

# Sensory-Motor Interaction in the Primate Auditory Cortex During Self-Initiated Vocalizations

Steven J. Eliades and Xiaoqin Wang

Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Submitted 31 July 2002; accepted in final form 6 December 2002

**Eliades, Steven J. and Xiaoqin Wang.** Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J Neurophysiol* 89: 2194–2207, 2003. First published December 11, 2002; 10.1152/jn.00627.2002. Little is known about sensory-motor interaction in the auditory cortex of primates at the level of single neurons and its role in supporting vocal communication. The present study investigated single-unit activities in the auditory cortex of a vocal primate, the common marmoset (*Callithrix jacchus*), during self-initiated vocalizations. We found that 1) self-initiated vocalizations resulted in suppression of neural discharges in a majority of auditory cortical neurons. The vocalization-induced inhibition suppressed both spontaneous and stimulus-driven discharges. Suppressed units responded poorly to external acoustic stimuli during vocalization. 2) Vocalization-induced suppression began several hundred milliseconds prior to the onset of vocalization. 3) The suppression of cortical discharges reduced neural firings to below the rates expected from a unit's rate-level function, adjusted for known subcortical attenuation, and therefore was likely not entirely caused by subcortical attenuation mechanisms. 4) A smaller population of auditory cortical neurons showed increased discharges during self-initiated vocalizations. This vocalization-related excitation began after the onset of vocalization and is likely the result of acoustic feedback. Units showing this excitation responded nearly normally to external stimuli during vocalization. Based on these findings, we propose that the suppression of auditory cortical neurons, possibly originating from cortical vocal production centers, acts to increase the dynamic range of cortical responses to vocalization feedback for self monitoring. The excitatory responses, on the other hand, likely play a role in maintaining hearing sensitivity to the external acoustic environment during vocalization.

## INTRODUCTION

Auditory perception of one's own vocalization is necessary to maintain the normal acoustic structure of speech, and perturbations in this feedback lead to alterations in vocal production. Alteration in this acoustic feedback has been demonstrated to directly affect human speech production, where shifts in perceived formant frequency elicit compensatory changes in vocalized frequency content (Houde and Jordan 1998, 2002). In songbirds, abnormal acoustic feedback leads to a degradation, or "decrystallization," of the highly stereotyped song-production sequence (Brainard and Doupe 2000; Leonardo and Konishi 1999). An understanding of auditory processing during vocalization is therefore essential to understanding both the auditory system and the vocal production mechanism. Such

interactions between speech production and auditory perception have long been suggested by psychophysical and perceptual studies (Liberman 1996). However, how such interaction takes place at the neuronal level in the primate brain is largely unclear. Although the primate auditory cortex on the superior temporal gyrus has long been studied for its sensory functions, little is known about the sensory-motor integration in this cortical region.

Attenuation of auditory responses during vocalization has been previously observed at several sites in the subcortical auditory system. To reduce the intensity of vocalization acoustics, the middle ear muscles in humans, cats, and bats contract synchronously with vocal production (Carmel and Starr 1963; Henson 1965; Salomon and Starr 1963; Suga and Jen 1975). More recently, cochlear microphonic potentials in bats have shown that the decay time of the dampened oscillation is decreased during, and sometimes immediately before, vocalization (Goldberg and Henson 1998). The brain stem of both bats and humans also acts as a site of additional vocalization-synchronized attenuation of auditory responses. Evoked potentials recorded from the human upper brain stem demonstrate decreased activation during speech production (Papanicolaou et al. 1986). The site of this neural attenuation has been localized in bats to the nucleus of the lateral lemniscus (Suga and Schlegel 1972; Suga and Shimozawa 1974). The amount of attenuation resulting from vocalization has been estimated to be 20–25 dB in the middle ear (Henson 1965) and an additional 15 dB in the brain stem (Suga and Shimozawa 1974). Single-unit recordings in the brain stem of primates (Kirzinger and Jurgens 1991) and bats (Metzner 1993) have shown a mix of both vocalization-related suppression and excitation in a small percentage of neurons in auditory structures from the cochlear nucleus to the lateral lemniscus and inferior colliculus.

At the level of the auditory cortex, scattered evidence from human experiments suggests auditory-vocal interaction during speech. Magnetoencephalogram (MEG) studies during phonation have recorded dampened responses to a subject's own voice compared with playback of recorded human speech (Curio et al. 2000; Gunji et al. 2001; Houde et al. 2002; Numminen and Curio 1999; Numminen et al. 1999). Positron emission tomography (PET) imaging studies in the human

Address for reprint requests: X. Wang, Dept. of Biomedical Engineering, Johns Hopkins University School of Medicine, 720 Rutland Ave., Ross 424, Baltimore, MD 21205 (E-mail: xwang@bme.jhu.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

auditory cortex have also shown a reduction in the level of cortical activation during speech production (Paus et al. 1996; Wise et al. 1999). Limited intra-operative multi-unit recordings have shown both weakly excitatory and inhibitory events observed during speech in the middle temporal gyrus and, to a lesser extent, the superior temporal gyrus (Creutzfeldt et al. 1989). However, the nature of auditory-vocal interaction in the human auditory cortex at the level of single neurons remains largely unknown.

The non-human primate literature contains only a single report, published more than 20 years ago (Müller-Preuss and Ploog 1981), that attempted to address the issue of cortical auditory-vocal interaction at the level of single cortical neurons, an area of research that has been largely untouched since. This study in squirrel monkeys showed reduced or absent response to voluntarily produced or electrically evoked vocalizations compared with playback of recorded vocalizations. The bulk of evidence was based on electrically stimulated vocalizations; only a few neurons were recorded during spontaneously emitted, or self-initiated, voluntary vocalizations. Because of the confounding factors associated with electrical stimulation, however, it was not possible to compare the observed neural activity during vocalization to the activity immediately before vocalization. The study also showed that many neurons had similar activity during electrically stimulated vocalization and vocal playback. The contribution of subcortical factors to these observations, however, remains unclear. Unfortunately, perhaps due to the limited scope of the reported observations and the difficulty of this kind of experiment, little follow-up has been given to this pioneer study in the past two decades.

In songbirds, despite extensive studies of neural processing in the song production circuits (see review, Margoliash 1997), there have been relatively few studies of the mechanisms of auditory-vocal interaction during phonation. Some suppression of auditory responses immediately following song production was reported in the vocal nuclei (area HVC) of songbirds; however, the vocal motor activity in this area prevented observation of any alterations in sensory response during phonation (McCasland and Konishi 1981). This phenomenon has not yet been systematically addressed in the motor and premotor song areas of the avian forebrain and has yet to be explored in the sensory processing pathway (e.g., field-L, the analogue of the mammalian auditory cortex).

Compared with the numerous studies of auditory-vocal processing in the song-production system of songbirds, there has been relatively little research in non-human primates. The slow progress in non-human primates may have resulted partially from difficulties in creating appropriate animal models that both maintain vocal activities in captivity and provide access to neural activities in the auditory cortex during self-initiated vocalizations under behaving conditions. We have attempted to address issues using a vocal primate model, the common marmoset (Wang 2000), and single-unit chronic recording techniques developed for this species (Lu et al. 2001a,b). The common marmoset (*Callithrix jacchus*) is a highly vocal primate with a rich vocal repertoire and remains vocal in captivity (Agamaite and Wang 1997; Epplé 1968; Wang 2000). Our findings showed that single-neuron activities in the auditory cortex of awake marmosets were modulated by inputs, presumably from brain structures involved in vocal production, both

prior to and during self-initiated vocalizations. Results of the present study provided clear evidence of sensory-motor integration at the neuronal level in the auditory cortex of non-human primates.

## METHODS

### *Electrophysiological recordings*

All recording sessions were conducted in a double-walled, sound-proof chamber (Industrial Acoustics, Bronx, NY) with an interior covered by 3-in acoustic absorption foam (Sonex, Illbruck). Marmoset monkeys (*Callithrix jacchus*) were adapted to sit quietly in a semi-restraint device within the soundproof chamber with their heads immobilized. We have developed a chronic recording preparation in awake marmoset monkeys to laterally approach the auditory cortex (Lu et al. 2001a), which lies largely on the surface of the superior temporal gyrus in the marmoset (Aitkin and Park 1993). Vocal activity and neural activity in the auditory cortex were recorded simultaneously onto two channels of a digital audio tape recorder (Panasonic SV-3700). Vocalizations were recorded from a microphone (AKG C1000S) placed at mouth level ~6 inches in front of the animal. Neural activities were recorded using tungsten microelectrodes (A-M Systems, Carlsborg, WA or Micro Probe, Potomac, MD) with impedance of 2–5 M $\Omega$ . Action potentials of single neurons were detected by a template-based spike sorter (MSD, Alpha Omega Engineering, Nazareth, Israel). For each neuron, its basic response properties (e.g., CF, latency, and rate-level characteristics) were characterized, and its responses to presentations of other auditory stimuli (e.g., click trains, amplitude- and frequency-modulated tones, wide and narrow band noises and prerecorded marmoset vocalizations) were also recorded (Liang et al. 2002; Lu et al. 2001b). Locations of the recordings included both primary and lateral and posterior secondary auditory fields in all cortical layers. Acoustic stimuli used in auditory stimulus experiments were delivered free-field through a speaker located ~1 m in front of the animal and were calibrated at a location near an animal's head. All experimental procedures have been approved by the Johns Hopkins University Animal Care and Use Committee.

### *Data analysis*

Results reported were based on responses recorded from 104 single units recorded from the auditory cortex of two awake marmosets while the animals voluntarily vocalized. The obtained vocalization examples were distributed over 134 h of recordings. Due to the inherent complexity and unpredictability of primate vocal behavior, significant time was required to obtain sufficient samples and led to a limitation on the control of the number of vocal responses collected from each unit. While some data from the first animal was part of a larger study with auditory stimuli, all data from the second animal was obtained solely for this study. All vocalizations from the first animal were phee calls, while the second animal vocalized a mix of phee, trill, peep, and tsik calls (Agamaite and Wang 1997; Epplé 1968). In total, 1,236 vocalizations were recorded (993 phee, 101 trill, 110 peep, and 32 tsik calls) during these experiments. Because spontaneous activities of auditory cortical neurons were generally low, it was not always possible to determine if a vocalization resulted in suppression of discharges, in particular for calls with short duration (trill, peep, and tsik). Quantitative analyses of cortical responses were therefore performed on 513 long-duration phee calls (~1 s) during which sufficient neural activities were available. The quantitative analyses were performed on 79 units for which phee-call responses were recorded.

