Current Biology

Volume 28
Number 5

March 5, 2018

www.cell.com

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

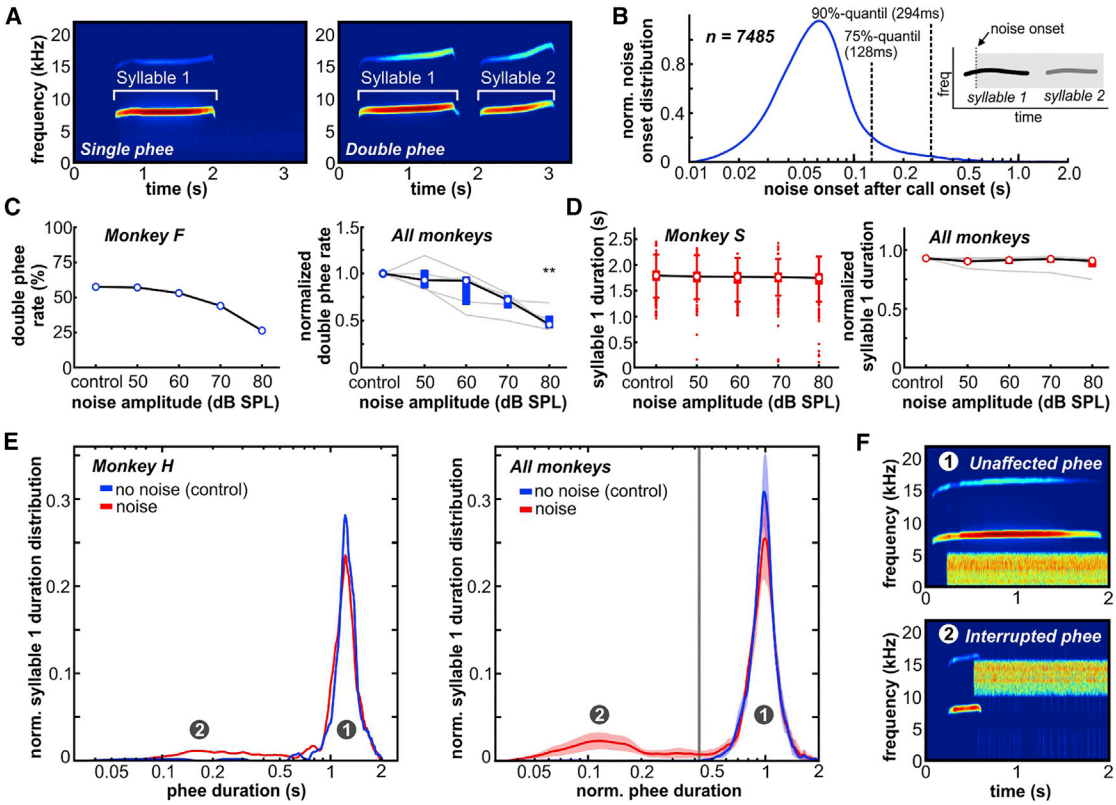

Figure 1. Marmoset Monkeys Interrupt Their Calls during Vocal Production as a Response to Perturbing Noise Playback

