

Cooperative vocal control in marmoset monkeys via vocal feedback

Jung Yoon Choi,^{1,2} Daniel Y. Takahashi,^{1,2*} and Asif A. Ghazanfar^{1,2,3*}

¹Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey; ²Department of Psychology, Princeton University, Princeton, New Jersey; and ³Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey

Submitted 6 March 2015; accepted in final form 28 April 2015

Choi JY, Takahashi DY, Ghazanfar AA. Cooperative vocal control in marmoset monkeys via vocal feedback. *J Neurophysiol* 114: 274–283, 2015. First published April 29, 2015; doi:10.1152/jn.00228.2015.—Humans adjust speech amplitude as a function of distance from a listener; we do so in a manner that would compensate for such distance. This ability is presumed to be the product of high-level sociocognitive skills. Nonhuman primates are thought to lack such socially related flexibility in vocal production. Using predictions from a simple arousal-based model whereby vocal feedback from a conspecific modulates the drive to produce a vocalization, we tested whether another primate exhibits this type of cooperative vocal control. We conducted a playback experiment with marmoset monkeys and simulated “far-away” and “nearby” conspecifics using contact calls that differed in sound intensity. We found that marmoset monkeys increased the amplitude of their contact calls and produced such calls with shorter response latencies toward more distant conspecifics. The same was not true in response to changing levels of background noise. To account for how simulated conspecific distance can change both the amplitude and timing of vocal responses, we developed a model that incorporates dynamic interactions between the auditory system and limbic “drive” systems. Overall, our data show that, like humans, marmoset monkeys cooperatively control the acoustics of their vocalizations according to changes in listener distance, increasing the likelihood that a conspecific will hear their call. However, we propose that such cooperative vocal control is a system property that does not necessitate any particularly advanced sociocognitive skill. At least in marmosets, this vocal control can be parsimoniously explained by the regulation of arousal states across two interacting individuals via vocal feedback.

arousal; auditory cortex; cingulate; primate vocalizations

DOUBLING PHYSICAL DISTANCE drops the intensity of acoustic signals by six decibels, and thus distant signals have lower amplitudes compared with proximal signals. Speakers seem to be tacitly aware that, as the distance between themselves and listeners is increased, vocal intensity must also be increased to maintain effective communication (Johnson et al. 1981; Pelegrín-García et al. 2011). Thus low-amplitude speech signals elicit high-amplitude responses from the listener. Under these conditions, adaptively adjusting speech amplitude is a cooperative act; it facilitates the reception of the signal by listener. It is presumed that the ability to change speech amplitude as a function of listener distance is a high-level social skill (Johnson et al. 1981; Pelegrín-García et al. 2011). The questions that motivated the present study are 1) is this cooperative, flexible vocal output peculiar to humans; and 2) is it necessarily linked to “higher” sociocognitive abilities, or can it be explained by a simpler mechanism?

We used marmoset monkeys to answer these questions. Marmosets readily engage in contact (“phee”) call turn-taking behavior with conspecifics, whether or not they are pair bonded (Takahashi et al. 2013). They can adjust (without any training) the timing of their vocalizations relative to the calls of conspecifics (Miller et al. 2009; Takahashi et al. 2013) and to avoid intermittent background noise (Egnor et al. 2007; Roy et al. 2011). The flexibility in their vocal output may derive in part from the robust interactions between the auditory cortex and neural circuits related to vocal production. These interactions occur under many different scenarios, including spontaneous vocalizations produced in isolation (Eliades and Wang 2003, 2005), across different vocalization types (Eliades and Wang 2013), and in response to altered vocal feedback (Eliades and Wang 2008). Nevertheless, there is no evidence to date that any monkey species can flexibly control their vocal output with reference to another conspecific, that is, in the same manner that humans adjust their speech amplitude as a function of listener distance. We hypothesized that vocal feedback may be used in such a scenario.

By broadcasting phee calls at two different amplitudes to marmoset monkeys, we simulated an out-of-sight conspecific at two different distances. Although there are multiple acoustic cues related to distance, intensity is the strongest cue and has been effectively used as the sole cue in distance-estimation experiments with humans (Ashmead et al. 1990; Rosenblum et al. 1987) and monkeys (Ghazanfar et al. 2002a; Maier et al. 2004). Figure 1 illustrates a simple model for cooperative vocal amplitude control. Increasing arousal levels “drive” the production of phee calls, whereas vocal feedback inhibits the drive. The strength of this inhibition varies according to the amplitude of the vocal signal. Higher amplitude/nearby vocal signals result in the emission of a lower-amplitude vocalization; lower-amplitude/far-away calls would not inhibit the drive as much, and, subsequently, the vocalization emitted by the listener would be higher in amplitude. Our behavioral data are consistent with the model predictions; marmoset monkeys produce louder phee calls in response to simulated conspecifics that sound farther away. Additionally, we observed that more distant-sounding phee calls elicit faster responses from marmosets. Thus “vocal cooperation” in marmosets is a system property that is the result of balancing the arousal states of vocally interacting individuals. We elaborate on our model to include both the amplitude and timing effects we observed.

MATERIALS AND METHODS

Subjects. The subjects used in the study were six (3 males) adult (3–9 yr), captive common marmosets (*Callithrix jacchus*) housed at Princeton University. Marmosets live with their pair-bonded mates in

* D. Y. Takahashi and A. A. Ghazanfar are cosenior authors.

Address for reprint requests and other correspondence: A. Ghazanfar, Princeton University, Princeton, NJ 08544 (e-mail: asifg@princeton.edu).

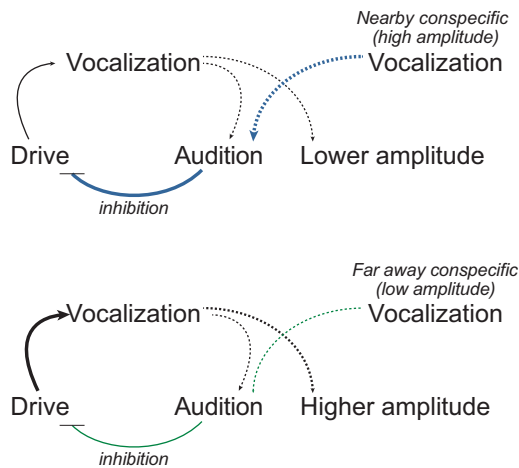


Fig. 1. A simple model of cooperative vocal-amplitude control. Each marmoset's call dynamics is represented by 3 components: audition, drive, and vocalization. All the interactions are excitatory, except the interaction from audition to the drive, which is inhibitory. *Top*: hypothetical neural dynamics for the response to nearby (high-amplitude) calls. *Bottom*: hypothetical neural dynamics for the responses to far-away (low-amplitude) calls.

family groups; all were born in captivity. They had ad libitum access to water and were fed once daily with standard commercial chow supplemented with fruits and vegetables. All experiments were performed in compliance with the guidelines of, and were approved by, the Princeton University Institutional Animal Care and Use Committee.

Experimental setup. Subjects were taken from their home cage to the experiment room in a transfer cage. In the experiment room (floor area, 2.5×2.5 m), the subject was then transferred to a testing box that rested on a table (0.66 m in height) in one corner (Fig. 2A). The testing box was made of Plexiglas and wire in the form of a triangular prism ($0.30 \times 0.30 \times 0.35$ m). In a diagonally opposite corner, a speaker was positioned on top of an identical table. An opaque curtain made of black cloth divided the room to visually occlude the subject from the speaker, thereby simulating an out-of-sight conspecific. Two digital recorders (ZOOM H4n Handy Recorder) were placed directly in front of the testing box at a distance of 0.76 m. One recorder was used to record the subject vocalizations, and the microphone output of the other recorder was connected to the computer to provide feedback parameters to the interactive playback software. Audio signals were

acquired at a sampling frequency of 96 kHz. Because of the design of the testing box, the direction of the marmoset's vocalizations was typically forward, in the direction of the microphone. That is, only one side of the testing box was visually clear and had a wire grid that the marmoset could grasp and look out toward the room. The microphone was directed toward this side of the testing box.

Interactive playback design. To capture the timing and contingency of vocal interactions most effectively, we employed a real-time interactive playback design. This design aims to simulate vocal exchanges between marmosets that are out of visual contact with conspecifics (Takahashi et al. 2013). Interactive playbacks can robustly elicit vocal output from marmosets (Miller et al. 2009), and we based the design of our MATLAB-based software on prototype software shared with us courtesy of Cory Miller, University of California, San Diego, CA.