Firing rates associated with each vocalization, based on discharges of well-isolated units, were calculated off-line from digitized neural activity before, during, and after self-initiated vocalization using a level-based spike detection method. Two response measures were used to quantify changes in discharge rates during a vocalization. A

percentage change in firing rate was calculated for each vocalization response as  $(R_{\text{vocal}} - R_{\text{prevocal}})/R_{\text{prevocal}}$ , where  $R_{\text{vocal}}$  and  $R_{\text{prevocal}}$  are discharge rates during vocalization and for the 4 s preceding vocalization responses, respectively. In addition, a normalized measure, the Vocalization Response Modulation Index ( $\text{RMI}_V$ ), was calculated as  $(R_{\text{vocal}} - R_{\text{prevocal}})/(R_{\text{vocal}} + R_{\text{prevocal}})$ . A  $\text{RMI}_V$  of 0 indicates that the firing rate was identical during vocalization and spontaneous periods, whereas a value of  $-1$  indicates a complete suppression of spontaneous firings. A  $\text{RMI}_V$  of  $+1$  indicates a unit with either very strongly driven vocalization response, a very low spontaneous rate, or both. Vocalizations with sufficient neural activity were classified as either suppressed or excited for later analysis based on the percent change in firing rate and  $\text{RMI}_V$ . Of 513 vocalizations, 421 were classified as suppressed, 92 as excited.

A number of recorded vocalizations coincided with the presentation of external acoustic stimuli. Because each stimulus was presented multiple times, the single trial response to the stimulus during vocalization was compared with the average response of stimulus trials when the animal was not vocalizing to quantify the effects of self-initiated vocalization on responses produced by external stimuli. A Stimulus Response Modulation Index ( $\text{RMI}_S$ ) was used to quantify alterations in stimulus-driven response. The  $\text{RMI}_S$  was calculated as  $(R_{\text{Stim+Vocal}} - R_{\text{Stim}})/(R_{\text{Stim+Vocal}} + R_{\text{Stim}})$ , where  $R_{\text{Stim+Vocal}}$  was the firing rate during concurrent stimulus with vocalization and  $R_{\text{Stim}}$  was the average firing rate during stimulus alone.

The onset of vocalization was determined by the detection of spectral energy in vocalization frequency bands (3–12 kHz). The duration of pre-vocalization suppression was measured from the onset of suppression to the beginning of vocalization. The onset of suppression was calculated from a cumulative peristimulus time histogram (PSTH, binwidth = 1 ms) of discharges by identifying the deflection point in the slope that indicated a reduction in firing rate. Each bin in the cumulative PSTH represented the total number of spikes up to that time.

The interval over which vocalization-related changes in neural activity were significant was determined from a population histogram (binwidth = 5 ms) by comparing a 1,000-ms period of spontaneous activity to a sliding window of activity (100-ms duration, 10-ms steps) before and during vocalization. The Wilcoxon rank-sum test was performed between the spontaneous firing rate and the firing rate within each individual window, and  $P$  values  $<0.05$  were considered statistically significant. The long duration of the sliding window was necessitated by the sparseness of cortical discharges.

In most neurons, multiple vocalization responses were recorded and the median and the inter-quartile range of the  $\text{RMI}_V$  were computed for each unit, including those vocalization examples that failed to elicit any observable change in neural response. Those units with sufficient vocal samples ( $\geq 3$ ) were tested statistically to determine the reliability of the observed responses. A PSTH of vocalization responses was calculated for each unit (binwidth = 20 ms). The activity during vocalization was compared with the spontaneous activity ( $>500$  ms preceding vocal onset) using the Wilcoxon test.

## RESULTS

We have studied single-unit activities in the auditory cortex of awake marmosets while the animals made self-initiated vocalizations. Simultaneous recordings of neural activities and vocalizations were made from 104 single units in two awake marmosets in which a large number (1,236) of self-initiated vocalizations were observed. These vocalizations occurred during both spontaneous discharges and in the presence of external auditory stimuli. The characteristic frequency (CF) and rate-level function as well as other response properties of the studied neurons were characterized (Liang et al. 2002; Lu et al. 2001a,b).

We will begin by separately describing and analyzing the two classes of responses to self-initiated vocalization, suppression

and excitation. We will then analyse these two classes of responses in a unit-by-unit manner to study sensory-motor interactions in the context of a population of auditory cortical neurons. Finally, we will analyze the direct contributions of the auditory cortex during vocalization-related sensory-motor interaction.

### *Vocalization induced suppression of single-unit activity in the auditory cortex*

In a majority of cases, a self-initiated vocalization caused a suppression of activity in auditory cortical neurons. Several representative examples of this suppression are given in Fig. 1. In each case shown, a well-isolated unit was firing spontaneously prior to the animal's vocalization. During vocalization, however, the units' spontaneous activities were either partially or completely inhibited. It was not uncommon to observe that all neural activity was completely suppressed for the entire duration of vocalization (Fig. 1A). In cases where more than one unit's activities were recorded by the same electrode, it was often observed that activities from all units were suppressed simultaneously (Fig. 1B). However, although suppression was the most frequently observed response to a self-initiated vocalization, not all recorded units showed suppression during vocalization. This response diversity was manifested in the example in Fig. 1C in which the unit with the larger action potentials exhibited the prominent vocalization-induced suppression, whereas the unit with the smaller action potentials was not suppressed; rather it maintained its activity throughout the vocalization. Another important aspect of vocalization-induced suppression was its timing. Individual examples in Fig. 1 suggest that the suppression began prior to the onset of vocal production (Fig. 1, A and C).

While many of the vocalizations occurred during periods of silence, and thus modulated spontaneous activities of auditory neurons, many also occurred during the presentation of acoustic stimuli. When these stimuli produced driven activity in neurons, self-initiated vocalizations were observed to suppress the stimulus-driven discharges in most cases, such as the example in Fig. 2A. The vocalization-induced suppression, therefore could alter both spontaneous and stimulus driven activity.

The presentation of previously recorded marmoset vocalizations was used to study differences between vocal production and perception. The same auditory cortical neuron showing suppression resulting from a self-initiated vocalization (Fig. 2A) responded, however, to a similar vocalization played back passively from a speaker at comparable sound level (Fig. 2B). This dichotomy, along with onset of suppression being prior to a vocalization, indicated that the suppression of neural discharges was unlikely induced by the acoustic characteristics of the self-initiated vocalization but rather by inhibitory mechanisms associated with the production of vocalization.

While most recorded vocalizations were phee calls as shown in Figs. 1 and 2, vocalization-induced suppression was also observed for other types of vocalizations. Figure 3 shows an example of the neural response during a trill call. The unit was being driven by a band-pass noise stimulus but did not fire during the time when the animal produced two short segments of vocalization (Fig. 3A). Closer observation (Fig. 3B) verifies that these brief vocalizations show the FM characteristic of the trill class of marmoset calls (Agamaite and Wang 1997; Epplé



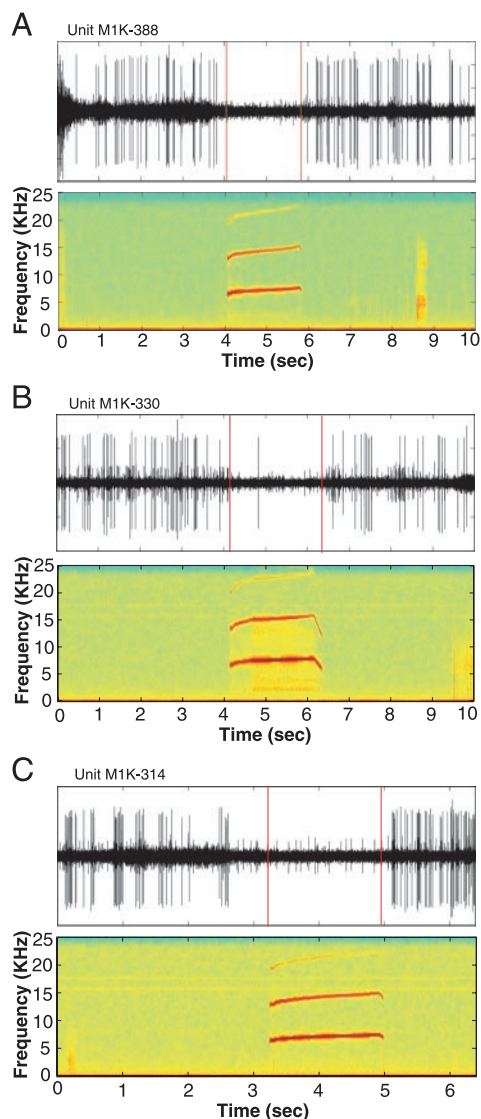


FIG. 1. Representative examples of the suppression of spontaneous neural activities in the auditory cortex by self-initiated vocalizations. *Top*: extracellular recording traces containing well-isolated single units; *bottom*: corresponding spectrograms of acoustic recording showing captured phee vocalizations. Two vertical red lines mark the onset and offset of each vocalization. *A*: a unit with high spontaneous activity is completely suppressed while the animal is vocalizing. The suppression appears to begin before the onset of the vocalization. *B*: a second example where 2 units are captured simultaneously both of which are nearly completely suppressed, though not completely, during vocalization. *C*: another example in which the unit with the larger action potentials is completely suppressed during, and prior to, vocalization. A second unit with smaller action potentials, recorded by the same electrode, does not appear to be affected by the vocalization.

1968). However, because there were other gaps in firing during the stimulus period, this example alone cannot positively determine whether the unit's firing was inhibited. This example illustrates the difficulty in assessing inhibition based on short vocalizations because of the sparseness in cortical discharges. The quantitative analyses described below are therefore based on longer-duration calls.

#### Magnitude and timing of vocalization-induced suppression

A large number of samples in which long-duration phee calls caused suppression of cortical activity were analyzed to quan-

titatively describe the modulatory effects of self-produced vocalizations. Spike trains from all suppressed samples were aligned by the onset times of corresponding vocalizations, based on which a population histogram was calculated (PSTH). The duration of phee calls included in the samples was typically  $\sim 1$  s, but could last up to 2 s. The resulting aggregate activity confirmed the discharge suppression revealed in individual samples. It also demonstrated that the suppression began *prior* to the onset of vocalization. The suppression became statistically significant  $\sim 220$  ms before vocal onset and remained significant for 1,730 ms. When the spike trains were aligned by the vocalization offset, the responses returned to the normal activity level at the completion of vocalization (Fig. 4, inset). These results indicate that vocalization-induced suppression was therefore an inhibition of neural activity that began prior to the onset, and persisted for the duration, of self-initiated vocal production.