(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

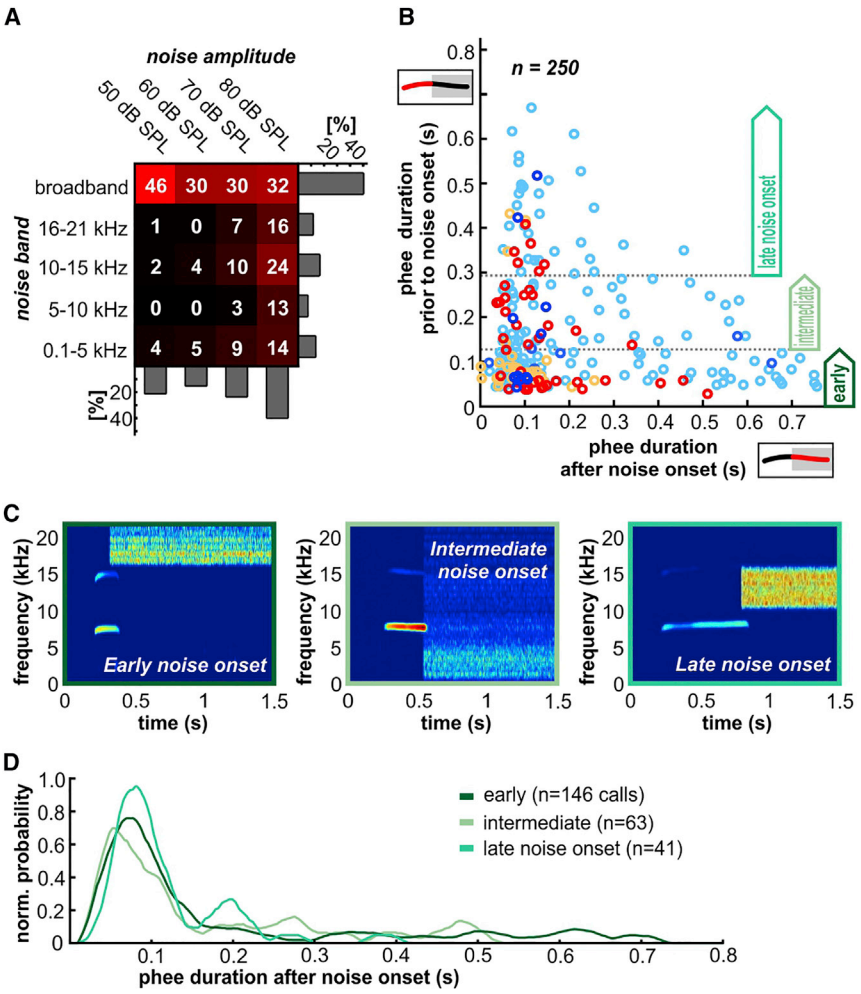

Figure 2. Occurrence of Phee Call Interruption Is Dependent on Noise Conditions and Directly Related to the Onset of Noise Perturbation