We used six phee calls as our stimulus set, three from each of two adult individuals (one female and the other male). These individuals were familiar to the subjects, as they were housed in the same room (but not the same home cage). Each phee call had two versions: high amplitude (50–60 dB at 2.04 m from the source) and low amplitude (20–35 dB at 2.04 m from the source). Call amplitude is known to be a reliable cue for sound-source distance in a number of species (Ashmead et al. 1990; Gamble 1909; Ghazanfar et al. 2002a; Maier et al. 2004; Naguib 1997; Nelson 2000; Rosenblum et al. 1987). Thus high-amplitude phee calls simulated a nearby conspecific, and low-amplitude phee calls simulated the same conspecific as being far away (Fig. 2B). Sounds produced at different distances are subject to frequency-amplitude-dependent changes in their spectral profiles and can be used as distance cues (Naguib and Wiley 2001). Consistent with this phenomenon, Fig. 2C shows that the stimuli are more attenuated in the high-frequency range, as can be seen by the absence of a visible second harmonic in the far-away stimulus. The exact distance between the source and the listener represented by different sound amplitudes depends on several factors such as geometry of the source, humidity, temperature, scattering factors, etc. If we assume that a source produces a sound pressure of 90 dB at 0.1 m, which is a typical sound pressure for adult marmoset phee call, and apply the inverse distance law for sound pressure and the damping of air at 7 kHz (ISO 9613-1:1993, 27°C, 50% relative humidity, standard pressure at sea level), the high- and low-amplitude stimuli simulate the phee calls of a conspecific located ~ 5 m and ~ 60 m from the subject (Fig. 2B). To mask acoustic disruptions external to the testing room, pink noise was broadcast at ~ 40 dB (-15 dB at 7 kHz) throughout each session through a separate speaker. Figure 2C shows examples of

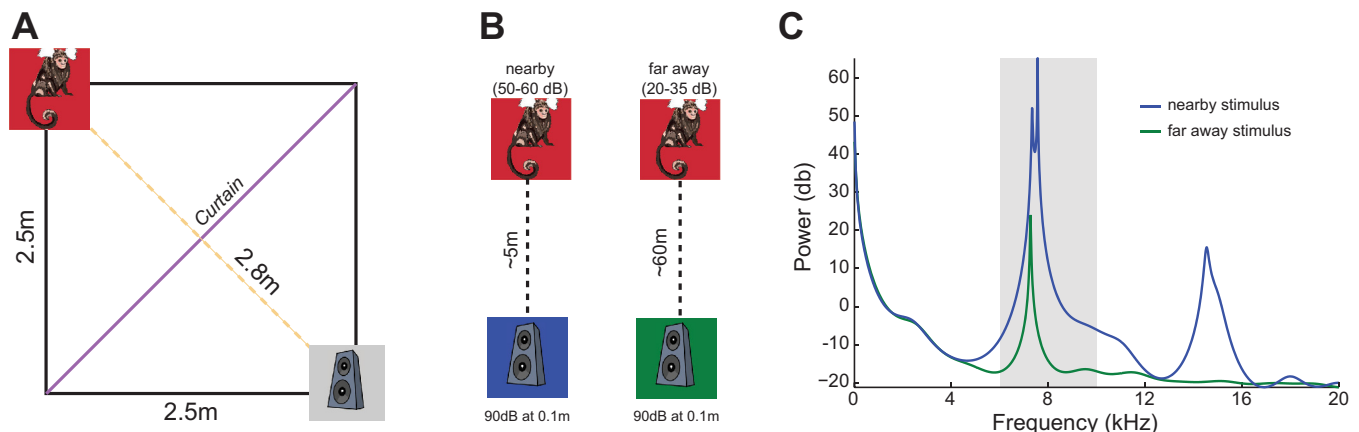


Fig. 2. Experimental setup. *A*: dimension of the experiment room and position of the subjects. The speaker was located in the opposite side of the room. *B*: approximate distance between a source and the listener, simulated by the 2 different sound-stimulus amplitudes. The high-amplitude sound simulated a distance of ~ 5 m from a source producing sound of 90 dB at 0.1 m. The low-amplitude sound simulated a distance of ~ 60 m from the same source. *C*: power spectrum of the stimulus sounds recorded by the microphone at 0.76 m from the subjects. The nearby/high-amplitude (blue line) and far-away/low-amplitude (green line) stimuli are salient compared with the background noise (room + masking noise). The gray region indicates the fundamental frequency range of the typical phee call (6–10 kHz).

the power spectra of typical low- and high-amplitude stimuli embedded in the background noise (room and pink noise), recorded by the microphone. This shows that both stimuli were audible to the subjects.

Testing began with a 6-min habituation period followed by the initiation of vocal interactions through the playback of a vocalization. Subsequently, the software played back eight vocalizations (4 high and 4 low amplitudes) randomly selected from a pool of three calls from one subject. Each vocalization was played contingent on detecting the vocalization of the subject and with response latency of 2–5 s. A vocalization by the subject following this playback was considered contingent if it occurred within 9 s from the end of the preceding playback stimulus. This is consistent with previous studies that demonstrate that the majority of marmoset call responses occur within 9 s and that the peak of response probability is around 3–5 s (Miller et al. 2009; Takahashi et al. 2013). If the subject failed to respond, then the software played another call 15–20 s after the preceding playback stimulus. The temporal parameters used by the playback software were consistent with naturalistic vocal interactions between marmosets (Takahashi et al. 2013).

We kept the total number of stimuli to eight per session and two sessions for each subject to avoid any habituation that could mask the effect of our experimental manipulation. After all eight stimuli were played, we recorded a 3-min baseline to be able to compare the rate of responses for different animals and sessions. Each subject participated in two sessions, one with familiar male calls and the other with the calls of a familiar female.

Data analysis. We manually validated and screened phee-call responses based on timing using Adobe Audition (version 1.5). The overwhelming majority of calls produced by marmosets, when visually occluded from their conspecifics, are phee calls (Takahashi et al. 2013). In the case that one playback stimulus elicited multiple consecutive phee calls before the next stimulus was presented, only the first phee was considered as the response. When there was an overlap between the stimulus and the vocal response, the response was counted as long as the stimulus onset preceded the response onset. The minimum interval from stimulus onset to vocal-response onset in our data was 1.76 s.

For each vocal response, we measured amplitude, response latency, rate, syllable number, overall call duration, and spectral structure (harmonic ratio, maximum, minimum, mean, and peak frequencies; DiMattina and Wang 2006). The amplitude of each call was the peak amplitude of recorded signal in decibels. The amplitude was calibrated using the room noise level (~40 dB), which was measured using a sound-level meter (BK Precision 732 A). Vocal-response latency was defined as the time delay between the offset of the preceding playback stimulus and the onset of the subsequent vocal response. Response rate was calculated as the total number of phee responses from the subject within a condition; the maximum number of responses in a condition was four, and the minimum was zero. The number of syllables was measured by inspection of spectrograms, and duration measures were made using cursor-based selection of vocalizations. The harmonic attenuation of each call was calculated as the average of 10 times the logarithm of the ratio between the amplitude of f_0 and the amplitude of first harmonic of the call. The maximum, minimum, and mean frequencies were calculated as the maximum, minimum, and mean values of the f_0 frequency (between 6–10 kHz) of the call. The peak frequency was calculated as the value of the frequency at which the amplitude achieved its peak. All the values were obtained calculating the spectrogram of the call (sampling rate = 96 kHz, fast Fourier transform size = 1,024 points, overlap = 50%, Hanning window).

For statistical analysis of amplitude, latency, syllable number, and duration, we only analyzed the sessions where a direct comparison between far-away and nearby conspecifics was available within a session ($n = 8$ sessions). This selectivity was used because sensitive measures of individual vocal acoustics across sessions could not be correctly represented in the dataset when the calls from one session

contribute to only one condition but not to the other condition. The overall response rate to playback stimuli was 21.6%, which is consistent with a previous playback experiment (Miller et al. 2009). The low response rate meant that there was a large number of missing data from our design. This precluded statistical analyses using ANOVA or mixed-effects models. ANOVA was not an option for the current analysis given the unbalanced and repeated-measure (repeated subjects) structure of our data set. A mixed-effects model was also not appropriate, as the small number of repetitions (only 2, to minimize habituation effects) and the large number of missing data (again, marmosets responded to 21.6% of stimuli) did not allow reliable modeling using this method.