We further quantified the time course of suppression by measuring, in each sample, the length of discharge suppression preceding the onset of vocalization (see METHODS). Figure 5 shows the distribution of the length of pre-vocalization suppression (open symbol). Suppression began as early as several hundred milliseconds before a vocalization was heard with a median length of 271 ms. This onset duration is similar to the duration of statistically significant suppression before vocal onset calculated on the basis of population PSTH (220 ms, Fig. 4). Overlaid is the inter-spike-interval (ISI) distribution measured from discharges over a period of 3–4 s prior to the

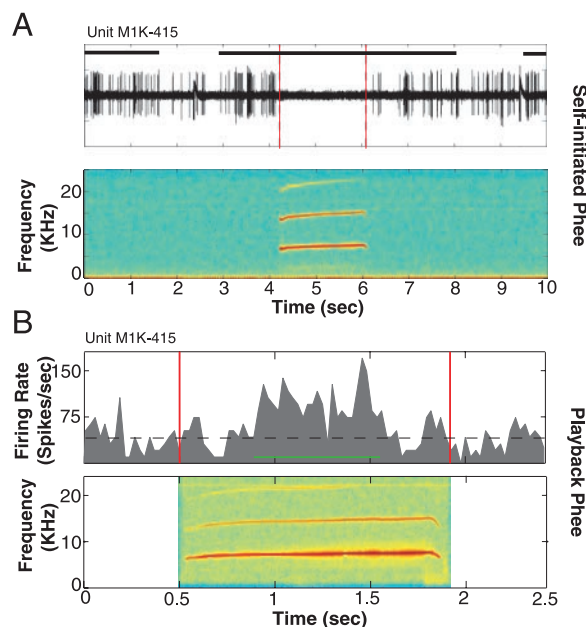


FIG. 2. *A*: an example in which a self-initiated vocalization completely suppresses stimulus-driven discharges. The presence of external acoustic stimuli (playback of vocalizations presented at 70 dB) is indicated by bars above the neural recording trace. *B*: a peristimulus time histogram (PSTH, *top*) is shown for the same unit as in *A* in response to passive playback of a previously recorded vocalization (*bottom*). The dashed line indicates the mean spontaneous firing rate and the green bar (*top*) specifies the duration of statistically significant response ( $P < 0.05$ , see METHODS). Although the self-produced vocalization (*A*) was spectrally similar to the playback vocalization (*B*), both with similar intensities (80 dB), they had opposite effects on the discharges of the unit shown.

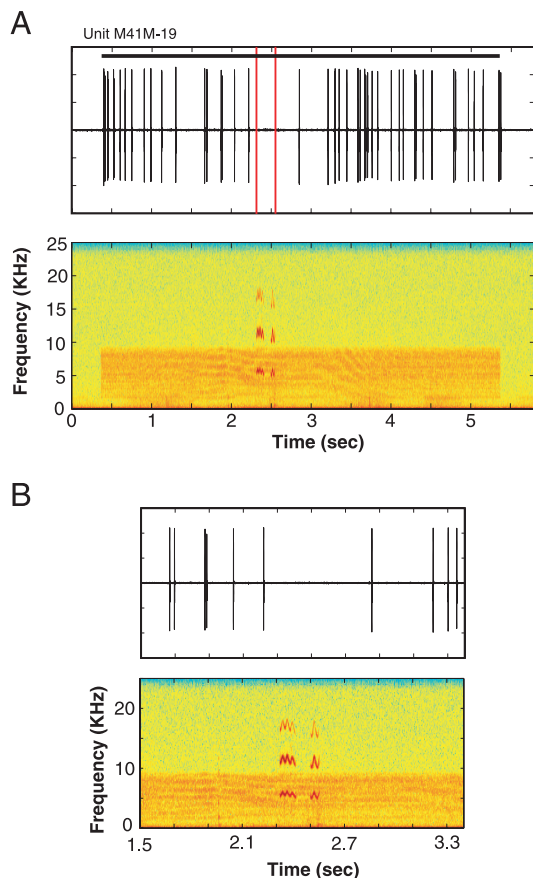


FIG. 3. An example of cortical responses during a brief vocalization. *A*, top: recording of a unit being driven by a band-pass noise (3.9 kHz, 2.5 octave bandwidth, 40 dB) during which the animal produced short trill calls. The spectrogram (*bottom*) shows sinusoidal FM characteristic of 2 brief trill calls as well as the spectral content of the band-pass noise stimulus. During vocalization there is an absence of unit activity. *B*: a magnified view of a portion of the unit activity and spectrogram shown in *A*.

vocalization (Fig. 5, green line). The two distributions are significantly different ( $P < 0.05$ ), demonstrating that the reduced discharge rates before vocal onset could not be attributed to irregularity in spontaneous neural firing.

Comparison of discharge rates during and in the absence of self-initiated vocalizations was used to quantify the magnitude of vocalization-induced suppression. Two different measures were used to reflect the change in firing rate. The distribution of the percent change in firing rate during vocalization (see METHODS) displays large reductions ( $>50\%$ ) in the firing rate in the majority of samples (Fig. 6A). The median reduction in firing rate caused by vocalization was 77%. In 50 samples, neural firing was completely suppressed during vocal production. A second quantification of suppression magnitude, the  $RMI_V$ , normalizes the changes in firing rate between  $-1$  and  $1$  (see METHODS) and was used to contrast firing rate changes under other conditions. Similar trends, including the peak indicating complete suppression, can be seen in Fig. 6, *A* (percent change) and *B* ( $RMI_V$ ). These observations clearly showed that vocal production by a marmoset resulted in significant suppressions of discharges of single units in the auditory cortex that began before a vocalization was acoustically produced.

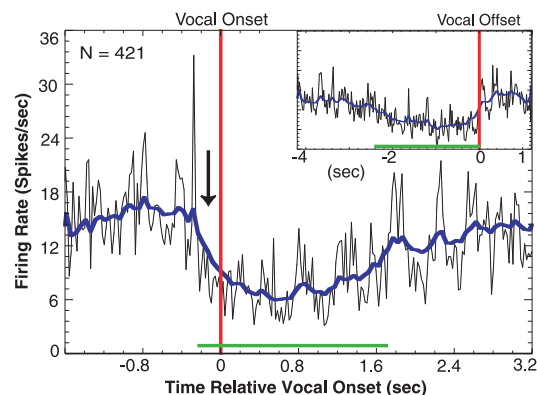


FIG. 4. Average suppressed vocalization response. A population histogram for all suppressed responses aligned by vocal onset is shown (binwidth = 20 ms). The vocal onset is indicated by a red line, and the time axis is referenced to this point. The blue line is a moving average (100-ms window) and shows that suppression begins prior to vocalization as indicated by an arrow. The green bar indicates the period over which suppression was continuously significant ( $P < 0.05$ ). Suppression was significant starting 220 ms before vocal onset and remained until 1,730 ms after vocal onset. *Inset*: a population histogram aligned by vocal offset is shown. The green line indicates significant suppression ( $P < 0.05$ ) that lasts until 20 ms after vocalization. The binwidth and the window size for the moving average (blue line) are the same in both plots.

#### Effects of vocalization-induced suppression on discharges driven by external acoustic stimuli

From time to time, an animal would produce a vocalization during which an acoustic signal was presented. As illustrated by an example in Fig. 2, discharges evoked by external auditory stimuli could be suppressed by self-initiated vocalization. Trials where a vocalization overlapped with an external stimulus were compared with other trials of the identical stimulus presented when the animal was not vocalizing. Figure 7A shows an example of a sinusoidal amplitude-modulated (sAM) tone that elicited strong, stimulus-locked discharges in a unit

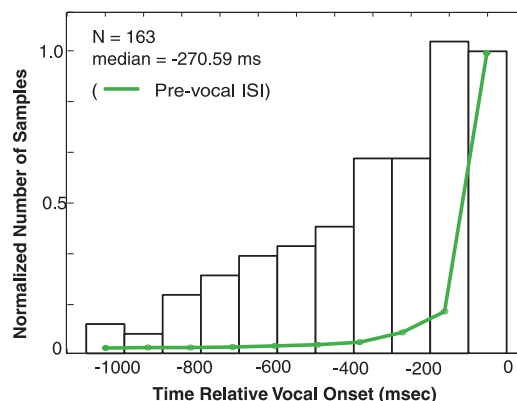


FIG. 5. Duration of prevocalization suppression. The duration of suppression prior to the onset of vocalization was calculated from 163 samples with sufficient spontaneous activities to permit estimation and that were free of potentially complicating external stimulus presentations. The distribution of this duration is shown and has a median of 271 ms. A prevocal inter-spike interval (ISI) distribution (green line) was also computed from spikes within a 3- to 4-s time window prior to onset of the pre-vocalization suppression period (in the same group of samples) and was used as a control to validate the duration of prevocal suppression as not resulting from discharge irregularity. Both distributions have been normalized to the same scale for comparison. The two distributions are significantly different ( $P < 0.05$ , Wilcoxon rank-sum test).

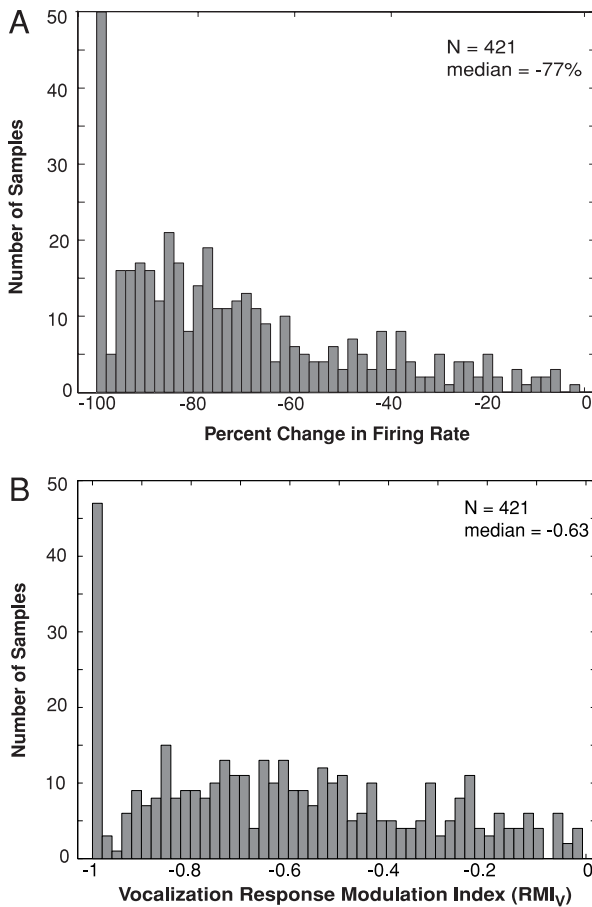


FIG. 6. Quantification of the magnitude of suppression. The firing rate during vocalization was compared with that before vocalization to determine the amount of vocalization-induced suppression. *A*: the distribution of the percent change in firing rate (see METHODS for definition) is shown and has a median change of  $-77\%$ . The large peak at  $-100\%$  indicates a complete suppression of neural discharges during many vocalizations. *B*: the distribution of the Vocalization Response Modulation Index ( $RMI_V$ , median  $-0.63$ ). This normalized index ranges from  $-1$  to  $1$ , with negative values indicating suppression, and shows the same complete suppression peak at  $-1$ .