(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

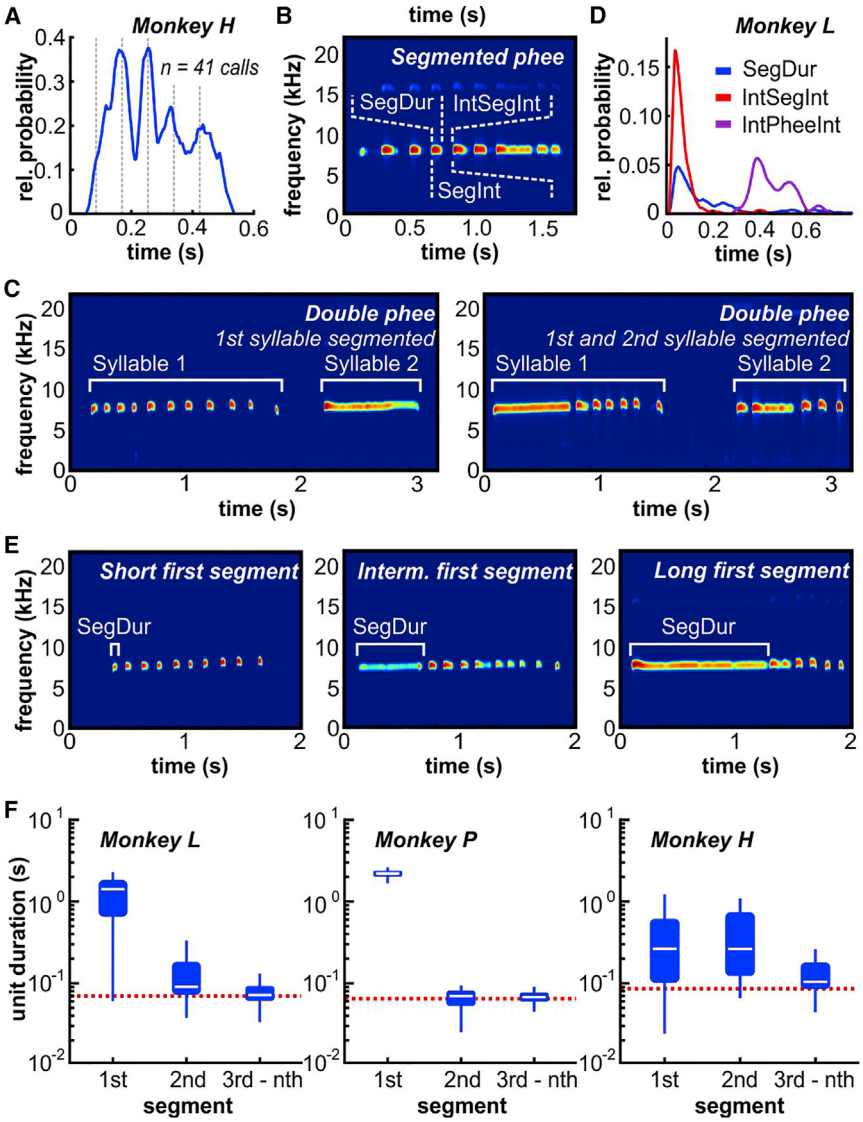

Figure 3. Segmented Phees of Marmoset Monkeys

(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

