

Motor planning for vocal production in common marmosets

Cory T. Miller*, Steven J. Eliades¹, Xiaoqin Wang¹

Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University, School of Medicine

ARTICLE INFO

Article history:

Received 6 May 2009

Initial acceptance 9 June 2009

Final acceptance 23 July 2009

Available online 29 September 2009

MS. number: A09-00291

Keywords:

Callithrix jacchus

common marmoset

primate vocal production

vocal motor plan

The vocal motor plan is one of the most fundamental and poorly understood elements of primate vocal production. Here we tested whether a single vocal motor plan comprises the full length of a vocalization. We hypothesized that if a single motor plan was determined at vocal onset, the acoustic features early in the call should be predictive of the subsequent call structure. Analyses were performed on two classes of features in marmoset phoe calls: continuous and discrete. We first generated correlation matrices of all the continuous features of phoe calls. Results showed that the start frequency of a phoe's first pulse significantly correlated with all subsequent spectral features. Moreover, significant correlations were evident within both the spectral and temporal features, but there was little relationship between these feature classes. Using a discrete feature, 'the number of pulses in the phoe call', a discriminant function was able to correctly classify the number of pulses in the calls well above chance based solely on the acoustic structure of the call's first pulse. Together, these data suggest that a vocal motor plan for the complete call structure is established at call onset. These findings provide a key insight into the mechanisms underlying vocal production in nonhuman primates.

© 2009 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

Motor control occurs through an interaction between the sensory and motor systems (Jordan & Wolpert 2000; Shadmehr & Wise 2005). At the initiation of a deliberate action, the motor system establishes a plan consisting of the sequential steps leading to an end goal. Sensory feedback about environmental perturbations occurring during the action are then integrated with the motor plan in order to produce the observable behaviour. Importantly, these two elements of control, the initiation of a planned motor action and modifications to that plan in response to events that occur during the action, are independent aspects of a motor act. When a monkey decides to reach for a fruit on a branch, for example, a motor plan is generated and executed with the end goal being to grasp the fruit on the branch. Wind might begin to shake the branch, forcing the monkey to adjust the timing and trajectory of the reach to achieve the goal. Similarly to visually guided behaviours, an analogous process presumably occurs for vocal motor actions (i.e. vocal production), but considerably less is known about this system in nonhuman primates.

Evidence suggests that deliberate motor actions involving a sequence of events are planned prior to the onset of the initiating act (Sternber et al. 1978; Sarlegna & Sainburg 2008). The notion of a motor plan was initially developed from the observation that in speech each individual motor act in a sequence is not produced independently, but as a single organized motor pattern (Lashley 1951). Since the elements of the action are produced as a single event, a plan for all components must be generated prior to the initial action. Some of the most elegant examples of this process come from oculomotor eye saccades (Zingale & Kowler 1987). When viewing the world, we constantly make saccades to encode the stream of information available in the visual field. Although some saccades are reflexive, occurring in response to peripheral events that drive shifts in attention, most are deliberate actions that involve planning prior to execution. Zingale & Kowler (1987), for example, showed that humans plan the full sequence of up to five saccades prior to the initiation of the first oculomotor act. These authors hypothesized that one benefit of planning all events of a sequential motor act at onset is that it frees resources, such as attention, that can then be used to detect and encode sensory events that occur during the motor action. This would potentially be important for the feedback critical to modifying the motor action in response to unexpected environmental disruptions. As many vocalizations consist of a series of acoustically and temporally distinct pulses, it is possible that a motor plan is also generated before the vocalization is produced. For the vocalization to be planned, however, it must be a deliberate motor act.

* Correspondence: C. T. Miller, Cortical Systems and Behavior Laboratory, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, #109, La Jolla, CA 92093, U.S.A.

E-mail address: corymiller@ucsd.edu (C.T. Miller).

¹ S. J. Eliades and X. Wang are at the Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University, School of Medicine, 720 Rutland Avenue, Traylor 412, Baltimore, MD 21205, U.S.A.

The question of vocal control in nonhuman primates is somewhat controversial. Early studies of squirrel monkeys showed limited vocal control and learning at the behavioural level, as well as neuronal responses in brain stem nuclei related to vocal production (Jurgens 1995, 2009; Deacon 1997; Hammerschmidt & Fischer 2008). Based on these data, many argued that vocal production was largely reflexive and, therefore, animals could exert little control over the timing and structure of the call. More recent behavioural studies of nonhuman primates, however, show evidence of vocal control both in long-term learning (Marshall et al. 1999; Weiss et al. 2001) and as a result of environmental perturbations that occur during call production (Mitani & Gros-Louis 1998; Miller et al. 2003; Brumm et al. 2004; Egnor et al. 2006). The extent of this vocal control is clearly constrained, particularly when compared to other taxonomic groups, such as songbirds, cetaceans and humans (Leonardo & Konishi 1999; Janik 2000; Hammerschmidt & Fischer 2008), but these results suggest that it is not entirely absent. Empirical evidence, for example, shows that primates are able to exert control over when a call is produced. During an experimental scenario in which white noise bursts and silence periods were alternated, tamarins (*Saguinus oedipus*) initiated vocal production only during the silence periods, suggesting that the timing of this motor act is deliberate (Egnor et al. 2007). Moreover, several recent studies suggest that at least some of the neural mechanisms underlying primate vocal production reside in cortical substrates (Gemba et al. 1995, 1999; Petrides et al. 2005; C. T. Miller, A. Dimauro, A. Pistorio, S. Hendry & X. Wang, unpublished data). Building on these studies, the question we address here is whether, like other motor systems, a motor plan for the entire structure of the call is established prior to vocal onset.

Vocalizations comprising a variable number of acoustically and temporally distinct pulses offer an excellent opportunity to investigate the structure of the vocal motor plan. Specifically, we can ask the following question. When initiating the vocalization, is a vocal motor plan established for the full duration of the call? Or is a new motor plan generated for each subsequent pulse? If the former is true, acoustic features early in the vocalization may be correlated with the length and global acoustic structure of the vocalization, making it empirically possible to predict the call's structure from its initial acoustic elements.

