

Dynamics of Auditory–Vocal Interaction in Monkey Auditory Cortex

Steven J. Eliades and Xiaoqin Wang

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

Single neurons in the primate auditory cortex exhibit vocalization-related modulations (excitatory or inhibitory) during self-initiated vocal production. Previous studies have shown that these modulations of cortical activity are variable in individual neurons' responses to multiple instances of vocalization and diverse between different cortical neurons. The present study investigated dynamic patterns of vocalization-related modulations and demonstrated that much of the variability in cortical modulations was related to the acoustic structures of self-produced vocalization. We found that suppression of single unit activity during multi-phrased vocalizations was temporally specific in that it was maintained during each phrase, but was released between phrases. Furthermore, the degree of suppression or excitation was correlated to the mean energy and frequency of the produced vocalizations, accounting for much of the response variability between multiple instances of vocalization. Simultaneous recordings of pairs of neurons from a single electrode revealed that the modulations by self-produced vocalizations in nearby neurons were largely uncorrelated. Additionally, vocalization-induced suppression was found to be preferentially distributed to upper cortical layers. Finally, we showed that the summation of all auditory cortical activity during vocalization, including both single and multi-unit responses, was weakly excitatory, consistent with observations from studies of the human brain during speech.

Keywords: feedback, marmoset, sensory–motor interaction, vocalization, vocal production

Introduction

Studies of sound coding in the auditory cortex have yielded a great deal of information about the neural representation of externally presented sound. In natural environments, however, humans and both animals commonly engage in both listening and speaking (or vocalizing), often simultaneously. Our knowledge of auditory processing related to these self-generated sounds is limited. The sensory inputs that result from these sounds, such as human speech, often have important behavioral effects. Humans continuously monitor their speech in order to compensate for any perturbations in acoustic structure. Shifts in the spectral profile of speech feedback, for example, result in compensatory changes in both the produced fundamental and formant frequencies (Burnett *et al.*, 1998; Houde and Jordan, 1998). Alteration in the perceived intensity of speech, whether due to amplification, attenuation or masking, leads to compensatory changes in the amplitude of production (Lane and Tranel, 1971; Siegel and Pick, 1974). Delays in auditory feedback lead to stuttering-like speech in normal subjects (Fairbanks, 1955) and have been suggested to be one possible mechanism underlying stuttering disorders (Lee, 1950). Human speech is not the only self-produced sound subject to feedback monitoring, as

animals also show similar feedback-dependant vocal control behavior, including temporal patterning in birdsong (Leonardo and Konishi, 1999), frequency in bat echolocation sounds (Schuller *et al.*, 1974) and amplitude in primate, cat and bird vocalizations (Sinnot *et al.*, 1975; Nonaka *et al.*, 1997; Cynx *et al.*, 1998; Brumm *et al.*, 2004). In order to understand neural mechanisms underlying this type of sensory–motor control, it is important to study sensory processing in the auditory system during vocal production.

Alterations of physiological properties and neural activities during vocalization have been found throughout the auditory pathway. The intensity of acoustic inputs is attenuated during vocalization by the contraction of the middle ear muscles (Carmel and Starr, 1963; Henson, 1965; Suga and Jen, 1975). Studies of the bat brainstem have revealed echolocation-related activity in lateral lemniscal and nearby nuclei (Suga and Shimozawa, 1974; Metzner, 1989) whose function is necessary for feedback-dependent alterations in echolocation call frequency (Smotherman *et al.*, 2003). Vocalization-related activity has also been found, though not systematically studied, in a small number of neurons in the primate auditory brainstem (Kirzinger and Jurgens, 1991; Tammer *et al.*, 2004). However, in both humans and non-human primates, most studies of auditory–vocal interactions have been primarily focused on the auditory cortex. A variety of measurement techniques, including magnetoencephalography (MEG; Numminen *et al.*, 1999; Curio *et al.*, 2000; Gunji *et al.*, 2001; Houde *et al.*, 2002), positron emission tomography (PET; Paus *et al.*, 1996; Wise *et al.*, 1999) and intra-operative electrocorticography (ECoG, Crone *et al.*, 2001), have shown a reduction in human auditory cortical responses during speech production, compared with passive listening, that is largely absent from auditory-brainstem recordings (Houde *et al.*, 2002). Limited intra-operative multi-unit recordings have also shown both weakly excitatory and inhibitory events in the human middle and, to a lesser extent, superior temporal gyri (Creutzfeldt *et al.*, 1989).

Reduction of neural activity at the cellular level has been previously observed in the auditory cortex of squirrel monkeys electrically stimulated to vocalize (Müller-Preuss and Ploog, 1981). Our recent study in the auditory cortex of spontaneously vocalizing (self-initiated) marmoset monkeys revealed two types of single-neuron responses (Eliades and Wang, 2003). The majority of cortical neurons showed vocalization-induced suppression beginning prior to the onset of vocal production, while the remaining minority showed vocalization-related excitation beginning after the vocal onset. The suppression was interpreted as inhibition originating from vocal production centers, whereas the excitation was thought to represent sensory responses to auditory feedback of the self-produced

vocalization. Neurons suppressed during vocal production were generally driven or unresponsive, but not inhibited, by similar vocalizations played from a speaker. The neurons showing suppression were also found to respond poorly to external sounds during vocalization, but neurons showing excitation responded to external sounds similarly during or in the absence of vocal production. Quantitative analysis of the magnitude of suppression also suggested that it was at least partly attributable to cortical inhibition. This study, however, raised a number of unanswered issues. First, was the variability of vocalization-related modulations in individual neurons random, or related to acoustic properties of vocalization? Second, what are the possible origins for the response diversity between different neurons? Finally, how can single-unit observations from the primate auditory cortex be reconciled with the weak excitation during speech recorded from human brains by imaging and other measures? The current study attempts to address these issues by examining dynamic changes in the vocalization-related modulations of individual cortical neurons as a function of the acoustic structure of self-produced vocalization, and by examining the diversity between neurons as a function of their anatomic locations, auditory response properties and quality of action potential isolation.

Materials and Methods

Experimental and Electrophysiological Recording Procedures

Recordings were obtained from two awake marmoset monkeys (*Callithrix jacchus*), a highly vocal New World primate (Aitkin and Park, 1993; Wang, 2000). Details of surgical and recording procedures have been previously published (Lu *et al.*, 2001b; Eliades and Wang, 2003). Because of the unpredictable timing of animal's vocalizations, the auditory responses of the cortical neurons were studied using external stimuli while awaiting the animals to vocalize. All experiments procedures were approved by the Johns Hopkins University Animal Care and Use Committee.

