments to investigate potential recurrent call structures across them (Figures 4A and S2A). Phee unit structure was extremely robust within segmented phees, with highly stable mean durations with little variance. These findings are supported by auto-correlograms of phee units (Figures 4B and S2B) with shortest unit intervals between 133 (monkey L and P) and 153 ms (monkey P), resulting in inter-individual unit rhythms between 6.5 to 7.5 Hz, and are also supported by distributions of phee unit durations and inter-unit intervals (Figures 4C and S2C). Finally, we compared the variability of these units with other brief and distinct marmoset call types such as twitter, tsik, and ekk vocalizations. We calculated the mode and standard deviation of phee units and investigated call type duration distributions. We observed that durations of both phee units and call types are short and sharply tuned (all < 100 ms; Figures 4D–4F), in contrast to the long and variable phee vocalizations (Figures 4E and 4F). These findings suggest that like other marmoset call types, phees, which are naturally produced with highly variable durations of ± 30% of their mean duration (Figures 1E and S1B), are built out of short, highly stereotyped units (Figure 4G).

Our findings indicate that phee vocalization—a prominent marmoset call type—is not a discrete call pattern itself but is built of many sequentially uttered stereotyped brief units. Therefore, phee duration is defined by the number of consecutively produced phee units rather than the duration of a single predefined, impartible pulse. Interestingly, durations of these brief units

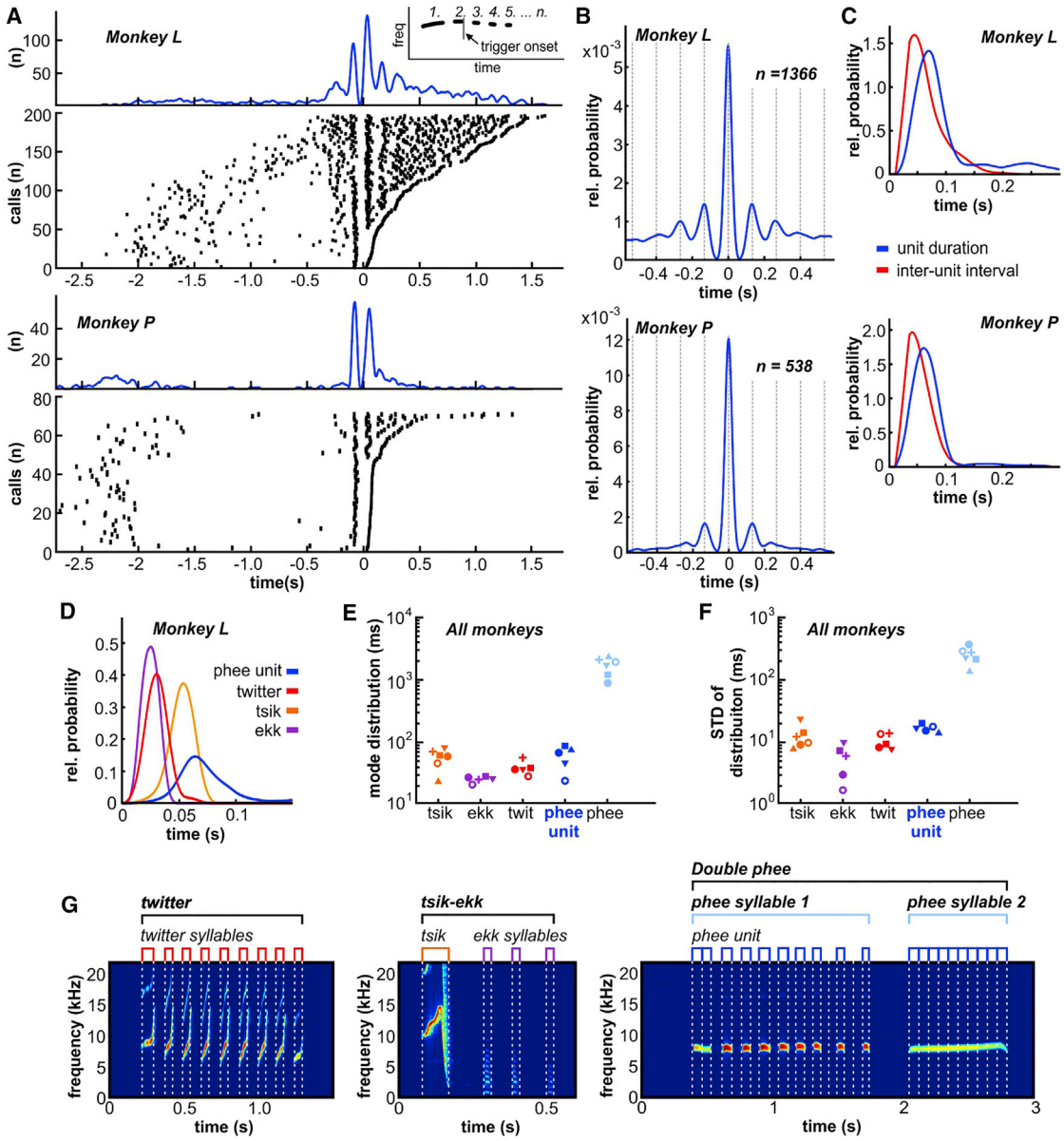


**Figure 4. Phee Vocalizations Consist of Consecutively Uttered Brief, Ultra-Precise Vocal Motor Units**

(A) Distribution of unit onsets triggered by the offset of the second unit for calls with at least three segments (as shown schematically in the upper right inset) for two monkeys. Lower panels show raster plots, upper panels the corresponding call unit density per monkey (monkey L: 196 calls, 1,142 units; monkey P: 71 calls, 262 units).

(B) Inter-unit interval histograms (auto-correlogram) for monkeys shown in (A). Vertical lines indicate multiples of first mode (monkey L: 308 calls, 1,366 units; monkey P: 209 calls, 538 units).

(C) Distribution of unit durations and inter-unit intervals < 300 ms for monkeys shown in (A) and (B).

(D) Example of an individual monkey's distribution of phee unit duration (n = 1,567 units) compared with distributions of twitter (n = 58), tsik (n = 69), and ekk call durations (n = 8).

(E and F) (E) Mode and (F) STDs of phee, phee unit, twitter, tsik, and ekk duration distributions of six monkeys. Individual monkeys are marked with different symbols (filled circles in [E] and [F] indicate duration modes and STDs of monkey L shown in [D]; triangles for P and squares for H)

(G) Example spectrograms for twitter, tsik-ekk, and double phee calls. Twitters consist of twitter syllables, tsik-ekks of tsik and ekk syllables, and phee calls of phee units.

See also Figure S2.

differs slightly between animals (Figure 4E) and also between siblings (monkeys P, L, and H are siblings), suggesting that each monkey has its "personal" vocal motor unit and that unit duration might not be inherited. Further studies should elucidate the basis for the observed differences in unit durations. Similarly, twitter calls or call-combinations such as tsik-ekks are characterized by the number of sequentially uttered concise syllables [14]. Our phee unit model challenges current theories on vocal

production and suggests that not only are defined calls, such as twitters and tsik-ekks, with clear observable interruptions built from consecutively produced brief units, but long phee calls are also built in this manner. The variable concatenation of units explains the high variability in call duration particularly noticeable in phee production [14]. Furthermore, duration distributions of these units also explain why phee-call durations do not exhibit multi-peaked distributions, as has been found for the short interrupted phee calls, even though they consist of concatenated units. Like other precise motor patterns, vocal motor units of phee calls exhibit a slight variation in duration. This "duration error" increases with the number of consecutively uttered units, resulting in a variation of phee-call durations for a distinct number of vocal motor units (Figure S2D). This is in accordance with call duration distributions of twitter calls, a multi-syllabic call type that consists of a variable number of concise syllables that also do not exhibit a multi-peaked duration distribution [19].

Our model can 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 phee units. Similar observations have been made in songs of passeriform birds. Song bouts consist of complex, distinct syllables that are learned during development [22]. Acoustic perturbation can interrupt ongoing song bouts only between, and not within, syllables [23, 24]. Similarly, learning processes induced by acoustic perturbation change acoustic features of the entire song syllable and not just from the initiation of acoustic perturbation [25].

Here, we present first evidence for such brief vocal motor units in monkey vocalization. The small fraction of interrupted and segmented phees in most animals indicates that marmosets may have stark neuronal and/or anatomical constraints in exhibiting such behavior. These constraints might be only barely to overcome by the marmosets, most likely because of the extrapyramidal nature of the primary vocal motor network [3]. However, it provides compelling evidence that the roots of precise vocal motor control mechanisms, a crucial preadaptation in the evolution of human speech in the primate lineage, can be investigated in marmoset monkeys. Human speech is defined by small, impartible vocal motor units produced at a stereotypical 3–8 Hz rhythm [26]. One theory of speech evolution posits that this rhythm may have evolved through the modification of rhythmic facial and/or laryngeal movements in the primate lineage [27]. Interestingly, unit intervals in segmented phee vocalizations exhibited a speech-like 7 Hz rhythm, supporting the idea that human speech rhythms may have evolved from such rhythmic movements in ancestral primates [28]. Further studies will have to elucidate how these segmented phees are produced biomechanically, e.g., whether they are composed of fast respiratory oscillations, so-called "mini-breaths" as have been shown to be present between twitter syllables in squirrel monkeys [29], or rather by fast, rhythmical movements of distinct laryngeal muscles (e.g., cricoarytenoid or thyroarytenoid) as during oscillatory vocal behavior in humans [30].

From a neurophysiological perspective, our phee unit model suggests a vocal pattern-generating network, which determines phee-call duration that might be directly inhibited in response to perturbing acoustic stimuli. Previous data indicate such a vocal pattern generating network situated in the lower brainstem receiving input from higher order structures [3, 4, 31]. One of these structures, the periaqueductal gray, exhibits call-duration-correlated activity and may be sufficient to determine phee-call duration [32]. However, considering the pre-vocal activity latencies within the PAG (≈100ms) [32, 33] together with the observed short latencies of call interruption after noise onset (< 100 ms) makes it unlikely that these inputs might be sufficient to produce the observed vocal behavior.

Our findings instead predict direct interactions between auditory input and a vocal pattern-generating network in the brainstem [3]. Deciphering how a vocal pattern-generating network is perturbed to interrupt call-pattern production with such short latencies is of great interest. Structures involved in the control of the observed behavior should contain neurons that exhibit vocal motor activity with short pre-vocal latencies that are inhibited in response to auditory stimulation. Structures containing such cells are the ventrolateral prefrontal cortex [34] and pontine and medullary reticular formation [35]. Therefore, we suggest two potential anatomically plausible audio-vocal loops, including auditory and premotor/prefrontal structures. First, a cortical audio-vocal loop from the auditory cortex to ventrolateral prefrontal cortex to premotor cortex to pontine reticular formation [3], all of which may serve as potential hubs in audio-vocal interaction [34–36]. Furthermore, a direct connection from the premotor cortex to single motoneuron pools might be sufficient, since the inhibition of single muscles, e.g., muscles involved in expiration, might be sufficient to interrupt vocal output. However, another anatomically plausible subcortical audio-vocal loop from the cochlear nucleus or superior olivary complex to the pontine reticular formation might be sufficient to mediate call interruption. 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 [37]. It would be interesting to elucidate whether cortical structures are crucial for such flexible vocal behavior or whether brainstem-based circuits are sufficient for the observed fast and precise behavioral responses.

## STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Marmosets
- METHOD DETAILS
  - Experimental Setup
  - Acoustic analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

### SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures and three audio files and can be found with this article online at https://doi.org/10.1016/j.cub.2018.01.070.

### ACKNOWLEDGMENTS

We thank John Holmes for proofreading and Cordula Gloge for her help in data analysis. 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 frame-work of the Excellence Initiative EXC 307).

## AUTHOR CONTRIBUTIONS

T.P. and S.R.H. designed the experiments; T.P. and J.L. conducted the noise playback experiments; T.P. and C.R.-S. conducted the vocal reinforcement experiments; T.P., J.L., and S.R.H. analyzed the noise playback experiment data; C.R.-S., J.L., and S.R.H. analyzed the vocal reinforcement experiment data; all authors interpreted the data and wrote the manuscript.

## DECLARATION OF INTEREST

The authors declare no competing interests.

Received: December 1, 2017
Revised: January 11, 2018
Accepted: January 23, 2018
Published: February 22, 2018

## REFERENCES

1. Balter, M. (2010). Evolution of language. Animal communication helps reveal roots of language. Science *328*, 969–971.
2. Hammerschmidt, K., and Fischer, J. (2008). Constraints in primate vocal production. In Evolution of Communicative Flexibility, D.K. Oller, and U. Griebel, eds. (Cambridge: MIT Press), pp. 93–121.
3. Hage, S.R., and Nieder, A. (2016). Dual Neural Network Model for the Evolution of Speech and Language. Trends Neurosci. *39*, 813–829.
4. Jürgens, U. (2002). Neural pathways underlying vocal control. Neurosci. Biobehav. Rev. *26*, 235–258.
5. Hammerschmidt, K., Jürgens, U., and Freudenstein, T. (2001). Vocal Development in Squirrel Monkeys. Behaviour *138*, 1179–1204.
6. Takahashi, D.Y., Fenley, A.R., Teramoto, Y., Narayanan, D.Z., Borjon, J.I., Holmes, P., and Ghazanfar, A.A. (2015). LANGUAGE DEVELOPMENT. The developmental dynamics of marmoset monkey vocal production. Science *349*, 734–738.
7. Takahashi, D.Y., Liao, D.A., and Ghazanfar, A.A. (2017). Vocal Learning via Social Reinforcement by Infant Marmoset Monkeys. Curr. Biol. *27*, 1844–1852.e6.
8. Gultekin, Y.B., and Hage, S.R. (2017). Limiting parental feedback disrupts vocal development in marmoset monkeys. Nat. Commun. *8*, 14046.
9. 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.
10. Hage, S.R., and Nieder, A. (2013). Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. Nat. Commun. *4*, 2409.
11. Choi, J.Y., Takahashi, D.Y., and Ghazanfar, A.A. (2015). Cooperative vocal control in marmoset monkeys via vocal feedback. J. Neurophysiol. *114*, 274–283.
12. Seyfarth, R.M., Cheney, D.L., and Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. Science *210*, 801–803.
13. Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., and Fischer, J. (2015). Vervets revisited: A quantitative analysis of alarm call structure and context specificity. Sci. Rep. *5*, 13220.
14. 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.
15. Sutton, D., Larson, C., Taylor, E.M., and Lindeman, R.C. (1973). Vocalization in rhesus monkeys: conditionability. Brain Res. *52*, 225–231.
16. Trachy, R.E., Sutton, D., and Lindeman, R.C. (1981). Primate phonation: Anterior cingulate lesion effects on response rate and acoustical structure. Am. J. Primatol. *1*, 43–55.
17. 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.
18. Egnor, S.E.R., Iguina, C.G., and 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.
19. Pistorio, A.L., Vintch, B., and 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.
20. Zürcher, Y., and Burkart, J.M. (2017). Evidence for Dialects in Three Captive Populations of Common Marmosets (Callithrix jacchus). Int. J. Primatol. *38*, 780–793.
21. Miller, C.T., Eliades, S.J., and Wang, X. (2009). Motor planning for vocal production in common marmosets. Anim. Behav. *78*, 1195–1203.
22. Brainard, M.S., and Doupe, A.J. (2002). What songbirds teach us about learning. Nature *417*, 351–358.
23. Cynx, J. (1990). Experimental determination of a unit of song production in the zebra finch (Taeniopygia guttata). J. Comp. Psychol. *104*, 3–10.
24. Hardman, S.I., Zollinger, S.A., Koselj, K., Leitner, S., Marshall, R.C., and Brumm, H. (2017). Correction: Lombard effect onset times reveal the speed of vocal plasticity in a songbird. J. Exp. Biol. *220*, 1541.
25. Sober, S.J., and Brainard, M.S. (2009). Adult birdsong is actively maintained by error correction. Nat. Neurosci. *12*, 927–931.
26. Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A.A. (2009). The natural statistics of audiovisual speech. PLoS Comput. Biol. *5*, e1000436.
27. MacNeilage, P.F. (1998). The frame/content theory of evolution of speech production. Behav. Brain Sci. *21*, 499–511, discussion 511–546.
28. Ghazanfar, A.A., Takahashi, D.Y., Mathur, N., and Fitch, W.T. (2012). Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. Curr. Biol. *22*, 1176–1182.
29. Häusler, U. (2000). Vocalization-correlated respiratory movements in the squirrel monkey. J. Acoust. Soc. Am. *108*, 1443–1450.
30. Titze, I.R., Finnegan, E.M., Laukkanen, A.M., Fuja, M., and Hoffman, H. (2008). Laryngeal muscle activity in giggle: a damped oscillation model. J. Voice *22*, 644–648.
31. Loh, K.K., Petrides, M., Hopkins, W.D., Procyk, E., and Amiez, C. (2017). Cognitive control of vocalizations in the primate ventrolateral-dorsomedial frontal (VLF-DMF) brain network. Neurosci. Biobehav. Rev. *82*, 32–44.
32. 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.
33. Düsterhöft, F., Häusler, U., and Jürgens, U. (2004). Neuronal activity in the periaqueductal gray and bordering structures during vocal communication in the squirrel monkey. Neuroscience *123*, 53–60.
34. Hage, S.R., and Nieder, A. (2015). Audio-vocal interaction in single neurons of the monkey ventrolateral prefrontal cortex. J. Neurosci. *35*, 7030–7040.
35. Hage, S.R., Jürgens, U., and Ehret, G. (2006). Audio-vocal interaction in the pontine brainstem during self-initiated vocalization in the squirrel monkey. Eur. J. Neurosci. *23*, 3297–3308.
36. Eliades, S.J., and Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature *453*, 1102–1106.
37. Jen, P.H.S., and Ostwald, J. (1977). Response of cricothyroid muscles to frequency-modulated sounds in FM bats, Myotis lucifugus. Nature *265*, 77–78.
38. Bezerra, B.M., and Souto, A. (2008). Structure and Usage of the Vocal Repertoire of Callithrix jacchus. Int. J. Primatol. *29*, 671–701.

# STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Experimental Models: Organisms/Strains | | |
| *Callithrix jacchus* | German Primate Center, Göttingen, Germany, and Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Germany | N/A |
| Software and Algorithms | | |
| MATLAB | MathWorks | R2014b |
| OpenEx | Tucker-Davis Technologies | N/A |
| Avisoft-Recorder | Avisoft Bioacoustics | version 4.2.22 |
| SASLab Pro | Avisoft Bioacoustics | version 5.2.09 |

## CONTACT FOR REAGENT AND RESOURCE SHARING

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).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Marmosets

We recorded 9185 calls produced by 6 adult common marmoset monkeys (*Callithrix jacchus*) housed at the University of Tübingen. Five animals participated in the noise playback experiment and three animals in the vocal conditioning experiment. Animals were usually kept in different sex pairs and were all born in captivity. The facility room was maintained at approximately 26°C, 40%–60% relative humidity, and with a 12h:12h light-dark cycle. They had *ad libitum* access to water and were fed daily with standard commercial chow and a selection of fruit, vegetables, mealworms, and locusts. Marshmallows and special fruit (e.g., banana, grapes) were used to transfer the animals from their home cages to a transfer box. Experimental procedures were approved by the local authorities of Tübingen (Regierungspräsidium) and are in agreement with the guidelines of the European Community for the care of laboratory animals.

## METHOD DETAILS

### Experimental Setup

#### *Noise-playback experiment*

The vocal behavior of five animals was recorded in response to noise playback that was initiated after vocal onset. Animals were transferred into a recording cage (0.6x0.6x0.8 m) that was placed in a soundproofed chamber, with *ad libitum* access to water and food pellets throughout the recording period. The vocal behavior of each individual monkey was recorded once a day with sessions ranging between 30 min to 2 hr. Recordings were performed for 10–28 days (mean: 17 ± 3 days) for each individual animal. The monkey's behavior was constantly monitored and observed with a video camera (ace acA1300-60 gc, Basler, Germany with 4.5–12.5 mm CS-Mount Objective H3Z4512CS-IR 1/2, Computar, Japan) placed on top of the cage and recorded with standard software (Ethovision XT version 4.2.22, Noldus, the Netherlands). Overall, we recorded 7999 vocalizations from five monkeys uttered in the noise-playback experiment. In this behavioral setup, 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%, L:90.8%, F:96.8%). Other call types such as trill-phees, twitter, trills, tsik-ekks [14] or segmented phees [20] were only rarely uttered (ratios were well below 2.5% for all other call types in all monkeys except segmented phees in monkey L [9.1%] and trill-phees in monkey H [4.6%]). Monkeys produced a mean 118 ± 9 (monkey S), 167 ± 31 (H), 117 ± 10 (W), 29 ± 4 (L), and 87 ± 7 (F) phee calls per session. We observed no systematic inter-individual differences in call duration between consecutively uttered segmented and unsegmented phees. While monkey W did not show any differences between segmented and unsegmented phees ($p = 0.831$, Wilcoxon sign rank test; median duration of 1.97 s for segmented versus 1.97 s for unsegmented phees), monkey L showed significantly longer segmented phees ($p = 5e\text{-}4$, Wilcoxon sign rank test; median duration: 1.82 s versus 1.51 s) and monkey H significantly shorter segmented phees ($p = 1.1e\text{-}4$, Wilcoxon sign rank test; median duration: 1.15 s versus 1.36 s) in comparison to unsegmented phees. Data were collected in sessions at various times during the day between 11 am and 5 pm.