Taking the average values of the measurements (amplitude, response time, etc.) per session was the best option for our data structure; this statistical approach is consistent with playback studies where one is measuring natural vocal responses that are highly variable in terms of rate of production. We calculated the mean of the call acoustic measures in each session and condition and performed paired t -tests to examine the statistical significance between comparable conditions. To calculate the effect size for the paired t -tests, we calculated the difference between the means and divided by the standard deviation of the differences between matching pairs of data (Cohen 1988). To verify the appropriateness of the use of paired t -tests on our data, for each comparison in our study, we applied a Jarque-Bera test of normality of our data distribution. In all cases, there was no evidence of significant deviation from normality ($P > 0.05$). To test whether the responses to nearby and far-away stimulus have significantly different characteristics, we applied a linear discriminant analysis using MATLAB. For the correlation analysis, we calculated the Cook's D for linear regression for each point. A data point with Cook's D value higher than 1 is considered an outlier (Cook and Weisberg 1982).

Call production in background noise experiment. We recorded spontaneous phee-call production in three different background noise levels: 40, 50, and 60 dB. In each session, each level of noise was played five times, each time for a period of 1 min. The order in which the different noise levels were presented was randomized. We used five subjects; all subjects participated in the interactive playback described above. Each subject was tested in four sessions. Before the experiment started, we recorded each subject for 5 min in silence (only room noise) to allow the subjects to habituate to the room. We measured the call amplitude for each condition and compared the averages. For robustness, we tested three different methods to correct for background noise, and all of them resulted in similar results. The first correction method was similar to the method described in Brumm et al. (2004). We first band passed the recorded signal at 5 to 9 kHz, and then the sound levels were converted to pressure units. A simple difference of the call pressure level and background pressure level gives the corrected pressure. The corrected pressure then was converted to sound level. In the second method, we played several calls at known sound levels and then played the three background noises. We then converted the sound levels to pressure and used a linear regression to estimate the change in pressure level caused by the introduction of noise at different levels. For the third method of correction, we used a Wiener filtering to denoise the signal using the spectral profile of the recorded background noise. We present the data based on the first method, corrected amplitude of recorded calls by the level of background noise (Brumm et al. 2004).

Model of vocal interactions. To elaborate on our findings, we developed a model to account for our findings that marmosets adjust both their vocal amplitude and timing according to the simulated distance from a conspecific. In the model, neural activity (excitatory-inhibitory activity) in the nodes "drive", "auditory", and "motor" are denoted by $D(t)$, $A(t)$, $M(t)$, respectively. Null neural activity implies balanced inhibitory and excitatory activities. Noise (or uncertainty) generated in this circuit is denoted by $e(t)$. The noise $e(t)$ was simulated as zero-mean independent Gaussian with a standard devi-

ation of 0.5. The initial conditions are $D(0) = A(0) = M(0) = 0$. Let α be a positive constant that represents how fast $D(t)$ can increase. The dynamics of the activities before drive reaches the threshold T , i.e., $D(t) < T$, is described by $D'(t) = \alpha|D(t)| - A(t) + e(t)$, $A(t) = 0$, $M(t) = 0$, where the primes on the functions [e.g., $D'(t)$] indicate time derivative and $| \cdot |$ indicates the absolute value. The above set of equations describes the dynamics before receiving or producing a call. The time when $D(t)$ reaches the threshold represents the onset of the call production. In our simulation, we set $\alpha = 0.0003$ and $T = 1,000$. Observe that because $\alpha|D(t)|$ is positive, when the activity in auditory, $A(t)$, is small, the activity in drive increases. This results in the generation of spontaneous calls. The larger the value of parameter α is, the faster the spontaneous call-production rate.

For the 1.5 s after reaching the threshold, the dynamics of the model is described by $D'(t) = \alpha|D(t)| - A(t) + e(t)$, $A(t) = 0$, $M(t) = D(t) - T$. The above set of equations describes the dynamics during call production. The magnitude of $M(t)$ describes the amplitude of the call.

After 1.5 s, the time when $D(t)$ reached the threshold, $D(t)$ is reset to 0, and the dynamics is started again. This reset represents the effect of self-vocalization on $D(t)$.

When a call input $I(t)$ is received, the auditory activity $A(t)$ increases proportionally to the magnitude (amplitude) of the input and then decays exponentially. Specifically, we have $D'(t) = \alpha|D(t)| - A(t) + e(t)$, $A'(t) = -bA(t) + I(t)$, $M(t) = 0$, where b is positive constant. We used $I(t) = 0.0002$ and $I(t) = 0.0001$ to simulate the nearby (loud) and far-away (soft) calls. The duration of the input was 3 s, and $b = 0.001$ in both cases. If b is small, the activity in $A(t)$ lasts for a longer duration. If both α and b are small, the model will simulate a marmoset with a low call-response rate. The Euler-Maruyama method was applied, with a step size equal to 0.01, to generate the activities in $D(t)$ and $A(t)$.

RESULTS

To determine whether or not cooperative control of vocal amplitude was unique to humans (Healey et al. 1997; Johnson et al. 1981; Markel et al. 1972; Michael et al. 1995; Pelegrín-García et al. 2011), we investigated whether marmoset monkeys could flexibly adjust their vocal amplitude in response to simulated conspecific distance, that is, as a function of high- vs. low-amplitude phee-call playbacks where amplitude is used as a reliable cue for distance (Ashmead et al. 1990; Gamble 1909; Ghazanfar et al. 2002a; Maier et al. 2004; Naguib 1997; Nelson 2000; Rosenblum et al. 1987) (Fig. 2). There were at least three possible outcomes to our experiment. First, marmosets may not change their phee response at all as a function of simulated conspecific distance. This may be because they either cannot accurately relate call amplitude with conspecific distance and/or because they cannot control their vocal amplitude. The latter is unlikely, however, as both marmosets and tamarins can raise their call amplitudes in noisy environments, the Lombard effect (Brumm et al. 2004; Egnor and Hauser 2006; Eliades and Wang 2012). Second, marmosets could raise the amplitude of their phee calls in response to nearer, louder conspecifics. This would be akin to the competitive vocal response observed in mammals, birds, and anurans during competitive encounters with conspecifics (Brumm and Todt 2004; Gerhardt and Huber 2002; Wyman et al. 2008). Finally, marmosets may cooperatively adjust their phee-call amplitude by calling more loudly to conspecifics that are perceived to be located at a greater distance. This would be consistent with cooperative vocal turn-taking behavior observed between any marmosets (whether they are related or unrelated, pair bonded

or not pair bonded) (Takahashi et al. 2013) and with our model of vocal-feedback inhibition (Fig. 1).

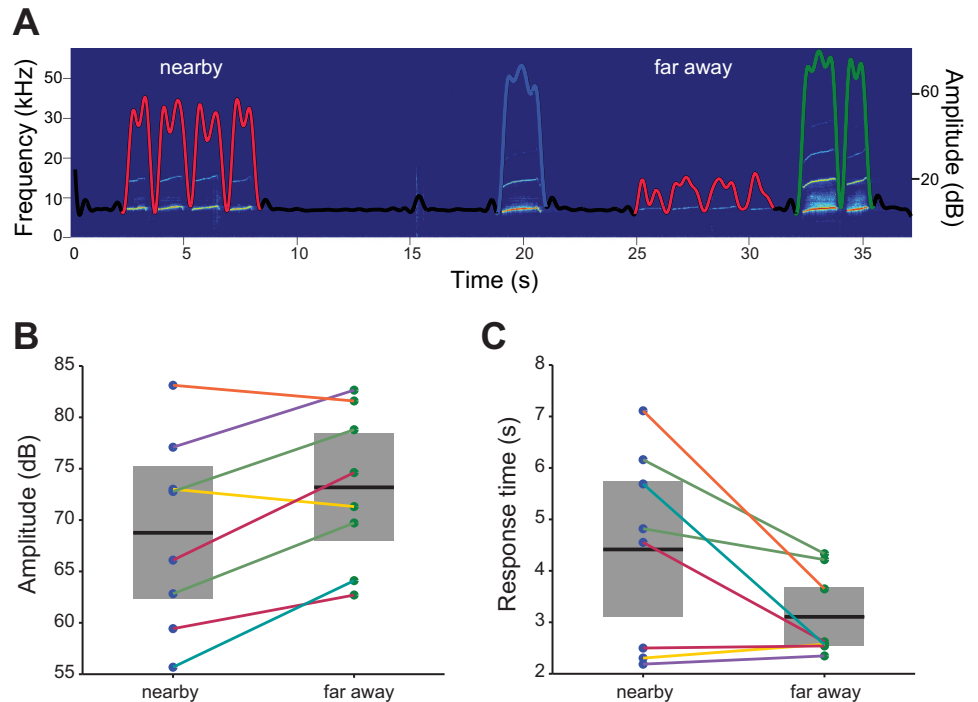
Consistent with playback studies where one is measuring natural vocal responses whose rates are highly variable, we averaged our measurements across sessions. In our study, the overall response rate to playback stimuli was 21.6%, which is consistent with a previous playback experiment (Miller et al. 2009) and with naturalistic vocal interactions in captivity (Takahashi et al. 2013). We only analyzed the sessions where the direct comparison between far-away and nearby conspecifics was available within a session ($n = 8$ sessions). The number of calls for each session was 2, 3, 1, 4, 1, 3, 3, and 1 for nearby and 3, 2, 3, 3, 2, 3, 1, and 3 for far away, respectively. We used paired t -tests to examine the statistical significance between comparable conditions because the low response rate precluded the use of ANOVA or mixed-effects models (see MATERIALS AND METHODS). To provide an accurate assessment of the robustness of our statistics, we report effect sizes (Cohen 1988).