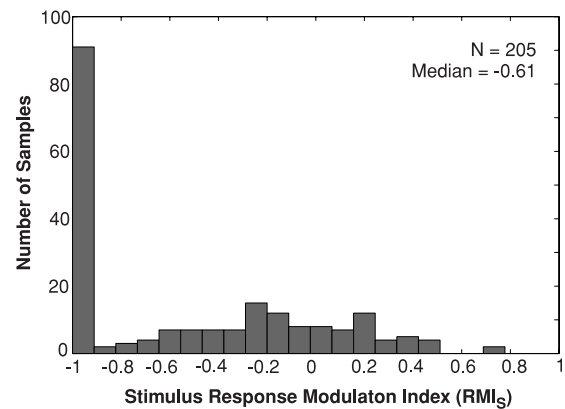


FIG. 8. A population summary of the suppression of auditory responses during vocalization. The Stimulus Response Modulation Index ( $RMI_S$ ) was calculated for 205 external auditory stimuli presented during vocalization (see METHODS). Negative indexes indicate suppression of auditory responses, and positive indexes indicate increased auditory responses due to self-initiated vocalization. The median index ( $-0.61$ ) was similar to that of vocalization-induced suppression in the absence of external stimuli ( $RMI_V$ , median  $-0.63$ , see Fig. 6*B*). The  $RMI_S$  shows a large peak at  $-1$ , indicating a complete suppression of all neural activity, including responses to external acoustic stimuli, in many samples.

(*top*), but failed to produce any response during a self-initiated vocalization (*bottom*). The suppression of acoustic responsiveness generally persisted throughout the duration of a vocalization. However, stimulus driven discharges returned rapidly once a vocalization ended as shown by the example in Fig. 7*B*.

A  $RMI_S$  was calculated to quantify the effects of vocalization on cortical neurons' responses to external acoustic stimuli presented during a suppressive vocalization (as determined by a negative  $RMI_V$ , see Fig. 6*B*). The distribution of this index is shown in Fig. 8. A positive  $RMI_S$  value indicates increased stimulus-driven responses during a vocalization, whereas a negative  $RMI_S$  represents a suppressed stimulus-driven response. Most vocalizations resulted in reduced stimulus-driven responses during vocalization compared with responses in the absence of vocalization. The large peak at  $-1$  demonstrates a set of stimuli that failed to evoke any neural activity during self-produced vocalization. Interestingly, a small number of

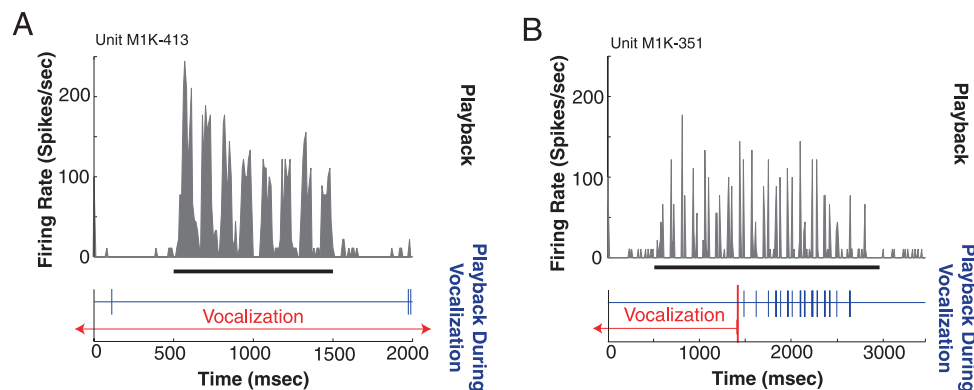


FIG. 7. Examples of responses to external acoustic stimuli during vocalization-induced suppression. *Top*: PSTH of the response to a passively presented acoustic stimulus; *bottom*: spike train showing the response to one presentation of the same stimulus during a self-initiated vocalization (blue). A red line indicates the duration of vocalization. The thick black bar marks the stimulus duration. The blue ticks are markers indicating the time of neural discharges. *A*: a unit responding strongly to a sinusoidal frequency-modulated sound (sFM, carrier frequency set at CF of 2.99 kHz, 70 dB, 256 Hz modulation depth, 8-Hz modulation frequency) played passively but unresponsive to the same stimulus during vocalization is shown. *B*: an example showing that responses to an external stimulus (a vocalization presented at 30 dB) during a vocalization are suppressed but return rapidly after the end of the vocalization. The same stimulus evoked sustained responses throughout its duration during passive presentation.



stimuli actually showed an increase in firing rate (positive  $\text{RMI}_S$ ), although this may be an artifact of the single-sample nature of the stimulus-vocalization overlap. Overall, however, auditory cortical neurons demonstrated a largely reduced responsiveness to acoustic stimuli during self-initiated vocalization. The median value of the  $\text{RMI}_S$  was  $-0.61$ , similar to the effect of vocalization alone (median  $\text{RMI}_V = -0.63$ ).

#### Vocalization induced excitation in the auditory cortex

While most cortical responses we observed during a self-initiated vocalization exhibited suppression ( $n = 421$ ), a smaller number of responses instead showed increased neural firing during vocalization ( $n = 92$ ). Examples of this second pattern of response to self-produced vocalizations are shown in Fig. 9. The well-isolated unit in Fig. 9A had low spontaneous activity, whereas the second unit in Fig. 9B had much higher spontaneous activity at rest. Both units, however, showed high firing rates when the animal vocalized. This vocalization-related excitation appeared to begin immediately following the start of vocalization in the first example (Fig. 9A), although the spontaneous activity obscured the onset of vocalization-related activity in the second example (Fig. 9B).

A number of the vocalization-related excited responses were aligned by their corresponding vocal production onsets to analyze their time courses. The aggregate activity of these responses showed a clear increase in firing rate during vocalization (Fig. 10). In contrast to the timing of vocalization-

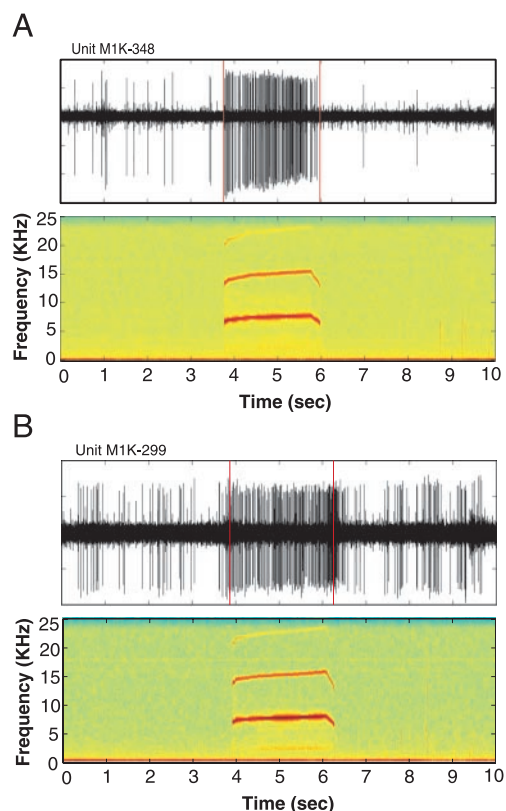


FIG. 9. Examples of excitatory responses during vocalization. A: an example showing a unit with low spontaneous rate that is strongly driven during vocalization. The driven discharges appear to begin immediately after the beginning of vocalization. B: a high spontaneous rate unit also showing excitation to a higher firing rate during vocalization.

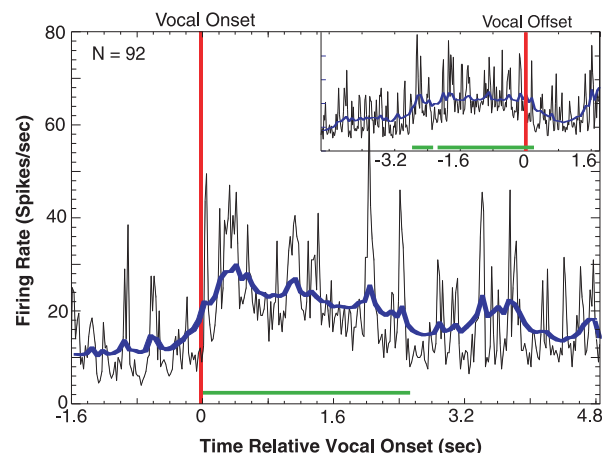


FIG. 10. Average excited vocalization response. The population histogram of excitatory responses to self-produced vocalizations, aligned by vocalization onset, is shown (binwidth = 20 ms). The vocalization onset is indicated by the red line. Excitation is seen to begin at the first bin following the vocal onset, in sharp contrast to the time course of vocalization-induced suppression (Fig. 4). The green bar indicates the duration over which excitation was significant ( $0-2,190$  ms,  $P < 0.05$ ). The moving average (blue line, 100-ms window) smooths out the rapid onset of activity. Inset: excitation persists throughout vocalization and is seen to cease shortly following the end of offset-aligned vocalizations.

induced suppression, excitation caused by self-initiated vocalizations did not occur until after the onset of vocalization. This increase in firing rate became statistically significant at the onset of vocalization and remained significant for 2,190 ms. The excitation then ceased following the end of vocalization (Fig. 10, inset). These excitatory responses are thus possibly the result of auditory feedback through the ascending auditory system.

The degree of increased neural activity during these vocalizations was quantified using the same measures as used for vocalization-induced suppression. The distribution of the percent change in firing rate showed a peak corresponding to an approximate 100% increase in firing rate (Fig. 11A). The increase in firing rate, however, was often much higher, extending up to 1,000% or greater. The normalized  $\text{RMI}_V$  measure displayed a broad range of positive values (indicating increased neural activities) with a median increase of 0.54 (approximately equivalent to a 200% increase in firing rate), in contrast to the negative  $\text{RMI}_V$  values calculated for suppressive vocalizations (Fig. 6B). Such large median increases in firing rate and  $\text{RMI}_V$  were indicative of strongly driven discharges during vocalization. These observations clearly showed that, in a smaller set of samples, self-produced vocalizations caused excitatory responses that began immediately after the onset of vocalization.

#### Effects of vocalization-related excitation on responses to external acoustic stimuli

Figure 12A shows an example of a stimulus (sAM) that drove a unit regardless of whether the animal was vocalizing. During overlapping stimulus presentation and vocalization, the firing rate increased, reflecting the summed responses to the vocalization and the stimulus. Although the sample size was much smaller, when an  $\text{RMI}_S$  was calculated for these concurrent presentations of external stimuli and excitatory vocaliza-

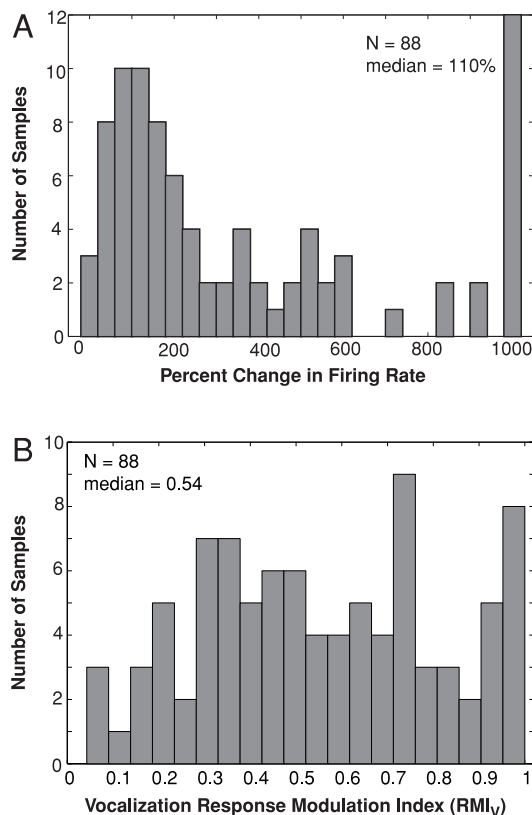


FIG. 11. Quantification of the magnitude of excited vocalization response. The firing rate during vocalization was compared with that before vocalization to calculate the amount of excitation during vocalization. *A*: distribution of the percent change in firing rate. There is a prominent peak centered around a 100% increase in firing (median = 110%). *B*: the distribution of the  $RMI_V$  has a median of 0.54. Positive values indicate excitation during vocalization.

tions, the distribution (Fig. 12*B*) was different from that of the suppressed responses (Fig. 8). The  $RMI_S$  distributions for the excitatory and inhibitory vocalizations have median indexes of 0.03 and  $-0.61$ , respectively. The distribution for excitatory vocalizations was more closely centered around zero. The distribution of the  $RMI_S$  in Fig. 12 indicated that the total response to concurrent presentation of external stimulus and an excitatory self-initiated vocalization was, on average, approximately the same as the presentation of the external stimulus alone.