The vocal behavior was collected with eight microphones (MKH 8020 microphone with MZX 8000 preamplifier, Sennheiser, Germany), which were positioned in an octagonal design around the cage to ensure optimal signal-to-noise ratio independent from the animals' body or head position, digitized using an A/D interface (Octacapture, Roland, Japan; sample rate: 96 kHz), and

recorded using standard software (Avisoft-Recorder, Avisoft Bioacoustics, Germany). A custom-written program (OpenEX, Tucker-Davis Technologies, U.S.A.) running on a work station (WS-X in combination with an RZ6D multi I/O processor, Tucker-Davis Technologies, U.S.A.) monitored the vocal behavior in real-time via an additional microphone (MKH 8020 microphone with MZX 8000 preamplifier, Sennheiser, Germany) placed on top of the cage, which automatically detected vocalizations through online calculation of several acoustic parameters, such as call intensity, minimal 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 128 ms after call onset (see Figure 1B).

For two out of three uttered vocalizations, we played back noise bursts of different frequency-bands and amplitude via a loudspeaker (MF1 Multi-Field Magnetic Speakers, Tucker-Davis Technologies, U.S.A.) positioned on top of the cage, immediately after vocal detection. Noise bursts had a duration of 4 s (including 10ms rise times) to ensure noise perturbation throughout the initiation of the second phee syllable (see Figure 1B). Five different noise-band conditions (broadband noise and bandpass filtered noise bands: 0.1–5kHz, 5–10kHz, 10–15kHz, and 16–21kHz) were played back at four different amplitudes (50dB, 60dB, 70dB, 80dB) each. These 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 was performed with the same custom-written software used for call detection. Avisoft and TDT recordings were clocked offline with custom-written software (MATLAB, Mathworks, U.S.A.). We did not find systematic differences in noise-related interruptions of phee vocalizations and the corresponding noise amplitude and frequency-band conditions. Therefore, we combined all noise amplitude and frequency-band conditions into one noise condition.

#### *Vocal reinforcement experiment*

Segmented phees were only occasionally observed in our monkeys during the noise-playback experiment. We therefore decided to investigate the acoustic structure of segmented phees by reinforcing three marmosets (two of which were examined in the noise playback experiment) to vocalize. The monkeys were trained to sit in a primate chair in a soundproof chamber. Vocalizations were recorded via a microphone (MKH 8020 microphone with MZX 8000 preamplifier, Sennheiser, Germany) positioned 10cm in front of the monkey's head. Each time the monkeys uttered a vocalization, regardless of call type, they received a liquid reward (mixture of water, marshmallow, fruit, marmoset gum, and curd cheese) provided by a small metal syrinx directly in front of the monkey's face. With this approach, we found that monkeys exhibited high numbers of calls, and fortunately uttered high numbers of segmented phees in some sessions. In the present study, we focused on these sessions with a considerable amount of segmented phees for in-depth analysis under a controlled experimental design. Since monkey H did not produce a sufficient number of segmented phees in the vocal reinforcement experiment, we included segmented phees produced in the noise-playback experiment (during control condition only) for this monkey as well. Monkey L and H were usually trained between 10 am and 12 am and monkey P between 11 am and 1 pm.

Vocal detection and reward presentation were synchronized and performed automatically with a custom-written program (OpenEX, Tucker-Davis Technologies, U.S.A.) running on a work station (WS-8 in combination with an RZ5 bioamp processor and RZ6D multi I/O processor, Tucker-Davis Technologies, U.S.A.). Vocalizations were recorded using the same system with a sampling rate of 100kHz. Vocal behavior of each individual monkey was recorded in daily 30min sessions (27 sessions for monkey L, 41 for monkey P, and 11 for monkey H) resulting in 308 segmented phees for monkey L, 209 for monkey P, and 41 for monkey H.

### Acoustic analysis

In the noise playback experiment, call onsets and offsets were manually detected using a custom-written MATLAB graphical user interface (Mathworks, U.S.A.) from the recording channel with best signal-to-noise ratio according to the position of the vocalizing animal in the cage. In the vocal conditioning experiment, call on- and offsets, as well as call unit on- and offsets, were manually flagged offline using standard software (SASLab Pro version 5.2, Avisoft Bioacoustics, Germany). Call duration was calculated as the difference between the beginning and end of the vocalization. In segmented phees (see below), phee unit duration was calculated as the difference between the beginning and end of the unit. Inter-unit interval was defined as the difference between the beginning of a segment and the end of the preceding one within the same phee syllable. The unit interval was defined as the difference between the beginning of a unit and beginning of the preceding unit. The spectrograms were calculated using a 1024-point FFT window, Hanning window (512 samples), and 125-sample overlap. We classified marmoset vocalizations into groups using previous definitions [8, 14, 19, 38]. Calls were manually classified as phee, twitter, tsik, and ekk calls based on their spectro-temporal profile and auditory playback. The four call types showed a very defined and distinct profile and could be easily classified manually. Phee is a tone-like long call with F0 around 7–10 kHz and is uttered individually as single phees or as two consecutive syllables, so-called "double phees." As previously reported [20], marmoset phees were occasionally segmented into two or more phee segments and were thus defined as segmented phees. This phee call variation was exhibited by three of our marmosets (monkey L = 88 segmented phees, W = 32, H = 28) and was defined by a segmentation of the phee syllable into two (monkey W) or more (monkeys L and H) brief phee segments, separated by silent inter-segment intervals (Figure 3B). Phees could be segmented in both the first and/or second phee syllable (Figure 3C). Initial phee units of phee syllables showed great variability (Figure 3E) and were significantly longer than the second and/or all other following units (Figure 3F). In the final experiment, we also compared the distribution of phee unit durations and syllabic structures of other call types such as twitters (n = 128 calls with a total of 812 twitter syllables), tsiks (n = 218), and ekks (n = 177). A twitter is a brief upward FM sweep that is usually uttered as a multi-syllabic call. A tsik is a broadband short call consisting of a linearly ascending FM sweep that merges directly into a sharply descending linear FM sweep. An ekk is a

brief call that is defined as one of the lowest frequency marmoset calls. Tsik and ekk calls are often produced consecutively as multi-syllabic tsik-ekk calls. Since not all animals produced all above-mentioned call types in the noise playback experiment and/or conditioning experiment, we included additional recordings from the animal facility from monkeys L and W to the underlying dataset of Figures 4E and 4F to get an appropriate number to compare syllable durations for most call types (ekks and twitters could be recorded from five monkeys).

For the noise playback experiment, double phee ratios were calculated for each individual monkey and noise amplitude by comparing the relative amount of double phees produced within all phee calls uttered. For inter-individual comparison, double phee ratios for all four noise amplitude conditions were normalized by dividing them by the double phee ratio in the control condition (no noise) for each individual monkey. Median durations of the first syllable of a phee call in the four noise amplitude conditions were normalized by the median duration of the first phee syllable in the control condition (no noise) for each individual monkey. Normalized call duration was calculated by dividing all call durations by the median call duration for each individual monkey. Call duration probabilities were normalized by the total amount of uttered vocalizations within each condition or call type. Probability distributions of call, syllable, and phee unit durations; call offsets; and phee unit auto-correlograms were smoothed with a moving average (bin widths, 20 [Figures 4A, 4C, 4D, S2A, and S2C], 50 [Figures 3A and 3D, 4B, S1C, and S2B] and 100ms [Figures 1B and 1E, 2D, and S1B]; step size, 1ms) for illustrative purposes only. We defined the border between phees with normal duration and interrupted phees in the noise condition as the point where the pooled normalized probability distribution reached zero on the left side in the control conditions. As a result of our dataset, phee-call durations shorter than 43.5% of the median phee duration were defined as interrupted phee vocalizations.

## QUANTIFICATION AND STATISTICAL ANALYSIS

A Kruskal-Wallis test with post hoc multiple comparison test and Bonferroni correction was used to test for significant differences between single/double phee ratios and between phee syllable length distributions with increasing noise amplitude. Differences between lengths of inter-segment intervals and inter-syllable interval lengths were tested using a two-sided Wilcoxon rank sum test. We used Fisher's exact test to check for differences in the occurrence of interrupted phees between the noise and control conditions and the occurrence of short and long interrupted phees within the noise condition. To evaluate differences in call interruption behavior between noise onset time groups we used a Kruskal-Wallis test. A Friedman test and Wilcoxon signed rank test were performed to test for differences between interrupted phee ratios within specific time ranges during the recording sessions. In all performed tests, significance was tested at an alpha = 0.05 level. Statistical analysis was performed using MATLAB (MathWorks, Natick, MA).

# Chapter 2: Compensatory Mechanisms Affect Sensorimotor Integration During Ongoing Vocal-Motor Acts in Marmoset Monkeys (under review)

Thomas Pomberger[1,2,3], Julia Löschner[1,3], Steffen R. Hage[1,*]

[1]Neurobiology of Vocal Communication, Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Str. 25, 72076 Tübingen, Germany

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

[3]These authors contributed equally

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

## Abstract

In vertebrates, 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 the 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 several 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 rapid vocal flexibility and various audio-vocal integration mechanisms in marmoset monkeys. Using acoustic perturbation triggered by vocal behavior, we investigated whether marmoset monkeys 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 marmoset monkeys are capable of rapidly modulating call amplitude and frequency in response to such perturbing noise bursts. Vocalizations swiftly increased call frequency after noise onset indicating a rapid effect of perturbing noise on vocal motor pattern production. Call amplitudes were also affected. Interestingly, however, the marmosets did not exhibit the Lombard effect as previously reported but decreased their call intensity in response to perturbing noise. Our findings indicate that marmosets possess a general avoidance strategy to call in the presences of ambient noise and suggest that these animals are capable of counteracting a previously thought involuntary audio-vocal mechanism, the Lombard effect, presumably via cognitive control processes.

## Keywords

Audio-vocal integration; Callithrix jacchus; primate communication; vocal communication; vocal flexibility

## Significance Statement

Vocal communication must constantly deal with acoustic perturbation caused by ambient noise. Several mechanisms have evolved to improve signal detectability during vocal communication in a noisy environment such as increasing call amplitude and frequency, and calling in silent gaps. Using acoustic perturbation triggered by vocal behavior, we investigated whether marmoset monkeys exhibit such changes in call structure when perturbing noise starts after call onset or only when starting prior to call onset, as previously reported. We show that marmosets rapidly modulate call amplitude and frequency in response to perturbing noise in ongoing calls. Interestingly, call amplitude was decreased, indicating a general avoidance strategy in marmoset vocal behavior during ambient noise, suggesting their ability to counteract previously thought involuntary audio-vocal mechanisms.

## Introduction

Communication between individuals is a crucial aspect for evolutionary success and appears in various forms in nature ranging from olfactory[1,2] to visual[3] to vocal signals[4]. For proper communication, the transmission of a signal sent out by a sender has to be detected and decoded by one or more receivers[5]. 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[6]. Such vocal modifications can happen involuntarily as well as under volitional control.

One of the most efficient mechanisms to increase signal-to-noise ratio in call production is the so-called Lombard effect, i.e., the involuntary increase in call amplitude in response to masking ambient noise[7]. It is often accompanied by a shift in call frequency[8,9] as well as a change in call duration[10,11] and has been shown in many vertebrate species from fish to frogs to birds to mammals including humans[12,13], suggesting that the Lombard effect is an evolutionary old behavior 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[10,14,15]. 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 signal-to-noise ratio.

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[16,17] or increasing the duration of specific calls[17], as well as the attempt to call in silent gaps[18], 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 marmoset tend to produce single calls instead of call sequences in response to perturbing noise stimuli[19]. Interestingly, marmoset monkeys are also capable of interrupting ongoing vocalizations rapidly after noise perturbation onset[19], overturning decades-old concepts regarding vocal pattern generation[20–22], 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 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. 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.

## Results

We measured vocal behavior 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 (**Fig. 1A and B**). 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[23] (**Fig. 1C**). Other call types such as trill-phees, twitters, trills, tsik-ekks[23,24], and segmented phees[25] 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 (**Fig. 1B**). To investigate whether perturbation of different frequency bands within the hearing range of the monkeys has different effects on their vocal behavior, 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, 80 dB] each). 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 = 1544 phees, H = 1471, S = 1631, W = 1652). 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% ± 9.8%, n = 4 monkeys) in the control condition.

We first investigated if 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 ($F(3,4904)=6.42$, $p=2.0e-04$ for amplitude, $F(4,4904)=20.68$, $p<0.0001$ for frequency, n=3180). 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=1733; **Fig. 2A**). 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, **Fig. 2A**). Second phee syllables showed no significant shift in fundamental frequencies when perturbed by noise (F(3,1343)=1.56, p=0.20 for amplitude, F(4,1343)=1.24, p=0.29 for frequency, n=761 , **Fig. 2B**).

Next, we quantified the magnitude of the observed frequency shifts by calculating population effect sizes (ES) of the factors frequency ($ES_{freq}$), amplitude ($ES_{ampl}$), and the combination of both conditions ($ES_{freq\ x\ ampl}$) 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 by Cohen (1992). We found $ES_{freq\ x\ ampl}$ values of 0.035 for first syllables and 0.019 for second syllables, indicating an effect for first syllables (**Fig. 2A and 2B**). $ES_{freq}$ for the first syllable was above the threshold ($ES_{freq}$=0.023), while $ES_{ampl}$ was below ($ES_{ampl}$=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 (**Fig. 2C and fig. S5**). 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 (**Fig. 3A and fig. 3B**). We found a significant decrease in call amplitude for first phee syllables ($F(3,3084)=1.01$, $p=0.39$ for amplitude, $F(4,3084)=5.3$, $p=0.0003$ for frequency, n=2019). 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=3093; **Fig. 3A**). Furthermore, the combined effect size ($ES_{freq\ x\ ampl}=0.024$) was above 0.02 while the effect size for the frequency ($ES_{freq}=0.014$) and amplitude ($ES_{ampl}=0.007$) factors were 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 ($F(3,350)=3.76$, $p=0.011$ for amplitude, $F(4,950)=1.71$, $p=0.15$ for frequency, n=554). 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; **fig. 3B**). 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 ($ES_{freq\ x\ ampl}=0.064$, $ES_{freq}=0.030$, $ES_{ampl}=0.024$) 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[17], 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 our animals are able to show a Lombard effect or suppress it in a noisy environment in general when producing phee calls, we modified our behavioral experiment scheme. 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 pseudorandomized playback conditions with a total duration of 1260 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; **fig. 3C and fig. S6**), 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=0.10, n=68) while monkeys F and H showed no significant amplitude change under noise perturbation (first syllable (H): p=0.89, second syllable (H): p=0.15, n=234 ; first syllable (F): p=0.91, second syllable (F): p=0.06, n=184). Taken 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.

## 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. 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[17,26]. 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.

**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 ambient noise in birds and bats[8,9,27] and only one study investigated the effect of different noise bands on call frequencies. 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[9]. In contrast, the present study shows that in marmosets, call frequency was predominantly only affected when we directly masked the calls fundamental frequency. As a result, the strongest rise in call frequencies were 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[28–30]. Here, it has been predicted that shifts in song frequencies of about 200 Hz increase call detectability by about 10 to 20%[29], which is mainly due to the fact that the spectrum of environmental noise generally shows a decay in amplitude with increasing frequency[29–32]. 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[33].

**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 signal to noise ratio in a noisy environment during vocal production is the Lombard effect, i.e., the involuntary rise in call amplitude as a response to masking noise[12,13]. In the present study, noise perturbation starting after phee call onset had no systematic effect on call amplitude of the first syllable, i.e., 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 a previous study[17]. This apparent discrepancy might be explained by the different call types that were investigated in both studies. While we focused on phee calls, a high amplitude call that is produced at the upper end of the amplitude scale[16], the earlier study investigated the twitter call, a vocalization that is produced at lower amplitude intensities[17].

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[34–37] and seems to be mainly driven by higher-order cognitive processes including cortical structures[13].

**Vocal flexibility during perturbing noise in marmoset monkeys.** Current studies have revealed a high degree of vocal flexibility in marmoset monkeys[38], allowing them to control when[14], where[39], and what to vocalize[40]. In addition, recent studies revealed that marmosets

are able to modulate distinct call parameters in response to acoustic feedback[19,41]. This vocal flexibility allows marmosets to avoid calling in the presence of environmental noise and predominantly initiate their vocalizations in silent periods[14]. In a previous study, we demonstrated that marmosets interrupt their vocalizations shortly after noise onset when perturbation starts after vocal onset[19], 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 behavior[19] 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 are lower.

We suggest that marmoset monkeys exhibit this vocal behavior 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.

**Mechanisms counteracting involuntary audio-vocal effects need cognitive control.** Based on the current work and earlier studies[14,19], we propose a hypothetical neuronal model suggesting various audio-vocal control mechanism involving cortical, subcortical, and corticofugal connections capable of modulating vocal behavior in a noisy environment (**Fig. 4**). In accordance to earlier work[42,43], 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[13]. The decision to initiate or suppress a call, as well as counteracting an involuntary effect (Lombard effect), needs cognitive control.

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. 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.

## Material and Methods

***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 12h:12h 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 analyzed in this study are a fraction of calls that have been recorded in a previous study (Pomberger et al. 2018).

***Experimental Setup and Procedure.*** The vocal behavior 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 x 0.6 x 0.8 m), which was placed in a soundproofed chamber, with *ad libitum* access to water and food pellets throughout the recording period. In this behavioral setup, 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-phees, twitter, trills, tsik-ekks[23], or segmented phees[25] were only rarely uttered (ratios were well below 2.5% for all other call types in all monkeys except trill-phees 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 behavior 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 am and 5 pm. Recordings were performed for 10–28 days (mean: 17±3 days) for each individual animal. The monkey's behavior was constantly monitored and observed with a video camera (ace acA1300-60gc, Basler, Germany with 4.5–

12.5 mm CS-Mount Objective H3Z4512CS-IR 1/2, Computar, Japan) placed on top of the cage and recorded with standard software (Ethovision XT version 4.2.22, Noldus, the Netherlands). The vocal behavior was collected with eight microphones (MKH 8020 microphone with MZX 8000 preamplifier, Sennheiser, Germany), which were positioned in an octagonal design around the cage (**Fig. 1A**), digitized using an A/D interface (Octacapture, Roland, Japan; sample rate: 96 kHz), and recorded using standard software (Avisoft-Recorder, Avisoft Bioacoustics, Germany). A custom-written program (OpenEX, 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.) monitored the vocal behavior in real-time via an additional microphone (MKH 8020 microphone with MZX 8000 preamplifier, Sennheiser, Germany) 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 (**Fig. 1B**).

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 detail 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, U.S.A.) 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 (**Fig. 1C**). 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, 80 dB) each. 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.

***Data Analysis.*** We programmed a custom-written GUI (Matlab, Mathworks, U.S.A.) to clock Avisoft, Noldus, and TDT recordings offline and to extract the detected calls from the recording channel with the best SNR. Vocal onset to offset were 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, 1024 FFT, overlap of 25%, and temporal resolution of one millisecond. 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[19] were excluded from further analysis. Second phee syllables were only analyzed 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. 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 (**Fig. S1)**.

Call amplitudes were calculated for all phee calls during which the animals did not move their heads during call production. For these calls, head positions were manually labelled by marking the two white ear tufts in the GUI (see **Fig. 1B**). Next, a perpendicular line starting at the center 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 (**Fig. 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.

From the recordings of the microphone foremost in front of the animal, call amplitude trajectories (in dB SPL) were calculated using a sliding window approach (window size: 25 ms; step length: 1 ms). 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[44]. 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.

$$D(w) = P_S(w) - \alpha P_n(w)$$

$$P_s'(w) = \begin{cases} D(w), & if\ D(w) > \beta P_n(w) \\ \beta P_n(w), & otherwise \end{cases} \quad (1)$$

$$\alpha \leq 1, \qquad 0 < \beta \ll 1$$

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.[44], 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[45]. 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; **Figs. S2A and B**) 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.