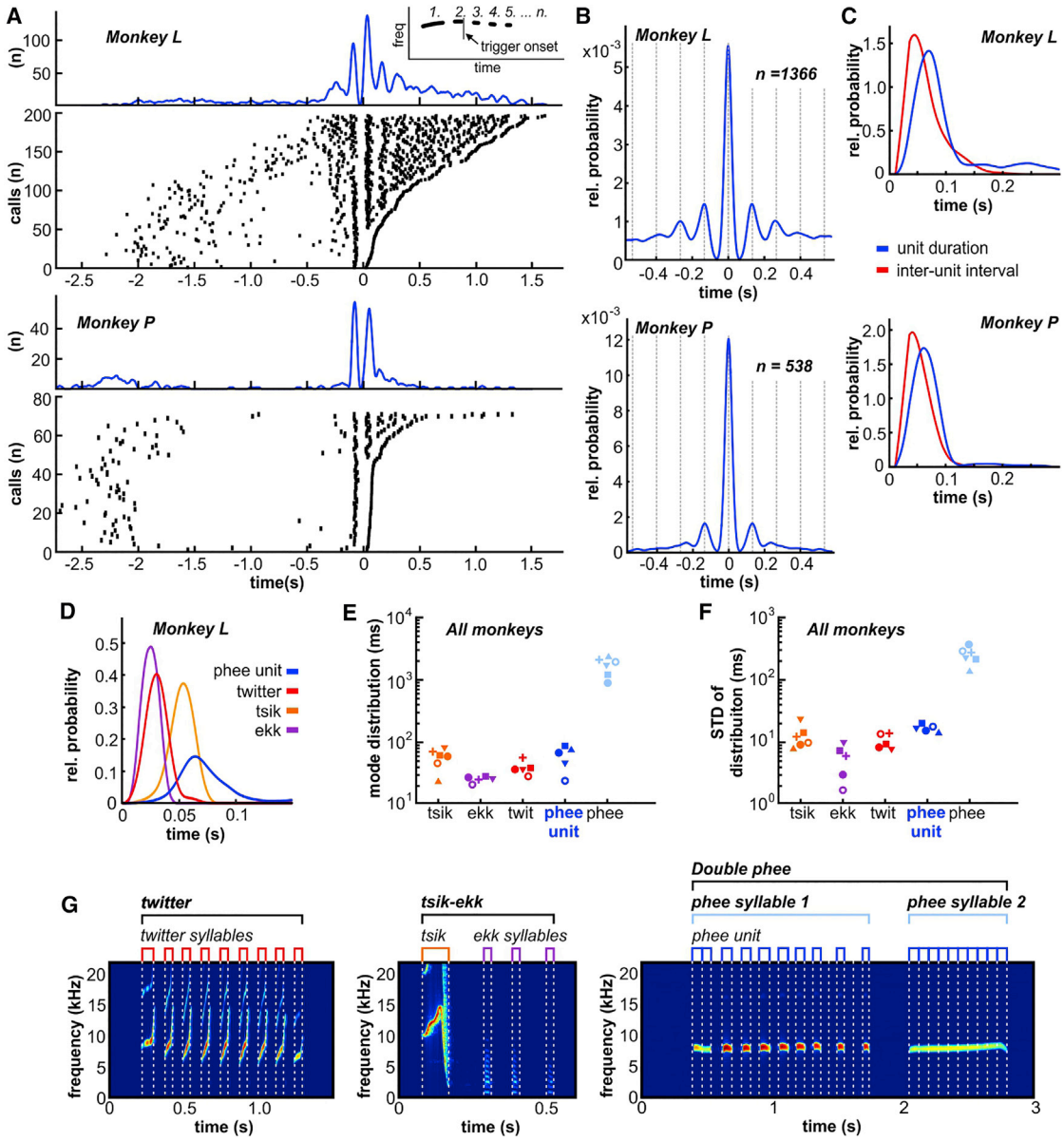

Figure 4. Phee Vocalizations Consist of Consecutively Uttered Brief, Ultra-Precise Vocal Motor Units

(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).