#### Distribution of units with different vocalization-induced response modulation

In previous sections, the analyses were based on each vocalization and its corresponding cortical responses. However, in most of the 79 phee-response containing single units we studied, more than one occurrence of self-initiated vocalization was captured. There was generally a consistency among vocalization-related responses in individual units. The median  $RMI_V$  was calculated from all vocalization responses recorded in each unit and plotted in Fig. 13*A*. The error bars represent the inter-quartile (25–75%) for each unit's  $RMI_V$ . The median  $RMI_V$ s for the units formed a continuous distribution extending from highly suppressed to excited. Units considered to have reliable vocalization responses ( $P < 0.01$ ,  $\geq 3$  vocal samples) were found for both negative and positive  $RMI_V$  values. The

number of vocalization samples obtained for each unit is shown for reference (Fig. 13*B*). Although Fig. 13*A* gives an impression of a continuous  $RMI_V$  distribution, many units showed observable, and significant, tendencies favoring either suppressed or excited vocalization responses. Only a small number of units displayed possible bimodal responses (i.e., samples of both suppressed and excited responses).

Both units with positive and negative  $RMI_V$ s were encountered in the primary auditory cortex as well as the lateral fields. The mean spontaneous firing rates were similar for both units showing positive and negative  $RMI_V$ s (negative: 10.62 spikes/s, positive: 10.54 spikes/s,  $P = 0.5$ , Wilcoxon), indicating that there was no bias due to spontaneous rates in determining the  $RMI_V$  of the sampled neurons. This is important because low spontaneous neurons would be biased against the measurement of suppression. There appeared to be no relationship between the  $RMI_V$  and unit's center frequencies (CFs, Fig. 13*C*). Similarly, there appeared to be no correlation between

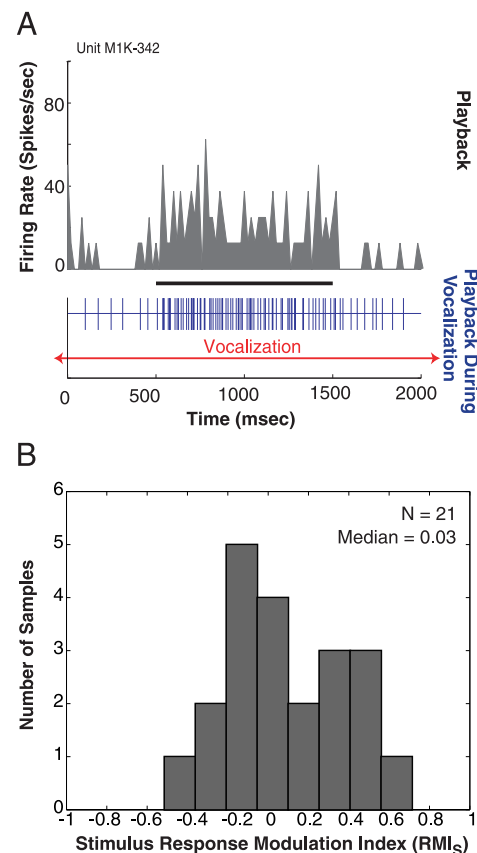


FIG. 12. Response to external acoustic stimuli during excitatory vocalization. *A*: a unit showing excitatory vocalization-related response is further excited by the presentation of an external auditory stimulus. *Top*: PSTH of the response to a passively presented auditory stimulus; *bottom*: spike train (blue) showing the response to one presentation of the same stimulus during a self-initiated vocalization. A red line indicates the duration of vocalization. The thick black bar marks the stimulus duration. The neuron responded to the stimulus, an sAM tone (22.16 kHz, 60 dB) played passively (firing rate: 21.5 spikes/s), and to the same stimulus during a self-initiated vocalization with a higher firing rate (70 spikes/s). *B*: distribution of the  $RMI_S$  calculated for the limited sample of stimuli presented during excitatory vocalization responses. The distribution has a median of 0.03 and shows a narrower range than the same index measured for suppressed responses (Fig. 8), indicating that, on average, units had similar responses to external acoustic stimuli with or without the presence of self-initiated vocalization.



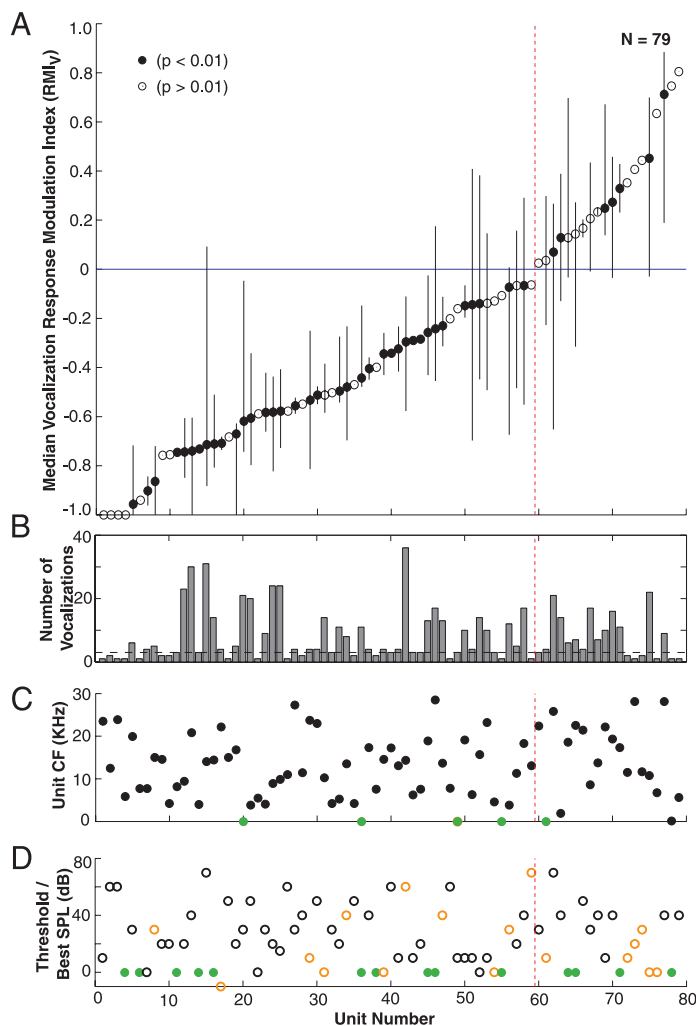


FIG. 13. Distribution of vocalization-related response modulations for individual units. *A*: the median RMI<sub>V</sub> is shown for 79 phee-response containing neurons. Circles, the median; and error bars, the inter-quartile range (25–75%), of the RMI<sub>V</sub>s for all vocalizations sampled for each neuron. Units are arranged by their increasing median RMI<sub>V</sub> value. Statistical tests were conducted in units where  $\geq 3$  vocal samples were available. Filled circles, units whose responses were considered statistically reliable ( $P < 0.01$ ). Dashed red line, separating line for units whose median RMI<sub>V</sub> was negative from those that were positive. *B*: the number of vocalization samples for each corresponding unit in *A* is plotted. Dashed line,  $n = 3$ . *C*: the center frequency (CF) is shown for each unit shown in *A*. The distributions of CFs shows no relationship to RMI<sub>V</sub>. Green dots, units that had no measurable frequency tuning. *D*: the rate-level properties for each unit are shown. For monotonic units (orange) the threshold SPL is plotted, and the preferred sound level (best SPL) is shown for nonmonotonic units (black). Both types of properties are seen across the RMI<sub>V</sub> distribution. Green dots, units that had no measurable rate-level properties.

RMI<sub>V</sub> and rate-level characteristics of a unit (Fig. 13*D*). Both monotonic and nonmonotonic rate-level functions were found for units throughout the RMI<sub>V</sub> distribution. The sampled units appeared to differ only largely in their responses to self-initiated vocalizations.

#### Source of vocalization induced suppression

Attenuation of auditory responses during vocalization has been previously observed in the middle ear and brain stem of both bats (Henson 1965; Metzner 1993; Suga and Jen 1975; Suga and Schlegel 1972; Suga and Shimozaawa 1974) and

humans (Papanicolaou et al. 1986; Salomon and Starr 1963). This attenuation has been estimated in bats to be equivalent to a 35- to 40-dB decrease in stimulus intensity during vocalization (Suga and Shimozaawa 1974). We have analyzed potential contributions of subcortical attenuation to the observed suppression of cortical activity (Figs. 14 and 15). Neurons in the auditory cortex of awake primates typically have two types of discharge rate versus sound level functions: monotonic or saturated and non-monotonic (Pfingst and O'Connor 1981; Wang et al. 1999). Expected discharge rates were estimated from the rate-level functions using the recorded vocalization based on the rate-level function recorded in that unit and taking into account the assumed 40 dB of subcortical attenuation. Figure 14 (*A* and *B*) illustrates this analysis with two representative examples. The discharge rate observed during vocalization was much smaller than the expected (after subcortical attenuation) rates for both monotonic (Fig. 14*A*) and non-monotonic (Fig. 14*B*) units.

A quantitative analysis of a population of suppressed vocalization-induced responses further substantiated this observation. The observed firing rates are plotted against the expected (after subcortical attenuation) firing rates (Fig. 15, blue). The columnar grouping seen at some expected rates was due to multiple occurrences of vocalization samples in particular units. For vocalization samples obtained, both monotonic and nonmonotonic units, the observed rate was lower than the expected rate. This indicates that the known subcortical attenuation may not fully account for the amount of suppression observed in the auditory cortex during self-initiated vocalization, although other differences between vocalization and playback were not examined.