***Frequency/amplitude calculation and normalization.*** Mean fundamental frequency values were obtained with a sliding window approach (window size: 10 ms, step size: 1 ms). We then calculated the mean of the fundamental frequency in a 20-5 ms time window prior to noise onset (for the noise conditions) and call detection (for the control condition) for each individual call and subtracted this value 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 20-5 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*.

***Phee call discrimination models.*** Marmoset monkeys tend to interrupt their phee calls after the first syllable in response to noise perturbation[19]. 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[21]. 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 and colleagues[21] that suggested that early peak frequencies and durations of phee calls are sufficient to predict whether a phee call consists out 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) 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 (**Fig. S3**). Since we observed that early amplitude values are also a good predictor (**Fig. S4**), we used a three-dimensional classifier with call amplitude values at 25 ms after call onset as the third measure for amplitude analyses (**Fig. S4**). Basically, in a first step the mean, $\mu_k$, and covariance matrix, $\Sigma_k$, of each class is calculated from all control values to obtain the density function of the multivariate normal at a point, x, using the following formula:

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

Where $|\Sigma_k|$ is the determinant of $\Sigma_k$, and $\Sigma_k^{-1}$ 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 $\hat{P}(k|x)$ that a point $x$ belongs to class $k$ based on:

$$\hat{P}(k|x) = \frac{P(x|k)\,P(k)}{P(x)}$$

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

$$\hat{y} = arg \min_{y=1,\ldots,K} \sum_{k=1}^{K} \hat{P}(k|x)\, C(y|k)$$

Where $\hat{y}$ is the predicted classification, $K$ is the number of classes, and $C(y|k)$ 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.

***Statistical analysis.*** Statistical analyses were performed with MATLAB (2016b, MathWorks, Natick, MA). We performed a two-way ANOVA to test for significant differences in shifts of fundamental call frequency and amplitude within all noise band conditions (alpha = 0.05, Bonferroni corrected). Effect sizes (ES) 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 $\boldsymbol{p}$ and $\eta_p^2$ is calculated as:

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

***Data availability.*** 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.

## 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. S.R.H. provided the animals and supervised the project.

## Acknowledgments

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).

## References


1. Poldrack, R. A. & Farah, M. J. Progress and challenges in probing the human brain. *Nature* **526**, 371–379 (2015).
2. Stockhorst, U. & Pietrowsky, R. Olfactory perception, communication, and the nose-to-brain pathway. *Physiol. Behav.* **83**, 3–11 (2004).
3. Osorio, D. & Vorobyev, M. A review of the evolution of animal colour vision and visual communication signals. *Vision Res.* **48**, 2042–2051 (2008).
4. Ackermann, H. & Hage, S. R. Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behav. Brain Sci.* **37**, 529–604 (2014).
5. Bradbury, J. W. & Vehrencamp, S. L. *Principles of Animal Communication*. *American Entomologist* (1998).
6. Brumm, H. & Slabbekoorn, H. Acoustic Communication in Noise. *Adv. Study Behav.* **35**, 151–209 (2005).
7. Lombard, E. Le signe de l'elevation de la voix. *Ann. Mal. L'Oreille du Larynx* 101–119 (1911).
8. Hage, S. R., Jiang, T., Berquist, S. W., Feng, J. & Metzner, W. Ambient noise induces independent shifts in call frequency and amplitude within the Lombard effect in echolocating bats. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 4063–8 (2013).
9. Osmanski, M. S. & Dooling, R. J. The effect of altered auditory feedback on control of vocal production in budgerigars (Melopsittacus undulatus). *J. Acoust. Soc. Am.* **126**, 911–9 (2009).
10. Brumm, H. Signalling through acoustic windows: Nightingales avoid interspecific competition by short-term adjustment of song timing. *J. Comp. Physiol. A Neuroethol. Sensory, Neural, Behav. Physiol.* **192**, 1279–1285 (2006).
11. Luo, J., Goerlitz, H. R., Brumm, H. & Wiegrebe, L. Linking the sender to the receiver: Vocal adjustments by bats to maintain signal detection in noise. *Sci. Rep.* **5**, 1–11 (2015).
12. Brumm, H. & Zollinger, S. The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour* **148**, 1173–1198 (2011).
13. Luo, J., Hage, S. R. & Moss, C. F. The Lombard Effect: From Acoustics to Neural Mechanisms. *Trends Neurosci.* **41**, 938–949 (2018).
14. Roy, S., Miller, C. T., Gottsch, D. & Wang, X. Vocal control by the common marmoset in the presence of interfering noise. *J. Exp. Biol.* **214**, 3619–3629 (2011).
15. Zelick, R. D. & Narins, P. M. Analysis of acoustically evoked call suppression behaviour in a neotropical treefrog. *Anim. Behav.* **30**, 728–733 (1982).

16. Eliades, S. J. & Wang, X. Neural correlates of the lombard effect in primate auditory cortex. *J. Neurosci.* **32**, 10737–48 (2012).
17. Brumm, H. Acoustic communication in noise: regulation of call characteristics in a New World monkey. *J. Exp. Biol.* **207**, 443–448 (2004).
18. Roy, S., Miller, C. T., Gottsch, D. & Wang, X. Vocal control by the common marmoset in the presence of interfering noise. *J. Exp. Biol.* **214**, 3619–3629 (2011).
19. Pomberger, T., Risueno-Segovia, C., Löschner, J. & Hage, S. R. Precise Motor Control Enables Rapid Flexibility in Vocal Behavior of Marmoset Monkeys. *Curr. Biol.* **28**, 788–794 (2018).
20. Miller, C. T., Flusberg, S. & Hauser, M. D. Interruptibility of long call production in tamarins: implications for vocal control. *J. Exp. Biol.* **206**, 2629–2639 (2003).
21. Miller, C. T., Eliades, S. J. & Wang, X. Motor planning for vocal production in common marmosets. *Anim. Behav.* **78**, 1195–1203 (2009).
22. Egnor, S. E. R., Iguina, C. G. & Hauser, M. D. Perturbation of auditory feedback causes systematic perturbation in vocal structure in adult cotton-top tamarins. *J. Exp. Biol.* **209**, 3652–3663 (2006).
23. Agamaite, J. A., Chang, C.-J., Osmanski, M. S. & Wang, X. A quantitative acoustic analysis of the vocal repertoire of the common marmoset (Callithrix jacchus). *J. Acoust. Soc. Am.* **138**, 2906–2928 (2015).
24. Pistorio, A. L., Vintch, B. & Wang, X. Acoustic analysis of vocal development in a New World primate, the common marmoset (Callithrix jacchus). *J. Acoust. Soc. Am.* **120**, 1655–1670 (2006).
25. Zürcher, Y. & Burkart, J. M. Evidence for Dialects in Three Captive Populations of Common Marmosets (Callithrix jacchus). *Int. J. Primatol.* **38**, 780–793 (2017).
26. Egnor, S. R. & Hauser, M. Noise-induced vocal modulation in cotton-top tamarins (Saguinus oedipus). *Am. J. Primatol.* **68**, 1183–1190 (2006).
27. Schuster, S., Zollinger, S. A., Lesku, J. A. & Brumm, H. On the evolution of noise-dependent vocal plasticity in birds. *Biol. Lett.* **8**, 913–916 (2012).
28. Bermúdez-Cuamatzin, E., Ríos-Chelén, A. A., Gil, D. & Garcia, C. M. Experimental evidence for real-time song frequency shift in response to urban noise in a passerine bird. *Biol. Lett.* **7**, 36–38 (2011).
29. Nemeth, E. & Brumm, H. Birds and Anthropogenic Noise: Are Urban Songs Adaptive? *Am. Nat.* **176**, 465–475 (2010).
30. Pohl, N. U., Leadbeater, E., Slabbekoorn, H., Klump, G. M. & Langemann, U. Great tits in urban noise benefit from high frequencies in song detection and discrimination. *Anim. Behav.* **83**, 711–721 (2012).
31. Halfwerk, W. & Slabbekoorn, H. A behavioural mechanism explaining noise-

dependent frequency use in urban birdsong. *Anim. Behav.* **78**, 1301–1307 (2009).

32. Pohl, N. U., Slabbekoorn, H., Klump, G. M. & Langemann, U. Effects of signal features and environmental noise on signal detection in the great tit, Parus major. *Anim. Behav.* **78**, 1293–1300 (2009).
33. Luo, J., Kothari, N. B. & Moss, C. F. Sensorimotor integration on a rapid time scale. *Proc. Natl. Acad. Sci.* **114**, 6605–6610 (2017).
34. Kobayasi, K. I. & Okanoya, K. Context-dependent song amplitude control in Bengalese finches. *Neuroreport* **14**, 521–524 (2003).
35. Pick, H. L., Siegel, G. M., Fox, P. W., Garber, S. R. & Kearney, J. K. Inhibiting the Lombard effect. *J. Acoust. Soc. Am.* **85**, 894–900 (1989).
36. Therrien, A. S., Lyons, J. & Balasubramaniam, R. Sensory Attenuation of Self-Produced Feedback: The Lombard Effect Revisited. *PLoS One* **7**, (2012).
37. Vinney, L. A., Mersbergen, M. Van, Connor, N. P. & Turkstra, L. S. Vocal Control : Is It Susceptible to the Negative Effects of Self-Regulatory Depletion? *J. Voice* **30**, 638.e21-638.e31 (2016).
38. Ghazanfar, A. A., Liao, D. A. & Takahashi, D. Y. Volition and learning in primate vocal behaviour. *Anim. Behav.* 1–9 (2019).
39. Choi, J. Y., Takahashi, D. Y. & Ghazanfar, A. A. Cooperative vocal control in marmoset monkeys via vocal feedback. *J. Neurophysiol.* **114**, 274–283 (2015).
40. Liao, D. A., Zhang, Y. S., Cai, L. X. & Ghazanfar, A. A. Internal states and extrinsic factors both determine monkey vocal production. *Proc. Natl. Acad. Sci.* **115**, 3978–3983 (2018).
41. Eliades, S. J. & Tsunada, J. Auditory cortical activity drives feedback-dependent vocal control in marmosets. *Nat. Commun.* (2018). doi:10.1038/s41467-018-04961-8
42. Hage, S. R. & Nieder, A. Dual Neural Network Model for the Evolution of Speech and Language. *Trends in Neurosciences* **39**, 813–829 (2016).
43. Hage, S. R. Precise vocal timing needs cortical control. *Science.* **363**, 926–928 (2019).
44. Berouti, M., Schwartz, R. & Makhoul, J. Enhancement of speech corrupted by acoustic noise. in *ICASSP '79. IEEE International Conference on Acoustics, Speech, and Signal Processing* **4**, 208–211 (1978).
45. Brumm, H., Schmidt, R. & Schrader, L. Noise-dependent vocal plasticity in domestic fowl. *Anim. Behav.* **78**, 741–746 (2009).

## Figures

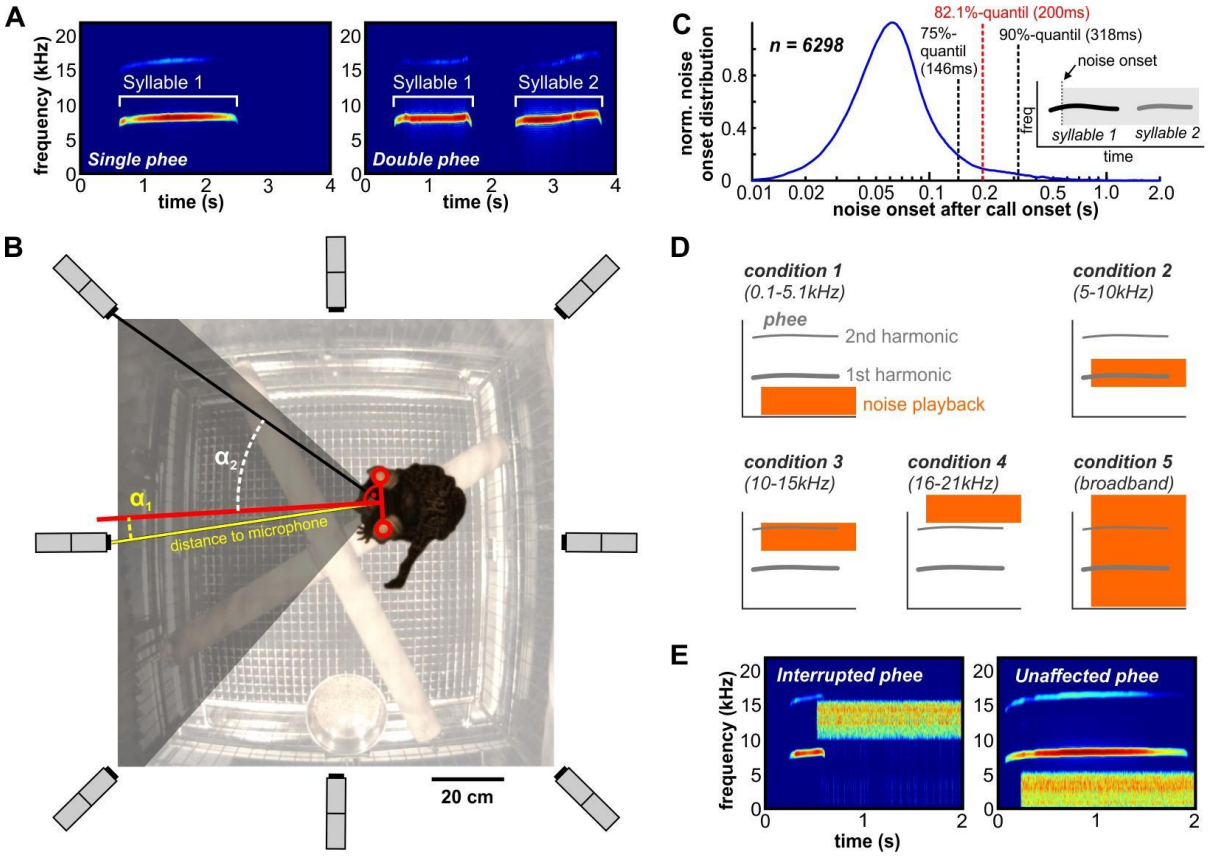


**Figure 1: Experimental setup and design.** (**A**) Exemplary spectrograms of single and double phee calls. (**B**) The vocal behavior of monkeys was recorded in a soundproof chamber. The behavior was continuously monitored and recorded. The red line shows the monkey's head position in relation to the two closest microphones (yellow and black line). 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. (**C**) Relative vocal detection distribution over time (s). (**D**) Noise condition overview with masking properties. (**E**) Exemplary spectrograms of an interrupted single phee (10–15 kHz noise condition) and unaffected phee (0.1–5 kHz noise condition).

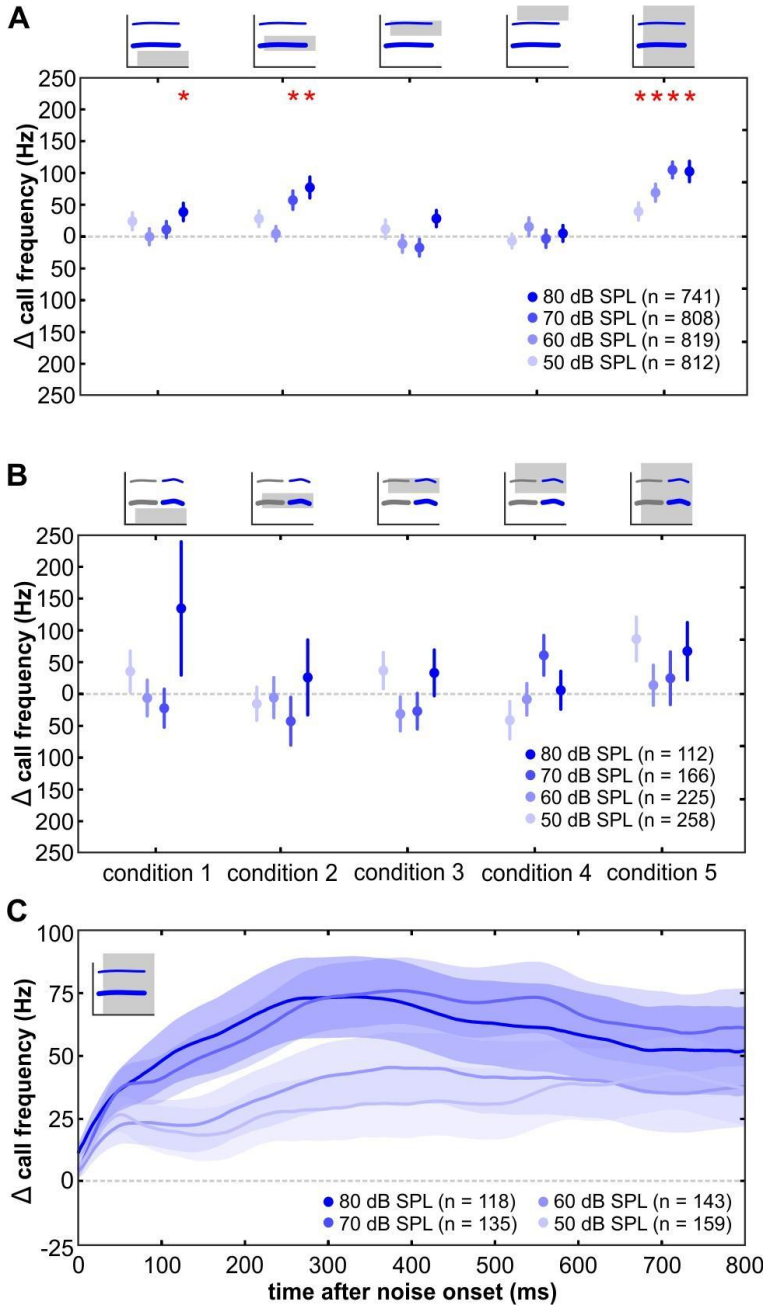


**Figure 2: Increasing frequency shifts in response to noise bursts.** Δ call frequency (Hz) per corresponding noise condition normalized to control data (dashed lines). Mean of median frequencies after noise onset of each call pooled over all monkeys ± SEM (**A**) for first phee syllables 0–800 ms after noise onset (**B**) for second phee syllables 100–400 ms after second syllable onset. (**C**): First syllables mean Δ call frequency courses (Hz) ± SEM of all amplitude conditions during broadband noise over time after noise onset (ms). Asterisks denote significant differences.

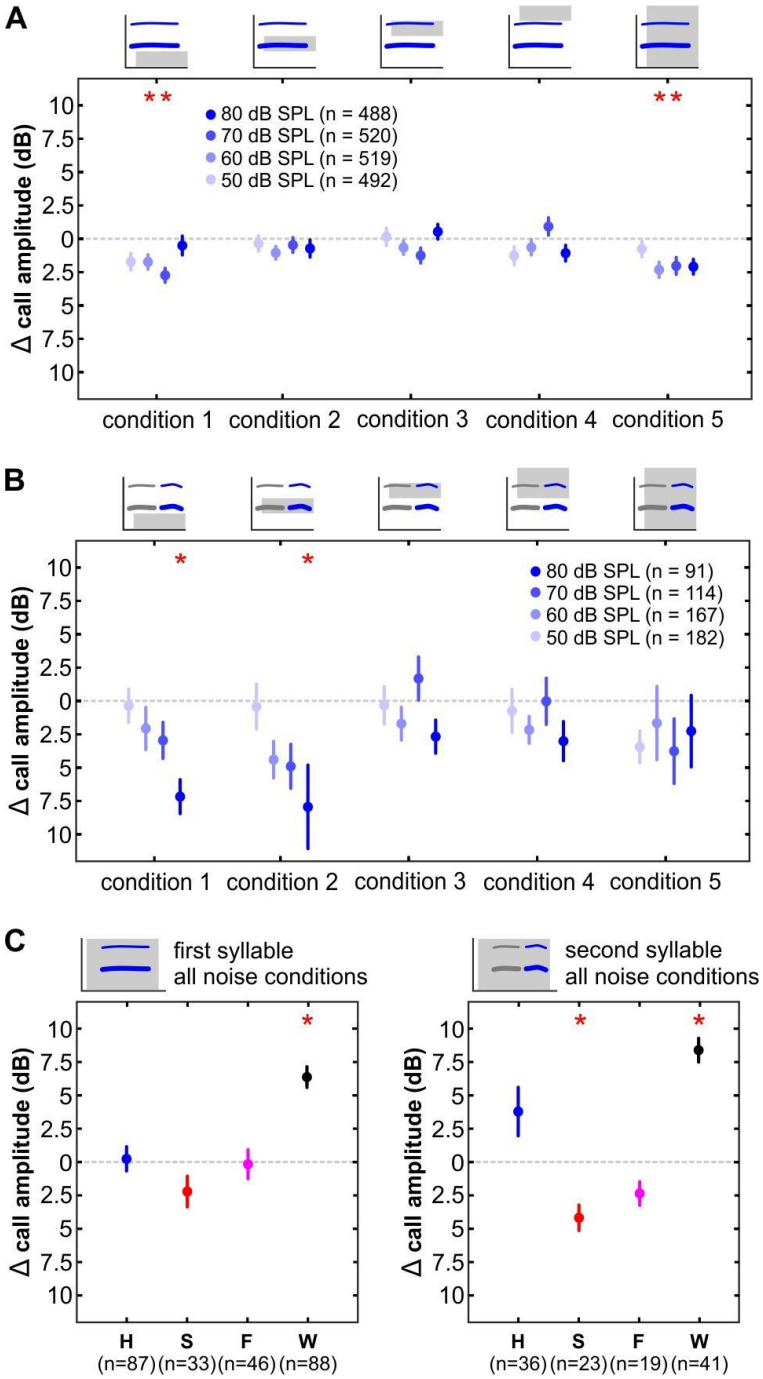


**Figure 3: Decreasing amplitude shifts in response to noise bursts.** Δ call amplitude (dB) per corresponding noise condition normalized to control data (dashed lines). Mean of median amplitudes after noise onset of each call pooled over all monkeys ± SEM (**A**) for first phee syllables and (**B**) for second phee syllables. Over all noise conditions pooled max Δ amplitudes (dB) ± SEM during 180 s noise per monkey (**C**) for first syllables and (**D**) for second syllables. Asterisks denote significant differences.