Current Biology, Volume *28*

Supplemental Information

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

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

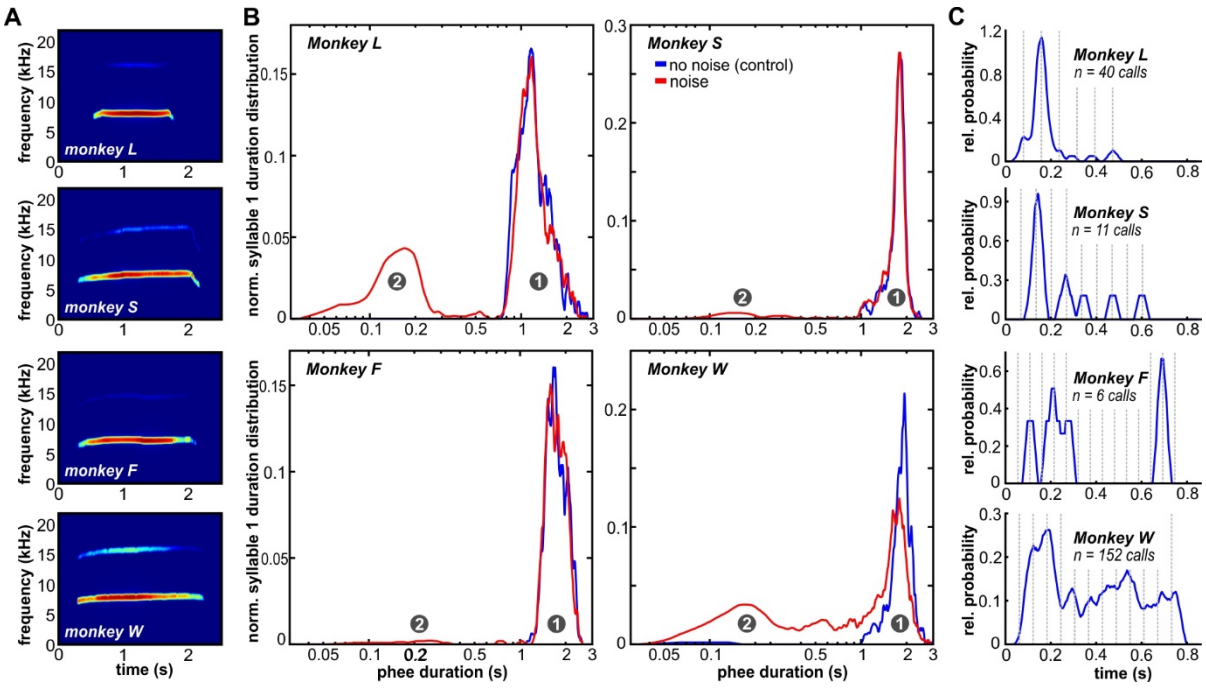

Figure S1. Additional spectrograms of phee calls, phee duration distributions, and multi-modal phee duration distributions of monkeys S, L, W, and F, related to Figure 1 and 3. (A) Example spectrograms of single phees. (B) Phee duration distributions for calls that were (noise) and were not perturbed (control). Distributions of perturbed calls (noise) show bimodal distributions with phees being unaffected (1) and interrupted (2) by noise perturbation in all monkeys. (C) Multi-modal phee duration distribution of interrupted phees.

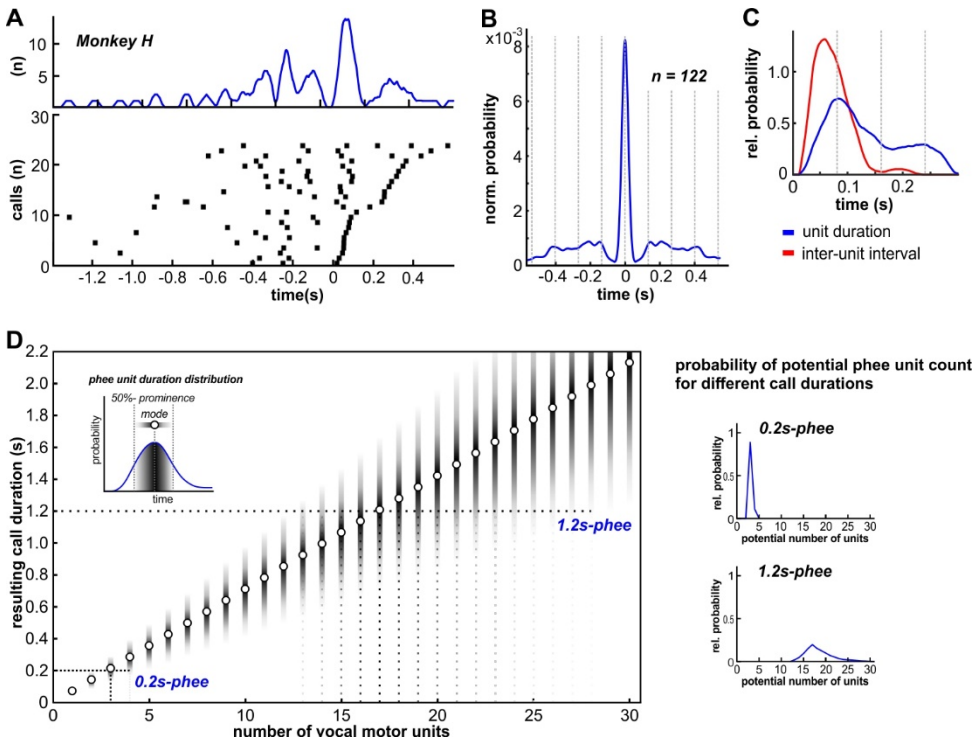

Figure S2. Phee segmentation in marmoset monkeys, related to Figure 4. (A) Distribution of segment onsets triggered by the offset of the second segment of each call for an additional individual monkey. Lower panel shows a raster plot; the higher panel represents the corresponding call segment density histograms (monkey H: 24 calls, 88 segments). (B) Inter-segment interval histogram for the additional monkey shown in (A). Vertical lines indicate multiples of the first mode (monkey H: 41 calls, 122 segments). (C) Distribution of segment durations and inter-segment intervals <300ms for the monkey shown in (A) and (B). (D) Schematic of the correlation between number of vocal motor units and the resulting phee call duration taking into account the unit distribution error (here defined as vertical distance equal to half the mode prominence [50%-prominence]). While the number of vocal motor units are clearly indicated for short phee call durations such as 0.2s, it is not possible to trace back the potential phee unit count for long phee call durations, such as 1.2s.