We applied the same analysis to vocalization-related excitatory responses in Fig. 14 (*C* and *D*). The observed firing rates during vocalization in these units were above the expected firing rates (after subcortical attenuation) for both monotonic (Fig. 14*C*) and nonmonotonic (Fig. 14*D*) units. In fact, the monotonic example shown in Fig. 14*C* displayed an observed activity much higher than highest rate of the unit's rate-level function. The difference between observed and expected remained in the case of a nonmonotonic example despite the effect of shifting the sound level 40 dB toward a higher point in the rate-level function (Fig. 14*D*). Across the population of excitatory units, the observed firing rate was usually greater than, or close to, the expected firing rate (Fig. 15, red). These differences may reflect a contribution of bone conduction, or may possibly suggest cortical compensation of subcortical attenuation. When compared with vocalization-induced suppression, there is minimal overlap of the observed activity, even when the expected activities were similar. This further suggests the mechanistic differences between suppression and excitation during vocalization.

#### DISCUSSION

Our observations of both suppressed and excited discharges during self-initiated vocalizations at the level of single neurons represent an advance in our knowledge of sensory-motor interactions in the auditory cortex of primates. The suppression of neuronal activities in the auditory cortex of humans and primates during speaking or vocalization has been suggested based on limited observations in the past several decades

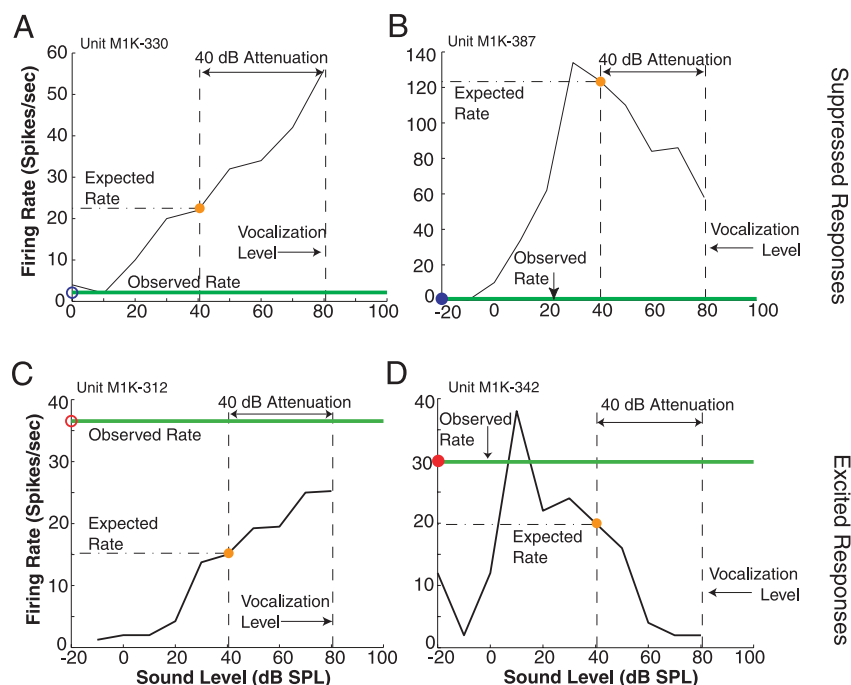


FIG. 14. Estimation of expected cortical responses from subcortical attenuation. Analyses are shown for 4 units with representative rate-level characteristics (A and C: monotonic; B and D: nonmonotonic). The rate-level curves (black) were obtained using CF tones. Rate-level functions obtained using playback vocalization closely mirrored that obtained by CF tones. The intensity of self-initiated vocalizations were, on average, 81 dB SPL, though in some instances they were as high as 105 dB SPL. The average vocalization sound level (81 dB) is marked by a vertical dashed line. Orange dot, the expected (estimated) firing rate corresponding to a sound level that is 40 dB lower than the vocalization sound level, an assumed attenuation attributable to subcortical mechanisms. Green bar, the observed firing rate during vocalization for both suppressed (A and B, blue marker) and excited (C and D, red marker) vocalization responses. A and B: the observed rate is lower than the expected rate in both suppressed examples. C: an example showing that the observed excitatory vocalization response was stronger than the expected response in a monotonic unit. In this monotonic example, subcortical attenuation would make the differences between observed and expected rates even larger. The observed activity in this case was greater than the entire rate-level curve of the unit. D: an example showing that the observed response was also stronger than the expected response for a non-monotonic unit.

(Creutzfeldt et al. 1989; Müller-Preuss and Ploog 1981). The present study provided more extensive evidence for interpreting vocalization-induced modulation in the auditory cortex of primates. We showed that self-initiated vocalizations suppressed spontaneous discharges, a direct indication of the inhibition on cortical neurons; that vocalization-induced inhibition in the auditory cortex began several hundreds of milliseconds prior to the vocal onset, indicating that it was related to the initiation of a vocalization; that suppression of cortical discharges cannot be fully accounted for by known subcortical attenuation (Henson 1965; Suga and Shimozawa 1974); and we

identified other neurons with excitatory responses during vocalization. These issues have not been thoroughly investigated in previous studies. Our findings provided clear indication that the observed response modulations in the auditory cortex were likely due to influence from vocal production systems rather than simply due to acoustic feedback via ascending auditory pathway during vocalization.

### Comparison with previous studies

There has been only one report in the non-human primate literature on the subject of sensory-motor interaction at the level of single cortical neurons during vocalization in monkeys (Müller-Preuss and Ploog 1981). That study showed reduced or absent responses to electrically evoked vocalizations in squirrel monkeys and demonstrated a difference between auditory cortical firings during vocal production and playback perception. Müller-Preuss and Ploog (1981) also reported a small number of spontaneously produced vocalizations that were observed to suppress spontaneous neural discharges. The findings of our current study confirmed the earlier observations of reduced neuronal activity during vocalization. On the basis of a large number of self-initiated vocalizations, we demonstrated quantitatively that, in many neurons, self-initiated vocalizations suppressed both spontaneous and stimulus-driven firings. Furthermore, we showed that this suppression begins prior to the onset of vocalization, an observation that was previously not possible with electrically evoked vocalizations. While the earlier study (Müller-Preuss and Ploog 1981) demonstrated reduced cortical responses during vocalization, the observed responses could not be categorically attributed to cortical inhibition. Alteration in feedback intensity and spectrum by subcortical events may have contributed to differences between playback and vocalization responses. Only by demonstrating suppression of neural activity in the absence of sound were we able to attribute the reduced cortical activity seen by the current study to neurally mediated inhibition associated with vocal

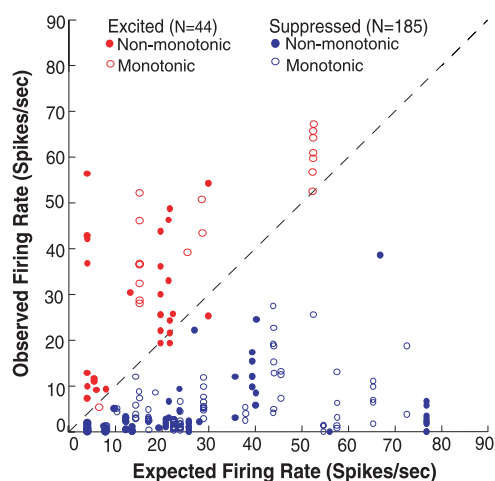


FIG. 15. Comparison between estimated and observed cortical responses. The relationship between observed and expected firing rates is shown for the population of suppressed and excited responses. These samples are from units with both monotonic (open) and nonmonotonic (filled) rate-level functions. Blue, suppressed responses; red, excited responses. Data points distributed along a vertical column represent multiple samples from the same unit (share the same expected firing rate). Units with no measurable rate-level properties were not included in this analysis. The diagonal line has a slope of 1. The 2 response types show minimal overlap. The observed firing rate for suppressed responses was always less than the expected rate, while the excited rate was either greater than or nearly equal to the expected firing rate.

production. The previous study also reported a group of neurons that showed no preference to playback or vocalization (Müller-Preuss and Ploog 1981). The excitatory responses we observed likely correspond to these neurons and reflect responses to auditory feedback (self-perception) of the produced vocalization.

Several studies have investigated effects of vocal production on auditory cortical responses using noninvasive techniques in humans. MEG studies during speaking have shown dampened responses to a subject's own voice compared with playback of recorded speech (Curio et al. 2000; Houde et al. 2002; Numminen and Curio 1999; Numminen et al. 1999). PET imaging studies in the human auditory cortex have also shown a reduction in cortical activity during speech production (Paus et al. 1996; Wise et al. 1999). While the reduction in the responsiveness of the auditory cortex revealed by these studies does not lead to the proof of inhibition in the auditory cortex during speaking, the observed phenomenon may be explainable by the findings of the present study. Two different types of vocalization-related neural response in the auditory cortex observed in our study suggest that the dampened responses during speaking in the human auditory cortex are possibly the combined responses of the two types of cortical vocalization responses, one inhibited and one excited. The MEG and PET techniques lack the spatial resolution necessary to separate groups of intermingled neurons and therefore only reflect their summed activity. Because of the proportion of vocalization-related responses showing excitation during self-initiated vocalizations is much smaller than those showing suppression (Fig. 13), the net effect of combining the two types of responses would be a dampened response as compared with playback sounds. Inhibitory sensory-motor interactions have been more extensively studied in the visual system where, for example, saccade-related inhibition of the visual cortex has been shown to be necessary to avoid incoherent inputs during eye movement (Judge et al. 1980).

It is important to point out the distinction between findings of the present study and those from songbird literature regarding gating of auditory information into nuclei involved in song production. It has been reported that neurons in RA, a motor structure, are responsive to playback of birdsongs under anesthetized conditions but are unresponsive when birds are awake (Dave et al. 1998). However, unlike the neurons in songbirds, neurons in marmoset auditory cortex that exhibit auditory-vocal interactions are highly responsive to playback of appropriate external acoustic stimuli when marmosets are *awake* (e.g., Fig. 7). In contrast, responsiveness of neurons in the auditory cortex of marmosets is weakened or diminished under anesthesia (Wang et al. 1995). The equivalent of the primate auditory cortex, a sensory cortical region, in songbirds is the field L not RA or HVC (Doupe and Kuhl 1999). Whether the auditory-vocal interactions reported here also occur in the field L of songbirds remains to be seen.

#### *Contribution of the auditory cortex to auditory-vocal interaction*

During vocalization, activity has been observed in the middle ear (Carmel and Starr 1963; Henson 1965; Saloman and Starr 1963; Suga and Jen 1975), cochlea (Goldberg and Henson 1998), and brain stem (Kirzinger and Jurgens 1991;

Metzner 1993; Papanicolaou et al. 1986; Suga and Schlegel 1972; Suga and Shimozaawa 1974) that results in a reduction of the auditory response to vocalization. Although this subcortical attenuation likely contributed to our observed suppression at the cortical level, it does not appear to fully account for the extent of the suppression we observed (Figs. 14 and 15). When the expected activity of cortical neurons was calculated from their rate-level characteristics, including an adjustment for the supposed 40 dB of subcortical attenuation, it was found to be much greater than the actually observed activity during vocalization. The difference suggests that cortical neurons are subject to additional inhibitory mechanisms during vocalization. These analyses cannot account for other subcortical difference between playback and vocalized sound perception nor were they intended to. Instead, they showed a quantitative difference between observed and expected cortical activity without breaking down the complex contributions of individual factors to subcortical mechanisms. Additionally, self-produced vocalization also induces suppression in other subcortical structures of bats including some neurons in the inferior colliculus (IC) (Metzner 1993). Whether vocalization also affects neural activities of the medial geniculate body (MGB) remains to be studied in the future.