Here we addressed this issue in the common marmoset, *Callithrix jacchus*, phoe call, the species' long-distance contact call (Miller & Wang 2006). This call consists of a variable number of temporally distinct pulses (Fig. 1). While the mode number of pulses is two, all individuals regularly produce phoe calls comprising a range of pulses (C. T. Miller, unpublished data). The variable structure of marmoset phoe calls makes this call type an ideal candidate to test whether a motor plan for the structure of the call is established at vocal onset or whether a new plan is generated for each subsequent pulse while the call is being produced. Although vocalizations consist of a diverse number of acoustic features, the measured variables can be divided into features along a discrete or continuous scale. Discretely scaled features can be grouped into distinct categories, while continuously scaled features occur along a continuum. The analyses performed here reflect these differences. In the first analyses, we performed cross-correlation analyses of the numerous continuously scaled spectral and temporal features in this call type. We predicted that if the sequence of pulses in the call is planned, these acoustic features should be highly correlated. For the second analysis, we focused on a discrete feature, the number of pulses in the phoe call, and compared the acoustic structure of the first pulse of phoe calls consisting of one to three pulses. Using discriminant function analysis, we tested whether the acoustic structure of the first pulse could be used to determine the total number of pulses in the call.

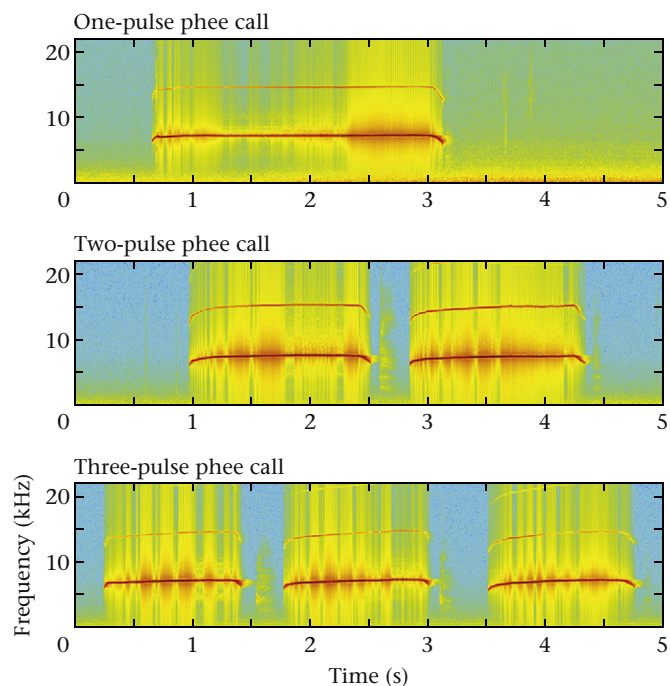


Figure 1. Spectrograms for exemplars of one-, two- and three-pulse phoe calls produced by an individual common marmoset (m9N).

We hypothesized that if the number of pulses were planned before initiating the vocalization, acoustic differences between phoe calls consisting of different numbers of pulses would be evident in the first pulse of the call.

METHODS

Subjects

We recorded 1701 phoe calls produced by 10 adult common marmosets (6 male, 4 female). The common marmoset is a small-bodied (~ 400 g), New World primate endemic to the rainforests of northeastern Brazil (Rylands 1993). Subjects are housed in social groups consisting of pair-bonded mates and up to two generations of offspring. This highly vocal primate has been the subject of several previous behavioural and neural studies of vocal communication (Norcross & Newman 1993; Wang & Kadia 2001; Eliades & Wang 2003; DiMattina & Wang 2006; Miller & Wang 2006; Pistorio et al. 2006). All experimental protocols were approved by the Johns Hopkins University Animal Care and Use Committee.

Recording Procedure

All phoe calls were recorded from unrestrained animals engaged in natural, species-typical vocal behaviours. For eight subjects, we recorded phoe calls directly to a computer hard drive in a testing room away from the colony using a Sennheiser directional microphone. Vocalizations produced by the two other subjects were recorded in the colony room. These subjects, however, were removed from their group cages and placed in individual cages during recording sessions. For these recordings session, we utilized an AKG directional microphone and recorded all calls directly to a computer hard drive. Each of these experimental set-ups allowed isolated recordings of individual marmoset phoe calls. No

differences were evident in the gross acoustic structure of phee calls produced by these two groups of individuals.

Acoustic Analysis

All phee calls recorded during the test recordings were digitized as individual files for analysis. Using custom MatLab code written by C.T.M. (Mathworks, Inc., Natick, MA, U.S.A.), we analysed the following spectro-temporal features of each phee call: call duration (s), pulse duration (s), interpulse interval (IPI) (s), duration from pulse onset to peak frequency (s), duration from peak frequency to pulse offset (s), pulse start frequency (Hz), pulse end frequency (Hz), pulse mean frequency (Hz), pulse minimum frequency (Hz), pulse peak frequency (Hz), pulse delta frequency (i.e. maximum change in frequency or bandwidth) (Hz), slope from pulse onset to peak frequency (Hz/s), slope from peak frequency to pulse offset (Hz/s).

Statistical Analysis

Analyses comprised two distinct sets based on whether the variable occurs on a discrete or continuous scale. Continuous variables are ones that are scaled along a continuum. In this analysis, all the measured spectro-temporal parameters were considered continuous variables. A discrete variable is one that can be clustered into finite amounts. Here we use the number of pulses as a discretely scaled variable of phee calls.

Continuous variable analysis

Sixty-two per cent of the vocalizations in our data set were two-pulsed phee calls. Since this represented the largest group of calls and these calls have multiple acoustic pulses, analyses of continuous variables were performed only on these calls. The primary statistical test performed on these features was Pearson product-moment correlation. We computed correlation coefficients for combinations of acoustic features to test the extent to which the features covaried. Because our data set was large and it included many features, statistical significance could be reached without a meaningful effect size. As such, we considered a correlation meaningful only if it exhibited a correlation coefficient ≥ 0.3 . All correlation coefficients of at this level also exhibited a P value of ≤ 0.0001 .

Discrete variable analysis

Three types of statistical tests were performed on these data. First, as the dependent variables in these acoustic analyses were likely to covary, we analysed the global acoustic structure of the phee call classes using a MANOVA. In this analysis, the individual acoustic features measured for the first pulse of the phee call served as the dependent variables and the number of pulses in the call served as the independent variable. Second, to test which individual acoustic features were best for distinguishing between phee calls consisting of different numbers of pulses, we used multivariate regression analysis. Third, the final set of tests involved performing discriminant function tests of the data set. This analysis utilizes the dimensions of the independent variables for predicting group membership for a categorical dependent variable. We implemented discriminant functions to test whether a model could be generated to correctly classify the phee call classes based on the acoustic structure the call's first pulse. All of the acoustic features measured for the first pulse of phee calls were used in this analysis. For cross-validation, we used half of the data set for a particular test to build the function; we then ran the second half of the data set through the original function to test how accurately these new data were classified.