Neural activities were recorded using tungsten microelectrodes (A-M Systems, Carlsborg, WA, or Microprobe, Potomac, MD) with impedances of 2–5 M Ω . Action potentials were initially sorted online, using a template-based spike sorter (MSD, Alpha-Omega Engineering, Nazareth, Israel), in order to maximize the quality of the recorded spikes. The raw neural data from the electrode were also simultaneously digitized onto one channel of a DAT recorder (Panasonic SV-3700). Action potentials were sorted a second time (off-line) from the raw data recordings using a feature-based manual clustering method (Lewicki, 1998). This allowed sorting of the well-isolated single-units separated during recording as well as simultaneously recorded single-units of smaller spike size. A multi-unit background activity was also obtained from crossings of a manually defined threshold, but excluding events already sorted as single-units. The majority of analyses presented were performed only on the large-size single-units, the same set of data used in our previous work (Eliades and Wang, 2003).

Well-isolated single-units were acoustically characterized using auditory stimuli delivered free field from a speaker located ~1 m in front of the animal. Basic auditory response characteristics were obtained for each neuron, including center frequency (CF), rate-level curve (RL), latency and responses to pure tone stimuli versus narrow and wideband noise. When a unit was unresponsive or weakly responsive to pure tone stimuli, narrow band noise was used instead to determine its CF and RL characteristics. Other stimuli were also presented as part of an ongoing study of complex sound representation, including click trains (Lu *et al.*, 2001a) and amplitude and frequency modulated tones (Liang *et al.*, 2002). Units were recorded from both primary auditory cortex (A1) and lateral belt areas, including all cortical layers. More extensive physiological recordings not included in this report allowed the separation of A1 from lateral belt areas based upon tone or noise preferences (Rauschecker *et al.*, 1995) and A1 from the

adjacent rostral core area (R) based upon reversal of the tonotopic gradient.

Vocal Recordings and Analysis

The monkeys used in the reported experiments produced spontaneous, self-initiated vocalizations during extracellular recording sessions, each lasting 3–5 h. Behavioral conditions were identical to those noted in our previous study (Eliades and Wang, 2003). Briefly, an animal sat in a primate chair, with its head immobilized, within a double-walled sound attenuation chamber (IAC-1024) while acoustic stimuli were being presented. The stimuli were played at regular intervals within a given stimulus protocol, generally between 600 and 5000 ms (depending upon the protocol). Between the end of one protocol and the start of the next was period of silence of variable length. There were also periods of time when a variety of search stimuli (tone, noise, vocalization samples, etc.) were played while the experimenter was advancing the electrode to isolate single-units. While animals produced some vocalizations during the presentation of acoustic stimuli, the majority of vocalizations were produced during periods of silence or search stimuli. No clear correlation between self-initiated vocalizations and any particular stimuli was noted. Animals vocalized at irregular intervals, sometimes with only a single call over several h and other times occurring in bouts with a call occurring regularly every 20–40 s over a 10 or 20 min period. Animals were not rewarded for vocalizing, and the regular presentation of auditory stimuli persisted regardless of whether or not the animal vocalized. A few attempts were made in later recording sessions to increase the number of vocalizations by introducing other animals or visual stimuli into the sound chamber, but without much success.

Vocal recordings were obtained using a directional microphone (AKG C100S) placed at mouth level ~15 cm in front of the animal and were digitized simultaneously with the neural signal onto the other channel of the DAT recorder. The vocalizations obtained were distributed over 134 h of recording. All vocalizations obtained from the first animal were isolation calls (phee calls; Eppele, 1968; Agamonte and Wang, 1997), while the second animal produced a mix of isolation and social calls. All together, 1236 vocalizations and their neural responses were recorded.

Recorded vocalizations were transferred to a PC for acoustic analysis. Because of their prevalence and long duration (simplifying analysis of the neural data), only phee calls were subjected to quantitative analyses. The onset and offset of vocalization were determined by the presence of spectral energy in vocal frequency bands (3–12 kHz). A short-time Fourier transform (spectrogram) was performed on each vocalization and the peak frequency (frequency with the maximum energy) obtained for each time bin (bin size = 2.6 ms). The mean frequency of the call was obtained by averaging the time-specific peak frequencies. The average energy of the vocalization was calculated from the logarithmic root-mean-square energy of the call, normalized by the energy of the quiet acoustic background, and divided by the duration of the call. When single and multi-phrase vocalizations were compared, only the first phrase results of the multi-phrase were used.

Analysis of Neural Data

Analysis was primarily performed on 109 well-isolated single-units whose recording included at least one instance of vocalization by the animal. These single-units were based on recorded large action potentials with signal-to-noise ratios (SNR) > 10 dB, the criterion of the data included in Eliades and Wang (2003). In the present study, we also analyzed additional single-units with smaller action potentials and multi-unit background activity that were recorded from the same electrode along with the large action potentials. For convenience, we will refer to these three sets of data as 'large unit', 'small unit' and 'multi-unit', respectively. The small unit and multi-unit data were only used in a few specific analyses, as indicated in the text, figures and legends. Small units were analyzed primarily for comparisons between multiple units recorded simultaneously from a single electrode. Multi-units were included only when comparing our primate neurophysiology to imaging studies of the human auditory cortex.

The effect of vocalization on neural firing was quantified using a normalized rate metric, the Response Modulation Index (RMI), defined

as $RMI = (R_{\text{vocal}} - R_{\text{prevocal}}) / (R_{\text{vocal}} + R_{\text{prevocal}})$, where R_{vocal} and R_{prevocal} are the firing rate during and before vocalization, respectively. Multi-phrase vocalization RMIs were calculated individually for each phrase and used the same prevocal activity period for both phrases. An RMI of 0 indicates no vocalization-related change in firing rate, while an RMI of -1 indicates a complete suppression of all spontaneous or driven activity. An RMI of +1 indicates a unit with strong vocalization-related excitation, a low spontaneous rate or both. For the purposes of classification, units were considered suppressed or excited based on a median RMI, calculated across all vocalization responses recorded from that unit, of either less or greater than 0. Temporal patterns of vocalization-related activity were studied by constructing peri-stimulus time histograms (PSTHs) using 25 ms bins, and aligning the responses by either the onset or offset of vocal production. Responses from single-trials were converted to spike-density functions, for display purposes only, by convolving single-unit spike trains with a gaussian function ($\sigma = 25$ ms) with unit area.