We found that marmoset monkeys cooperatively adjust their vocal amplitude in accordance with simulated conspecific distance and our model. Figure 3A shows two phee-call responses from a single session, the first to a nearby stimulus (blue line represents the response amplitude) and the second to a far-away stimulus (green line represents the response amplitude). The response to the far-away stimulus was louder than the response to the nearby one. This can be seen by the difference in heights between the blue and green lines and also by the salient third harmonic observed for the response to far-away stimulus but not for the nearby stimulus. This pattern was statistically significant across the population and with a very large effect size ($n = 8$ sessions, paired t -test, $P = 0.018$, effect size = 1.08); responses to far away were on average 4.43 dB louder (far away 73.19 ± 2.68 dB vs. nearby 68.76 ± 3.30 dB) (Fig. 3B). The difference in response-call amplitudes can be further appreciated when the amplitude is measured relative to the baseline amplitude, i.e., when we subtract the baseline amplitude. The mean relative response amplitude to nearby stimuli is 2.96 ± 3.93 dB and to far-away stimuli is -1.47 ± 4.28 dB ($n = 8$ sessions, paired t -test, $P = 0.018$, effect size = 1.08).

The phee-call responses in Fig. 3A also have different latencies; the marmoset subject responded more quickly following the far-away call relative to the nearby call. This pattern also held across the group (Fig. 3C). Responses to far-away calls (3.11 ± 0.29 s) were, on average, 1.31 s faster than responses to nearby calls (4.42 ± 0.67 s). This difference was statistically significant and, again, with a large effect size ($n = 8$ sessions, paired t -test, $P = 0.042$, effect size = 0.88). These faster and slower response latencies are within the species-typical range reported for naturalistic interactions (Takahashi et al. 2013).

To determine whether one could predict conspecific distance based on the amplitude and latency of subject responses, we performed a linear discriminant analysis. Figure 4A shows that higher amplitude and faster vocal output, i.e., data points above the magenta line, tend to be in response to far-away conspecifics (and vice versa for nearby conspecifics); this distinction can be correctly classified at an 81.25% level. In other words, the subjects' responses have enough information to recover the distance of the conspecific. Figure 4B shows that the changes

Fig. 3. Responses to nearby and far-away stimuli. *A*: exemplar of spectrogram and amplitude of responses to the nearby and far-away stimuli; x-axis, time in seconds; y-axis, frequency in kilohertz. The response call to nearby stimulus (blue) had smaller amplitude than the response call to far-away stimulus (green). *B*: amplitude of the response calls to the nearby (blue circles) and far-away (green circles) stimuli; y-axis, average amplitude in decibels. Different sessions are indicated by lines with different colors joining 2 dots. Sessions from the same subject have the same color. The gray rectangles indicate the 95% confidence intervals. The horizontal black lines indicate the mean amplitude values. *C*: response time to loud and soft stimuli; y-axis, average response time in seconds. The color convention is as in *B*.



in response amplitudes and latencies are highly correlated [following the removal of one outlier (orange plus sign)] ($n = 7$ sessions, Spearman correlation = -0.89 , $P = 0.012$). The outlier has Cook's $D = 1.53$ for linear regression, and all other data points have Cook's $D < 0.5$.

We also measured response rates, number of syllables, and call durations. One prediction is that, given that they respond more loudly and quickly to far-away conspecifics, marmosets may also increase their rate of calling, as that would also facilitate auditory contact. To compare different animals and sessions, we calculated the response rate as the number of responses divided by the amount of call produced during the baseline period. Figure 5A shows that there were no systematic differences in the rate of vocal responses between nearby and

far-away conditions (0.66 ± 0.19 vs. 0.76 ± 0.24 ; $n = 8$, $P = 0.604$, effect size = 0.19). Moreover, this result eliminates the potential confound that marmosets may have had trouble hearing the low-amplitude, far-away calls because of the background noise. If that were the case, then response rates should have been lower for that condition; they were not. To facilitate signal transmission over greater distances, marmosets could have also increased the number of syllables in their calls or the overall duration of their calls. For example, in two studies of the Lombard effect (whereby subjects increased the amplitude of their calls as function of background noise), both marmosets (Brumm et al. 2004) and tamarins (Egnor and Hauser 2006) also increased the syllable duration, thereby effectively increasing the overall duration of their calls. Figure 5B shows

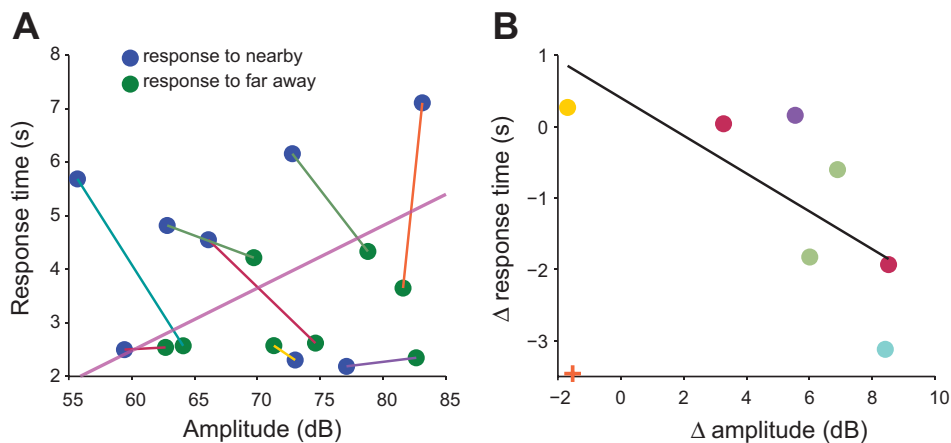


Fig. 4. Relationship between amplitude and timing of the responses to nearby and far-away stimuli. *A*: discriminant analysis for the responses to nearby and far away; x-axis, amplitude in decibels; y-axis, response time in seconds. The responses to nearby stimuli (blue circles) were concentrated on lower-amplitude/longer-response-time region, and the responses to far-away stimuli (green circles) concentrated on high-amplitude/shorter-response-time region. The different line colors indicate different animals. The magenta line is the linear discrimination line separating the responses to nearby from far-away stimuli. *B*: correlation between the magnitude change in amplitude and response time; x-axis, magnitude change in decibels; y-axis, magnitude change of response time in seconds. The circles indicate different sessions, and the different colors indicate different animals. The orange plus sign indicates the outlier data point. The linear regression line (black) was obtained by applying a robust regression to the data set excluding the outlier.