The timing of suppression also supports the conclusion of a cortical role in vocalization-induced suppression. Auditory cortical neurons were inhibited several hundred milliseconds before a vocalization was produced. Such suppression of spontaneous neural discharges at the cortical level in the absence of acoustic stimuli was more likely to be observed if the cortical neurons themselves were the target of inhibition rather than the suppression of a subcortical site whose outputs drive the spontaneous activity of the auditory cortex. This suggests that cortical neurons are subjected to inhibition beginning before vocalization and that, once vocalization begins, attenuation in subcortical structures is added, both contributing to differences between responses to playback stimuli and self-initiated vocalization. Additionally, cochlear and lateral lemniscal attenuation mechanisms have been observed to occur either immediately (a few ms) before, or synchronized with, vocal production (Goldberg and Henson 1998; Metzner 1993; Suga and Shimozaawa 1974). In the auditory cortex, the suppression begins far earlier than in these subcortical structures. We would therefore argue that these cochlear and brain stem effects could be the result of descending cortical efferent control. Recent work in the cortifugal system has demonstrated that activity in the auditory cortex acts to alter the response properties of neurons in the colliculus and thalamus (Sakai and Suga 2001; Yan and Suga 1998; Zhang and Suga 2000) as well as the properties of cochlear hair cells (Xiao and Suga 2002). It is therefore conceivable that the modulation of the auditory cortex before and during vocalization alters the cortical efferent controls of the middle ear, cochlea, and brain stem causing the subcortical attenuation of self-produced vocal feedback.

The role of subcortical structures during vocalization is likely more complex than simple attenuation. While evoked potentials studies (Papanicolaou et al. 1986; Suga and Shimozaawa 1974) demonstrate overall attenuation, single-unit recordings in structures from the cochlear nucleus to the colliculus have revealed a mix of inhibited and excited vocalization-related activities (Kirzinger and Jurgens 1991; Metzner 1993). Excited responses have been attributed to acoustic feedback



perception by the ascending auditory system (Kirzinger and Jurgens 1991), although, intriguingly, some excitation has been seen before vocal onset (Metzner 1993) that may suggest a more complex mechanism. Although we did not observe consistent onsets of cortical excitation before vocalization in the present study, the presence of noninhibited brain stem neurons may serve as inputs to noninhibited cortical neurons allowing the observed vocalization-related excitations.

#### *Possible origin of modulatory effects in the auditory cortex during self-produced vocalization*

Findings of this study suggest that the suppression of neural activity in the auditory cortex during vocalization results, at least partially, from an internal modulatory neural mechanism rather than depending entirely on the acoustic feedback of the produced vocalization. Neurons that are suppressed during vocalization are often excited by playback of recorded vocalizations. Furthermore, the suppression actually begins prior to the onset of vocalization and therefore cannot be purely a result of auditory feedback. Possible sources of modulatory inputs responsible for the vocalization-induced suppression of the auditory cortex are brain structures involved in vocal production. Pre-vocalization activity in monkeys has been observed in the prefrontal, premotor, and other cortical areas  $\leq 1,000$  ms preceding vocalizations (Gemba et al. 1995). The interval of this prevocal activity is comparable to the distribution of the timing of prevocalization suppression in the auditory cortex (Fig. 5). In non-human primates, anatomical and physiological connections between the frontal motor areas and the superior temporal gyrus have been demonstrated in a number of species, including the Old World and New World monkeys (Alexander et al. 1976; Chavis and Pandya 1976; Hackett et al. 1999; Jones and Powell 1970; Morel and Kaas 1992; Pandya and Kuypers 1969; Petrides and Pandya 1988; Romanski et al. 1999a). It is therefore possible that sensory-motor interactions in the auditory cortex observed in our experiments originate in frontal vocal-production areas and act via frontal-temporal connections in the form of direct inhibition of auditory cortical neurons both before and during vocalization. An earlier study has also shown inhibitory effects on the auditory cortex produced by electrically stimulating the cingulate cortex (Müller-Preuss et al. 1980), evidence that other brain regions can cause inhibitory modulations of this sensory area. If vocalization-related suppression were to originate from such cortical-cortical connections, one might expect to find suppressed-responding neurons in the upper cortical layers, whereas excited-responding neurons, lacking such connections, might reside in middle and deeper cortical layers. At this point, however, there is insufficient evidence to support such separation.

#### *Contribution of bone conduction to vocalization-related neural activity*

While the observed cortical responses during vocalization are likely due to neural mechanisms, the contribution of bone conduction cannot be excluded. Vocalization produces vibrations in the skull that are subjected to alteration in both frequency and phase (Békésy 1949; Stenfelt et al. 2000) before being transduced in the cochlea through a complex set of mechanisms (see review by Tonndorf 1976). In general, how-

ever, bone-conducted signals show reduced intensity at higher frequencies. Bone-conducted vocalizations provide a stimulus of equal intensity to air-conducted feedback (Békésy 1949) and are subject to the same magnitude of vocalization-associated changes, such as attenuation by middle-ear muscle contraction (Irvine and Wester 1974).

While we do not know the characteristics of bone conduction in the marmoset, the studied phee calls contain high-frequency energy (more than  $\sim 5$  kHz), and thus bone-conducted signals are likely attenuated. We can discount bone conduction as the primary cause of cortical inhibition because suppression was observed to begin before any vocalization was produced, although contribution to further modulation of inhibited firings during vocalization cannot be excluded.

Contribution of bone conduction to the analysis of observed and expected vocalization-related responses was largely ignored because, presumably, the previously reported subcortical attenuation included any modulation by bone conduction.

#### *Responses to external acoustic stimuli during vocalization*

During vocalization-induced inhibition, presentation of external acoustic stimuli resulted in decreased stimulus responses compared with stimuli presented when the animal was quiet. Such suppression of acoustic responses was reported by Müller-Preuss and Ploog (1981) for a small number of samples. Because many units had monotonic rate-level functions, this suppression of acoustic responsiveness was likely not the results of an acoustic factor such as the increased sound level caused by the vocalization. The difference in frequency spectra between vocalizations ( $>6$ – $7$  kHz) and external acoustic stimuli (e.g., Fig. 7A, carrier frequency at 2.99 kHz) reduces the likelihood of two-tone inhibition that often occurs for specific frequency combinations (Kadia and Wang 2003). Additionally, the nearly normal stimulus responses, during vocalization-related excitation, to similar stimuli argues against suppression resulting from acoustic interference. It is therefore likely that the suppression of responses to external acoustic stimuli results from the same neural mechanisms that suppresses the spontaneous activity of cortical neurons.

#### *Potential roles of neurons with inhibitory and excitatory responses to self-produced vocalization*

Because of the proximity between the mouth and the ears and the conduction of sound by the skull bones, one's own voice represents a high-intensity input to the auditory system. In marmosets, the intensity of self-produced phee calls can be as high as 105 dB SPL. Such an acoustic stimulus would likely be beyond the saturation point for most neurons with monotonic rate-level characteristics and far above the peak response level for neurons with nonmonotonic rate-level characteristics. At high sound levels, most neurons would be unable to effectively encode details of the vocalization either due to saturated firing rates or inhibited discharges. One possible role of the inhibitory interactions between vocal production and perception systems could be to elevate response thresholds, or shift peak response levels, of neurons in the auditory cortex, effectively attenuating auditory feedback and inputs from the periphery, so that neurons can respond to loud self-produced vocalizations with a greater dynamic range. Without such

adjustment, the auditory system could be overwhelmed and unable to effectively encode self-produced sounds during vocalization.

Perceiving the acoustic environment during vocalization is also essential to both humans and animals to take proper actions (such as avoiding a predator or picking up another person's conversation in a party). The extent of vocalization-induced suppression suggests that the sensitivity of much of the auditory cortex would be reduced as a result of vocalization. In fact, many neurons show poor, and often absent, responses to the presentation of external auditory stimuli during vocalization. The existence of a subpopulation of neurons showing vocalization-related excitation provides a possible means for the auditory cortex to maintain hearing sensitivity during vocalization. During excitatory responses, neurons respond nearly normally to external acoustic stimuli regardless of the animal's vocal production.

One possibility is therefore that neurons showing suppression play a specialized role to encode auditory feedback during vocalization, whereas those exhibiting excitation act to monitor external acoustic environment without much loss of sensitivity. Maintaining auditory perception during vocalization may play a role in providing the vocal production system with feedback for the purpose of self-monitoring during speaking and possibly during vocalization.

#### *Broader implications of auditory-vocal interactions in primate auditory cortex*

Although the primate auditory cortex has long been studied for its sensory functions, our findings suggest that this cortical region is modulated by inputs from brain structures involved in vocal production. These modulatory inputs may be indicative of an efference-copy system at work. The feed-forward neural inputs from the vocal production system to the auditory cortex may provide a neural representation of the expected vocalization that is compared with the auditory cortical representation of acoustic feedback during vocalization. The difference between these two representations (an error signal) may be fed back to the vocal production system and used to modify or refine vocal production. The presence of some neurons with complete suppression during vocalization, which does not fit well with the threshold-shift representation discussed in the preceding text, could be considered equivalent to a perfect feedback-efferent match if the outputs of these neurons are used as error signals for the vocal production system.

Delays in auditory feedback and decreased frontal-temporal activity have been implicated as possible mechanisms underlying human stuttering (Fox et al. 1996). Alterations in auditory feedback are also responsible for long-term alteration of bird song (Brainard and Doupe 2000; Leonardo and Konishi 1999) and short-term alterations in human speech (Houde and Jordan 1998). It is therefore essential that a feedback-efferent comparison takes place to detect and correct perturbations in vocal production, even if corrections do not occur on a moment-to-moment basis. We do not yet know that over what time scales acoustic feedback affects marmoset vocal production.

The findings of the present study also have important implications for the segregation of auditory information in the superior temporal gyrus. It has been suggested that the primate

auditory cortex and its projections are segregated into "what" and "where" pathways (Kaas et al. 1999; Romanski et al. 1999b; Tian et al. 2001). The evidence of sensory-motor interactions in the auditory cortex suggests that cortical processing of auditory information must include additional pathways. The auditory system needs to know not only where a sound comes from and what is the meaning of a sound but also *who* produces it. An important distinction between the auditory system and other sensory systems is that, in vocal species such as human and non-human primates, self-produced vocal communication sounds (i.e., speech and vocalizations) represent a large proportion of behaviorally important inputs. A "vocal" pathway may have evolved to process vocal communication sounds. As we suggested earlier (Wang 2000), such a pathway is likely a mutual communication channel between the vocal production system and the auditory cortex, possibly via reciprocal connections between the frontal lobe and the superior temporal gyrus. We speculate that the neural circuitry mediating sensory-motor interactions in the auditory cortex may also play a role in cortical coding of vocal communication sounds in primates, in terms of providing a channel for memory retrieval. This pathway may not be involved when nonvocal acoustic signals are processed by the cerebral cortex. Findings from the present study should encourage further studies along this direction that will likely lead to a better understanding of cortical mechanisms underlying vocal communication.