An attempt was made to discover the relationship between acoustic parameters of the produced vocalization (frequency and amplitude) and the modulation of neuronal firings. The limited number of samples available and the inability to control the acoustics thereof prevented the construction of a parametric tuning curve. The relationship between vocalization acoustics and neural responses was quantified using a Pearson's linear correlation coefficient (r) and tested using the t -distribution. This analysis was limited to those units in which four or

more vocalization responses were recorded. Only the responses during the first phrase were used from multi-phrase vocalizations.

Results

Temporal Patterns of Vocalization-related Modulation during Multi-phrase Vocalizations

The majority of vocalizations produced by the animals under our experimental conditions were phee (isolation) calls. While most of these consisted of a single phrase, 39% of the phee calls contained two or more phrases (Fig. 1A). Our previous analysis of this set of data combined both single and multi-phrase vocalizations (Eliades and Wang, 2003). In the present study, we examined the multi-phrase cases independently. Similar to what we described in our earlier report, there were two types of responses during multi-phrase vocalizations, suppression and excitation (portayed in Fig. 1A-C and D-F, respectively). In the suppression case, the neuron's spontaneous activity was either partially or completely eliminated during the first vocal phrase, as illustrated by an example in Figure 1A. The end of the first phrase was usually followed by a burst of spikes during the inter-phrase

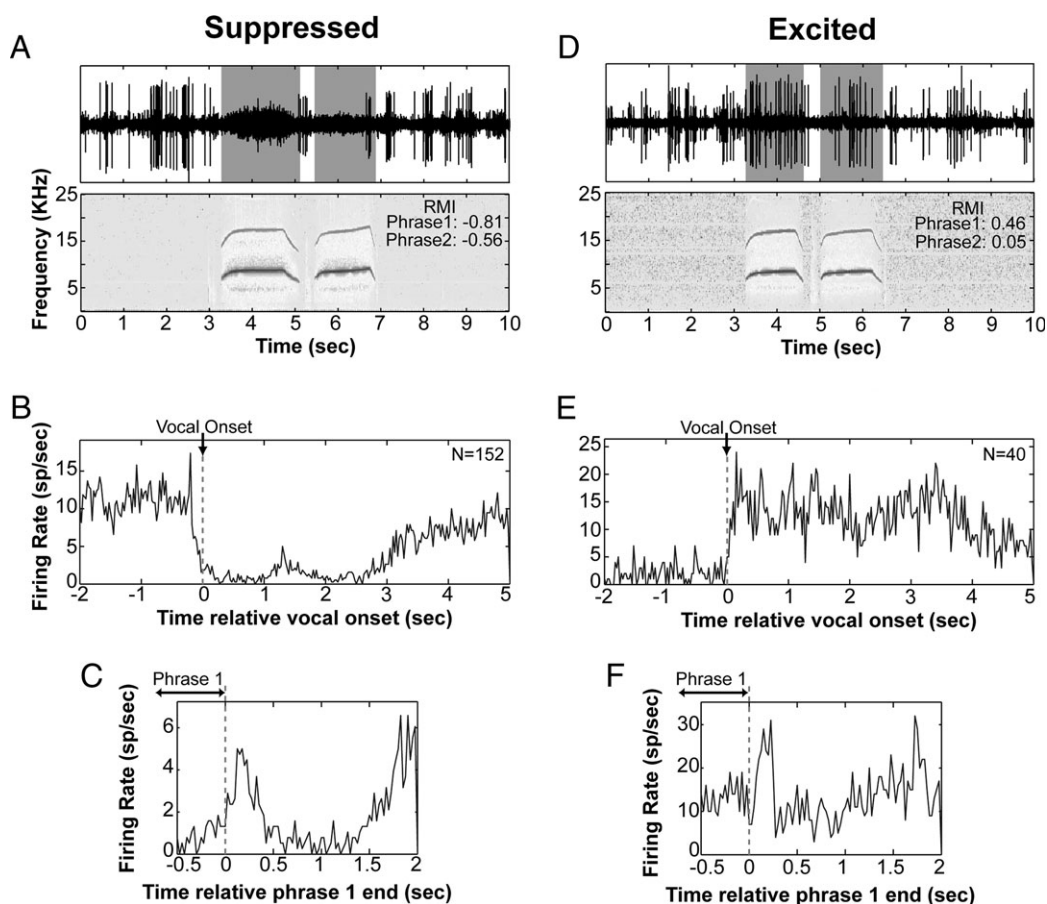


Figure 1. Responses to multi-phrase vocalizations. (A, D) Examples of suppression (A) and excitation (D) during multiple phrases of vocalization. Top: digitized extracellular recording traces of well isolated neurons; bottom: corresponding spectrograms of simultaneously recorded vocalizations. Shaded regions indicate the period of vocalization. (B, E) Population PSTHs for suppressed (B) and excited (E) vocalization responses created by aligning spike trains to the onset of vocal production (dashed line). Suppressed responses show an onset before vocalization, while excitation begins after. A small return of activity occurs during the inter-phrase period. (C, F) PSTHs showing the inter-phrase period calculated by aligning responses to the end of the first phrase of vocalization. This alignment reveals a transient increase in firing just after the end of the first vocal phrase in both suppressed and excited responses.

period and then a second period of suppression during the second phrase of the phee. In this example, the degree of suppression was slightly less for the second phrase than for the first (phrase 1 RMI = -0.81 , phrase 2 RMI = -0.56). Figure 1D shows an example of an excitation case. The excited response showed an increase in the number of spikes during the first phrase, followed by lag in firing during the inter-phrase period, and finally by a second increase in firing during the second phrase (Fig. 1D). The degree of excitation was decreased from the first to the second phrase for this particular example (phrase 1 RMI = 0.46 , phrase 2 RMI = 0.05), to the point where the second phrase response was only marginally above spontaneous firing.