RESULTS

Continuous Feature Analysis

We analysed 1701 naturally produced phee calls from 10 adult common marmosets. The data set consisted of 507 one-pulse phees, 1052 two-pulse phees and 142 three-pulse phees. As discussed above, we used only two-pulse phees in the continuous feature analysis. To test the extent to which the spectral and temporal features of phee calls were correlated across phee calls, we performed the following cross-correlation analyses. For this analysis, we only tested those measured acoustic features that were solely temporal ($N = 4$) or spectral ($N = 12$). Features that involved a combination of spectral and temporal parameters (i.e. duration to peak frequency) were not included.

As the first measured feature of a phee call is the start frequency of the first pulse, we first tested how well this feature correlated with the other spectral and temporal features of phee calls (Fig. 2). Interestingly, this spectral feature was not significantly correlated with any of the four temporal features, but was significantly correlated with all but two of the spectral features. This suggests that the general spectral structure of the phee call can be predicted from the initiation of the first pulse.

To further test the extent to which whether spectral features are correlated with temporal features, we generated a correlation matrix comparing all the spectral and temporal features of phee calls. Figure 3 shows the results of this analysis. The results show that only two spectral features (p2 peak frequency, p2 delta frequency) were significantly correlated with a temporal feature (interpulse interval). This analysis suggests that overall there is little correlation between the temporal and spectral acoustic features of marmoset phee calls.

Because of the distinction between spectral and temporal feature correlations observed in the first set of analyses, we next computed cross-correlation matrices to test for relationships within the temporal and spectral features alone. Figure 4 shows the correlations observed between each of the four primary temporal features measured here. Significant correlations were observed for all feature pairs with one exception. Figure 5 shows the correlations between all 11 spectral features; data for the p1 start frequency are not shown in this figure as the data are already shown in Fig. 2. Similarly to the temporal features, analyses showed that nearly every spectral feature cross-correlation was statistically significant. In fact, the only features that were not significantly correlated were those related to the delta frequency. Together, these analyses showed high correlations between the spectral and temporal dimensions individually, but virtually no relation between these two acoustic dimensions.

Discrete Variable Analysis

To test whether global spectro-temporal differences were evident in the first pulse of one-, two- and three-pulse phee calls, we analysed the data using a MANOVA. Results showed a significant differences across the phee call classes ($F_{11,1689} = 80.43$, $P < 0.0001$), suggesting that the first pulse of one-, two- and three-pulse phee calls are acoustically distinguishable.

Multivariate regression analyses showed that it was not only the relative global acoustic structure of the first pulse that differed between these three phee call classes. Nine individual acoustic features also differed statistically between phee classes (pulse duration: $F_{1,1699} = 805.1$, $P < 0.0001$; duration to peak frequency: $F_{1,1699} = 369.8$, $P < 0.0001$; duration from peak frequency to end: $F_{1,1699} = 19.75$, $P < 0.0001$; end frequency: $F_{1,1699} = 36.1$, $P < 0.0001$; mean frequency: $F_{1,1699} = 39.1$, $P < 0.0001$; peak frequency:

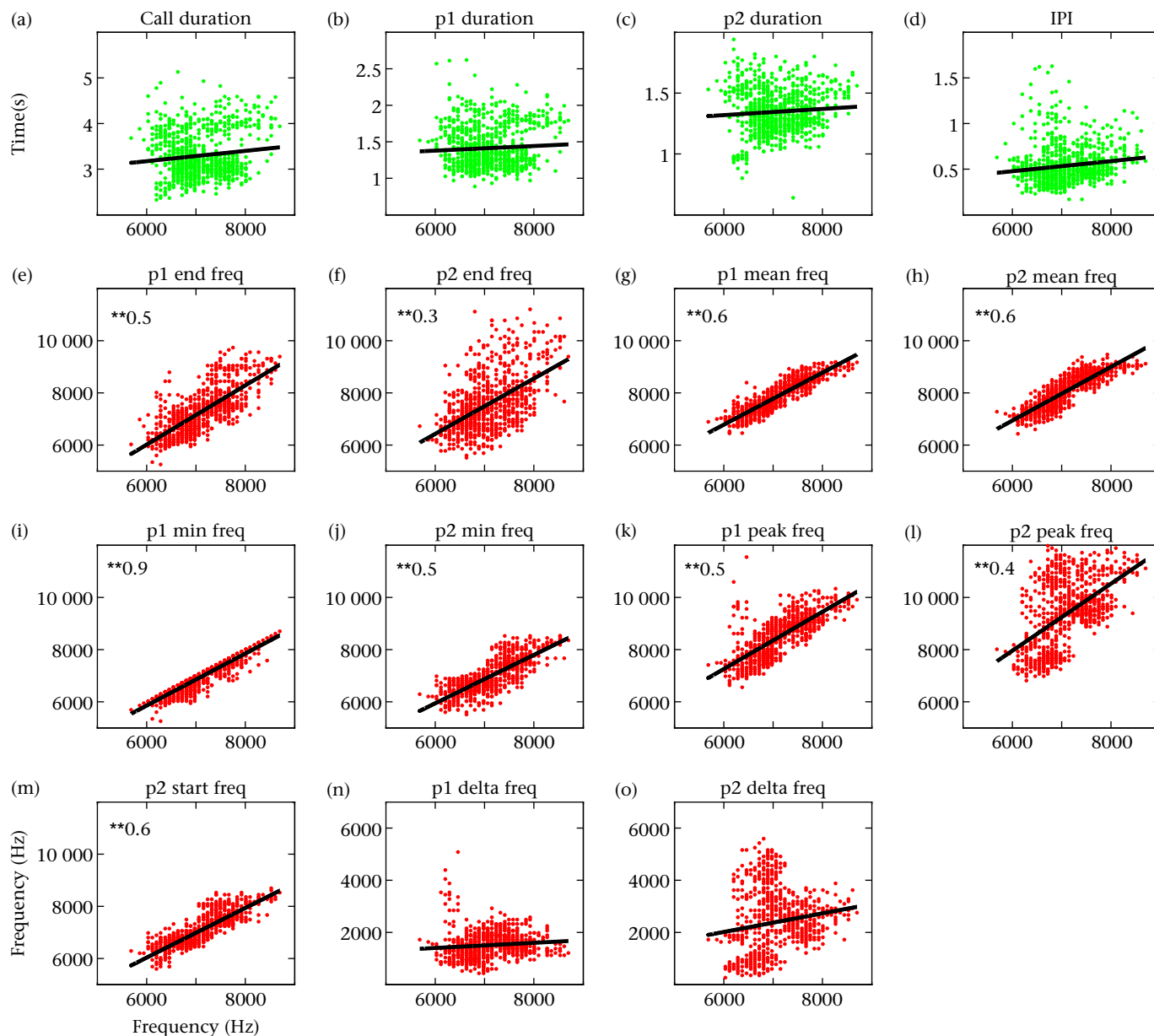


Figure 2. Cross-correlations between the start frequency of the first pulse in common marmoset phlee calls and (a–d) all four temporal features and (e–o) 11 other spectral features. Data points (dots) and mean correlations (solid lines) are shown. **Statistically significant correlation coefficients. p1 = pulse one, p2 = pulse two.