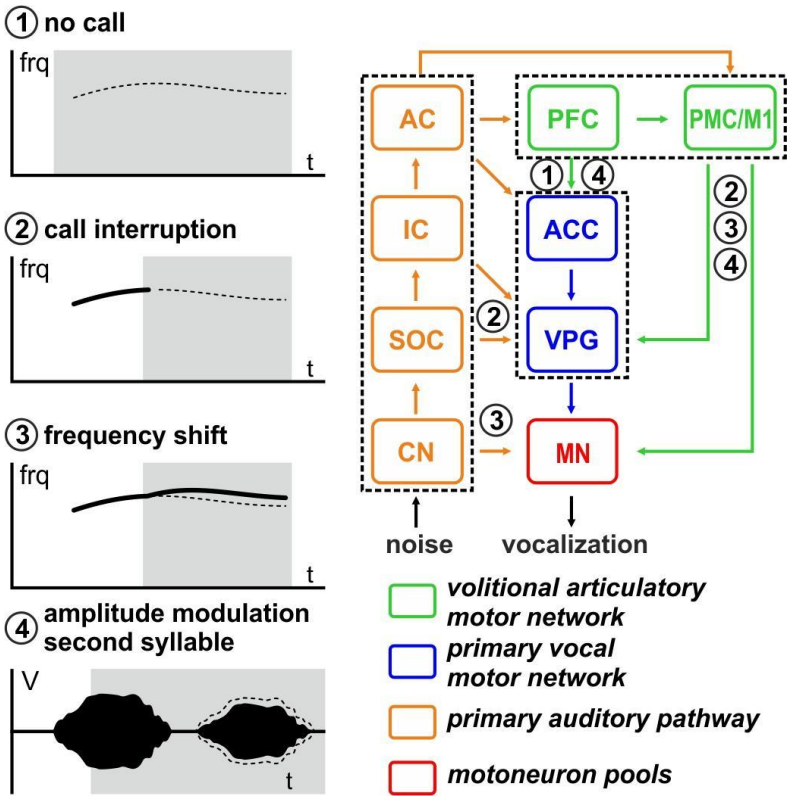


**Figure 4: Hypothetical neuronal model for audio-vocal interaction.** Call production might be affected by ambient noise at different brain levels. Audio-vocal integration mechanisms are known to happen between cortical and subcortical structures as well as via corticofugal projections. See text for further explanation.

## Supplemental Figures

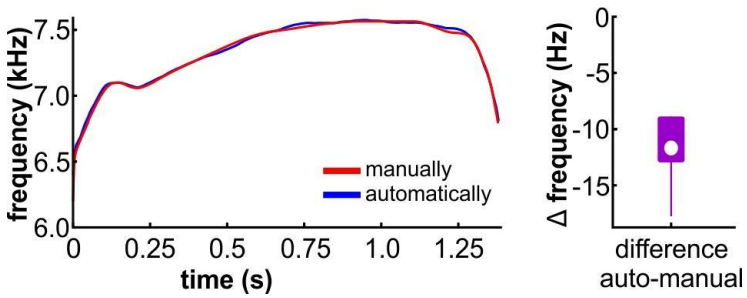


**Figure S1: Comparison of automatically vs manually marked calls.** (**A**) Frequency course (Hz) over time (ms) of a manually (blue) and automatically (red) marked example phee call. (**B**) Mean Δ frequencies of automatically-manually marked calls.

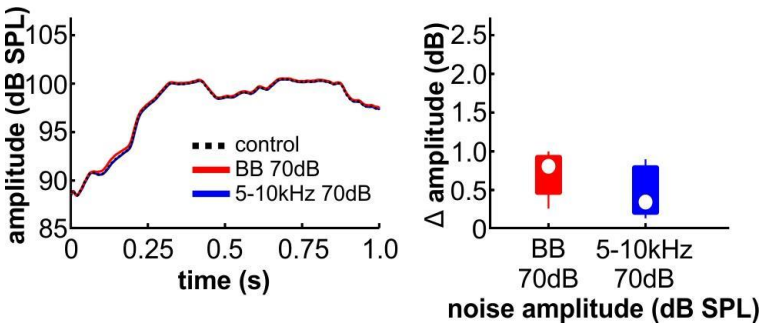


**Figure S2: Test for noise subtraction accuracy.** (**A**) Mean amplitude courses (dB SPL) for both 70 dB overlapping noise conditions (5–10 kHz, blue; broadband, red) as well as for the control (no noise, black dashed) of 10 test phees over time (ms). (**B**) Maximum Δ amplitude (dB) compared to control calls of 70 dB broadband and 5–10 kHz noise conditions.

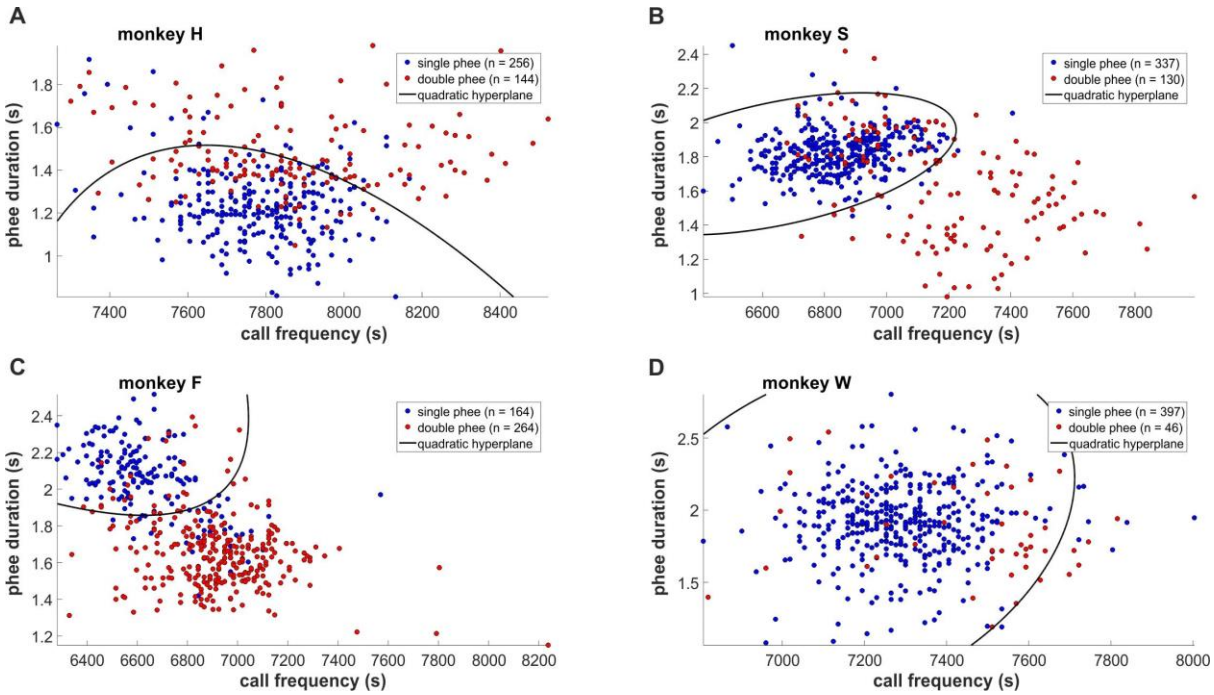


**Figure S3: Single-double phee discrimination with a two-dimensional classifier.** For single-double phee discrimination, first syllable fundamental peak frequencies (Hz) at 25 ms after call onset and corresponding phee durations (s) were used.

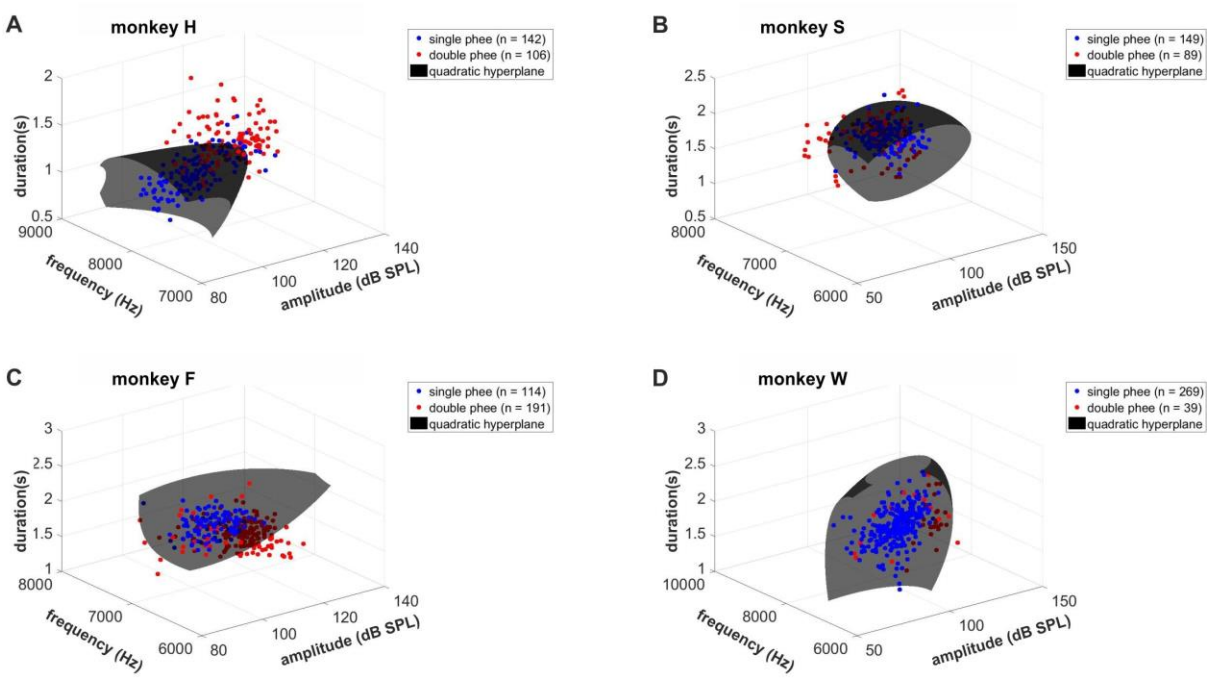


**Figure S4: Single-double phee discrimination with a three-dimensional classifier.** For single-double phee discrimination, first syllable fundamental peak frequencies (Hz), as well as amplitudes (dB SPL) at 25 ms after call onset and corresponding phee durations (s), were used.

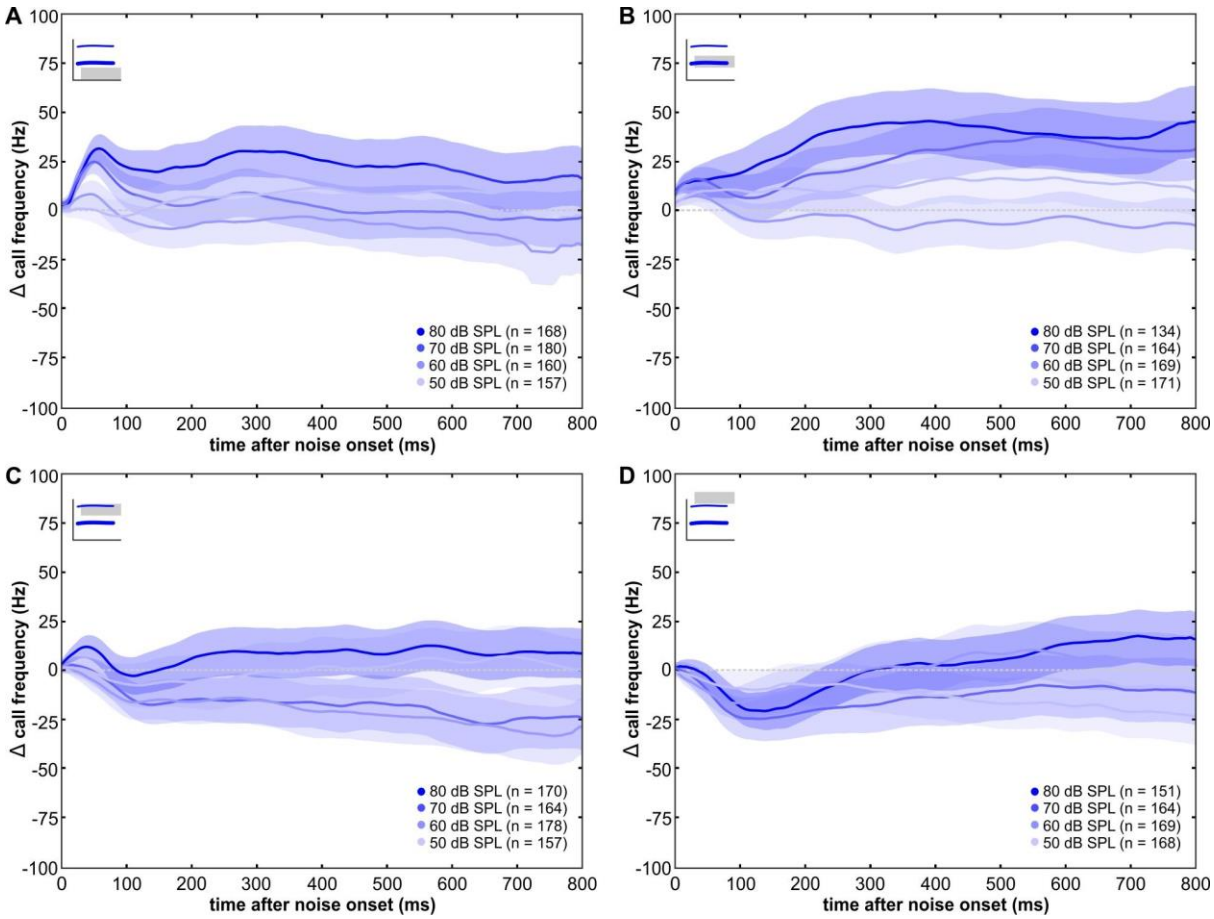


**Figure S5: Frequency courses in response to noise bursts.** First syllables mean Δ frequency courses (Hz) ± SEM of all amplitude conditions normalized to control data (dashed lines) during (**A**) 0.1–5.1 kHz, (**B**) 5–10 kHz, (**C**) 10–15 kHz, and (**D**) 16–21 kHz noise conditions over time after noise onset (ms).

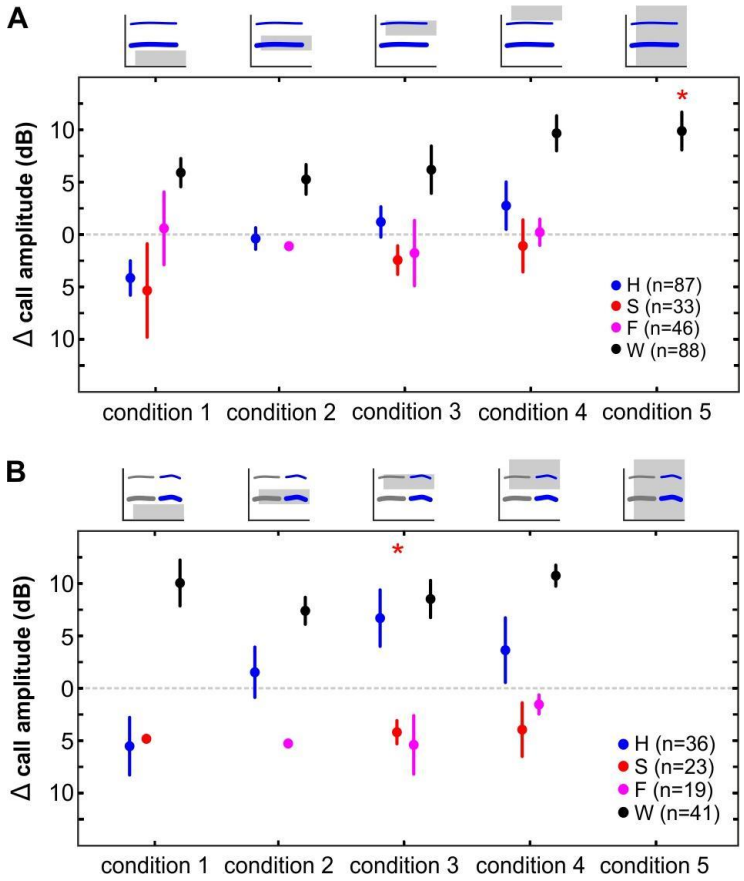


**Figure S6: Amplitude shifts in response to noise bursts.** Pooled maximum Δ amplitudes (dB) ± SEM during 180 s noise per monkey per noise condition normalized to control data (dashed lines) (**A**) for first syllables and (**B**) for second syllables. Asterisks denote significant differences.

# Chapter 3: Cognitive control of complex motor behavior in the marmoset monkey (under review)

Thomas Pomberger[1,2,†], Cristina Risueno-Segovia[1,2,†], Yasemin B. Gultekin[1,2,†], Deniz Dohmen[1,2,†], Steffen R. Hage[1,*]

[1]Neurobiology of Vocal Communication, Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Str. 25, 72076 Tübingen, Germany

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

[†]These authors contributed equally

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

## Abstract

Marmosets have attracted significant interest in the life sciences. Similarities with human brain anatomy and physiology, such as the granular frontal cortex, as well as the development of transgenic lines and potential for transferring rodent neuroscientific techniques to small primates make them a promising neurodegenerative and neuropsychiatric model system. However, whether marmosets can exhibit complex motor tasks in highly controlled experimental designs, one of the prerequisites for investigating higher-order control mechanisms underlying cognitive motor behavior, has not been demonstrated. We show that marmosets can be trained to perform complex vocal behaviors in response to arbitrary visual cues in highly controlled operant conditioning tasks. Our results emphasize the marmoset as a suitable model to study complex motor behavior and the evolution of cognitive control underlying speech.

## Main Text

The marmoset, a small New World primate, has recently garnered considerable interest as a suitable model organism in the life sciences (*1*). Similarities to humans in terms of genetic and physiological features, in combination with high fertility, short life span, and the ease to keep them under captivity makes them an especially efficient model for biomedical research and genetics (*2*). The prospect of developing primate transgenic lines (*3*), their granular frontal cortex, and potential for transferring a number of rodent neuroscientific techniques to a small primate with a lissencephalic brain (*1*) position the marmoset as a promising neurodegenerative and neuropsychiatric model system of prefrontal cortex dysfunctions. As an example, marmosets are highly social and vocal animals that use vocal signals for acoustic communication (*4*–*6*), a behavior that is severely affected by neurodegenerative diseases such as Parkinson's and Alzheimer's disease in humans (*7*, *8*).

To date, a variety of neurophysiological methods, as well as different brain imaging techniques have been successfully developed and established in marmosets, highlighting the potential for using these animals to study cognitive processes and their underlying neural network in different conditions and contexts (*1*, *2*). Currently, these methods are used in anesthetized (*9*, *10*), freely moving (*11*), and restrained marmosets (*12*), as well as in animals that have been trained to perform basic motor tasks such as licking (*13*, *14*), saccadic eye movements (*15*, *16*), or arm reaching (*17*, *18*) in response to visual or auditory stimuli. However, neuroscience needs complex behaviors to learn more about certain brain-behavior relationships (*19*). Unfortunately, marmosets have not yet been trained to perform complex behavioral tasks in well-controlled experimental designs, a prerequisite for the investigation of frontal neural networks underlying intricate cognitive processes, as shown in the canonical macaque model (*20*–*22*). Therefore, providing evidence that marmoset monkeys can be successfully trained to perform complex behavioral

tasks would bridge the gap and make them a suitable model system for investigating the neural network underlying cognitive processes in health and disease.