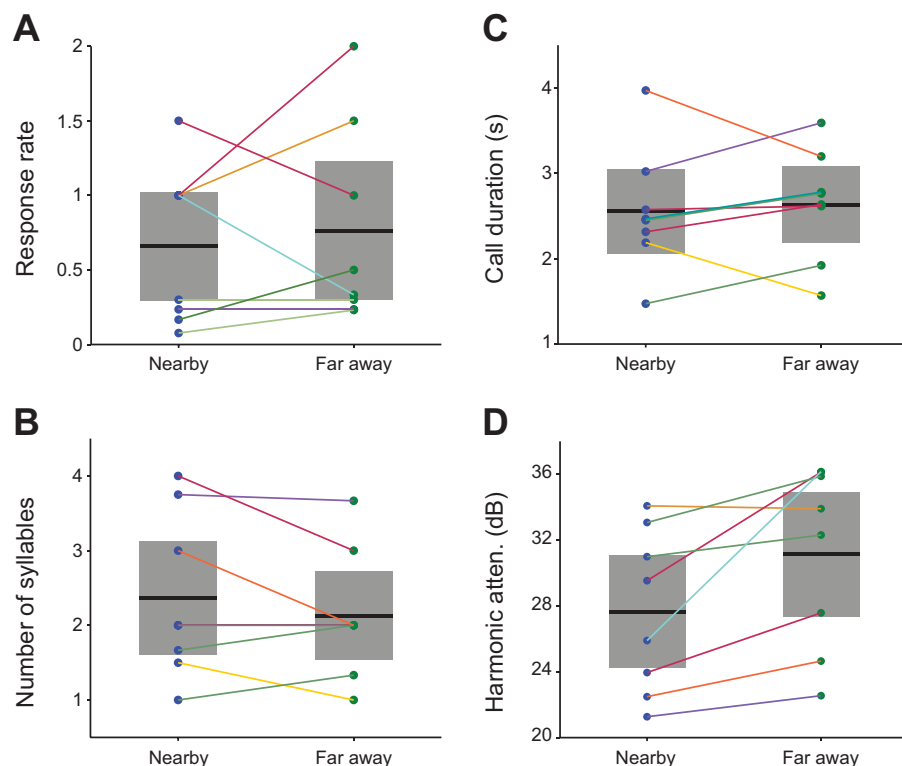


Fig. 5. Responses to nearby and far-away stimuli measured by other acoustic parameters. *A*: response rate to nearby and far-away stimuli; y-axis, response rate = number of responses divided by number of calls produced during baseline. The color conventions are the same as in Fig. 3*B*. *B*: number of syllables of the response calls for nearby and far-away stimuli; y-axis, average number of syllables of the response calls. The color conventions are the same as in *A*. *C*: call duration for nearby and far-away stimuli; y-axis, average call duration in seconds. Color conventions are the same as in *A*. *D*: harmonic attenuation of response calls to nearby and far-away stimuli; y-axis, harmonic attenuation = $10 \cdot \log(\text{amplitude of } f_0) - 10 \cdot \log(\text{amplitude of first harmonic})$. Color conventions are the same as in *A*.

that the number of syllables in phee-call responses is the same in nearby and far-away conditions (2.36 ± 0.39 vs. 2.12 ± 0.30 syllables; $n = 8$ sessions, $P = 0.248$, effect size = 0.45). Similarly, Fig. 5*C* shows that the duration of the response calls are the same in nearby (2.56 ± 0.25 s) vs. far away (2.63 ± 0.23 s) conditions ($n = 8$ sessions, $P = 0.677$, effect size = 0.15).

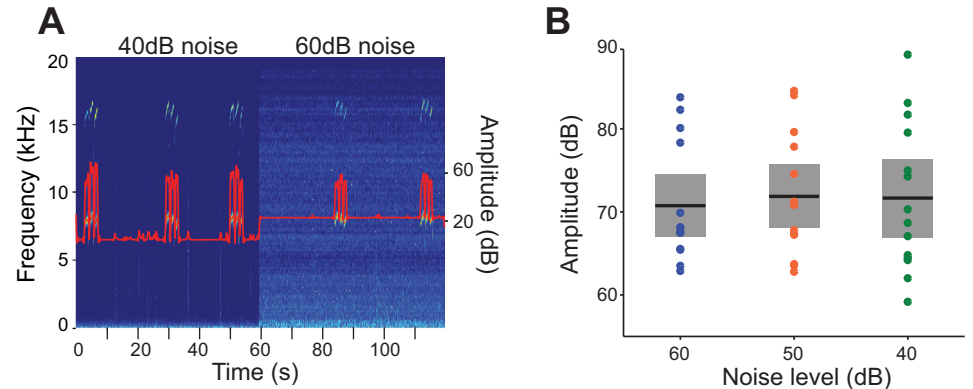
We also investigated whether or not the marmosets changed the spectral structure of their vocal responses as a function of simulated distance from listeners. We measured five acoustic features previously used to characterize marmoset calls (DiMattina and Wang 2006): harmonic attenuation, the maximum, minimum, mean, and peak frequencies. Figure 5*D* shows that the harmonic attenuation was weaker for nearby (27.66 ± 1.74 dB) than for far-away (31.14 ± 1.94 dB) conditions ($n = 8$ sessions, $P = 0.0229$, effect size = 1.03). No differences between the nearby vs. far-away conditions were found for measures of max frequency (8.44 ± 0.23 kHz vs. 8.31 ± 0.26 kHz; $n = 8$ sessions, $P = 0.628$, effect size = 0.18), mean frequency (7.42 ± 0.17 vs. 7.37 ± 0.15 kHz; $n = 8$ sessions, $P = 0.368$, effect size = 0.34), minimum frequency (6.10 ± 0.20 vs. 6.10 ± 0.12 kHz; $n = 8$ sessions, $P = 1$, effect size = 0), and peak frequency (7.61 ± 0.17 vs. 7.50 ± 0.14 kHz; $n = 8$ sessions, $P = 0.225$, effect size = 0.47).

Our hypothesized vocal-feedback mechanism for cooperative vocal control is only valid if it operates with some specificity within the sensory statistics appropriate for vocal communication (Sober and Brainard 2012). To test this, we investigated whether or not, in the context of social isolation, marmoset phee-call amplitude was modulated by different environmental noise levels in the same manner as it is by conspecific vocalization levels. There are three possible outcomes for this experiment. First, if it is simply the level of auditory input (regardless of signal type: phee call vs. noise)

that modulates the vocal amplitude, then higher noise levels should elicit lower amplitude calls (and vice versa). This would indicate a vocal-feedback mechanism but one that is more general and not specific to cooperative vocal exchanges. A second possibility is that marmosets raise the amplitude of their calls as background noise levels increase (the Lombard effect). This would be the opposite pattern of vocal output modulation than what we observed during vocal exchanges and would thus suggest that two different vocal-feedback mechanisms are at play. A final possibility is that, in the context of social isolation, marmosets are in a maximally aroused state (high drive; Fig. 1) regardless of background noise levels and they cannot modulate their vocal amplitude at all (i.e., they are at the “ceiling” of their vocal-output intensity). This would be broadly consistent with our proposed vocal-feedback mechanism as one that is specific for phee calls vs. background noise (Fig. 1). Thus the second and third putative outcomes would support our hypothesis that the proposed vocal-feedback mechanism is for cooperative vocal exchanges, whereas the first possibility would preclude it in favor of a more general mechanism.

We recorded marmoset phee-call production in three different background noise levels: 40, 50, and 60 dB. Our data support the third outcome; in the context of social isolation, marmosets are unable to modulate their phee-call amplitude as a function of background noise. The average call amplitudes were 70.76 ± 1.91 dB for the 60-dB condition, 71.87 ± 1.95 dB for the 50-dB condition, and 71.67 ± 2.40 dB for the 40-dB condition (Fig. 6). There was no statistical difference between the call amplitudes ($n = 14$ for each condition, repeated-measure ANOVA, $P = 0.836$, eta squared = 0.004). Under these conditions, the marmosets are calling at the ceiling of their ability to change vocal intensity; they are in a maximally aroused state. Consistent with this notion (and our model; Fig.

Fig. 6. Vocal production in noisy background. *A*: exemplar of spectrogram and amplitude (red line) of calls produced during 40-dB and 60-dB background noise; *x*-axis, time in seconds; *y*-axis left, frequency in kilohertz; *y*-axis right, amplitude in decibels. *B*: Amplitude of calls produced during 60-dB (blue circles), 50-dB (orange circles), and 40-dB (green circles) background noise; *x*-axis, noise level in decibels; *y*-axis, amplitude in decibels. The gray rectangles indicate the 95% confidence intervals, and the horizontal black lines indicate the mean relative amplitude values.