$F_{1,1699} = 42.0, P < 0.0001$; delta frequency: $F_{1,1699} = 25.1, P < 0.0001$; slope 1: $F_{1,1699} = 20.6, P < 0.0001$; slope 2: $F_{1,1699} = 54.7, P < 0.0001$; Table 1). We observed that many of these features showed a graded change from one- to three-pulse phlee calls. Pulse duration, for example, showed an inverse relationship between the number of pulses in the phlee and the duration of the first pulse (Table 1). In other words, the fewer the number of pulses in the call, the longer the duration of the first pulse of one-pulse phlees (mean = 1.87 s) was longer than both two-pulse (mean = 1.43 s) and three-pulse (mean = 1.16 s) phlees. The spectral content of the phlee showed a similar pattern.

To test whether we could predict the number of pulses in the phlee based on the structure of the first pulse, we performed a series of discriminant function analyses. Our first set of tests examined how well calls could be classified as consisting of a particular number of pulses for each individual caller. By analysing the data

for each individual, we avoided the inherent problems with using a multifactorial design for discriminant function analyses (Mundry & Sommer 2007). We also computed a population analysis to test the extent to which a discriminant function could classify calls with a greater range of variability. Although this may potentially create statistical error, as discussed by Mundry & Sommer (2007), in all cases, population analyses were consistent with our individual subject tests.

The first analysis tested how well a discriminant function could classify calls as being either one-pulse or two-pulse phlee calls (Fig. 6). All individuals produced calls that could be classified correctly as consisting of one or two pulses well above chance (i.e. 50%: marmoset-10n = 75.0%; marmoset-17p = 84.5%; marmoset-18r = 95.6%, 29o = 88.0%; marmoset-38m = 81.2%; marmoset-3o = 82.7%; marmoset-42n = 84.3%; marmoset-49p = 92.5%; marmoset-49r = 72.5%; marmoset-9n = 90.4%). The

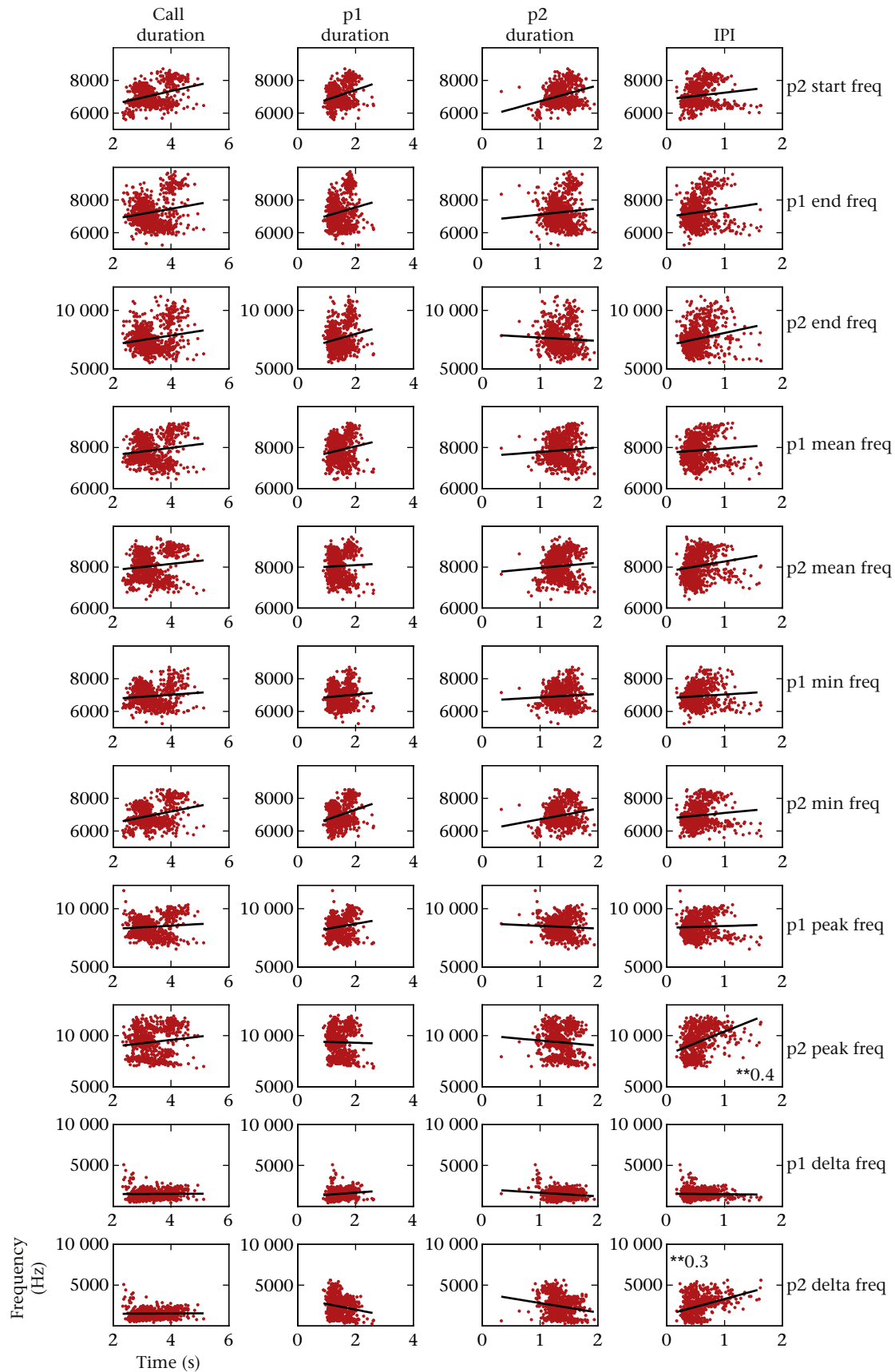


Figure 3. Cross-correlations between all four temporal features and 11 spectral features of common marmoset phoe calls. Data points (dots) and mean correlations (solid lines) are shown. **Statistically significant correlation coefficients. p1 = pulse one, p2 = pulse two.