We trained four adult marmoset monkeys (*Callithrix jacchus*) to perform a computer-controlled go/nogo detection task by using their vocal behavior as a response. We show that they are able to volitionally control their vocal output and use it as an immediate response to a learned, abstract visual cue, thus demonstrating the ability to instrumentalize their vocal output to perform a task successfully. Furthermore, we trained one monkey to switch between distinct call types from trial to trial in response to different visual cues in a discrimination task. Our findings show that marmoset monkeys can be trained to perform a complex motor behavior in a highly controlled experimental design suggesting their suitability as an innovative non-human primate model to decipher higher-order cognitive motor control mechanisms.

We recorded 10,619 vocalizations from four marmoset monkeys that where uttered in 80 daily sessions while performing either a detection or discrimination task (Table 1). In the visual detection task, monkeys were trained to sit in a monkey-chair in front of a monitor in a soundproof chamber (Fig. 1A) and required to vocalize after cueing by an arbitrary visual stimulus (red square) to receive a reward (see Fig. 1A and B for experimental setup and design and Methods for details). The data were obtained from 15 consecutive daily sessions per monkey. Within sessions, all monkeys produced a variety of different call types (Fig. 1C) with short call types such as chirps (40.6±20.8%) and tsiks (25.4±14.6%) occurring most frequently (Fig. 1D). All monkeys exhibited mean call counts between 55 and 232 calls/session resulting in high call rates of between 2 and 11 calls per minute (Table 1). Throughout self-initiated trials, all monkeys produced significantly more calls in “hit” than “catch” trials ($p=6.1e-05$ for all individual monkeys, $n = 30$, Wilcoxon signed rank test). Mean values were high for all monkeys for “hit” (monkey E: 72.3±3.3%, monkey H: 85.2±3.5%, monkey L: 77.1±3.2%, monkey P: 20.8±4.8%) and low for “false alarm” rates (monkey E: 27.4±3.5%, monkey H: 20.2±4.1%, monkey L: 32.2±2.8%, monkey P: 4.8±0.5%), resulting in

a mean population “hit” rate of 63.9±3.6% and a mean population “false alarm” rate of 21.2±2.0% (Fig. 2A). These data show that the monkeys reliably produced calls in response to the visual go-cue. All monkeys showed similar and distinct response patterns during “hit trials” with median latencies between 828 and 1219 ms, resulting in a median population response latency of 903 ms (Fig. 2B). Next, we investigated whether the vocal response latency was dependent on the cue delay, i.e., the waiting period between self-induced trial initiation and “go”-cue onset. Therefore, we tested the relationship between the vocal response latency and duration of the corresponding cue delay for all monkeys. Vocal response latencies significantly decreased with longer cue delays in two monkeys (p=8.53e-06, r=-0.11, n=976 for monkey H and p=8.7e-03, r=-0.08 n=1745 for monkey L; Pearson’s correlation, Fig. 2C). Differences in median response latencies of hit calls uttered during go-signals after short (1–1.5 s) and long pre-cue durations (2.5–3 s for monkey L and 4.5–5 s for monkey H) were small in both of these monkeys (monkey H: 121 ms, monkey L: 192 ms). We then investigated whether animals exhibited a different ratio of call types within the three phases of the visual detection task (go, catch, and precue) and outside of the self-initiated trials. Even though call type ratios differed between animals, we observed a significant difference in call type distribution between these four phases (p=1.7e-02, df=12, sum square=2650.2, 3-Way ANOVA, Fig. 2D). This difference could be explained by a higher occurrence of long call types, such as trills and phees, outside of trials and a higher occurrence of short call types, such as chirps, tsiks, and ekks, in trials.

Within trials, monkeys predominantly produced calls within the “go” phase. However, we also observed that monkeys produced a substantial number of calls during the pre-cue phase, resulting in trial abortion (Table 1, Fig. 2E). Furthermore, we measured a significant correlation between the number of such precue calls and the corresponding total number of calls per session for all monkeys (p=4.9e-05, r=0.86 for monkey E; p=2.7e-26, r=0.93 for H; p=2.7e-10, r=0.98 for L; p=1.2e-09, r=0.97 for P; n=30, Pearson’s correlation, Fig. 2F). We hypothesized that the

number of precue calls and, more generally, the corresponding total number of calls per session are directly related to the arousal state of the animal and that animals in a higher arousal state are capable of inhibiting call production to a lesser extent than monkeys in a low arousal state. To test this, we compared the number of precue calls with the corresponding sessions' call rate, which has been shown to directly correlate with the arousal state in monkeys (*23*, *24*). We observed a significant correlation between the number of precue calls and corresponding call rate for the session, suggesting a significant role of the animals' arousal in the overall calling behavior ($p=4.7e-03$, $r=0.86$ for monkey E; $p=2.5e-17$, $r=0.84$ for H; $p=9.5e-06$, $r=0.89$ for L; $p=9.6e-08$, $r=0.95$ for P; $n=30$, Pearson's correlation, Fig. 2F). We also tested whether the arousal state affected the performance of the monkeys in the detection task. We computed d′-sensitivity-values by subtracting z-scores (normal deviates) of median "hit"-rates from z-scores of median "false alarm"-rates (see Methods). No significant correlations were found between d′ values and the precue calls of the corresponding sessions, suggesting that there was no influence of the state of arousal on task performance ($p=0.69$, $r=0.11$ for monkey E; $p=0.28$, $r=0.14$ for H; $p=0.19$, $r=0.36$ for L; $p=0.21$, $r=0.35$ for P; $n=30$, Pearson's correlation, Fig. 2F).

Our findings show that marmoset monkeys possess the capability to volitionally control vocal output in general. As a next step, we wanted to investigate whether these animals are able to utter different call types on command. We, therefore, trained one of our animals (monkey H) to perform a visual discrimination task (Fig. 1B and Methods for details). Here, the animal had to produce two different types of vocalizations in response to distinct visual cues. As during the visual detection task, the monkey was required to vocalize in response to arbitrary visual cues (red and blue squares). However, here the monkey was trained to utter brief "chirp"-vocalizations or "chirp sequences" in response to the blue square and to emit long "trill calls" or call combinations, such as "chirp-trill" and "trill-phee" sequences, in response to the red square (Fig. 3A).

In the first 10 sessions after the discrimination task had been introduced (initial training phase), we observed that the animal showed a significantly higher vocal response during the red go signal (go1) than during the new blue go signal (go2; 91.9±1.8% vs. 62.8±3.8%; p=2e-03, n=20, Wilcoxon signed-rank test; Fig. 3B). However, the animal produced significantly more correct call types in the go2 than in the go1 phase (22.7±1.7% vs. 94.1±1.2%; p=2e-03, n=20, Wilcoxon signed-rank test; Fig. 3C and D). The finding that the monkey showed a higher response to the red go signal might be explained by the use of this signal during the preceding detection task and, therefore, that the animal was still in the state of generalizing to the blue cue as a go signal in this phase. The better performance during the go2 signal was in accordance with the finding that monkeys predominantly produced chirp vocalizations during the preceding detection task (Fig. 1D). As a result, the yet untrained monkey automatically exhibited a higher probability of uttering a correct rather than wrong call type (Fig. 1B) during the go2 signal and a low probability for uttering a correct call type and a high probability for uttering a wrong call type during the go1 signal (Fig. 3C and D), respectively. We then recorded 10 additional sessions after six months of training (final training phase). We observed that the monkey significantly increased vocal responses during go2 trials (p=3.71e-02, n=20, Wilcoxon signed-rank test) to a similar level to its performance during go1 trials (87.4±4.9% vs. 81.5±6.4%; p=0.16, n=20, Wilcoxon signed-rank test; Fig. 3B). Hit rates were still significantly lower during go1 than go2 trials (57.6±3.1% vs. 84.8±2.9%; p=2e-03, n=20, Wilcoxon signed-rank test; Fig. 3C). In comparison to the initial training phase, hit rates significantly increased for go1 trials (2e-03, n = 20, Wilcoxon signed-rank test), while they slightly yet significantly decreased for go2 trials (1.95e-02, n=20, Wilcoxon signed-rank test; Fig. 3C). Call rate significantly decreased between the initial and final training phase (p=5.8e-03, n=20, Wilcoxon rank sum test). Response latencies for correct go1 vocalizations were significantly shorter than for correct go2 vocalization in the initial training phase (median latencies: 646 ms vs. 966 ms; p=1.69e-21, n=652, Wilcoxon rank sum test; Fig. 3D). In the final training phase, median response latencies for correct go2 calls decreased, resulting in

similar response latencies between correct go1 and go2 calls (median latencies: 693 ms vs. 722 ms; p=0.2, n=332, Wilcoxon rank sum test; Fig. 3E).

Our findings demonstrate that marmoset monkeys are capable of volitionally initiating complex vocal-motor behavior in a highly controlled experimental design. In contrast to other studies reporting low performance for simple eye movement or lever pressing tasks (*13*), we show for the first time that marmoset monkeys can be trained to vocalize on command in response to arbitrary visual cues in a go/no-go detection task. Additionally, we report that marmosets are able to learn to switch between two distinct call types from trial to trial in response to different visual cues in a discrimination task. Earlier studies showed that rhesus monkeys are capable of vocal control and producing calls on command in a goal-directed way (*21*, *25*, *26*). From an evolutionary perspective, our data suggest that the origins of the ability to volitionally control vocal output is much older than previously thought and that the last common ancestor of Old and New World monkeys, which lived more than 35 million years ago (*1*), probably had the ability to volitionally control its vocal output.

Volitional control of vocal output is a crucial preadaptation for the evolution of human speech in the primate lineage (*27*, *28*). Recent neurophysiological studies in rhesus monkeys found similar activity in brain structures underlying volitional vocal output in monkeys and their homologous structures in the human brain that are crucial for speech production (*29*, *30*). Both structures exhibited similar neural activity related to vocalizations and speech signals, respectively, supporting a continuous evolution of vocal communication systems in the primate lineage ultimately giving rise to speech in humans (*27*, *28*). However, we are just starting to understand the underlying neural mechanisms responsible for the cognitive control of vocal production in primates. We present that marmoset monkeys can be trained to perform complex behavioral tasks, i.e., cognitive control of vocal behavior, in a controlled experimental environment, a prerequisite for being able to pinpoint underlying brain mechanisms. These findings, in

combination with the other recently established neurophysiological and genetical tools, make them a suitable primate model to study complex motor behavior in general and the evolutionary aspects of cognitive control underlying human speech in health and disease.

## Methods

### Experimental animals

We used four marmosets (*Callithrix jacchus*; two males, two females) housed at the University of Tübingen, Germany. Animals were usually kept in different sex pairs and were all born in captivity. Monkey H was hand-raised by an animal caretaker from the third postnatal day and reunited with its siblings after three months (for details see (*31*)). The facility room was maintained at approximately 26°C, 40–60% relative humidity, and a 12h:12h light-dark cycle. They had ad libitum access to water and were fed daily with standard commercial chow and a selection of fruit, vegetables, mealworms, and locusts. Marshmallows and special fruit (e.g., banana, grapes) were used to transfer the animals from their home cages to a transfer box. Experimental procedures were approved by the local authorities of Tübingen (Regierungspräsidium) and were in agreement with the guidelines of the European Community for the care of laboratory animals.

### Data acquisition

Stimulus presentation, behavioral monitoring, and reward presentation were synchronized and performed automatically using a custom-written program (OpenEX and Synapse, Tucker-Davis Technologies, USA) running on a workstation (WS-8 in combination with an RZ2 bioamp processor and RZ6D multi I/O processor, Tucker-Davis Technologies, USA) and a custom-written MATLAB program running on another PC, which was connected via an A/D interface card (PCIe 6321, National Instruments) with the workstation (Fig. 1A). A monitor screen connected to the desktop PC was positioned in front of the animal's head at a distance of 40 cm for visual stimulus presentation. Vocalizations were recorded using a microphone (MKH 8020 microphone with MZX 8000 preamplifier, Sennheiser, Germany in combination with a phantom power, PAN 48.2,

Palmer) positioned 10 cm in front of the monkey's head and connected to a multi I/O processor (RZ6D, Tucker-Davis Technologies, USA). Vocalizations were recorded using the same system at a sampling rate of 100 kHz. Vocal onset times were detected offline using software (Avisoft-SASLab Pro 5.2.13, Avisoft Bioacoustics) to ensure precise timing for data analysis. The monkey's behavior was constantly monitored using a USB video camera (Brio, Logitech) placed in front of the monkey.

Behavioral protocol

The monkeys were trained to sit in a primate chair in a soundproof chamber. In the first part of the study, we trained all monkeys to perform a visual go/nogo detection protocol. A trial began when the monkey initiated a "ready"-response by pushing down on a lever (see Fig. 1A). A visual cue, indicating the "nogo"-signal ("pre-cue"; white square, width: 14° of visual angle) appeared for a randomized time from 1 to 5 s for one monkey (monkey H) and 1 to 3 s for the other monkeys (monkey E, L, and P); vocal output had to be withheld during this period. Next, in 80% of trials the visual cue was changed to a colored "go"-signal (red square; width: 14° of visual angle) lasting for 3000 ms. During this time, the monkey had to emit a vocalization to receive a liquid reward (mixture of water, marshmallows, fruit, marmoset gum, and curd cheese) provided by a small metal syringe directly in front of the monkey's face. In 20% of trials, the cue remained unchanged for another 3000 ms ("catch"-trial). In this period, the monkey had to withhold call production. "Catch"-trials were not rewarded. Calls during "catch"-trials were defined as "false alarms". For monkey E and L, we played back audio recordings from the animal facility to maintain their motivational state during the session.

In the second part of the study, we trained one monkey (monkey H) to perform a visual discrimination protocol, where the animal had to produce two different types of vocalizations in response to distinct visual cues. As in during the visual detection task, the monkey initiated a trial by pushing down a lever and the "nogo"-signal appeared for a randomized time from 1 to 5 s (see

Fig. 1B). Next, the visual cue was changed to either a red or blue square. Both “go”-signals appeared pseudo-randomly with equal probability (p=0.5). The monkey was trained to utter brief “chirp”-vocalizations in response to the blue square and long “trill” calls or call combinations such as “chirp-trill” and “trill-phee” sequences in response to the red square. During the visual discrimination protocol, the monkey had to keep the the bar pressed throughout the pre-cue phase to indicate its alertness and bar releases aborted the trial. One session was recorded per individual per day.

Data analysis

Fifteen consecutive sessions per individual during the visual detection task and 10 consecutive sessions during the visual discrimination task were used in the data analysis. In accordance with the go/nogo detection paradigm, successful “go”-trials were defined as “hits”, unsuccessful “catch”-trials as “false alarms” in the visual detection paradigm. For the visual discrimination protocol, the utterance of the correct vocalization in response to a specific visual cue was defined as a “hit” a vocal response with the wrong call type as a “false alarm”. A recent study reported mean response latencies for a simple motor task, namely saccadic eye movements, of 200 ms in marmoset monkeys (*16*). Consequently, we counted vocalizations in the first 200 ms following precue onset as calls outside trials, calls in the first 200 ms following “go” and “catch” trial onset as precue calls, and in the first 200 ms following “go” and “catch” trial offset as “hit” and “false alarm” calls, respectively.

In the current study, we did not aim to classify calls within one call type into subtypes. We classified marmoset vocalizations into previously defined main groups (*32*–*34*). Calls were manually classified as chirp, trill, phee, peep, twitter, tsik, or ekk calls based on their spectro-temporal profile and auditory playback. The eight call types show a very defined and distinct profile and could be easily classified manually (*31*, *33*–*36*). Chirps are calls consisting of a short and descending FM sweep; trill calls are defined by sinusoidal-like FM throughout the call; phee is a

tone-like long call with F0 around 7-10kHz; peeps are short duration tone-like calls that have a sharply ascending or sharply descending FM; twitter is a short upward FM sweep; tsik is a broadband short call consisting of a linearly ascending FM sweep that merges directly into a sharply descending linear FM sweep, and ekk is a short call that is defined as one of the lowest-frequency marmoset calls. Other call types were rarely uttered and defined as “others”. In cases where animals produced call sequences during the vocal detection task, the first call uttered was taken into account for call classification. Call probability distributions were calculated using a moving average (bin width, 500 ms, step size, 1 ms) and smoothed using a Gaussian kernel (bin width, 100 ms; step size, 1 ms) for illustrative purposes only.

Data normalization

Probability distribution for hit, false alarm, and pre-cue call latencies calculated in the visual detection task were normalized with regard to the hit rate of every single recording session. Probability distributions in the visual discrimination task were normalized for both go-signals with regard to the absolute number of calls uttered within the respective go-trials (go1 and go2) of every single recording session.

Statistical analysis

Statistical analyses were performed using MATLAB (MathWorks, Natick, MA). We computed d′ sensitivity values by subtracting z-scores (normal deviates) of median “hit”-rates from z-scores of median “false alarm”-rates. Extreme values of “hit”-rates and “false alarm”-rates were corrected as performed previously (*37*). A one-way analysis of variance (Kruskal-Wallis test) was performed to test for significant differences in call response latency according to the duration of the precue delay. Wilcoxon sign rank tests with Bonferroni correction were calculated to test for significant differences in the vocal performance with respect to the two go signals (go1 and go2) and the two training phases (initial and final) during the visual discrimination task. We used a 3-Way ANOVA to test whether animals exhibited different call type ratios at different time points during sessions.

Pearson's correlations were performed to identify relationships between several parameters of vocal behavior. In all performed tests, significance was tested at an alpha = 0.05 level.

## References


1. C. T. Miller *et al.*, Marmosets: A Neuroscientific Model of Human Social Behavior. *Neuron*. **90**, 219–233 (2016).

2. V. Marx, Neurobiology: learning from marmosets. *Nat. Methods*. **13**, 911–916 (2016).

3. E. Sasaki *et al.*, Generation of transgenic non-human primates with germline transmission. *Nature*. **459**, 523–527 (2009).

4. J. I. Borjon, A. A. Ghazanfar, Convergent evolution of vocal cooperation without convergent evolution of brain size. *Brain. Behav. Evol.* **84**, 93–102 (2014).

5. J. M. Burkart, C. Finkenwirth, Marmosets as model species in neuroscience and evolutionary anthropology. *Neurosci. Res.* **93**, 8–19 (2015).

6. T. Pomberger, C. Risueno-Segovia, J. Löschner, S. R. Hage, Precise Motor Control Enables Rapid Flexibility in Vocal Behavior of Marmoset Monkeys. *Curr. Biol.* **28**, 788–794 (2018).

7. M. Mesulam *et al.*, Alzheimer and frontotemporal pathology in subsets of primary progressive aphasia. *Ann. Neurol.* **63**, 709–719 (2008).

8. S. Sapir, L. Ramig, C. Fox, Speech and swallowing disorders in Parkinson disease. *Curr. Opin. Otolaryngol. Head Neck Surg.* **16**, 205–210 (2008).

9. H. Zeng *et al.*, Local Homogeneity of Tonotopic Organization in the Primary Auditory Cortex of Marmosets. *Proc Natl Acad Sci U S A*. **in press** (2019), doi:10.1101/398677.

10. D. J. Schaeffer, K. M. Gilbert, J. S. Gati, R. S. Menon, S. Everling, Intrinsic Functional Boundaries ofLateral Frontal Cortex in the Common Marmoset Monkey. *J. Neurosci.* **39**, 1020–1029 (2019).

11. S. Roy, L. Zhao, X. Wang, Distinct Neural Activities in Premotor Cortex during Natural Vocal Behaviors in a New World Primate, the Common Marmoset (Callithrix jacchus). *J. Neurosci.* **36**, 12168–12179 (2016).

12. C.-C. Hung *et al.*, Functional Mapping of Face-Selective Regions in the Extrastriate Visual Cortex of the Marmoset. *J. Neurosci.* **35**, 1160–1172 (2015).

13. E. D. Remington, M. S. Osmanski, X. Wang, An operant conditioning method for studying auditory behaviors in marmoset monkeys. *PLoS One*. **7**, e47895 (2012).

14. X. Song, M. S. Osmanski, Y. Guo, X. Wang, Complex pitch perception mechanisms are shared by humans and a New World monkey. *Proc. Natl. Acad. Sci.* **113**, 781–786 (2016).