1), the average call amplitudes were roughly the same for all background noise levels (range: ~ 71 – 72 dB), which are very similar to the sound level of vocal responses to simulated far-away conspecifics (~ 73 dB; Fig. 3*B*). On the basis of these data, we conclude that the vocal-amplitude changes in response to differences in social distance are specific to vocal exchanges.

Having established this specificity, we elaborated on our simple model (Fig. 1) to present a version that accounts for both the vocal-amplitude and timing changes exhibited by marmosets as a function of simulated conspecific distance. On the basis of what is known about the neural circuitry related to primate vocal production (Jürgens 2002), we assumed that there are three basic nodes that each represent different groups of brain areas related to 1) motor production of vocalizations (e.g., periaqueductal gray area, reticular formation, nucleus retroambiguus, etc.), 2) the drive to produce such vocalizations (e.g., limbic structures such as the anterior cingulate cortex, amygdala, hypothalamus, etc.), and 3) the vocal feedback necessary to monitor one's own and others' vocalizations. There are known feedforward and feedback anatomical connections between all three groups of areas, but the dynamics of interactions between the areas are not understood at all. Here we provide model predictions that account for our behavioral results.

To show how the conceptual model in Fig. 1 can explain the modulation of the vocal amplitude and timing as a function of simulated conspecific distance, we put forth the following simple model for the dynamics of interaction (Fig. 7). The

activity (a balance of excitation and inhibition) in the drive node varies through time depending on both internal (e.g., stress levels) and external factors (e.g., social context). Whenever the activity in drive reaches a threshold, the motor node produces a vocalization. The loudness of the vocalization depends on the magnitude of activity in the drive when it exceeds this threshold. The auditory node, upon reception of a vocal signal from a conspecific, inhibits the activity in drive, and as such it controls how high the drive activity can go. Therefore, if the auditory node receives a high-amplitude call, it will strongly inhibit the drive and, as a result, the vocal response will be lower in amplitude. The opposite effect is observed if the auditory node receives a low-amplitude call. This model also accounts for the differences in the timing of vocal responses (Fig. 7). When the inhibition from the auditory node onto the drive is strong (e.g., following a high-amplitude, nearby vocalization), the activity in drive will take longer to reach threshold. By contrast, when that inhibition is weaker (e.g., following a lower amplitude, far away vocalization), the drive will reach threshold faster.

DISCUSSION

Humans can cooperatively adjust the amplitude of their speech signals in accordance with the perceived distance of the listener. They produce louder speech signals toward listeners perceived to be at longer distances, thereby coordinating communicative exchanges (Healey et al. 1997; Johnson et al. 1981;

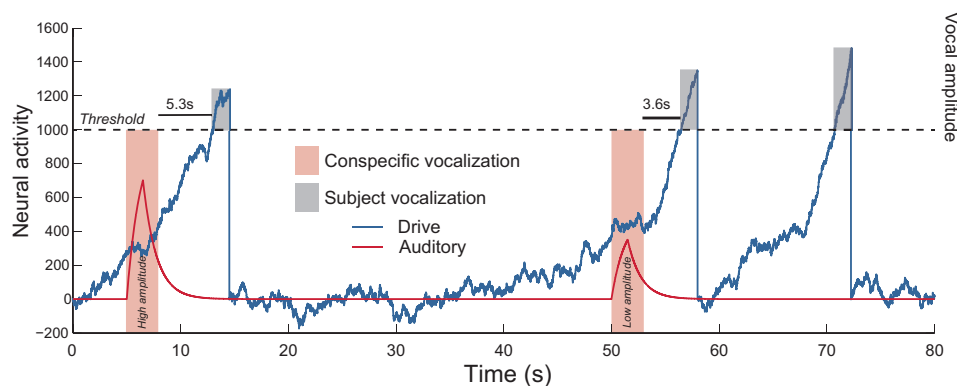


Fig. 7. An elaborated model of marmoset vocal interactions as a function of vocal amplitude. The dynamics of activities in drive, auditory, and motor nodes are simulated. In the *x*-axis, time is in seconds; in the *y*-axis, neural activity is represented in arbitrary units. The blue line represents the activity in the drive node, and the red line represents the activity in auditory node. For visualization purposes, we multiplied the activity in auditory by a factor of 2,000. The threshold is indicated by the dashed line. The gray rectangles represent the motor node output (a vocalization). Their height represents the amplitude. The pink regions represent the heard conspecific calls.

Markel et al. 1972; Michael et al. 1995; Pelegrín-García et al. 2011). We investigated whether marmoset monkeys could similarly control the amplitude of their vocalizations toward conspecifics simulated (via playback) to be calling at two different distances. We found that marmosets, not only increase the amplitude of their phee calls toward more distant conspecifics, but also produce these calls at a shorter latency. Thus we conclude that marmosets can cooperatively control their phee-call production to coordinate and facilitate communicative exchanges. This is consistent with their vocal turn-taking behavior, which is strikingly similar to human conversations (albeit on a much slower timescale) (Takahashi et al. 2013).

It is often presumed that nonhuman primates (hereafter, primates) lack flexibility in vocal production (Seyfarth and Cheney 2010). This presumption is consistent with developmental studies of Old World monkeys that report that experience-dependent changes in the spectral or temporal features of vocalizations are not as dramatic as those seen in humans (Hammerschmidt et al. 2000; Owren and Dieter 1989; Seyfarth and Cheney 1986). However, measures of flexible vocal output can take many other forms, including flexibility as to what events elicit a vocalization and when to produce a vocalization. A number of studies demonstrate that macaque monkeys (*Macaca mulatta*) can be conditioned to vocalize in response to experimenter cues, suggesting that they have some control over what to vocalize toward and when to do so (Aitken 1981; Coudé et al. 2011; Hage et al. 2013; Hihara et al. 2003; Sutton et al. 1973). These effects require considerable training. Our data are consistent with other demonstrations of vocal control (without training or reward) by marmoset and tamarin monkeys. Both can, not only adjust the timing of their vocalizations relative to the calls of conspecifics (Ghazanfar et al. 2001, 2002b; Miller and Wang 2006; Takahashi et al. 2013) and engage in extended series of vocal exchanges that are conversation like (Takahashi et al. 2013), but also adjust the timing of their calls to minimize interference with intermittent background noise (Egnor et al. 2007; Roy et al. 2011). We found that marmosets respond with a shorter delay when the listener is simulated to be far away vs. nearby. Therefore, marmosets could be using the response latency as a cue to localize conspecifics. This is a strategy that would be beneficial in a noisy environment.

We demonstrated that marmosets can control the amplitude and timing of their vocalizations according to changes in social context (as opposed to noise levels) and do so in a cooperative fashion (to increase the likelihood that a conspecific will hear the call). We note that the observed change in call amplitude (~4.5 dB) does not compensate fully for the change in stimulus amplitude (Fig. 3A). Therefore, it is possible that the marmosets are using other cues in addition to amplitude to compensate for the perceived distance. In addition to amplitude and timing, it is possible that marmosets could modulate additional features of their response calls as a function of simulated listener distance (Naguib and Wiley 2001). We observed a significant change in harmonic attenuation, which indicates that marmosets can modulate fine spectral features of their vocalizations. However, we did not observe changes in the numbers of syllables, call durations, or other types of spectral structure in their vocal responses. This stability in certain call features is consistent with the idea that some

acoustic characteristics are used to convey different types of information other than distance, e.g., group identities and sex (Miller et al. 2010b). Taken together, this suggests that marmosets can flexibly manipulate certain acoustic features while holding others constant to communicate separate channels of information concurrently.

We investigated the possibility that any loudness change could be equivalent to the changes in vocal loudness as a function of conspecific distance. If this were the case, then our claim and proposed vocal-feedback mechanism would not be specific to vocal exchanges; it would instead reflect a more reflexive mechanism. This also relates to the nature of vocal feedback in general; it is only helpful if it is accurate and within the sensory statistics appropriate for vocal communication (Sober and Brainard 2012). If the feedback system cannot distinguish between background noise and a conspecific vocal signal, then it is not particularly useful for communication. To be effective, the vocal-feedback mechanism must facilitate the decision of whether or not to modify vocal behavior based on sensory feedback (and risk adapting to signals that do not accurately reflect performance) or ignore sensory input (and risk leaving errors uncorrected) (Sober and Brainard 2012). We investigated this by measuring phee-call amplitude levels in background noise of different intensities. We found that, in the context of social isolation, marmosets did not modulate their phee-call amplitude as a function of background noise. This is consistent with our proposed vocal-feedback mechanism as one that is specific for phee calls vs. background noise. We therefore hypothesize that the vocal-amplitude changes in response to differences in social distance are specific to vocal exchanges; it is cooperative vocal control.