The authors thank Drs. T. Lu, L. Liang, and R. Snider for technical and experimental assistance, Drs. E. Bartlett and T. Lu for helpful comments on the manuscript, and A. Pistorio for assistance with animal training.

This work was supported by a National Institute on Deafness and Other Communication Disorders grant and by a Presidential Early Career Award for Scientists and Engineers (X. Wang).

#### REFERENCES

- Agamaite JA and Wang X. Quantitative classification of the vocal repertoire of the common marmoset (*Callithrix jacchus jacchus*). *Assoc Res Otolaryngol Abstr* 20: 573, 1997.
- Aitkin L and Park V. Audition and the auditory pathway of a vocal new world primate, the common marmoset. *Prog Neurobiol* 41: 345–367, 1993.
- Alexander GE, Newman JD, and Symmes D. Convergence of prefrontal and acoustic inputs upon neurons in the superior temporal gyrus of the awake squirrel monkey. *Brain Res* 116: 334–338, 1976.
- Békésy G von. The structure of the middle ear and the hearing of one's own voice by bone conduction. *J Acoust Soc Am* 21: 217–232, 1949.
- Brainard MS and Doupe AJ. Auditory feedback in learning and maintenance of vocal behavior. *Nat Rev Neurosci* 1: 31–40, 2000.
- Carmel PW and Starr A. Acoustic and nonacoustic factors modifying middle ear muscle activity in waking cats. *J Neurophysiol* 26: 598–616, 1963.
- Chavis DA and Pandya DN. Further observations on corticofrontal connections in the rhesus monkey. *Brain Res* 117: 369–386, 1976.
- Creutzfeldt O, Ojemann G, and Lettich E. Neuronal activity in the human lateral temporal lobe. II. Responses to the subjects own voice. *Exp Brain Res* 77: 476–489, 1989.
- Curio G, Neuloh G, Numminen J, Jousmaki V, and Hari R. Speaking modifies voice-evoked activity in the human auditory cortex. *Hum Brain Mapp* 9: 183–191, 2000.
- Dave AS, Yu AC, and Margoliash D. Behavioral state modulation of auditory activity in a vocal motor system. *Science* 282: 2250–2254, 1998.
- Doupe AJ and Kuhl PK. Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22: 567–631, 1999.
- Epplé G. Comparative studies on vocalization in marmoset monkeys (*Haplorhina*). *Folia Primatol* 8: 1–40, 1968.
- Fox PT, Ingham RJ, Ingham JC, Hirsch TB, Downs JH, Martin C, Jerabek P, Glass T, and Lancaster JL. A PET study of the neural systems of stuttering. *Nature* 382: 158–161, 1996.

- Gemba H, Miki N, and Sasaki K. Cortical field potentials preceding vocalization and influences of cerebellar hemispherectomy upon them in monkeys. *Brain Res* 697: 143–151, 1995.
- Goldberg RL and Henson OW, Jr. Changes in cochlear mechanics during vocalization: evidence for a phasic medial efferent effect. *Hear Res* 122: 71–81, 1998.
- Gunji A, Hoshiyama M, and Kakigi R. Auditory response following vocalization: a magnetoencephalographic study. *Clin Neurophysiol* 112: 514–520, 2001.
- Hackett T, Stepniewska I, and Kaas JH. Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res* 817: 45–58, 1999.
- Henson OW. The activity and function of the middle-ear muscles in echolocating bats. *J Physiol* 180: 871–887, 1965.
- Houde JF and Jordan MI. Sensorimotor adaptation in speech production. *Science* 279: 1213–1216, 1998.
- Houde JF and Jordan MI. Sensorimotor adaptation of speech I: Compensation and adaptation. *J Speech Lang Hear Res* 45: 295–310, 2002.
- Houde JF, Nagarajan SS, Sekihara K, and Merzenich MM. Modulation of the auditory cortex during speech: an MEG study. *J Cog Neurosci* 14: 1125–1138, 2002.
- Irvine DR and Wester KG. Middle ear muscle effects on cochlear responses to bone-conducted sound. *Acta Physiol Scand* 91: 482–496, 1974.
- Jones EG and Powell TPS. An anatomical study of converging sensory pathways within the cerebral cortex of monkey. *Brain* 93: 793–820, 1970.
- Judge SJ, Wurtz RH, and Richmond BJ. Vision during saccadic eye movements. I. Visual interactions in striate cortex. *J Neurophysiol* 43: 1133–1155, 1980.
- Kaas JH, Hackett TA, and Tramo MJ. Auditory processing in primate cerebral cortex. *Curr Opin Neurobiol* 9: 164–170, 1999.
- Kadia SC and Wang X. Spectral integration in A<sub>1</sub> of awake primates: neurons with single- and multi-peaked tuning characteristics. *J Neurophysiol* 89: 1603–1622, 2003.
- Kirzinger A and Jurgens U. Vocalization-correlated single-unit activity in the brain stem of the squirrel monkey. *Exp Brain Res* 84: 545–560, 1991.
- Leonardo A and Konishi M. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature* 399: 466–470, 1999.
- Liang L, Lu T, and Wang X. Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. *J Neurophysiol* 87: 2237–2261, 2002.
- Liberman AM. *Speech: A Special Code*. Cambridge, MA: MIT Press, 1996.
- Liberman AM and Mattingly IG. A specialization for speech perception. *Science* 243: 489–494, 1989.
- Lu T, Liang L, and Wang X. Neural representation of temporally asymmetric stimuli in the auditory cortex of awake primates. *J Neurophysiol* 85: 2364–2380, 2001a.
- Lu T, Liang L, and Wang X. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci* 4: 1131–1138, 2001b.
- Mann VA and Liberman AM. Some differences between phonetic and auditory modes of perception. *Cognition* 14: 211–235, 1983.
- Margoliash D. Functional organization of forebrain pathways for song production and perception. *J Neurobiol* 33: 671–693, 1997.
- McCasland JS and Konishi M. Interaction between auditory and motor activities in an avian song control nucleus. *Proc Natl Acad Sci USA* 78: 7815–7819, 1981.
- Metzner W. An audio-vocal interface in echolocating horseshoe bats. *J Neurosci* 13: 1899–1915, 1993.
- Morel A and Kaas JH. Subdivisions and connections of auditory cortex in owl monkeys. *J Comp Neurol* 318: 27–63, 1992.
- Müller-Preuss P, Newman JD, and Jurgens U. Anatomical and physiological evidence for a relationship between the “cingular” vocalization area and the auditory cortex in the squirrel monkey. *Brain Res* 202: 307–315, 1980.
- Müller-Preuss P and Ploog D. Inhibition of auditory cortical neurons during phonation. *Brain Res* 215: 61–76, 1981.
- Numminen J and Curio G. Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neurosci Lett* 272: 29–32, 1999.
- Numminen J, Salmelin R, and Hari R. Subject’s own speech reduces reactivity of the human auditory cortex. *Neurosci Lett* 265: 119–122, 1999.
- Pandya DN and Kuypers HG. Cortico-cortical connections in the rhesus monkey. *Brain Res* 13: 13–36, 1969.
- Papanicolaou AC, Raz N, Loring DW, and Eisenberg HM. Brain stem evoked response suppression during speech production. *Brain Lang* 27: 50–55, 1986.
- Paus T, Perry DW, Zatorre RJ, Worsley KJ, and Evans AC. Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur J Neurosci* 8: 2236–2246, 1996.
- Perry DW, Zatorre RJ, Petrides M, Alivisatos B, Meyer E, and Evans AC. Localization of cerebral activity during simple singing. *Neuroreport* 10: 3979–3984, 1999.
- Petrides M and Pandya DN. Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkeys. *J Comp Neurol* 283: 52–66, 1988.
- Pfingst BE and O’Connor TA. Characteristics of neurons in auditory cortex of monkeys performing a simple auditory task. *J Neurophysiol* 45: 16–34, 1981.
- Romanski LM, Bates JF, and Goldman-Rakic PS. Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkeys. *J Comp Neurol* 403: 141–157, 1999a.
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, and Rauschecker JP. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2: 1131–1136, 1999b.
- Sakai M and Suga N. Plasticity of the cochleotopic (frequency) map in specialized and nonspecialized auditory cortices. *Proc Natl Acad Sci USA* 98: 3507–3512, 2001.
- Salomon B and Starr A. Electromyography of middle ear muscles in man during motor activities. *Acta Neurol Scand* 39: 161–168, 1963.
- Stenfelt S, Hakansson B, and Tjellstrom A. Vibration characteristics of bone conducted sound in vitro. *J Acoust Soc Am* 107: 422–431, 2000.
- Suga N and Jan PH. Peripheral control of acoustic signals in the auditory system of echolocating bats. *J Exp Biol* 62: 277–311, 1975.
- Suga N and Schlegel P. Neural attenuation of responses to emitted sounds in echolocating bats. *Science* 177: 82–84, 1972.
- Suga N and Shimoza T. Site of neural attenuation of responses to self-vocalized sounds in echolocating bats. *Science* 183: 1211–1213, 1974.
- Tian B, Reser D, Durham A, Kustov A, and Rauschecker JP. Functional specialization in rhesus monkey auditory cortex. *Science* 292: 290–293, 2001.
- Tonndorf J. Bone Conduction. In: *Handbook of Sensory Physiology*, edited by Keidel WD and Neff WD. New York: Springer-Verlag, 1976, p. 37–84.
- Wang X. On cortical coding of vocal communication sounds in primates. *Proc Natl Acad Sci USA* 97: 11843–11849, 2000.
- Wang X, Liang L, Eliades S, Snider RK, Lu T, Kadia SC, and Barbour D. Characteristics of single-unit responses in the auditory cortex of awake marmoset monkeys. *Assoc Res Otolaryngol Abstr* 22: 35, 1999.
- Wang X, Merzenich MM, Beitel R, and Schreiner CE. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol* 74: 2685–2706, 1995.
- Wise RJ, Greene J, Buchel C, and Scott SK. Brain regions involved in articulation. *Lancet* 353: 1057–1061, 1999.
- Xiao Z and Suga N. Modulation of cochlear hair cells by the auditory cortex in the mustached bat. *Nat Neurosci* 5: 57–63, 2002.
- Yan W and Suga N. Corticofugal modulation of the midbrain frequency map in the bat auditory system. *Nat Neurosci* 1: 54–58, 1998.
- Zhang Y and Suga N. Modulation of responses and frequency tuning of thalamic and collicular neurons by cortical activation in mustached bats. *J Neurophysiol* 84: 325–333, 2000.