15. J. F. Mitchell, J. H. Reynolds, C. T. Miller, Active vision in marmosets: a model system for visual neuroscience. *J. Neurosci.* **34**, 1183–94 (2014).

16. K. Johnston, L. Ma, L. Schaeffer, S. Everling, Alpha-oscillations modulate preparatory activity in marmoset area 8Ad Alpha-oscillations modulate preparatory activity in marmoset area. *J. Neurosci.* **in press** (2019).

17. N. W. Prins *et al.*, Common marmoset (Callithrix jacchus) as a primate model for behavioral neuroscience studies. *J. Neurosci. Methods*. **284**, 35–46 (2017).

18. A. Takemoto *et al.*, Individual variability in visual discrimination and reversal learning performance in common marmosets. *Neurosci. Res.* **93**, 136–143 (2015).

19. J. W. Krakauer, A. A. Ghazanfar, A. Gomez-Marin, M. A. MacIver, D. Poeppel, Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*. **93**, 480–490 (2017).

20. J. A. Michaels, B. Dann, R. W. Intveld, H. Scherberger, Neural Dynamics of Variable Grasp-Movement Preparation in the Macaque Frontoparietal Network. *J. Neurosci.* **38**, 5759–5773 (2018).

21. S. R. Hage, A. Nieder, Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat. Commun.* **4**, 3409 (2013).

22. J. Wessberg *et al.*, Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*. **408**, 361–365 (2000).

23. K. Hammerschmidt, J. Fischer, in *Evolution of Communicative Flexibility* (2008), pp. 92–119.

24. D. A. Liao, Y. S. Zhang, L. X. Cai, A. A. Ghazanfar, Internal states and extrinsic factors both determine monkey vocal production. *Proc. Natl. Acad. Sci.* **115**, 201722426 (2018).

25. D. Sutton, C. Larson, E. M. Taylor, R. C. Lindeman, Vocalization in rhesus monkeys: Conditionability. **52**, 225–231 (1973).

26. P. G. Aitken, W. A. Wilson, Discriminative Vocal Evidence Conditioning for Volitional in Rhesus Control? *Brain Lang*. **8**, 227–240 (1979).

27. S. R. Hage, A. Nieder, Dual Neural Network Model for the Evolution of Speech and Language. *Trends Neurosci.* **39**, 813–829 (2016).

28. K. K. Loh, M. Petrides, W. D. Hopkins, E. Procyk, C. Amiez, Cognitive control of vocalizations in the primate ventrolateral-dorsomedial frontal (VLF-DMF) brain network. *Neurosci. Biobehav. Rev.* **82**, 32–44 (2017).

29. A. Flinker *et al.*, Redefining the role of Broca's area in speech. *Proc. Natl. Acad. Sci.* **112**, 2871–2875 (2015).

30. N. Gavrilov, S. Hage, A. Nieder, Functional specialization of the primate frontal lobe during cognitive control of vocalizations. *Cell Rep.*, 2393–2406 (2017).

31. Y. B. Gultekin, S. R. Hage, Limiting parental feedback disrupts vocal development in marmoset monkeys. *Nat. Commun.* **8**, 14046 (2017).

32. A. L. Pistorio, B. Vintch, X. Wang, Acoustic analysis of vocal development in a New World primate, the common marmoset (Callithrix jacchus). *J. Acoust. Soc. Am.* **120**, 1655 (2006).

33. B. M. Bezerra, A. Souto, Structure and Usage of the Vocal Repertoire of Callithrix jacchus. *Int. J. Primatol.* **29**, 671–701 (2008).

34. J. A. Agamaite, C.-J. Chang, M. S. Osmanski, X. Wang, A quantitative acoustic analysis of the vocal repertoire of the common marmoset ( Callithrix jacchus ). *J. Acoust. Soc. Am.* **138**, 2906–2928 (2015).

35. D. Y. Takahashi *et al.*, The developmental dynamics of marmoset monkey vocal production. *Science.* **349**, 734–738 (2015).

36. Y. B. Gultekin, S. R. Hage, Limiting parental interaction during vocal development affects acoustic call structure in marmoset monkeys. *Sci. Adv.* **4**, eaar4012 (2018).

37. M. J. Hautus, Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behav. Res. Methods, Instruments, Comput.* **27**, 46–51 (1995).

## Acknowledgments

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 frame-work of the Excellence Initiative EXC 307) and the Deutsche Forschungsgemeinschaft Grant HA5400/3-1.

## Author contributions

S.R.H. conceived the study and designed the experiments; T.P., C.R.-S., Y.G., and D.D. conducted the visual detection experiments; T.P. conducted the visual discrimination experiment; S.R.H., T.P., C.R.-S., and Y.G. performed data analyses; S.R.H. created the visualizations; all authors interpreted the data and wrote the manuscript; S.R.H. provided the animals, acquired funding, and supervised the project.

## Competing interests

Authors declare no competing interests.

## Data availability

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.

## Figures

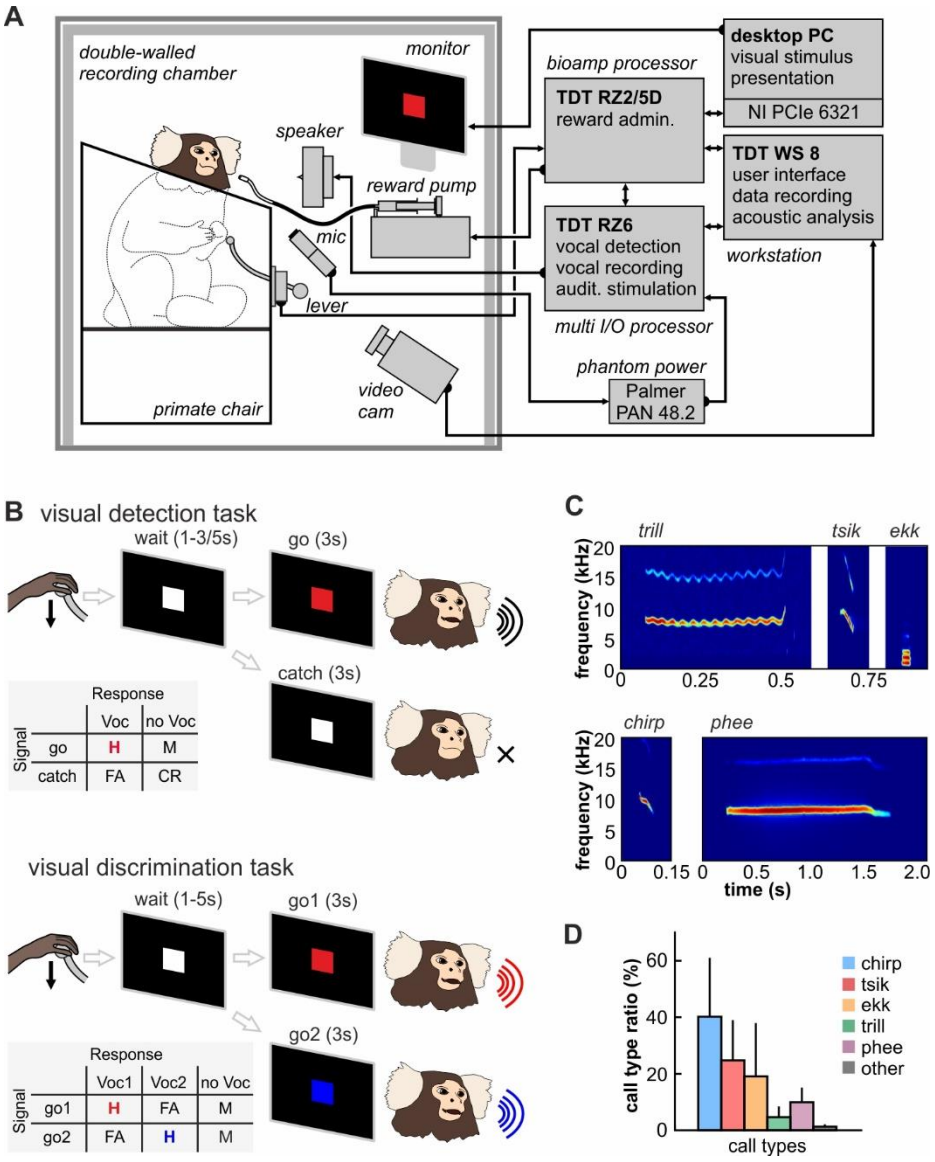


**Fig. 1. Training marmoset monkeys to call on command. A** Experimental setup. Animals were trained in a double-walled soundproof recording chamber. The schematic block diagram depicts the system used for stimulus presentation, behavioral monitoring, and reward presentation. **B** Training paradigms. All monkeys were trained in a go/no-go protocol to vocalize whenever a red cue appeared (detection task; upper depiction). One monkey was trained in a successive training period to utter distinct vocalizations in response to red and blue visual cues, respectively (discrimination task; lower depiction). H = hit; M = miss; FA = false alarm; CR = correct rejection. **C** Spectrograms of representative vocalizations uttered by the experimental animals. **D** Mean call type distribution of the four monkeys in visual detection sessions.

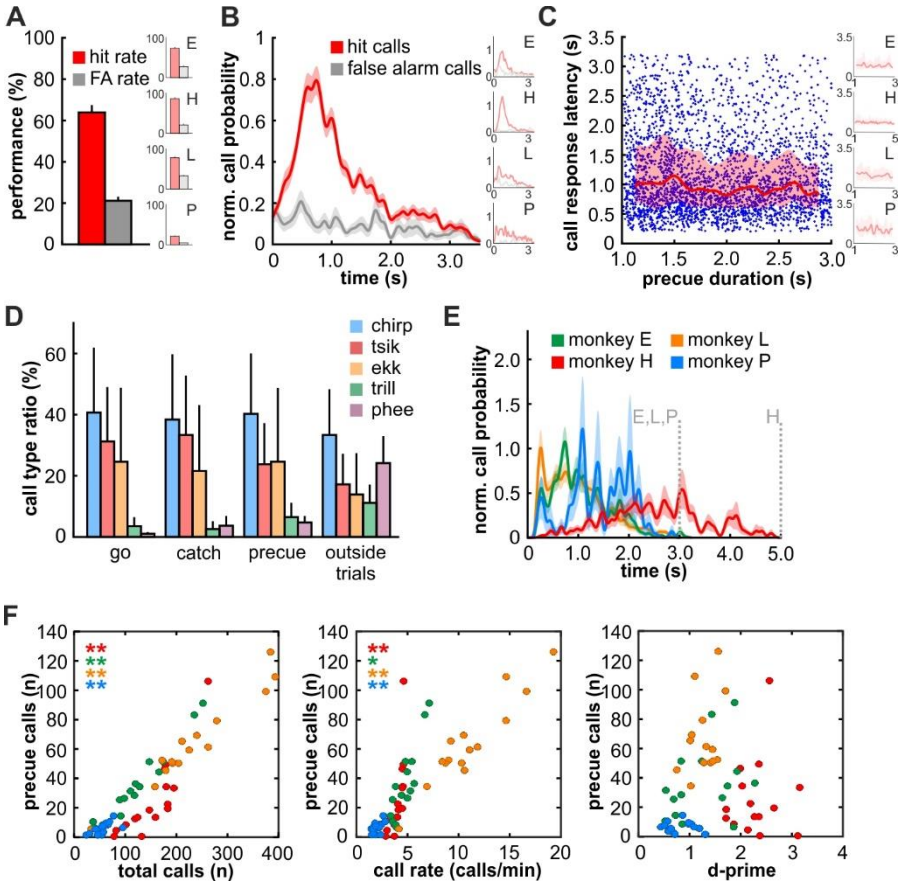


**Fig. 2. Marmosets call on command during visual detection task. A** Mean distribution of hit and false alarm rates of 15 sessions in four animals exhibit significantly higher call probabilities during "go" trials (hits) than during catch trials (false alarms). The main plot shows the group average; small plots show the mean of the four individual animals. Whiskers indicate standard error (SE). **B** Call probabilities during "go" trials and "catch" trials. The main plot shows the group average; small plots show the mean of the four individual animals. Data were normalized for 15 sessions per animals. Shaded areas indicate SE. **C** Correlation between call response latencies after go-cue onset and the preceding waiting period (cue delay). The main plot shows the individual response latency with the corresponding cue delay for all calls uttered by the four animals in the 15 sessions (blue dots). The red line indicates the median and the shaded area the 1st to 3rd quartile of the response latencies as a function of the pre-cue duration (bin width: 100 ms, step size: 1 ms) for the population (main plot) and each individual (small plots). **D** Mean call type distribution for four monkeys during three time periods for self-initiated trials (precue, go, and catch phase) and outside trials. **E** Mean call probabilities during the pre-cue phase for the four animals. Note that monkeys E, L, and P were trained with 1–3-s and monkey H with 1–5-s pre-cue latency, respectively (grey, dashed vertical line). Shaded areas indicate SE. **F** Correlations between precue calls and total number of calls, call rate (as a measure for arousal), and d′ values for each session for each animal. Subjects are colored according to (D). *p<0.01, **p<0.001Pearson's correlation.

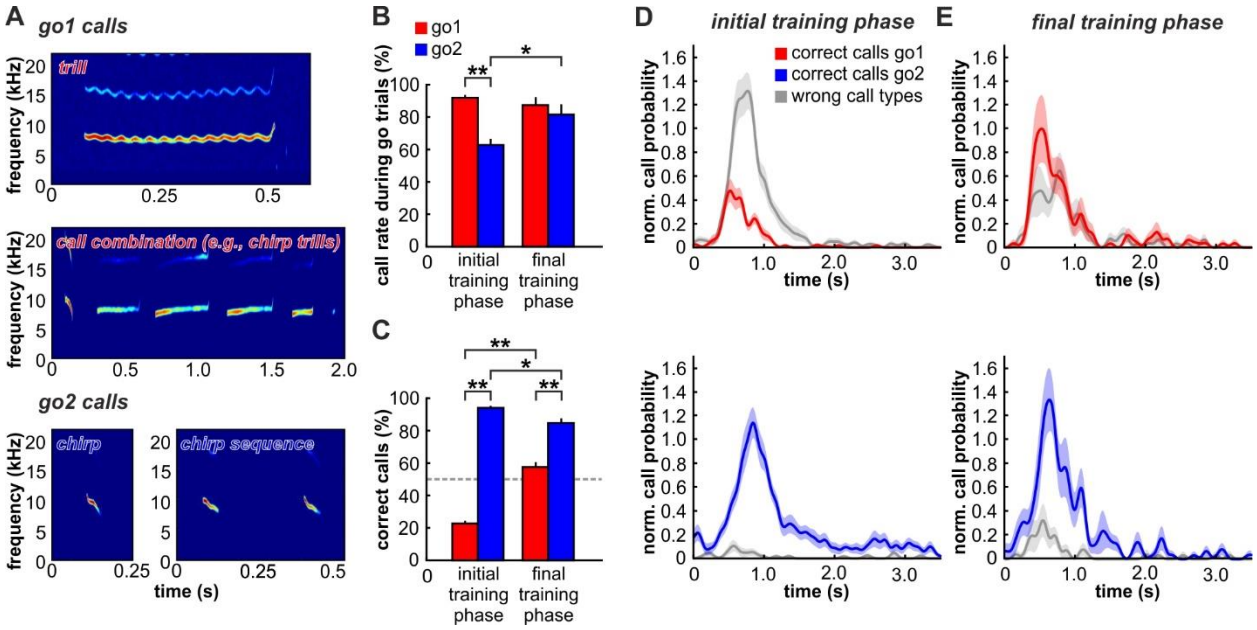


**Fig. 3. Volitional control of call type in a discrimination task shown in monkey H.** **A** Spectrograms of vocalization types. Trills and call combinations had to be uttered in response to the red visual cue (go1); chirps and chirp sequences had to be uttered in response to the blue visual cue (go2). **B** Distribution of go rates for 10 sessions in the initial and final training phase of the visual discrimination task. **C** Distribution of correct call types uttered during go1 and go2 cues in the initial and final training phase. (A and B) *p<0.05, **p<0.01, Wilcoxon signed-rank test. **D and E** Call probabilities during “go1” and “go2” trials in the initial (D) and final phase (E) of visual discrimination training (E).

**Table 1.** Vocal performance of monkeys during visual detection and discrimination tasks.

| **Detection task (4 monkeys)** | | | | | | |
|---|---|---|---|---|---|---|
| *Monkey ID* | *Calls in total* | *Calls per session* | *Call rate (calls/min)* | *Hit calls per session* | *Catch calls per session* | *Pre-cue calls per session* |
| Monkey E | 1,845 | 123 ± 16 | 4.8 ± 0.3 | 53 ± 6 | 5 ± 1 | 36 ± 6 |
| Monkey H | 2,286 | 152 ± 13 | 4.0 ± 0.1 | 118 ± 7 | 7 ± 1 | 25 ± 7 |
| Monkey L | 3,483 | 232 ± 25 | 11.1 ± 1.0 | 70 ± 4 | 7 ± 1 | 63 ± 8 |
| Monkey P | 822 | 55 ± 5 | 2.0 ± 0.1 | 18 ± 1 | 1 ± 0 | 6 ± 1 |
| | | | | | | |
| **Discrimination task (1 monkey)** | | | | | | |
| *Time of recording* | *Calls in total* | *Calls per session* | *Call rate (calls/min)* | *go1 calls per session* | *go2 calls per session* | |
| Pre-training | 1530 | 153 ± 8 | 4.5 ± 0.1 | 66 ± 5 | 54 ± 2 | |
| Post-training | 653 | 65 ± 4 | 3.6 ± 0.2 | 24 ± 1 | 23 ± 1 | |

Journal of Neuroscience Methods 311 (2019) 186–192

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

# Semi-chronic laminar recordings in the brainstem of behaving marmoset monkeys

Thomas Pomberger[a,b], Steffen R. Hage[a,*]

[a] Neurobiology of Vocal Communication, Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Str. 25, 72076 Tübingen, Germany
[b] Graduate School of Neural & Behavioural Sciences – International Max Planck Research School, University of Tübingen, Österberg-Str. 3, 72074 Tübingen, Germany

ARTICLE INFO

*Keywords:*
*Callithrix jacchus*
Semi-chronic recordings
Implanted electrodes
Laminar multi-site microprobes
Deep brain recordings
Single-unit recordings

ABSTRACT

*Background:* Chronic recordings with multi-electrode arrays are widely used to study neural networks underlying complex primate behaviors. Most of these systems are designed for studying neural activity in the cortical hemispheres resulting in a lack of devices being capable of simultaneously recording from ensembles of neurons in deep brainstem structures. However, to fully understand complex behavior, it is fundamental to also decipher the intrinsic mechanisms of the underlying motor pattern generating circuits in the brainstem.

*New method:* We report a light-weight system that simultaneously measures single-unit activity from a large number of recording sites in the brainstem of marmoset monkeys. It includes a base chamber fixed to the animal's skull and a removable upper chamber that can be semi-chronically mounted to the base chamber to flexibly position an embedded micro-drive containing a 32-channel laminar probe to record from various positions within the brainstem for several weeks.

*Results:* The current system is capable of simultaneously recording stable single-unit activity from a large number of recording sites in the brainstem of vocalizing marmoset monkeys.

*Comparison with existing methods:* To the best of our knowledge, chronic systems to record from deep brainstem structures with multi-site laminar probes in awake, behaving monkeys do not yet exist.

*Conclusions:* The semi-chronic implantation of laminar electrodes into the brainstem of behaving marmoset monkeys opens new research possibilities in fully understanding the neural mechanisms underlying complex behaviors in marmoset monkeys.

## 1. Introduction

The marmoset monkey (*Callithrix jacchus*) has recently garnered considerable interest as a neuroscientific model organism (Miller et al., 2016). The renewed focus on this already established animal model species (Eliades and Wang, 2008a, 2003; Fritsches and Rosa, 1996; Roberts and Wallis, 2000) has primarily been driven by the prospect of developing primate transgenic lines (Sasaki et al., 2009), but also by the potential for transferring a number of rodent neuroscientific techniques to a small primate model system (Miller et al., 2016) that can be used in controlled experimental designs (Song et al., 2016). Furthermore, marmoset monkeys are social and highly vocal New World monkeys making them an ideal model system to investigate cognition and social communicative behavior (Borjon and Ghazanfar, 2014; Burkart and Finkenwirth, 2015).