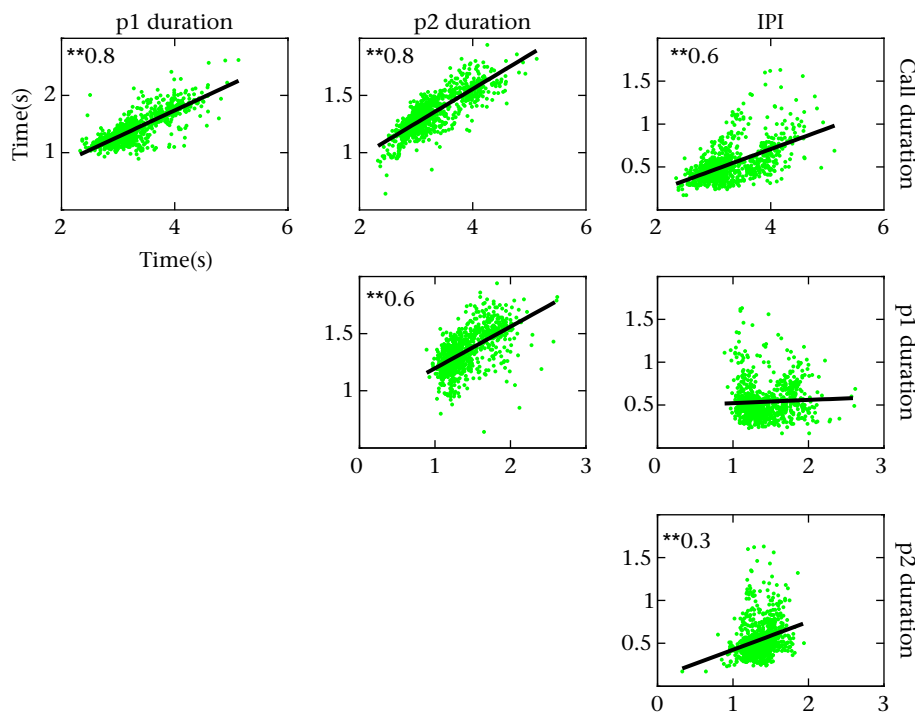


Figure 4. Cross-correlations between temporal features of common marmoset phee calls. Data points (dots) and mean correlations (solid lines) are shown. **Statistically significant correlation coefficients. p1 = pulse one, p2 = pulse two.

mean percentage correctly classified was 84.6% across the individuals. The subsequent cross-validation tests for all individuals were within 5% of the initial function. Pooling all calls together for a population analysis, a discriminant function was able to correctly classify 76.3% (76.0% cross-validation) of the calls as being either one- or two-pulse phees. Although this value was lower than the mean of the individual analysis, it was still above chance.

The next function tested classification of phee calls consisting of one or three pulses (Fig. 6). Here only three individuals produced sufficient three-pulse phee calls for the analysis. Discriminant functions were able to correctly classify these calls for each individual over 90% of the time (marmoset-17p = 92.7%; marmoset-49r = 95.5%; marmoset-9n = 98.1; mean = 95.4%). Cross-validation tests were all within 5% of the initial test. We next performed a discriminant function using all one- and three-pulse phee calls produced by all subjects. Here the discriminant function was able to correctly classify calls 89.4% (88.2% cross-validation) of the time.

The final discriminant function analysed one-, two- and three-pulse phee calls (Fig. 6). The same three animals used in the preceding analysis were used here for the individual analyses. As in the previous test, a discriminant function was able to correctly classify phee calls well above chance (i.e. 33.3%: marmoset-17p = 77.6%; marmoset-49r = 71.2%; marmoset-9n = 75.8%; mean = 73.9%). When combining all calls produced by all subjects, the population analysis was still able to correctly classify phee calls 60.8% (57.3% cross-validation) of the time. Interestingly, the acoustic feature with the highest correlation to the function was pulse duration ($r = 0.9$), suggesting that this feature contributed significantly to discriminating between the classes of phee calls.

DISCUSSION

The aim of this study was to test the structure of one of the most fundamental aspects of vocal production, the vocal motor plan. Motor plans are common in deliberate motor acts for determining

the sequence of actions prior to their initiation (Sternbert et al. 1978; Zingale & Kowler 1987; Sarlegna & Sainburg 2008). Its existence in the vocal motor system, however, has not been addressed. We asked the following question. When a vocalization is produced, does the motor plan comprise all acoustic components for the full duration of the vocalization, or is a new motor plan produced for each successive temporal element? We hypothesized that if the vocal motor plan comprises the entire length of the call, it should be possible to predict the structure of a vocalization based on features produced early in the call. To address this issue, we analysed the structure of common marmoset phee calls.

Evidence of a motor plan for vocal production was based on analyses of phee calls showing that its acoustic structure can be predicted from features early in the call. Continuously scaled acoustic features in two-pulse phee calls showed high correlations within spectral and temporal parameters (Figs 4, 5), although there was little correlation across these acoustic dimensions (Fig. 3). Interestingly, even the first feature measured, the start frequency of pulse one, was highly correlated with all subsequent spectral features (Fig. 2), suggesting that one could probably reconstruct the general spectral structure of a phee from this one feature alone. As phee calls vary in the number of pulses (Fig. 1), we next extended our analysis to test whether it would be possible to predict the number of pulses in a phee call based on the acoustic structure of the first pulse. Using discriminant function analyses, we found that calls could be classified correctly as one-, two- or three-pulse calls based on the acoustic structure of the first pulse well above chance (Fig. 6). Returning to the initial question of this study, these analyses suggest that the vocal motor plan at call onset comprises all acoustic elements for the duration of the call.