The onset of modulation in the multi-phrase vocalizations was also consistent with that seen in single-phrase cases. A PSTH was constructed from all suppressed multi-phrase vocalization responses and aligned by the onset of vocal production (Fig. 1B), and showed that the beginning of suppressive modulation began prior to the onset of vocalization. In contrast, a PSTH constructed from excitatory multi-phrase vocalization responses showed an increase in firing rate beginning shortly after the onset of vocalization (Fig. 1E).

It is interesting to note the burst of spikes observed during the inter-phrase period of suppressed vocalization responses (Fig. 1A). This phenomenon was observed in many samples of vocalization-induced suppression ($80/172$), and is reflected in the averaged activity by a transient increase in firing rate after the first period of suppression (Fig. 1B). In order to examine this short increase in spiking activity, the PSTH of suppressed cases was recalculated, this time aligning spike trains to the end of the first vocalization phrase (Fig. 1C). Under this alignment, the transient return of activity becomes more pronounced. Following the offset of the first phrase, the firing rate increased from <1 to 5 spikes/s before decreasing to a suppressed level after 500 ms (presumably due to onset of the second vocalization phrase). It is also interesting to note that this increase in firing actually begins shortly (several hundred ms) before the end of the first phrase (Fig. 1C). An increase in firing rate following the end of the first phrase was also observed between phrases of the averaged excitatory vocalization responses (Fig. 1F). This increased activity following the first phrase may be indicative of the end of suppression and an attempt to gradually return to baseline neural firing between phrases. Alternatively, it could reflect an auditory response to the sound made when the animal inhaled between the two phrases (see the spectrograms in Fig. 1A,D); however, the inhalation sound tended to occur late during the inter-phrase period and the return of spiking activity began nearly synchronously with the end of the first phrase. Another possibility is that this burst could represent a rebound following a release from inhibition (unexpectedly seen for both suppressed and some excited modulations).

Another aspect of the temporal dynamics of response during multi-phrase vocalizations was the relative degree of suppression or excitation between phrases. The average activity, reflected in the respective PSTHs, was generally consistent across both phrases (Fig. 1B,E). This relationship was examined quantitatively by comparing the firing rate change deviation, from the spontaneous rate, between the first two phrases in each of the multi-phrase samples (Fig. 2). There was a strong correlation between the vocalization-related activity of the first and second phrase for both suppressed and excited responses ($r = 0.93$, $t = 38.68$, $df = 246$, $P < 0.001$). Overall, there was a slightly stronger modulation (suppressive or excitatory) caused

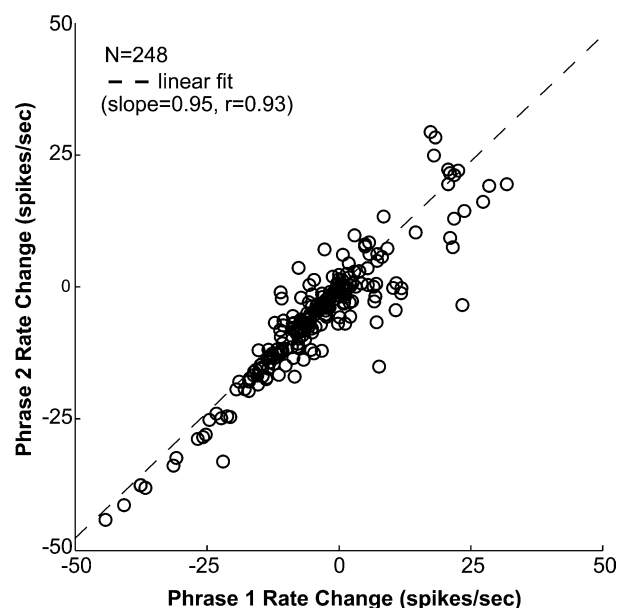


Figure 2. Response modulation consistency between phrases. The firing rate change from spontaneous is plotted comparing the first and second phrases of vocalization. Responses are seen to be similar during the two phrases. The diagonal line is a linear regression fit to the data (slope = 0.95).

by the first phrase than by the second. A linear regression comparing the vocalization-related activity of the second phrase to the first had a slope of 0.95 (95% confidence interval: 0.90 – 1.00). Excitatory responses tended to have more variability than suppressed ones, though this may reflect the larger dynamic range available for increased over decreased rates (decreased rates are bounded by zero). Vocalizations with three or more phrases also showed consistency between the responses during earlier and later phrases, though the correlation was much weaker between first and third phrases ($r = 0.22$, N.S.) than it was between the first and second phrases ($r = 0.43$, $t = 2.01$, $df = 20$, $P < 0.05$) or between second and third ($r = 0.45$, $t = 2.24$, $df = 20$, $P < 0.05$).

Relationship between Vocalization-related Neural Activity and the Acoustic Characteristics of Vocalization

In addition to being either single or multi-phrased, phee vocalizations also varied in both amplitude and frequency. Repeated vocalizations exhibited a degree of variability in acoustic parameters that can be exploited when studying neural modulations, although this could also be a source of variability in the responses of individual units. Two parameters were measured to characterize acoustic variability of the vocalizations, the mean energy and mean frequency. These parameters varied both within and between phrase categories.

The first phrase of multi-phrase phee's were, in general, of both higher energy and frequency than single phrase ones, though a small overlap was observed, particularly for frequency (Fig. 3A). The mean energy and frequency of vocalization were highly correlated (Fig. 3B), with vocalizations of higher frequency containing more energy ($r = 0.91$, $t = 56.94$, $df = 673$, $P < 0.001$). The second phrase of multi-phrase phee's was generally similar to the first phrase, though it had slightly lower energy (Fig. 3C, upper; difference -3.8 ± 5.9 dB, mean \pm STD) and a slightly higher frequency (Fig. 3C, lower; 0.06 ± 0.11 kHz).