In humans, adjusting speech amplitude in accordance with listener distance is thought to be a high-level sociocognitive skill (Johnson et al. 1981; Pelegrín-García et al. 2011). It is thought that the speaker must know something about the physical relation between sound and distance and must also have the communication disposition to make compensations that will serve the listener's needs. However, it need not be the result of a high-level process. We and others argued that the similar cooperative breeding strategies of humans and marmosets may have conferred upon them a greater tendency to be prosocial than other primates (Borjon and Ghazanfar 2014; Burkart and van Schaik 2010; Burkart et al. 2009; Snowdon and Cronin 2007). We hypothesize that this may have led to a greater capacity for cooperative communication (Borjon and Ghazanfar 2014; Takahashi et al. 2013). However, cooperative communication need not require human-like sociocognitive skills or the big brains that such skills seemingly require (Barrett and Rendall 2010; Barrett et al. 2007; Borjon and Ghazanfar 2014). That said, our model cannot fully account for the complexity of human communication. For example, the length of utterances is highly variable in human conversations (but not in marmoset phee-call exchanges), and this relates to the duration of inspiration (Whalen and Kinsella-Shaw 1997). Moreover, semantics, syntax, and a variety of other content-related cues (that are absent in marmoset calls) can influence human vocal exchanges and affect their timing (Winkworth et al. 1995).

We propose that the vocal-amplitude control exhibited by marmosets in our study is property of arousal-based "drive-

auditory” interactions across conspecifics (Owren and Rendall 2001; Owren et al. 2011). A simple model can account for this whereby vocal feedback inhibits the drive to produce a vocalization; the strength of this inhibition is inversely proportional to the amplitude of the call produced (Fig. 1). Our data were consistent with this model. However, our data also showed that marmosets were quicker to vocally respond to lower-amplitude, far-away sounds (Fig. 3). To account for both the amplitude and timing effects we observed, we elaborated on our model (Fig. 7). We demonstrated that changes in vocal amplitude could be mediated by the magnitude of drive activity upon exceeding the threshold to produce a vocalization; changes in vocal-response latency could be mediated by the time taken to reach that threshold. Vocal-feedback inhibition determines both the magnitude and speed of the drive activity. It is important to note that, because our model is a dynamic-system model, the change in activity in the auditory node can cause a perturbation in the trajectory of the entire system, and the effect of this change can be observed even after the cessation of the activity in auditory node, as can be seen in Fig. 7. The model also predicts that the exact amplitude and timing of response call will depend on the level of drive activity at the time of auditory input, which explains the variability of response-call amplitude and timing even within the same subject. Consistent with this model, higher arousal level (drive) has been associated with louder calls (Briefer 2012; Wyman et al. 2008) and faster responses (Christensen-Dalsgaard et al. 2002) in other species.

The exact neural mechanisms by which such vocal amplitude and timing control may be mediated by vocal feedback are not known but are certain to be complex. Our model is a first step toward understanding such mechanisms. The production and perception of vocalizations by marmosets engage a number of different brain areas (Miller et al. 2010a; Simoes et al. 2010). There is already considerable evidence for neural circuits that could support the proposed model. Neurophysiological, experimental lesion, and microstimulation in primates studies all implicate the anterior cingulate cortex as an important node for driving the production of vocalizations in accordance with arousal levels (Jürgens and Kirzinger 1982; Jürgens and Pratt 1979; Jürgens and von Cramon 1982) (for reviews, see Jürgens 1998; Paus 2001). There are direct connections between the anterior cingulate cortex and auditory association cortex in primates (Galaburda and Pandya 1983; Müller-Preuss and Jürgens 1976), auditory cortex and amygdala (Stefanacci and Amaral 2002), as well as between auditory association cortex and other frontal areas that may be involved in vocal exchanges (Miller et al. 2010a; Simoes et al. 2010) including the prefrontal cortex (Hackett et al. 1999; Romanski et al. 1999a, 1999b). Importantly, neurophysiological studies in marmosets, in particular, reveal that there are robust interactions between the auditory cortex and vocal-production systems under many different scenarios (Eliades and Wang 2003, 2005, 2008, 2013). An interesting prediction of the model is that the effect of hearing a vocalization should last even after its cessation. This is consistent with data from the marmoset auditory cortex (Bartlett and Wang 2005). Our data and model extend the role of vocal feedback to cooperative vocal communication.

ACKNOWLEDGMENTS

We thank Lauren Kelly for providing comments on a draft of this manuscript.

GRANTS

This work was supported by a Pew Latin American Fellowship (D. Y. Takahashi), a Brazilian Science without Borders Fellowship (D. Y. Takahashi), NIH R01NS054898 (A. A. Ghazanfar), and the James S. McDonnell Scholar Award (A. A. Ghazanfar).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

Author contributions: J.Y.C. and A.A.G. conception and design of research; J.Y.C. performed experiments; J.Y.C. and D.Y.T. analyzed data; J.Y.C., D.Y.T., and A.A.G. interpreted results of experiments; J.Y.C. and A.A.G. drafted manuscript; J.Y.C., D.Y.T., and A.A.G. approved final version of manuscript; D.Y.T. prepared figures; A.A.G. edited and revised manuscript.

REFERENCES

- Aitken PG. Cortical control of conditioned and spontaneous vocal behavior in rhesus monkeys. *Brain Lang* 13: 171–184, 1981.
- Ashmead DH, LeRoy D, Odom RD. Perception of the relative distances of nearby sound sources. *Percept Psychophys* 47: 326–331, 1990.
- Barrett L, Rendall D. Out of our minds: The neuroethology of primate strategic behavior. In: *Primate Neuroethology*, edited by Platt ML and Ghazanfar AA. New York, NY: Oxford University, 2010, p. 570–586.
- Barrett L, Henzi SP, Rendall D. Social brains, simple minds: Does social complexity really require cognitive complexity? *Philos T R Soc B* 362: 561–575, 2007.
- Bartlett EL, Wang X. Long-lasting modulation by stimulus context in primate auditory cortex. *J Neurophysiol* 94: 83–104, 2005.
- Borjon JI, Ghazanfar AA. Convergent evolution of vocal cooperation without convergent evolution of brain size. *Brain Behav Evol* 84: 93–102, 2014.
- Briefer EF. Vocal expression of emotions in mammals: mechanisms of production and evidence. *J Zool Lond* 288: 1–20, 2012.
- Brumm H, Todt D. Male-male vocal interactions and the adjustment of song amplitude in a territorial bird. *Anim Behav* 67: 281–286, 2004.
- Brumm H, Voss K, Kollmer I, Todt D. Acoustic communication in noise: regulation of call characteristics in a New World monkey. *J Exp Biol* 207: 443–448, 2004.
- Burkart JM, van Schaik CP. Cognitive consequences of cooperative breeding in primates? *Anim Cogn* 13: 1–19, 2010.
- Burkart JM, Hrdy SB, van Schaik CP. Cooperative breeding and human cognitive evolution. *Evol Anthropol* 18: 175–186, 2009.
- Christensen-Dalsgaard J, Ludwig TA, Narins PM. Call diversity in an Old World treefrog: level dependence and latency of acoustic responses. *Bioacoustics* 13: 21–35, 2002.
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum, 1988.
- Cook RD, Weisberg S. *Residuals and Influence in Regression*. New York, NY: Chapman & Hall, 1982.
- Coudé G, Ferrari PF, Rodà F, Maranesi M, Borelli E, Veroni V, Monti F, Rozzi S, Fogassi L. Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLoS One* 6: e26822, 2011.
- DiMattina C, Wang X. Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations. *J Neurophysiol* 95: 1244–1262, 2006.
- Egnor SER, Hauser MD. Noise-induced vocal modulation in cotton-top tamarins (*Saguinus oedipus*). *Am J Primatol* 68: 1183–1190, 2006.
- Egnor SER, Wickelgren JG, Hauser MD. Tracking silence: adjusting vocal production to avoid acoustic interference. *J Comp Physiol A* 193: 477–483, 2007.
- Eliades SJ, Wang X. Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J Neurophysiol* 89: 2194–2207, 2003.
- Eliades SJ, Wang X. Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cereb Cortex* 15: 1510–1523, 2005.