Importantly, these results do not suggest that primate vocal production is fixed once the call is initiated. Rather, the results presented here suggest that, like other motor actions, a plan is generated at the onset of vocal production. Once initiated, that motor plan can be modified in response to environmental perturbations occurring

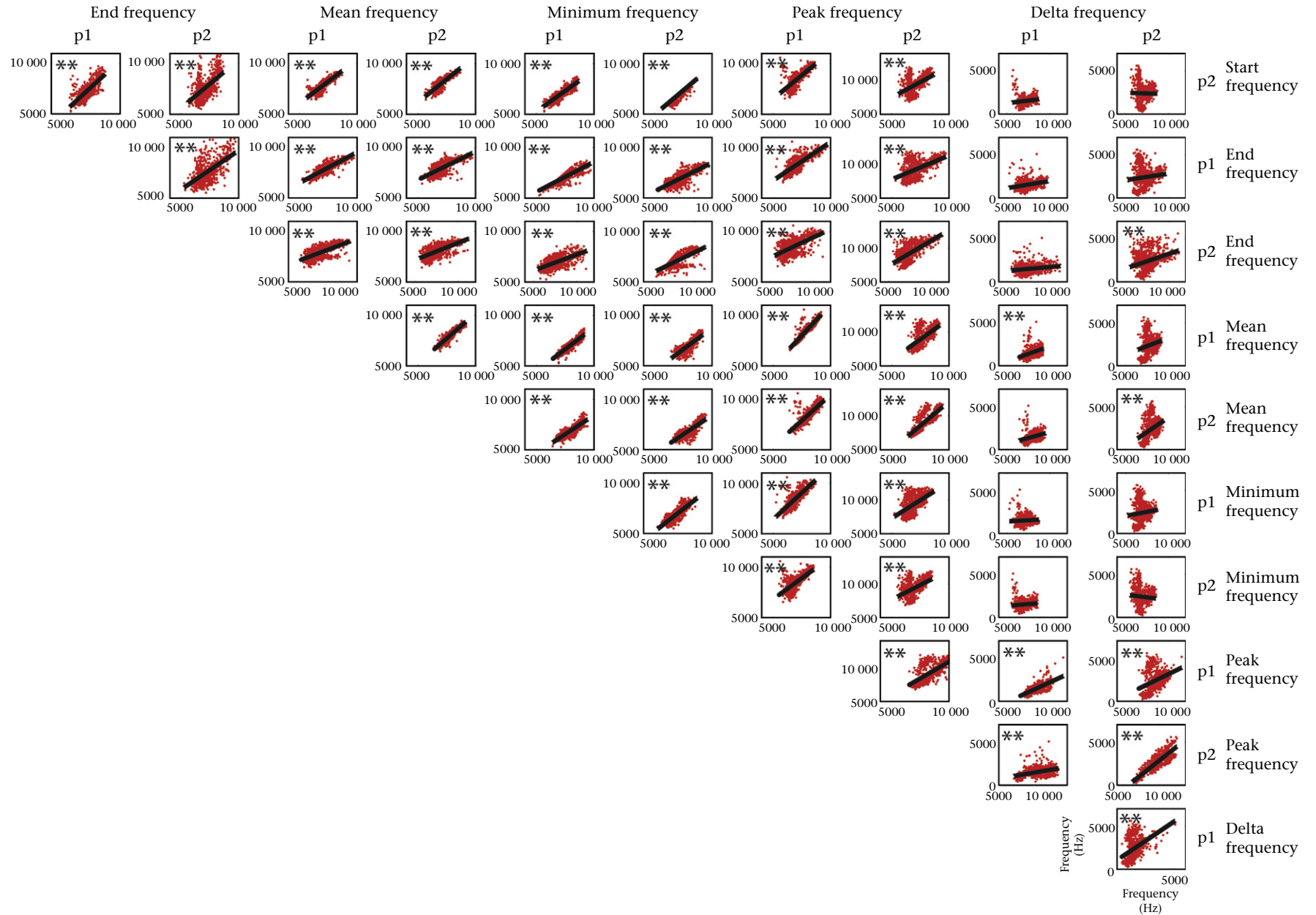


Figure 5. Cross-correlations between 11 spectral features of common marmoset phee calls. Data points (dots) and mean correlations (solid lines) are shown. **Statistically significant correlation coefficients. p1 = pulse one, p2 = pulse two.

Table 1

Acoustic differences for individual features in the first pulse of one-, two- and three-pulse marmoset phee calls

	One-pulse	Two-pulse	Three-pulse
	Mean±SD	Mean±SD	Mean±SD
Pulse duration (s)	1.78±0.29	1.43±0.27	1.17±0.33
Duration to peak frequency (s)	1.44±0.39	1.15±0.31	0.93±0.22
Duration from peak frequency to end (s)	0.34±0.36	0.27±0.26	0.29±0.29
End frequency (Hz)	7499.5±892.6	7265.5±812.4	7127.7±477.5
Mean frequency (Hz)	7927.6±720.7	7816.7±560.5	7552.5±357.2
Peak frequency (Hz)	8555.1±1064.1	8373.4±760.2	8030.1±566.6
Overall change in frequency (Hz)	1685.1±883.2	1536.1±740.6	1343.4±571.2
Slope from pulse onset to peak frequency, slope 1 (Hz/s)	899.3±510.9	986.9±470.5	1090.2±509.2
Slope from peak frequency to pulse offset, slope 2 (Hz/s)	605.2±397.1	803.2±435.5	784.4±446.8

during the action. They are separable, but obviously related components of the motor system for vocal production nonhuman primates. A remaining question, however, is the extent to which the motor plan is actually deliberate in primate vocal production. It is difficult to disambiguate this issue from only the data presented here. Evidence from Japanese macaque, *Macaca fuscata*, however, provides some insight into this question. Data show that these monkeys match the acoustic structure of the initiating vocalization during vocal exchanges of their coo calls (Suguira 1998). At some level, therefore, these monkeys must have generated a plan for the structure of their vocalization when uttering their vocal response. While data presented in the current paper do not allow us to determine the extent to which the vocal motor plan for marmoset phee is under complete volitional control, the results of this macaque study suggest that the animals are deliberately producing a call type as a response. More explicit experimental tests are needed to determine the extent to which marmosets also control the structure of their vocal motor plan.

These data may provide an insight into the mechanisms underlying vocal production and control in primates. Consider the correlation analyses performed on the continuously scaled features of phee calls. High correlations were evident within spectral and temporal parameters, but not between them. This may suggest that independent mechanisms underlie control over spectral and temporal dimensions of vocal production. At least two studies of tamarins found that changes in the temporal structure of vocalizations occur when animals are experimentally presented with white noise (Miller et al. 2003; Egnor et al. 2006). Although there is no evidence of spectral modifications in response to external stimuli, there has not been a direct systematic effort to experimentally test this parameter space. Future work is needed to test the prediction that the mechanisms governing control over spectral and temporal parameters during vocal production are separable.