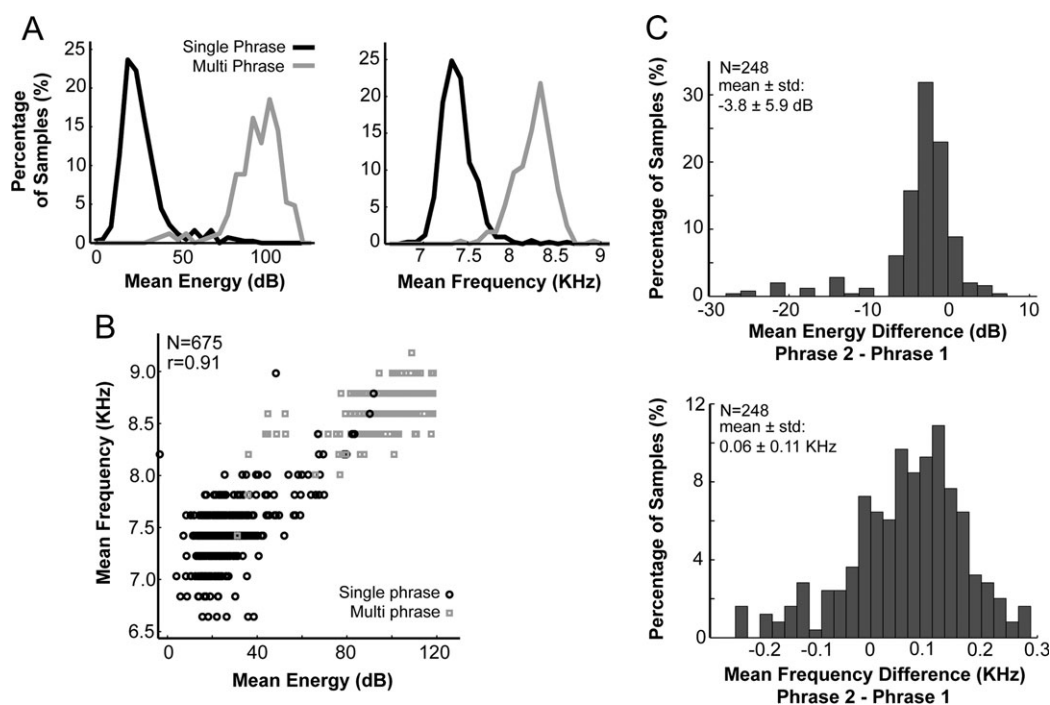


Figure 3. Distribution of acoustic parameters of the produced vocalizations. (A) The distributions of the mean vocal energy (left) and mean frequency (right) are shown for both single ($n = 427$, black) and multi-phased ($n = 248$, grey) vocalizations. Only the parameters of the first phase of multi-phased vocalizations are shown. Acoustic parameters were bimodally distributed between single and multi-phased with relatively few intermediate samples. (B) Scatter plot showing the correlation between mean frequency and mean energy during single and multi-phased vocalizations. Black and grey markers correspond to single and multi-phased vocalizations, respectively. (C) Histograms showing the change in mean energy (top) and frequency (bottom) between the first and second phases of multi-phased vocalizations ($n = 248$). Positive values indicate an increase in the acoustic parameter during the second phase. Mean and standard deviation values for the differences are indicated on the plots. Acoustics were similar between phases, as the observed changes were small.

Relative to the first phase, these were very small (-4.0 and 0.7% , respectively) but significant differences ($P < 0.01$ and $P < 0.001$, Wilcoxon rank-sum test).

Despite the variation both within and between phrase categories, the acoustic parameters did not change randomly from vocalization to vocalization when a sequence of vocalizations was produced by an animal. It was not uncommon for an animal to produce a bout of vocalizations where it began with single-phrase phee, transitioned to two-phrase phee, and then returned back to a single-phrase. An example of such a sequence of vocalizations recorded while holding a single unit is depicted in Figure 4B. The acoustic parameters of these sequentially produced phee are analyzed in Figure 4B. As can be seen, those calls produced with two phrases showed both higher mean frequency and higher energy than the single-phrase ones. However a gradual increase in frequency, as well as energy, was seen that preceded the transition from single to multi-phrase phee (marked by a vertical dashed line in Fig. 4B). The mean energy and frequency remained high even after a return to single phrase (sample 14 in Fig. 4B) before decreasing back to the single-phrase values. This data suggests that the switch from single to multi-phrase followed, rather than caused, the switch from low to high vocal parameter values.

We further examined the effects of such changes in vocalization acoustics on the modulated responses of individual auditory cortical units. A sample of one unit and its vocalizations is shown in Figure 4. The first five vocalizations (all single-phrase) elicited little changes in the pattern of neural firing (Fig. 4A). However, the next eight vocalizations (multi-phased) either completely or nearly completely suppressed the unit

(vocal samples 6–13). Subsequent single phrase vocalizations resulted in mixed, but generally less modulation of the unit's firing. This pattern of changes in response modulation corresponded well to the changes in the vocalization acoustic parameters (Fig. 4B). The set of vocalizations during which the unit was most strongly suppressed were those that contained multiple phrases and, correspondingly, had both higher mean energy and frequency. The degree of response modulation was quantified by the RMI measure (see Materials and Methods) and compared with vocalization acoustics (Fig. 4B, bottom). The RMI was strongly correlated with both the mean energy ($r = -0.81$, $t = 6.33$, $df = 21$, $P < 0.001$) and mean frequency ($r = -0.79$, $t = 5.90$, $df = 21$, $P < 0.001$) of vocalization (Fig. 4C). These analyses show that acoustic parameters of vocalization, rather than phrase category (single versus multi-phrase), determined the modulation of cortical responses. It is important to note that, because of the correlation between mean energy and frequency of phee, we could not determine which of the two characteristics was primarily responsible for the correlation between acoustics and neural responses. The apparent sensitivity of modulations to vocalization acoustics, curiously, was not easily explained by the unit's responses to pure tone stimuli. Both the ranges of frequencies and amplitudes of the produced phee were outside the tuning measured by tone stimuli (CF = 25.8 KHz, RL was non-monotonic with a peak at 70 dB SPL), although the much weaker second harmonic of vocalization would be closer to this range.