- Eliades SJ, Wang X.** Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* 453: 1102–1106, 2008.
- Eliades SJ, Wang X.** Neural correlates of the Lombard effect in primate auditory cortex. *J Neurosci* 32: 10737–10748, 2012.
- Eliades SJ, Wang X.** Comparison of auditory-vocal interactions across multiple types of vocalizations in marmoset auditory cortex. *J Neurophysiol* 109: 1638–1657, 2013.
- Galaburda AM, Pandya DN.** The intrinsic architectonic and connective organization of the superior temporal region of the rhesus monkey. *J Comp Neurol* 221: 169–184, 1983.
- Gamble EA.** Intensity as a criterion in estimating the distance of sounds. *Psychol Rev* 16: 416–426, 1909.
- Gerhardt HC, Huber F.** *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions*. Chicago, IL: University of Chicago, 2002.
- Ghazanfar AA, Flombaum JI, Miller CT, Hauser MD.** The units of perception in the antiphonal calling behavior of cotton-top tamarins (*Saguinus oedipus*): playback experiments with long calls. *J Comp Physiol A* 187: 27–35, 2001.
- Ghazanfar AA, Neuhoff JG, Logothetis NK.** Auditory looming perception in rhesus monkeys. *Proc Natl Acad Sci USA* 99: 15755–15757, 2002a.
- Ghazanfar AA, Smith-Rohrberg D, Pollen AA, Hauser MD.** Temporal cues in the antiphonal long-calling behaviour of cottontop tamarins. *Anim Behav* 64: 427–438, 2002b.
- Hackett TA, Stepniewska I, Kaas JH.** Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res* 817: 45–58, 1999.
- Hage SR, Gavrilov N, Nieder A.** Cognitive control of distinct vocalizations in rhesus monkeys. *J Cogn Neurosci* 25: 1692–1701, 2013.
- Hammerschmidt K, Newman JD, Champoux M, Suomi SJ.** Changes in rhesus macaque ‘coo’ vocalizations during early development. *Ethology* 106: 873–886, 2000.
- Healey EC, Jones R, Berky R.** Effects of perceived listeners on speakers’ vocal intensity. *J Voice* 11: 67–73, 1997.
- Hihara S, Yamada H, Iriki A, Okanoya K.** Spontaneous vocal differentiation of coo-calls for tools and food in Japanese monkeys. *Neurosci Res* 45: 383–389, 2003.
- Johnson CJ, Pick HL, Siegel GM, Cicciarelli AW, Garber SR.** Effects of interpersonal distance on children’s vocal intensity. *Child Devel* 52: 721–723, 1981.
- Jürgens U.** Neuronal control of mammalian vocalization, with special reference to the squirrel monkey. *Naturwissenschaften* 85: 376–388, 1998.
- Jürgens U.** Neural pathways underlying vocal control. *Neurosci Biobehav Rev* 26: 235–258, 2002.
- Jürgens U, Kirzinger A.** The effects of lesions in the cortical face area, supplementary motor cortex and anterior cingulate cortex on vocalization in the monkey. *Behav Brain Res* 5: 105, 1982.
- Jürgens U, Pratt R.** The cingulate vocalization pathway in the squirrel monkey. *Exp Brain Res* 34: 499–510, 1979.
- Jürgens U, von Cramon D.** On the role of the anterior cingulate cortex in phonation: a case-report. *Brain Lang* 15: 234–248, 1982.
- Maier JX, Neuhoff JG, Logothetis NK, Ghazanfar AA.** Multisensory integration of looming signals by Rhesus monkeys. *Neuron* 43: 177–181, 2004.
- Markel NN, Prebor LD, Brandt JF.** Biosocial factors in dyadic communication: sex and speaking intensity. *J Person Social Psychol* 23: 11–13, 1972.
- Michael DD, Siegel GM, Pick HL.** Effects of distance on vocal intensity. *J Speech Lang Hearing Res* 38: 1176–1183, 1995.
- Miller CT, Wang XQ.** Sensory-motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. *J Comp Physiol A* 192: 27–38, 2006.
- Miller CT, Beck K, Meade B, Wang X.** Antiphonal call timing in marmosets is behaviorally significant: interactive playback experiments. *J Comp Physiol A* 195: 783–789, 2009.
- Miller CT, Dimauro A, Pistorio A, Hendry S, Wang X.** Vocalization induced cFos expression in marmoset cortex. *Front Integr Neurosci* 4: 128, 2010a.
- Miller CT, Mandel K, Wang X.** The communicative content of the common marmoset phoe call during antiphonal calling. *Am J Primatol* 72: 974–980, 2010b.
- Müller-Preuss P, Jürgens U.** Projections from the ‘cingular’ vocalization area in the squirrel monkey. *Brain Res* 103: 29–43, 1976.
- Naguib M.** Use of song amplitude for ranging in Carolina wrens (*Thryothorus ludovicianus*). *Ethology* 103: 723–731, 1997.
- Naguib M, Wiley R.** Estimating the distance to a source of sound: mechanisms and adaptations for long-range communication. *Anim Behav* 62: 825–837, 2001.
- Nelson BS.** Avian dependence on sound pressure level as an auditory distance cue. *Anim Behav* 59: 57–67, 2000.
- Owren MJ, Dieter JA.** Infant cross-fostering between Japanese (Macaca-Fuscata) and rhesus macaques (M-Mulatta). *Am J Primatol* 18: 245–250, 1989.
- Owren MJ, Rendall D.** Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evol Anthropol* 10: 58–71, 2001.
- Owren MJ, Amoss RT, Rendall D.** Two organizing principles of vocal production: Implications for nonhuman and human primates. *Am J Primatol* 73: 530–544, 2011.
- Paus T.** Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nat Rev Neurosci* 2: 417–424, 2001.
- Pelegri-García D, Smits B, Brunskog J, Jeong CH.** Vocal effort with changing talker-to-listener distance in different acoustic environments. *J Acoust Soc Am* 129: 1981–1990, 2011.
- Romanski LM, Bates JF, Goldman-Rakic PS.** Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J Comp Neurol* 403: 141–157, 1999a.
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP.** Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2: 1131–1136, 1999b.
- Rosenblum LD, Carello C, Pastore RE.** Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception* 16: 175–186, 1987.
- Roy S, Miller CT, Gottsch D, Wang X.** Vocal control by the common marmoset in the presence of interfering noise. *J Exp Biol* 214: 3619–3629, 2011.
- Seyfarth RM, Cheney DL.** Vocal development in vervet monkeys. *Anim Behav* 34: 1640–1658, 1986.
- Seyfarth RM, Cheney DL.** Production, usage, and comprehension in animal vocalizations. *Brain Lang* 115: 92–100, 2010.
- Simoes CS, Vianney PV, de Moura MM, Freire MA, Mello LE, Sameshima K, Araujo JF, Nicolelis MA, Mello CV, Ribeiro S.** Activation of frontal neocortical areas by vocal production in marmosets. *Front Integr Neurosci* 4: 123, 2010.
- Snowdon CT, Cronin KA.** Cooperative breeders do cooperate. *Behav Process* 76: 138–141, 2007.
- Sober SJ, Brainard MS.** Vocal learning is constrained by the statistics of sensorimotor experience. *Proc Natl Acad Sci USA* 109: 21099–21103, 2012.
- Stefanacci L, Amaral DG.** Some observations on cortical inputs to the macaque monkey amygdala: an anterograde tracing study. *J Comp Neurol* 451: 301–323, 2002.
- Sutton D, Larson C, Taylor EM, Lindeman RC.** Vocalization in rhesus monkeys—conditionability. *Brain Res* 52: 225–231, 1973.
- Takahashi DY, Narayanan DZ, Ghazanfar AA.** Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr Biol* 23: 2162–2168, 2013.
- Whalen DH, Kinsella-Shaw JM.** Exploring the relationship of inspiration duration to utterance duration. *Phonetica* 54: 138–152, 1997.
- Winkworth AL, Davis PJ, Adams RD, Ellis E.** Breathing patterns during spontaneous speech. *J Speech Lang Hearing Res* 38: 124–144, 1995.
- Wyman MT, Mooring MS, McCowan B, Penedo MCT, Hart LA.** Amplitude of bison bellows reflects male quality, physical condition and motivation. *Anim Behav* 76: 1625–1639, 2008.