Combined with previous studies, a growing body of evidence suggests that control can be exerted over nonhuman primate vocal production in at least three ways. First, the individual decides when to produce a vocalization (Miller et al. 2009). Egnor et al. (2007) showed that when presented with alternating periods of noise and silence, cottontop tamarins, *Saguinus oedipus*, initiate a vocalization only during a silence period. Furthermore, the decision of when to produce an antiphonal call response in common marmosets is modulated by the social relationship of the two animals, suggesting that these monkeys are able to regulate the timing of call production dependent on the specific social context (Miller & Wang 2006). Second, as shown here, as well as in Japanese macaques (Suguira

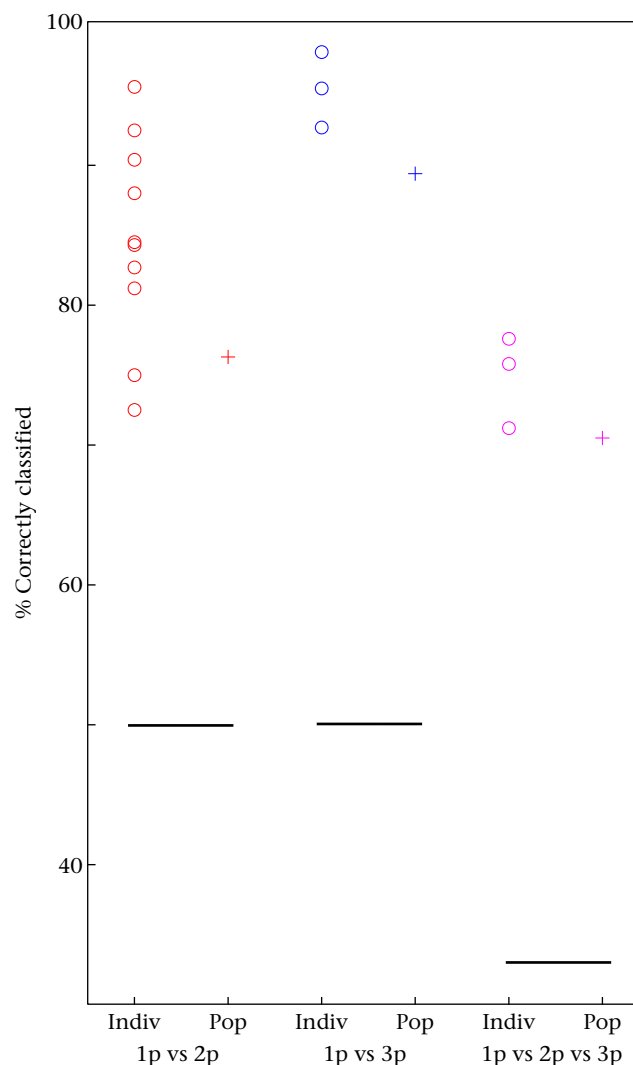


Figure 6. Percentage of one-pulse (1p), two-pulse (2p) and three-pulse (3p) common marmoset phee calls correctly classified using discriminant function analysis. Open circles: individual data ('Indiv'); +: population data ('Pop'); solid line: chance level for each analysis.

1998), once the decision to call is made, a motor plan for the vocalization is generated. And third, primates are able to affect changes over the call during production in response to external environmental events through auditory feedback (Miller et al. 2003; Brumm et al. 2004; Egnor et al. 2006). In addition to lower-level changes in vocal behaviour in response to external noise, such as the Lombard Effect (Brumm et al. 2004; Egnor & Hauser 2006), exerting top-down changes as a result of external perturbations is also evident (Mitani & Gros-Louis 1998; Miller et al. 2003; Egnor et al. 2006). A recent neurophysiology study provided evidence of a neural mechanism in auditory cortex that is likely to underlie the auditory feedback necessary for vocal control (Eliades & Wang 2008). Although vocal control and learning is certainly more constrained in nonhuman than human primates (Hammerschmidt & Fischer 2008), it is not altogether absent. Our task is to elucidate the mechanisms that both contribute to and limit vocal control in this taxonomic group.

The neural systems underlying the various motor actions all exhibit idiosyncratic features, although certainly commonalities do exist. Historically, the study of primate vocal production has typically been isolated from work on other motor systems in part

because many argued that primate vocal production is largely reflexive and mitigated by subcortical structures (Jurgens 1995; Deacon 1997). Recent data, however, suggest a more sophisticated system of vocal control that is probably mediated by neural mechanisms in the frontal cortex (Jurgens et al. 2002; Petrides et al. 2005; C. T. Miller, A. Dimauero, A. Pistorio, S. Hendry & X. Wang, unpublished data) similarly to many other motor control systems (Schall & Boucher 2007). Results presented here show evidence of another common feature across motor systems (Sternbert et al. 1978; Zingale & Kowler 1987; Sarlegna & Sainburg 2008); analyses show that a vocal motor plan is established prior to vocal production for the sequential structure of the vocalization. Interestingly, the limitations on vocal control is a unique feature of the primate vocal motor system that has both isolated the study of this system and may ultimately provide the most significant insights into the sensory–motor interactions that govern motor control more broadly.

Nonhuman primates are readily able to move parts of their bodies (e.g. arms, legs) freely and adjust to any number of environmental interferences, but the same is not true of vocal motor actions. Exactly why is not well understood, but it probably results from a disruption to some aspect of the sensory–motor pathway. Studies of nonhuman primate auditory cortex show that neurons inhibit activity prior to the onset of vocal production, presumably due to a command signal from the motor system (Eliades & Wang 2003). Moreover, auditory cortex neurons' sensitivity to acoustic perturbations actually increases during vocal production (Eliades & Wang 2008). These studies suggest both an anatomical and a functional relationship between the motor and auditory systems and that the sensory information necessary to guide changes in vocal production is available at the cortical level. We must presume, therefore, that the limited modification of vocal production must be related to the mechanisms that underlie how that sensory information integrates with the motor output. As the neural mechanisms for utilizing sensory feedback to modify actions exists in other motor systems, the vocal motor system provides a special case to test the precise neural mechanism underlying this key sensory–motor interaction and how specific disruptions of network result in degradation of feedback mediated motor control. Rather than only approach primate vocal production from the perspective of communication signalling, future work may benefit from conceptualizing the issues in relation to more general motor systems in primate cortex.

Acknowledgments

We thank Mark Bee, Asif Ghazanfar, Judith Scarl, Yi Zhou and two referees for helpful comments on this manuscript. This work was supported by grants from the National Institutes of Health to C.T.M. (K99 DC009007) and X.W. (R01 DC005808).