Recently, several chronic multi-electrode systems have been developed to record from several cortical brain regions, such as the premotor and auditory cortex, in behaving marmoset monkeys (Eliades and Wang, 2008b; Roy and Wang, 2012). These chronic systems have numerous advantages over acute recordings, such as increased yield and improved recording stability, as well as the ability to simultaneously record from a large number of neurons during complex behavior (Eliades and Wang, 2008a). In contrast, there have only been a few approaches to record from brainstem structures in awake and behaving monkeys, and mammals in general (Jürgens and Hage, 2006) – and these systems only measured neuronal activity at most from two positions. However, in light of recent work indicating that complex microcircuits are involved in motor behaviors such as respiration (Anderson et al., 2016; Del Negro et al., 2018; Harris et al., 2017) and vocalization (Hage and Jurgens, 2006; Hage and Nieder, 2016; Jürgens, 2002), as well as in audio-vocal integration mechanisms (Hage et al., 2006; Luo et al., 2018), it is necessary to simultaneously record from an ensemble of neurons within such circuits to fully understand the intrinsic mechanisms of these motor pattern generating brainstem

* Corresponding author.
*E-mail address:* steffen.hage@uni-tuebingen.de (S.R. Hage).

https://doi.org/10.1016/j.jneumeth.2018.10.026
Received 22 August 2018; Received in revised form 17 October 2018; Accepted 17 October 2018
Available online 21 October 2018

structures.

Here, we report a light-weight system to simultaneously measure single-unit activity from a large number of recording sites in the brainstem of marmoset monkeys. It includes (1) a base chamber fixed to the animal's skull and (2) a removable upper chamber that can be semi-chronically mounted to the base chamber to flexibly position an embedded micro-drive containing a 32-channel laminar probe to record from various positions within the brainstem for several weeks. The upper chamber can be removed and repositioned to record from new brainstem positions. This newly developed semi-chronic recording device combines the advantages of chronic recording systems, including stable recordings and short preparation times, with those of acute approaches, such as the flexibility in the choice of recording sites.

## 2. Material and methods

### 2.1. Animals

The system has been designed for the use in common marmoset monkeys (*Callithrix jacchus*). The animals used in this study are housed at the University of Tübingen and were all born in captivity. The facility room was maintained at approximately 26 °C, 40–60% relative humidity, and with a 12 h:12 h light-dark cycle. The marmosets had ad libitum access to water and were fed daily with standard commercial chow and a selection of fruit, vegetables, mealworms, and locusts. Marshmallows and special fruit (e.g., banana, grapes) were used as positive reinforcement. Experimental procedures were approved by the local authorities of Tübingen (Regierungspräsidium) and are in agreement with the guidelines of the European Community for the care of laboratory animals.

### 2.2. Animal preparation and laminar probe implantation

All surgical procedures were performed under aseptic conditions and general endotracheal anesthesia and were in accordance with the guidelines for animal experimentation and authorized by the Regierungspräsidium Tübingen. For the attachment of the base chamber (Fig. 1A–C), the animal was held using a stereotaxic apparatus (Kopf Instruments). The skin and underlying muscles were excised from the top of the skull. Four small elongated trepanations were drilled (Piezosurgery touch, Mectron, Germany) for the placement of the titanium anchor screws (Fig. 1A and D) around the desired position of the light-weight titanium base chamber (material: Ti-6Al-4V, weight: 1.4 g) as shown in Fig. 1E. The titanium anchor screws (M 1.6) were inserted into the holes, with the flattened head pushed into a position between the skull and dura, turned by 90° to ensure fixation, and fixed with nuts. The remaining openings were closed with bone wax and the screws and skull below the desired chamber position were covered with a thin layer of dental acrylic (Superbond, Sun Medical Co. Ltd., Japan). Next, the base chamber was positioned using the aid of a robotic stereotaxic micromanipulator (Neurostar, Dettingen, Germany) with the center of the chamber being stereotaxically positioned above the center of the area of interest in the brainstem (here: AP = 0, ML = 0 according to the stereotaxic brain atlas coordinates (Paxinos et al., 2012)). The stereotaxic placement of the base chamber in combination with the ability to precisely position the microdrive within the base chamber then enables stereotaxic positioning of the laminar probe in the brainstem. The chamber was then fixed to the skull with self-adhesive resin cement (RelyX Unicem, 3M Germany) covering the anchor screws and outer side of the chamber (Fig. 1B and E). Between chamber fixation and subsequent probe implantation, the base chamber was covered with a titanium protective cap (Fig. 1F and G) that was screwed to the chamber.

The light-weight semi-chronic laminar probe device ("upper chamber", Fig. 1A and B; Neuronexus, USA; material: 3d printed plastic (VisiJet M3 Crystal), weight: 5.3 g) was implanted a few weeks after chamber fixation. In a stereotaxic surgery, a small trepanation was performed within the base chamber at a position just above the stereotaxic coordinates at which the laminar probe was to be implanted (Fig. 1B), while the dura was maintained intact. For our experiments, we decided to use 32-channel laminar probes (V1 × 32-Edge, Neuronexus, USA; impedance: ~1 MΩ, probe dimensions at base: 50 × 175 µm) to be able to simultaneously record from several positions. The upper chamber was then screwed on to the base chamber and the laminar probe was lowered into the brain via the micro-drive until the tip of the probe reached the upper rim of the brainstem. Hereby, the laminar probes could be precisely positioned by turning the thread of the micro-drive in full or partial turns (one full turn lowers the electrode by 250 µm). At the end of the surgery the lower part of the chamber was filled with artificial dural sealant (Dura-Gel, Cambridge Neurotech, UK) to seal the trepanation and the chamber was closed with a removable protective cap (material: Ti-6Al-4V, weight: 2.6 g; Fig. 1A, B, F and G). Following surgery, animals underwent analgesic and antibiotic treatment for three to five days, which were given orally via fruit or marshmallow pieces, to allow optimal recovery. Additionally, and in accordance with a recent study, the behavior of each animal was monitored daily for 20 min for 5 days post-surgery using ethograms (Hage et al., 2014).

Special effort was put into the development of the upper chamber to allow maximum flexibility in laminar probe placement, i.e., positioning of the micro-drive within the chamber. We therefore designed the upper chamber in such a way that the micro-drive could be flexibly positioned within the upper chamber to position the laminar probe across almost the entire range from the center to the outer rim of the base chamber. This was accomplished by mounting the micro-drive to the upper chamber at three different positions, with the additional possibility to slightly shift the micro-drive back and forth via oblong screw-holes (Fig. 2A). In addition, the upper chamber could be mounted in twelve possible positions to the base chamber with three screws (Fig. 2B). The combination of flexibly positioning the micro-drive and, therefore, the laminar probe within the upper chamber and mounting the latter on the base chamber in multiple positions enables neural recordings from positions encompassing the entire lower brainstem and most of the upper brainstem with a single chamber implantation (Fig. 2B and C).

### 2.3. Neural and vocal recording setup

Prior to implantation, monkeys were trained to sit in a primate chair in a soundproof chamber. Vocalizations were recorded via a microphone (MKH 8020 microphone with MZX 8000 preamplifier, Sennheiser, Germany; phantom power for microphone by PAN 48.2, Palmer, Germany) positioned 10 cm in front of the monkey's head (Fig. 3). Each time the monkey uttered a vocalization, regardless of call type, they received a liquid reward (mixture of water, marshmallow, fruit, *Gummi arabicum*, and curd cheese) provided by a small metal syrinx directly in front of the monkey's face that was fed by a computer-controlled syringe pump (Pump 11 Elite, Harvard Apparatus, USA). With this approach, we found that monkeys produced a high number of calls. Vocal detection and reward presentation were synchronized and performed automatically with a custom-written program (OpenEX and Synapse, Tucker-Davis Technologies, USA) running on a workstation (WS-8 in combination with an RZ2 bioamp processor and RZ6D multi I/O processor, Tucker-Davis Technologies, USA). Vocalizations were recorded using the same system with a sampling rate of 100 kHz (Fig. 3). Additionally, a loud speaker and a monitor screen connected to a desktop PC were positioned in front of the animal's head for potential visual and acoustic stimulus presentations in later training stages.

At the beginning of a daily session monkeys were transferred from the animal facility to the experimental setup. For this, animals were trained with positive reinforcement to directly enter the primate chair when attached to their home cage. In the soundproof chamber, the monkey was placed in front of the microphone and a metal syrinx was

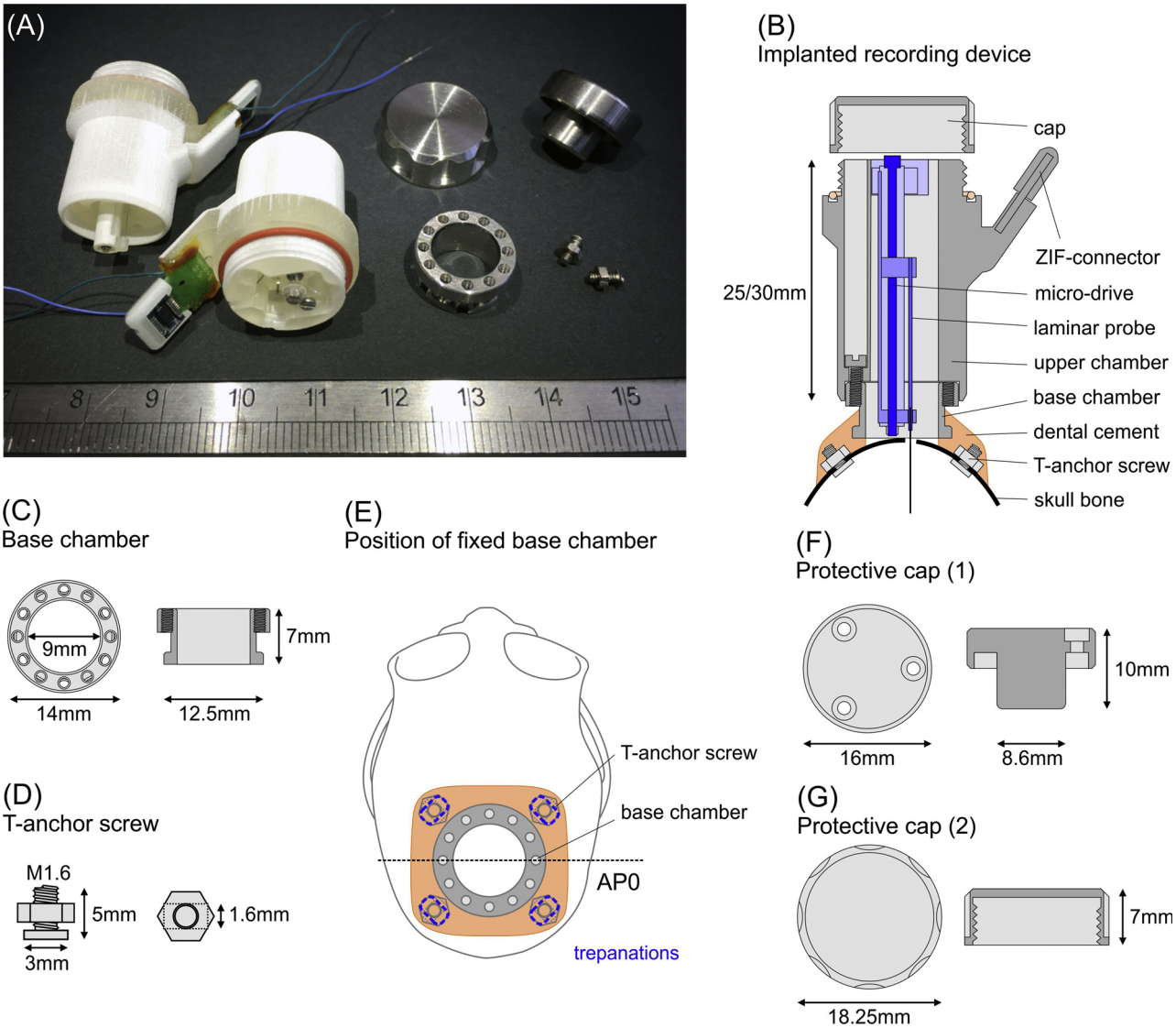


**Fig. 1.** Design of semi-chronic recording device. (A) Photograph of all assembly parts (scale in cm). (B) Illustration of the implanted recording device. The lower base chamber is fixed to the skull with the upper chamber (including the micro-drive with the laminar probe) attached to it. (C) Schematics of the base chamber that is fixed to the animal's head. (D) Schematics of the T-anchor screws that are used for firm hold of the implant. (E) Illustration of the chamber's position on the head of the animal. (F) Schematics of the protective cap, which is anchored to the base chamber during periods when no upper chamber is mounted. (G) Schematics of the protective cap that is screwed on the upper chamber between experiment sessions.

placed in front of the animal's head. The laminar probe was lowered to a new recording position. The probe was connected to a neural preamplifier (PZ2-32, Tucker-Davis Technologies, USA) via a motorized commutator (ACO32, Tucker-Davis Technologies, USA; Fig. 3) that was connected to the neural recording system (RZ2, Tucker-Davis Technologies, USA) and synchronized with the vocal recording system (RZ6D, Tucker-Davis Technologies, USA). With this approach animals were able to freely move their head during recording sessions. Sessions lasted approximately 20 min, i.e., the period in which the animals vocalized with high call rates. Monkeys were monitored via a high-resolution webcam (Brio, Logitech, Switzerland) during the entire session. The video signal was also stored on the workstation in synchronization with the vocal and neural data (Fig. 3).

### 2.4. Neural and acoustic data analyses

Signal acquisition, amplification, and filtering were performed using the Tucker-Davis Technologies system. Spike sorting was performed using standard software (Offline Sorter, Plexon, USA). Data analysis was accomplished using MATLAB (MathWorks, Natick, MA).

Call on- and offsets were manually flagged offline using standard software (SASLab Pro version 5.2, Avisoft Bioacoustics, Germany). Call duration was calculated as the difference between the beginning and end of the vocalization. Call types were manually classified as reported earlier (Gultekin and Hage, 2018, 2017) and in accordance to the previous literature (Agamaite et al., 2015; Bezerra and Souto, 2008; Takahashi et al., 2015).

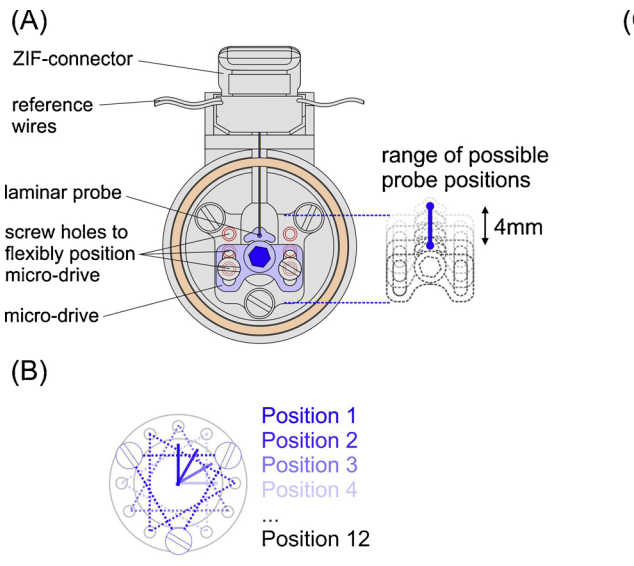


**Fig. 2.** Positioning possibilities of micro-drive within the upper recording chamber. (A) Top view on the upper chamber depicting the ability to flexibly position the micro-drive with a range of 4 mm. (B) Possible positions of upper chamber on the fixed base chamber further increases the range of potential electrode positions. (C) A combination of the flexible positioning of the micro-drive within the upper chamber and the twelve possible positions of the upper chamber on the fixed base chamber enables semi-chronic neural recordings with laminar probes. This method spans most of the latero-lateral dimension of the marmoset brainstem with the possibility of dense coverage in the rostro-caudal dimension. Solid and sketched lines depict possible recording sites with respect to the orientation of the base chamber. *ROI* region of interest.

In addition to their response to self-generated vocalizations, all recorded neurons were also tested with acoustic stimuli to determine their auditory response properties and distinguish vocal-motor, purely auditory, and audio-vocal neurons. Auditory stimuli included white noise and representative samples of seven marmoset vocalization types (phee, trill, chirp, tsik, ek, twitter, chatter). A neuron was determined to be auditory responsive if it responded to at least one of the above stimuli.

## 3. Results

The semi-chronic recording system reported here has been tested in three adult marmoset monkeys (one female: 5 years old, weighing 520 g; and two males: 2.5 and 6 years old, weighing 360 g and 420 g at day of base chamber implantation). The base chambers have been implanted for several months in all three monkeys (monkey S: 502 days, monkey H: 391 days, monkey F: 175 days). Neural activity has been recorded while the animals were vocalizing and chair-restrained in a soundproof chamber (Fig. 3). In all three monkeys, 32-channel laminar probes were semi-chronically implanted enabling stable neural recordings for an extended time period of several months (monkey S: 82 days; monkey H: 90 days, monkey F: 15 days (experiment still in progress)).

### 3.1. Recording quality

One of the significant advantages of our newly developed system is the capability to semi-chronically record neural activity simultaneously from multiple linear sites in the brainstem of behaving marmoset monkeys. In our approach, we use laminar probes with 32 contact sites, 60 µm apart, allowing neuronal recordings across an almost 2-mm linear range. Fig. 4A and B shows exemplar recordings with the laminar probes in the upper brainstem (Fig. 4C) highlighting several neighboring channels showing stable neuronal activity with well-isolated single unit activity. Due to the close proximity of recording sites, it is possible to record the same neuron at more than one recording site. The

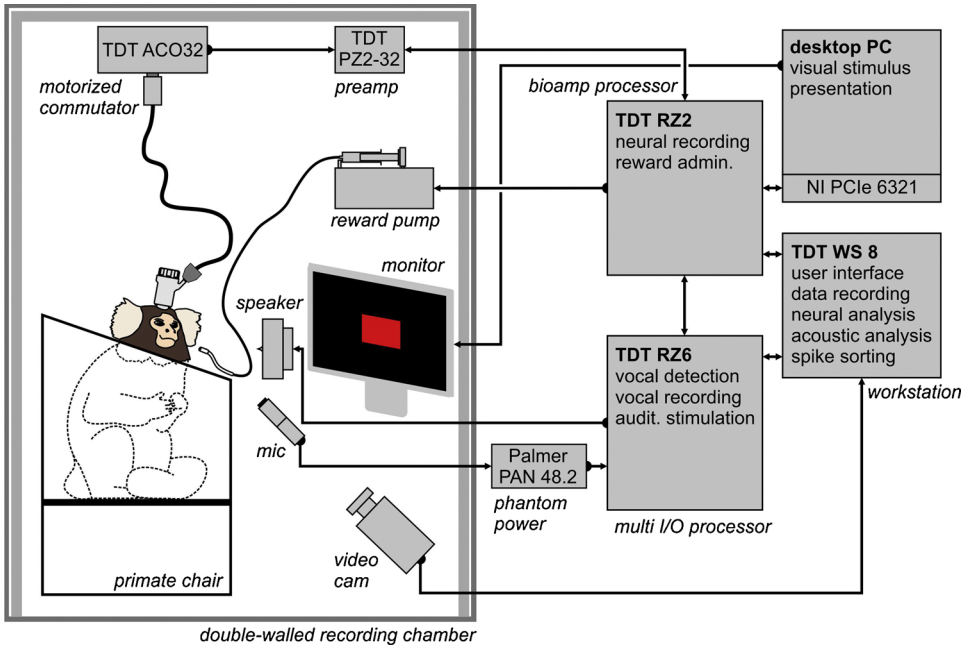


**Fig. 3.** Setup for recording multi-channel single-unit activity in behaving marmoset monkeys. Animals are trained to sit in a monkey chair in a double-walled soundproof recording chamber. The schematic block diagram depicts the system used for neural and acoustic recording as well as acoustic playback including a microphone, speaker, computer monitor, video camera, reward pump, motorized commutator, and neural preamplifier.