Most units analyzed (38/61) had an inverse relationship between RMI and mean vocal energy or frequency, as shown in Figure 4C, particularly those that were suppressed during

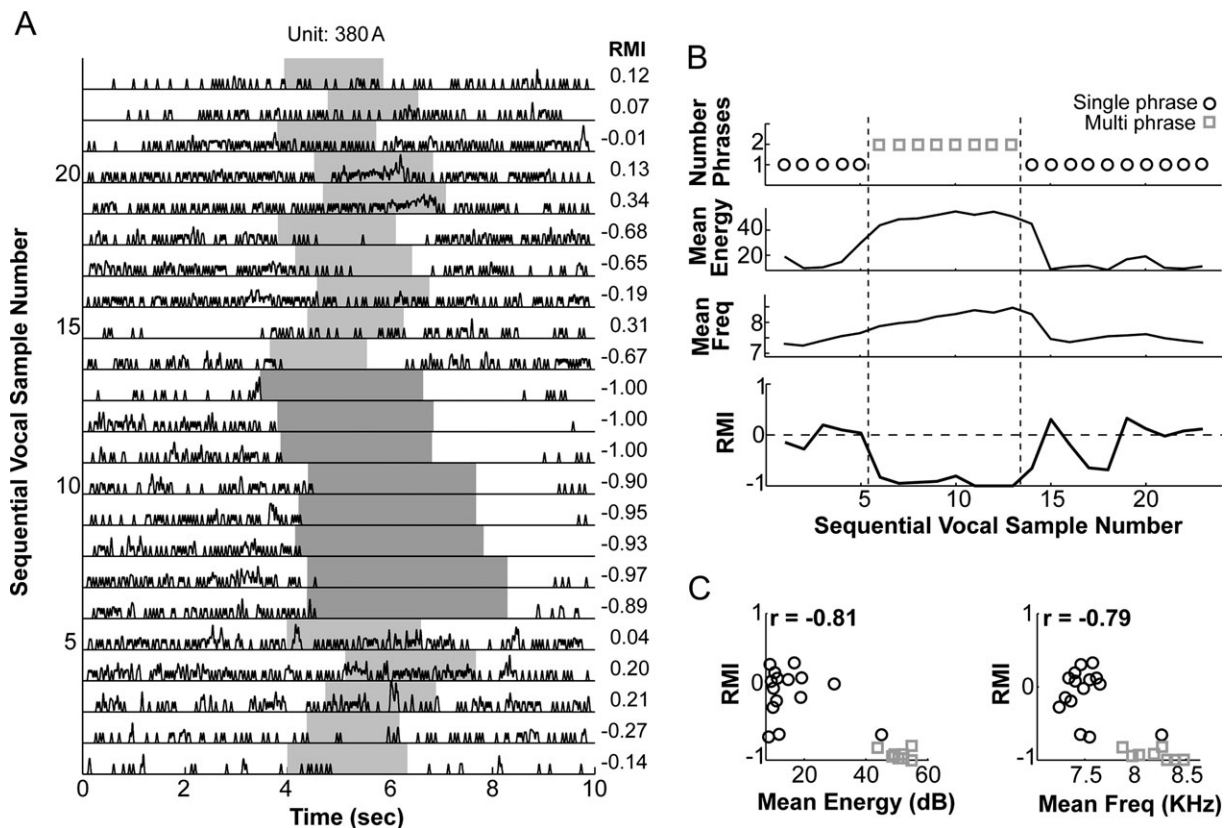


Figure 4. Variations in vocalization-induced modulations from a single unit. (A) Responses to multiple instances of vocalization recorded from a single unit are shown as single trial spike density functions (SDF, $\sigma = 25$ ms) and are ordered by the sequential order in which they were recorded. The RMI for each response is indicated at the right, and a shaded region shows the period of vocalization (the inter-phase period is kept shaded in this plot). Darker shading indicates multi-phased vocalizations. Each SDF has a vertical scale of 4 spikes/s. (B) The vocalization-induced RMI (bottom) is plotted along with the mean energy, mean frequency, and the number of vocal phrases (top) as a function of sequentially ordered sample number. Only the RMI of the first phrase of multi-phase vocalizations was used. Black and grey markers correspond to single and multi-phase vocalizations, respectively. Vertical dashed lines separate single from multi-phased phee vocalizations. (C) Scatter plots showing the negative correlation between RMI and mean vocal energy (left) and frequency (right) in this unit. The unit featured has a CF of 25.8 kHz and a non-monotonic RL with a peak at 70 dB SPL.

vocalization. However, units with positive correlations were also observed (Fig. 5A,B). The example unit in Figure 5A had greater excitation, or less suppression, at higher mean energies and frequencies (Fig. 5A,B). The unit exhibited significant correlation between RMI and both energy ($r = 0.92$, $t = 10.50$, $df = 20$, $P < 0.001$) and frequency ($r = 0.91$, $t = 9.82$, $df = 20$, $P < 0.001$). Another example unit (Fig. 5B) showed no modulation by phee of low frequency or energy, but became driven when the vocalizations entered higher energy states. The response modulation was strongly correlated with both the mean energy ($r = 0.91$, $t = 5.81$, $df = 7$, $P < 0.001$) and mean frequency ($r = 0.90$, $t = 5.46$, $df = 7$, $P < 0.001$) in this unit. The example shown in Figure 5A is of particular interest in that it switched vocalization-related modulation from suppressed to excited depending upon the acoustics of the produced vocalization. This analysis provides an explanation for previous observations that some units appeared to show bimodal modulations during vocalization (Eliades and Wang, 2003).

The excitatory modulations in both of the sample units shown in Figure 5 occurred at vocalization frequencies outside their tone-measured tunings (Fig. 5A: CF = 18.9 kHz; Fig. 5B: 18.6 kHz). Other units showed a range of correlation coefficients, both positive and negative, between RMI and vocalization energy of frequency (Fig. 6A). Overall, however, negative correlations were more prevalent than positive ones. A comparison between correlation coefficients and unit CF showed no consistent

relationship (Fig. 6B). Interestingly, no particular pattern was found for units whose CF was close to the mean vocal frequency (or one of its harmonics), as one might expect frequency tuning to underlie correlations between modulation and vocal acoustics. We cannot, however, discount the possible role of inputs outside the classic frequency receptive field (i.e. multiple frequency peaks) in this correlation. Comparison between amplitude correlation and unit RL tuning yielded similar results.