References

- Brumm, H., Voss, K., Kollmer, I. & Todt, D. 2004. Acoustic communication in noise: regulation of call characteristics in a New World monkey. *Journal of Experimental Biology*, **207**, 443–448.
- Deacon, T. W. 1997. *The Symbolic Species: the Co-evolution of Language and the Brain*. New York: W.W. Norton.
- DiMattina, C. & Wang, X. 2006. Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations. *Journal of Neurophysiology*, **95**, 1244–1262.
- Egnor, S. E. R. & Hauser, M. D. 2006. Noise-induced vocal modulation in cotton-top tamarins. *American Journal of Primatology*, **68**, 1183–1190.
- Egnor, S. E. R., Iguina, C. & Hauser, M. D. 2006. Perturbation of auditory feedback causes systematic perturbation in vocal structure in adult cotton-top tamarins. *Journal of Experimental Biology*, **209**, 3652–3663.
- Egnor, S. E. R., Wickelgren, J. G. & Hauser, M. D. 2007. Tracking silence: adjusting vocal production to avoid acoustic interference. *Journal of Comparative Physiology A*, **193**, 477–483.
- Eliades, S. J. & Wang, X. 2003. Sensory–motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology*, **89**, 2185–2207.
- Eliades, S. J. & Wang, X. 2008. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*, **453**, 1102–1106.
- Gemba, H., Miki, N. & Sasaki, K. 1995. Cortical field potentials preceding vocalization and influences of cerebellar hemispherectomy upon them in monkeys. *Brain Research*, **697**, 143–151.
- Gemba, H., Kyuhou, S., Matsuzaki, R. & Amino, Y. 1999. Cortical field potentials with audio-initiated vocalization in monkeys. *Neuroscience Letters*, **272**, 49–52.
- Hammerschmidt, K. & Fischer, J. 2008. Constraints in primate vocal production. In: *Evolution of Communicative Flexibility* (Ed. by D. K. Oller & U. Griebel), pp. 93–120. Cambridge, Massachusetts: MIT Press.
- Janik, V. M. 2000. Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*). *Science*, **289**, 1355–1357.
- Jordan, M. I. & Wolpert, D. M. 2000. Computational motor control. In: *The New Cognitive Neurosciences* (Ed. by M. Gazzaniga), pp. 601–618. Cambridge, Massachusetts: MIT Press.
- Jurgens, U. 1995. Neuronal control of vocal production in nonhuman and human primates. In: *Current Topics in Primate Vocal Communication* (Ed. by E. Zimmerman, J. D. Newman & U. Jurgens), pp. 199–206. New York: Plenum.
- Jurgens, U. 2009. The neural control of vocalization in mammals: a review. *Journal of Voice*, **23**, 1–10.
- Jurgens, U., Ehrenreich, L. & De Lanerolle, N. C. 2002. 2-Deoxyglucose uptake during vocalization in the squirrel monkey brain. *Behavioural Brain Research*, **136**, 605–610.
- Lashley, K. S. 1951. The problem of serial order in behavior. In: *Cerebral Mechanisms in Behavior: the Hixon Symposium* (Ed. by W. A. Jeffress), pp. 112–131. New York: J. Wiley.
- Leonardo, A. & Konishi, M. 1999. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature*, **399**, 466–470.
- Marshall, A. J., Wrangham, R. W. & Clark, A. P. 1999. Does learning affect the structure of vocalizations in chimpanzees? *Animal Behaviour*, **58**, 825–830.
- Miller, C. T. & Wang, X. 2006. Sensory–motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. *Journal of Comparative Physiology A*, **192**, 27–38.
- Miller, C. T., Flusberg, S. & Hauser, M. D. 2003. Interruptibility of cotton-top tamarin long calls: implications for vocal control. *Journal of Experimental Biology*, **206**, 2629–2639.
- Miller, C. T., Beck, K., Meade, B. & Wang, X. 2009. Antiphonal call timing in marmosets is behaviorally significant: interactive playback experiments. *Journal of Comparative Physiology A*, **195**, 783–789.
- Mitani, J. & Gros-Louis, J. 1998. Chorusing and convergence in chimpanzees: tests of three hypotheses. *Behaviour*, **135**, 1041–1064.
- Mundry, R. & Sommer, C. 2007. Discriminant function analysis with non-independent data: consequences and an alternative. *Animal Behaviour*, **74**, 965–976.
- Norcross, J. L. & Newman, J. D. 1993. Context and gender specific differences in the acoustic structure of common marmoset (*Callithrix jacchus*) phee calls. *American Journal of Primatology*, **30**, 37–54.
- Petrides, M., Cadoret, G. & Mackey, S. 2005. Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature*, **435**, 1235–1238.
- Pistorio, A., Vintch, B. & Wang, X. 2006. Acoustic analyses of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). *Journal of the Acoustical Society of America*, **120**, 1655–1670.
- Rylands, A. B. 1993. *Marmosets and Tamarins: Systematics, Behaviour, and Ecology*. Oxford: Oxford University Press.
- Sarlegna, F. R. & Sainburg, R. L. 2008. The roles of vision and proprioception in the planning of reach movements. In: *Progress in Motor Control* (Ed. by D. Sternad), pp. 317–335. New York: Springer.
- Schall, J. D. & Boucher, L. 2007. Executive control of gaze by the frontal lobes. *Cognitive Affective and Behavioral Neuroscience*, **7**, 396–412.
- Shadmehr, R. & Wise, S. P. 2005. *Computational Neurobiology of Reaching and Pointing: a Foundation for Motor Learning*. Cambridge, Massachusetts: MIT Press.
- Sternbert, S., Monsell, S., Knoll, R. & Wright, C. 1978. The latency and duration of rapid movement sequences: comparisons of speech and type writing. In: *Information Processing in Motor Control and Learning* (Ed. by G. E. Stelmach), pp. 117–152. New York: Academic Press.
- Sugriva, H. 1998. Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. *Animal Behaviour*, **55**, 673–687.
- Wang, X. & Kadia, S. C. 2001. Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *Journal of Neurophysiology*, **86**, 2616–2620.
- Weiss, D. J., Garibaldi, B. T. & Hauser, M. D. 2001. The production and perception of long calls by cotton-top tamarins (*Saguinus oedipus*): acoustic analyses and playback experiments. *Journal of Comparative Psychology*, **11**, 258–271.
- Zingale, C. M. & Kowler, E. 1987. Planning sequences of saccades. *Vision Research*, **27**, 1327–1341.