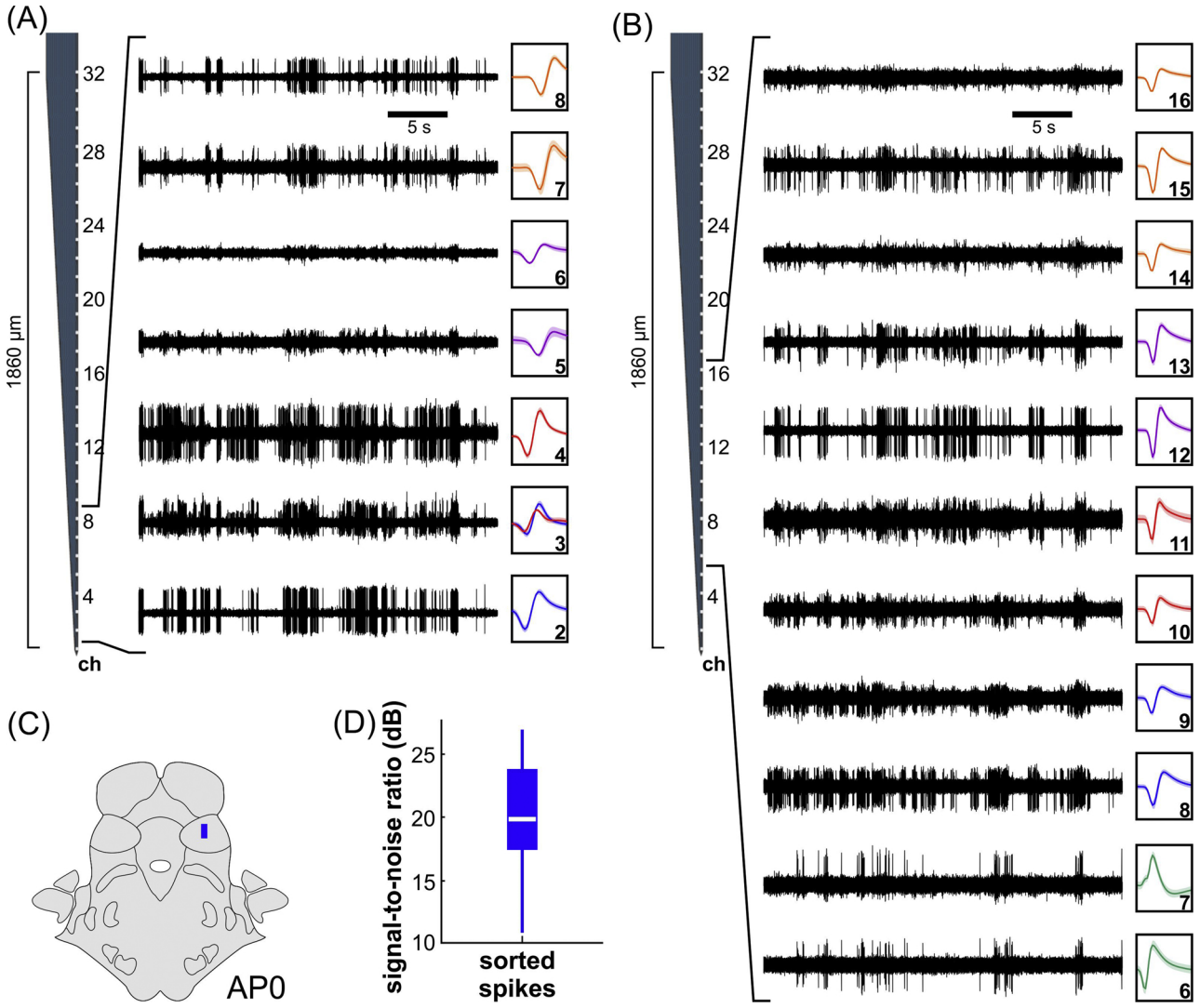


**Fig. 4.** Simultaneous recording of neural activity with laminar probes in the marmoset brainstem. (A) Exemplar simultaneous recording of seven recording sites. (B) Exemplar recording of eleven recording sites. Different colors of waveforms indicate the individual neurons recorded from neighboring recording sites. Identical neurons that have been recorded at two recording sites are labeled with the same color in (A) and (B). Please note that recordings shown in (B) were made from the same recording tract as (A) with the probe tip 250 µm lower in (B) than in (A). Therefore, recording sites have shifted between four to five sites from (B) to (A). (C) Calculated position of recording sites shown in (A) and (B). (D) Mean signal-to-noise ratio of all sorted waveforms (n = 19) shown in (A) and (B). Upper and lower margins of boxes: first and third quartiles, respectively; end of whiskers above and below boxes: 0.4% and 99.6% quantile, respectively.

signal-to-noise ratio (SNR) of the spikes with respect to the root mean square (RMS) of the background voltage level was calculated for all recorded neurons with a median value of 19.9 dB (Fig. 4A and B). This enables proper spike sorting as indicated by the small standard deviations shown for all sorted spike waveforms (Fig. 4A and B).

### 3.2. Recording stability

To quantify the stability of the neural recording, we evaluated the change in the SNR of the spikes with respect to the RMS of the background voltage level over the period of an entire recording session. Fig. 5A shows the raw trace of an exemplar recording in the upper brainstem of a vocalizing marmoset monkey restrained in a monkey chair with a freely moving head. Fig. 5B shows the development of the spike form over the session, Fig. 5C depicts the change in SNR over time, and Fig. 5D the calculated recording position of the neuron shown in Fig. 5A. No noticeable changes were observed in the raw data trace as well as the spike waveform during the recording session. Consequently, the SNR of the spikes varies only in a range of ± 0.3 dB (STD), reflecting small SNR variations throughout the recording session.

### 3.3. Measurement of event-related neural activity

In our lab, we use the newly developed semi-chronic recording system to record from multiple sites in the brainstem of vocalizing marmoset monkeys. Fig. 6A–C shows an exemplar recording of two neurons at the same recording site in the ventrolateral pontine brainstem. Fig. 6A shows an example of a neuron exhibiting vocalization-correlated activity. The neuron shows an increased firing rate just before the onset of chirp vocalizations that decreases during call production. During chirp-trill call combinations the neuron shows a similar activity with an increase in activity prior to the onset of the call sequence and inhibition during every single chirp and trill call following in the sequence. At the same contact of the laminar probe a second neuron was isolated showing a decrease in activity during the playback of phee vocalizations (Fig. 6B).

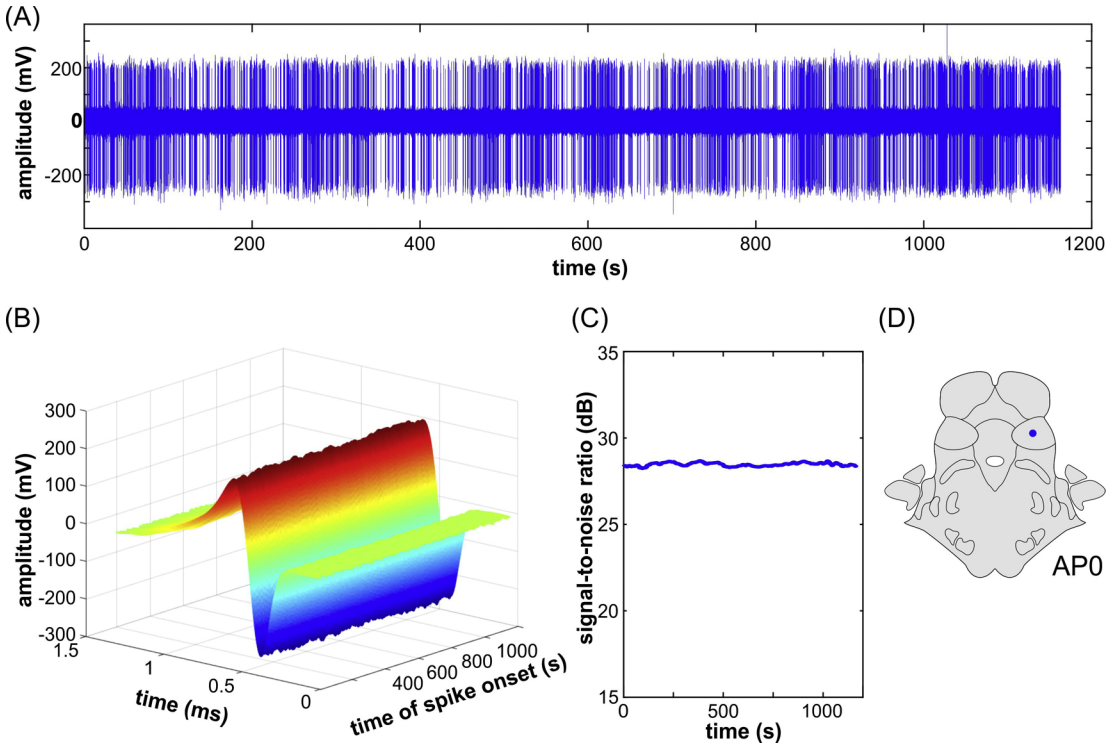


**Fig. 5.** Neural recording stability. (A) Raw neural data trace. (B) Spike waveforms from the session shown in (A). (C) Spike signal-to-noise ratio as a function of time. (D) Calculated position of the recording site shown in (A).

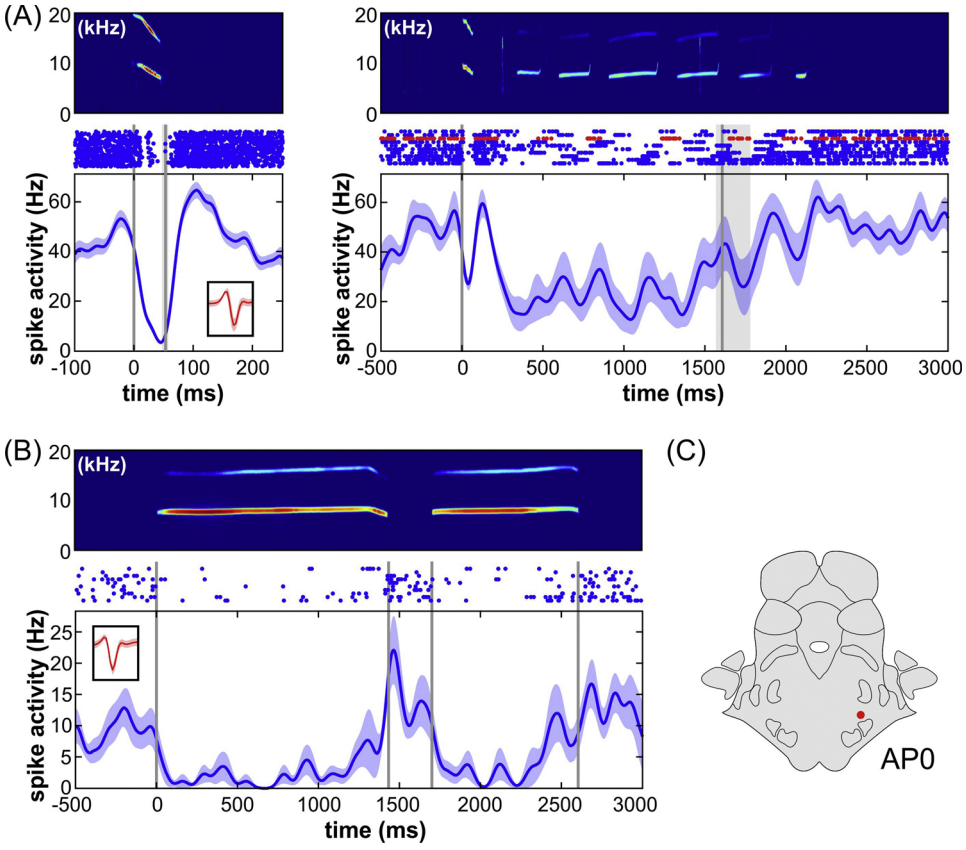


**Fig. 6.** Vocalization-correlated activity of single neurons in the lower marmoset brainstem. (A) Exemplar neuron showing a significant increase in neural activity prior to and a significant decrease in activity during the production of self-initiated chirp vocalizations (left; n = 100 calls) and chirp-trill sequences (right; n = 10 sequences). Please note that sequences were produced with a different number of trill calls and different inter syllable interval durations. The neural activity, which was directly correlated to the exemplar chirp-trill sequence shown in the upper right spectrogram is highlighted in red in the raster plot on the right. (B) Exemplar neuron showing a significant decrease in response to double phee playbacks (n = 10 playbacks). Upper panels in (A) and (B) show spectrograms of the produced vocalizations and playbacks, respectively. Middle panels show raster plots, lower panels represent the corresponding mean spike density histograms ( ± standard error) averaged and smoothed trial-wise with window sizes of 25 ms (for left panel in (A)) and 100 ms (for right panel in (A) and (B)). The vertical dark gray lines indicate the onsets and offsets of the produced calls in (A) and playback in (B). The vertical dark gray bars at the end of the calls in (A) indicate the median call/sequence duration and the upper and lower borders of the light gray bars the first and third quartile of the call/sequence distributions, respectively. Insets in (A) and (B) show the mean waveform of the sorted units ( ± standard deviation). (C) Calculated position of the recorded neurons shown in (A) and (B).

## 4. Discussion

We developed a semi-chronic recording system with laminar multi-electrodes to record from deep subcortical brain structures of vocalizing marmoset monkeys in a controlled experimental design. The system enables recording from various positions within an implanted titanium chamber using laminar probes attached to a micro-drive. The recording system is therefore capable of flexibly recording from large fractions of upper and lower brainstem regions. This device will help decipher brainstem-based neural networks underlying complex social behavior in marmoset monkeys. Recently, we showed that acoustic perturbation rapidly interrupts ongoing vocal behavior (Pomberger et al., 2018). With the developed system we are able to elucidate if and how vocal motor brainstem circuits are involved in such audio-vocal integration mechanisms. Furthermore, future research is now able to investigate ensembles of neurons embedded in complex neural microcircuits underlying other motor behaviors such as respiration (Anderson et al., 2016; Del Negro et al., 2018; Harris et al., 2017) that are predominantly generated by brainstem-based neural networks.

Up to now, only a few systems with chronically implanted multi-electrode arrays have been developed for the use in small primates (Eliades and Wang, 2008b; Roy and Wang, 2012). These systems allow stable recordings from up to 16 electrodes in the cortical hemispheres from behaving marmoset monkeys. However, chronic systems to record from deep brainstem structures in awake, behaving monkeys are virtually absent. To the knowledge of the authors there is only a single study that performed recordings with chronically implanted electrodes in the lower brainstem of behaving squirrel monkeys (Jürgens and Hage, 2006). In this study they were able to simultaneously record from up to two electrodes at a time which were mounted to a micro-drive and the electrodes could be reversibly implanted with the aid of a grid that was attached to the head of the animals. Inspired by the micro-drive design of this study, we used a similar approach in our current study. For more variability in electrode positioning, however, we chose to use a chamber in which the electrodes could be flexibly positioned and implanted.

The developed system combines a number of important features. First, it is lightweight, which is essential because of the small size of the marmoset monkey. Furthermore, with a height of up to 30 mm, it does not unbalance the monkey's head enabling the animal to freely move around in its home cage with no limitations in between recording sessions. Second, it can perform stable recordings simultaneously from 32 channels via laminar probes enabling dense recording to disentangle the potential intrinsic properties of microcircuits. Third, the probes can be flexibly positioned within the implanted chamber repetitively enabling dense recordings from most brainstem regions within one animal. Fourth, the design of the system is flexible enough to accommodate a wide range of probe designs (e.g., number of recording sites, arrangement of sites on probe) to flexibly adapt the system to the needs of specific experimental approaches. Finally, the developed system will be commercially available (Neuronexus) and useful for research groups lacking sophisticated machining and electronics expertise.

## Acknowledgments

We thank the entire team at Neuronexus who developed the recording systems for us in close collaboration and continue to adapt the design to our ever-changing needs. We are grateful to Peter Kronen for his exquisite support in anesthesia and Peter Dicke for assistance during surgeries. 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 (DFG) within the frame-work of the Excellence Initiative EXC 307) and the Deutsche Forschungsgemeinschaft (Grant HA5400/3-1).

## References

Agamaite, J.A., Chang, C.-J., Osmanski, M.S., Wang, X., 2015. A quantitative acoustic analysis of the vocal repertoire of the common marmoset (Callithrix jacchus). J. Acoust. Soc. Am. 138, 2906–2928.
Anderson, T.M., Garcia, A.J., Baertsch, Na., Pollak, J., Bloom, J.C., Wei, A.D., Rai, K.G., Ramirez, J.-M., 2016. A novel excitatory network for the control of breathing. Nature 536, 76–80.
Bezerra, B.M., Souto, A., 2008. Structure and usage of the vocal repertoire of Callithrix jacchus. Int. J. Primatol. 29, 671–701.
Borjon, J.I., Ghazanfar, A.A., 2014. Convergent evolution of vocal cooperation without convergent evolution of brain size. Brain Behav. Evol. 84, 93–102.
Burkart, J.M., Finkenwirth, C., 2015. Marmosets as model species in neuroscience and evolutionary anthropology. Neurosci. Res. 93, 8–19.
Del Negro, C.A., Funk, G.D., Feldman, J.L., 2018. Breathing matters. Nat. Rev. Neurosci. 19, 351–367.
Eliades, S.J., Wang, X., 2008a. Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature 453, 1102–1106.
Eliades, S.J., Wang, X., 2008b. Chronic multi-electrode neural recording in free-roaming monkeys. J. Neurosci. Methods 172, 201–214.
Eliades, S.J., Wang, X., 2003. Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. J. Neurophysiol. 2194–2207.
Fritsches, K.A., Rosa, M.G.P., 1996. Visuotopic organisation of striate cortex in the marmoset monkey (Callithrix jacchus). J. Comp. Neurol. 372, 264–282.
Gultekin, Y.B., Hage, S.R., 2018. Limiting parental interaction during vocal development affects acoustic call structure in marmoset monkeys. Sci. Adv. 4 eaar4012.
Gultekin, Y.B., Hage, S.R., 2017. Limiting parental feedback disrupts vocal development in marmoset monkeys. Nat. Commun. 8, 14046.
Hage, S.R., Jurgens, U., 2006. On the role of the Pontine Brainstem in vocal pattern generation: a telemetric single-unit recording study in the Squirrel Monkey. J. Neurosci. 26, 7105–7115.
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.
Hage, S.R., Nieder, A., 2016. Dual neural network model for the evolution of speech and language. Trends Neurosci. 39, 813–829.
Hage, S.R., Ott, T., Eiselt, A.-K., Jacob, S.N., Nieder, A., 2014. Ethograms indicate stable well-being during prolonged training phases in rhesus monkeys used in neurophysiological research. Lab. Anim. 48, 82–87.
Harris, K.D., Dashevskiy, T., Mendoza, J., Garcia, A.J., Ramirez, J.-M., Shea-Brown, E., 2017. Different roles for inhibition in the rhythm-generating respiratory network. J. Neurophysiol. 118, 2070–2088.
Jürgens, U., 2002. Neural pathways underlying vocal control. Neurosci. Biobehav. Rev. 26, 235–258.
Jürgens, U., Hage, S.R., 2006. Telemetric recording of neuronal activity. Methods 38, 195–201.
Luo, J., Hage, S.R., Moss, C.F., 2018. The lombard effect: from acoustics to neural mechanisms. Trends Neurosci. https://doi.org/10.1016/j.tins.2018.07.011.
Miller, C.T., Freiwald, W.A., Leopold, D.A., Mitchell, J.F., Silva, A.C., Wang, X., 2016. Marmosets: a neuroscientific model of human social behavior. Neuron 90, 219–233.
Paxinos, G., Watson, C., Petrides, M., Rosa, M., Tokuno, H., 2012. The Marmoset Brain in Stereotaxic Coordinates. Elsevier, London.
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.
Roberts, A.C., Wallis, J.D., 2000. Inhibitory control and affective processing in the prefrontal cortex: neuropsychological studies in the common marmoset. Cereb. Cortex 10, 252–262.
Roy, S., Wang, X., 2012. Wireless multi-channel single unit recording in freely moving and vocalizing primates. J. Neurosci. Methods 203, 28–40.
Sasaki, E., Suemizu, H., Shimada, A., Hanazawa, K., Oiwa, R., Kamioka, M., Tomioka, I., Sotomaru, Y., Hirakawa, R., Eto, T., Shiozawa, S., Maeda, T., Ito, M., Ito, R., Kito, C., Yagihashi, C., Kawai, K., Miyoshi, H., Tanioka, Y., Tamaoki, N., Habu, S., Okano, H., Nomura, T., 2009. Generation of transgenic non-human primates with germline transmission. Nature 459, 523–527.
Song, X., Osmanski, M.S., Guo, Y., Wang, X., 2016. Complex pitch perception mechanisms are shared by humans and a New World monkey. Proc. Natl. Acad. Sci. 113, 781–786.
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.

## 12. Acknowledgments

I want to thank everyone who helped me performing this project:

First, I want to thank my PI Steffen R. Hage for his guidance, advice and scientific expertise. Furthermore, I am especially thankful to Julia Löschner, Cristina Risueno-Segovia for discussing data and scientific questions as well as combining data leading to our successful joint projects. I also want to thank Deniz Dohmen for helping me building up and programming the setup for the operant conditioning project. Here, I also want to thank the members of my advisory board, Uwe Ilg and Andreas Nieder for their advice.

Working with marmoset monkeys also requires people from outside the lab, namely animal caretakers and veterinarians which participating in making every day work possible. Here, I especially want to thank Katrin Joksch for her excellent work in the animal facility.

I am, furthermore, very thankful to my family and Thomas for their support and love.

My last thanks are for the monkeys participating in my projects: Host Alberich, Sissi, Franz, Willi, Leonard, Eva and Penny.