Properties of Vocalization-related Modulations within Cortical Columns and across Cortical Layers

In another set of analyses, we examined vocalization-related modulations as a function of units' anatomic location, both within columns and at different cortical depths. Columnar organization was studied using the responses of units recorded simultaneously on a single electrode. Frequently more than one single-unit was evident during recording and we examined the relationship between vocalization-related modulations in pairs of single-units. One such pair (Fig. 7A) showed a unit with large action potentials that was completely suppressed during vocalization, while a second unit, with smaller action potentials, increased its firing instead. The RMI values of simultaneously recorded pairs of units are plotted in Figure 7B. The response properties of these pairs were heterogeneous. Nearly complete suppression of responses, for example, was often recorded along

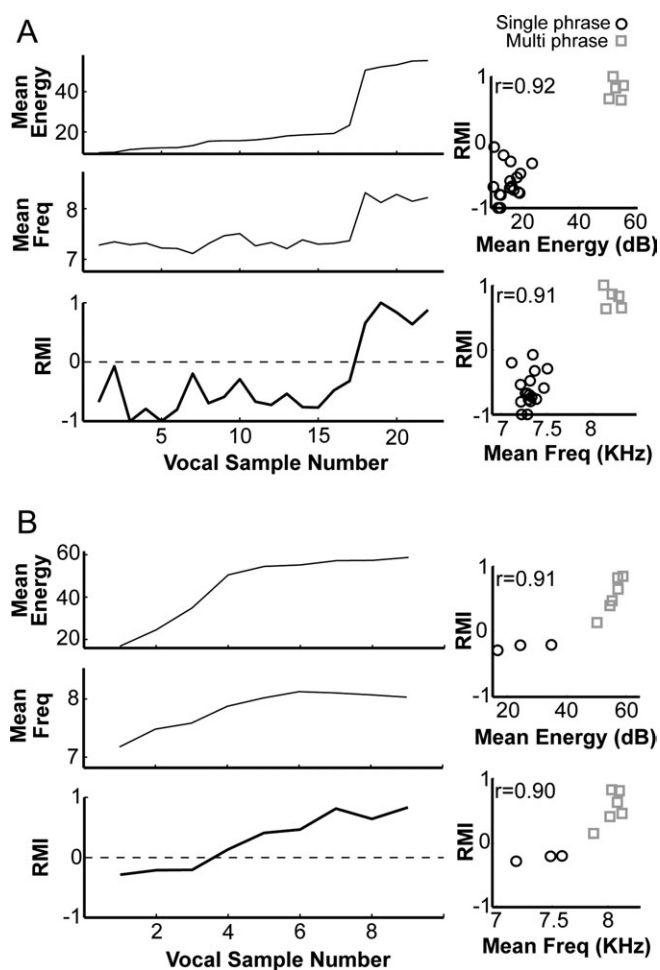


Figure 5. Relationship between response modulation and vocalization acoustics. Two representative example units are shown which have positive correlations between vocal parameters and responses. In each unit, responses to multiple instances of vocalization were recorded. One example (B) represents units with excitatory modulations, while the other (A) represents a small number of units with bimodal modulation (either excited or suppressed depending on the vocalization acoustics). Vocalization parameters (left column: top, middle plots) and corresponding RMI measure (left column: bottom plot) obtained from each unit have been reordered by the mean energy of vocalizations (left column: top plot), and only the first phrase of multi-phrase vocalizations is included. Scatter plots showing the correlation between the RMI measure and vocalization parameters are shown for both mean energy (right column: top plot) and mean frequency (right column: bottom plot). Black and grey markers correspond to single and multi-phrase vocalizations, respectively. Note that the RMI varies with changes in acoustic parameters.

with responses ranging from suppression (Fig. 7Ba) to excitation (Fig. 7Bb). Similarly, strongly excited vocalization responses were often paired with both suppressed (Fig. 7Bc) and excited (Fig. 7Bd) modulations. No discernable pattern could be found in the responses of these pairs of units ($r = 0.026$). The use of sharp, high impedance microelectrodes in recording these pairs limits their spatial distribution to a small volume of cortex, presumably from either the same or adjacent columns, and therefore suggests the involvement of a complex microarchitecture in generating vocalization-related modulations.

We now examine properties of vocalization-related modulation as a function of depth from the cortical surface. We categorized each unit as either suppressed or excited, based on the median RMI of all vocal responses recorded from that unit. It was found that vocalization-related suppression was unevenly distributed across recording depths (Fig. 8A). In the

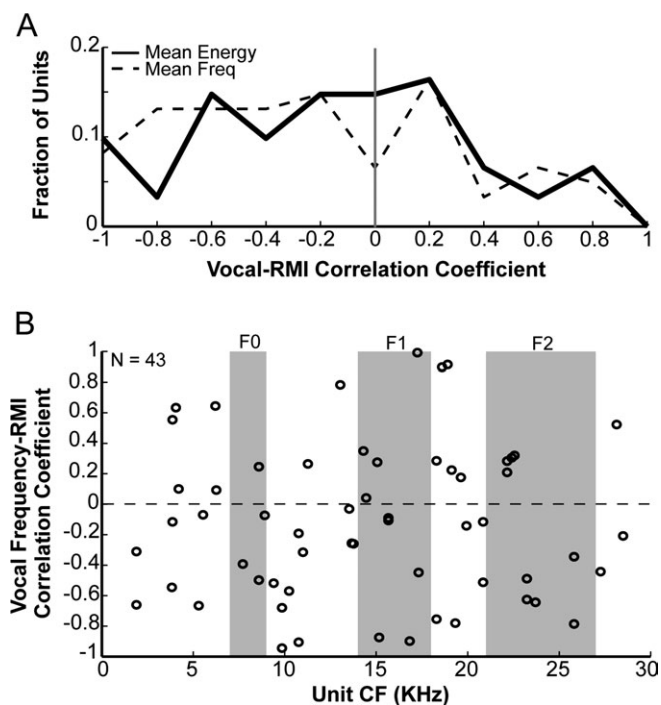


Figure 6. Vocal acoustic-responses correlation across the population of cortical units. (A) The population distribution of vocalization-modulation (vocal-RMI) correlation coefficients for the recorded units, including both energy (solid line) and frequency correlation (dashed line), shows a bias towards negative correlations. Coefficients were only calculated for units sampled with at least four vocalization responses. (B) Scatter plot comparing vocal mean frequency-RMI correlation coefficients with unit CFs. The ranges of mean vocalization frequency (F0) and its harmonics (F1, F2) are shaded to demonstrate the absence of a relationship between correlation and CF, even for units whose spectral tuning include vocal frequencies.

upper cortical layers, units favoring suppression accounted for 75–80% of sampled units. In contrast, deeper layers exhibited more equal fractions of suppressed and excited units. In order to compensate for bias by units that might be only weakly suppressed or excited (i.e. units with a median RMI close to zero), we also calculated the mean RMI as function of depth by averaging the unit median RMI of all units in each depth bin (Fig. 8B). The pattern of mean responses (Fig. 8B) mirrors the depth distribution based on unit classification (Fig. 8A). Suppression was dominant in the upper layers, strongest at 400 and 800 μm ($P < 0.001$, t -test) but still significant in adjacent depth at 200 and 1000 μm ($P < 0.05$). Outside the upper layers, the strength of excitation balanced, or even exceeded, that of suppression. A one-way analysis of variance showed significant dependence of the mean RMI on the cortical depth ($F = 2.1$, $df = 8$, $P < 0.05$).

Units were sampled from both primary and lateral belt auditory areas, determined by response strength to tone and noise stimuli (Rauschecker *et al.*, 1995). Fewer neurons were sampled in the lateral belt than in A1. Neither area was systematically mapped while animals vocalized because in the present study we could not know when animals would vocalize during our extracellular recording procedures. Both suppressed and excited vocalization responses were observed in these cortical areas. However, the limited sample size and the lack of systematic sampling across these areas prevented us from determining any differences in modulation patterns between different areas. Potential differences between primary and non-primary auditory cortical areas remains an important issue awaiting future study.

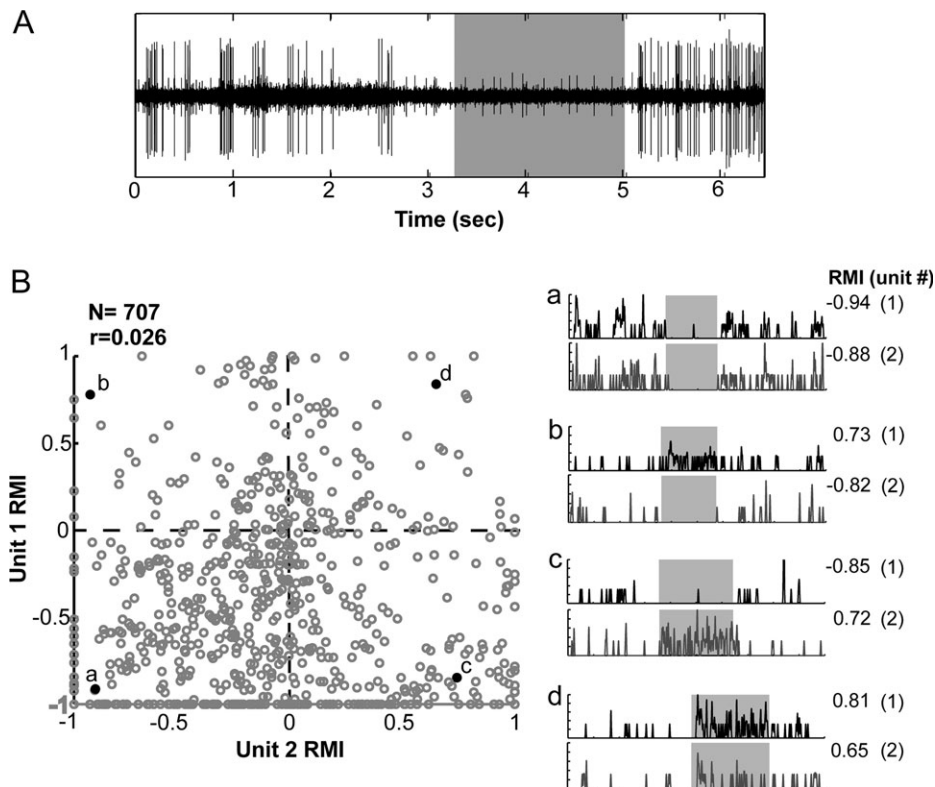


Figure 7. Vocalization responses properties of simultaneously recorded unit pairs. (A) Example of an extracellular trace containing two units, one with a large and the other with a smaller action potential size, showing different modulation during vocalization. (B) Scatter plot containing the RMIs of a large sample of unit pairs recorded simultaneously from a single electrode. No correlation in the RMI of such pairs was found. Unit 1 in each pair was always a large single-unit, while unit 2 was either a large or a small unit (see Materials and Methods for criteria). Specific examples of single-trial SDFs are shown (a–d) featuring pairs of similar (a, d) and opposite (b, c) directions of modulation. The RMI is indicated to the right of each sample in the pair, and their positions are marked on the scatter plot. Shaded areas correspond to the period of vocal production. The vertical scales for each SDF ($\sigma = 25\text{ms}$) is 4 spikes/s.

Comparison between Single Neuron Responses and Global Cortical Activity

Although we have found that single cortical units in a non-human primate show two types of vocalization-related responses, suppressed and excited, most human studies using imaging and other measurement techniques have thus far reported only dampened activation by self-produced speech. These results show that there is increased activity in auditory areas during speaking, but that the level of activity is smaller than that observed when the same speech sound is played back to the listener through a speaker. The methods used to record from the human brain lack the spatial specificity of microelectrode techniques, which suggests that the observations in these studies may reflect globally summed neural activity. In contrast, our previous findings have been based on well-isolated single units. In order to reconcile our neurophysiological findings with observations in humans, we further analyzed the population properties of both single units and simultaneously recorded multi-unit clusters (Fig. 9). Well-isolated single-unit recordings, based on large-action potentials, showed suppressed responses in 75% of units (Fig. 9A, solid line), as we reported previously (Eliades and Wang, 2003). Units of smaller action potential size, likely more distant from the recording electrode, exhibited a similar ratio of suppression and excitation but with a smaller magnitude of suppression (Fig. 9A, dashed line). Multi-unit

background activity, however, showed both a further reduction in suppression and an increased proportion and magnitude of excitation (Fig. 9A, dotted line).

We approximated the global activity in the auditory cortex by summing together PSTHs of all unit responses, including both suppressed and excited responses, recorded during vocalization (Fig. 9B). The sum of activities from only large, well-isolated single units (those used in our other analyses) showed an overall suppression of firing during vocalization (Fig. 9B, black line). On the other hand, a broader measure of auditory cortical activity obtained by summing the responses in all three categories (large, small and multi-units) showed an excitatory response pattern (Fig. 9B, grey line). This demonstrates that suppressive modulations during vocalization can be masked from an observer if a wider range of neural activities is included in the measurement.

Discussion

We have described the dynamics of vocalization-related modulations in the firing of auditory cortical neurons during self-initiated vocalizations. The temporal dynamics of neural firing indicate that units are modulated specifically during vocal periods and rebound during quiet intervals. The variability of modulation from one vocalization to another was found to be

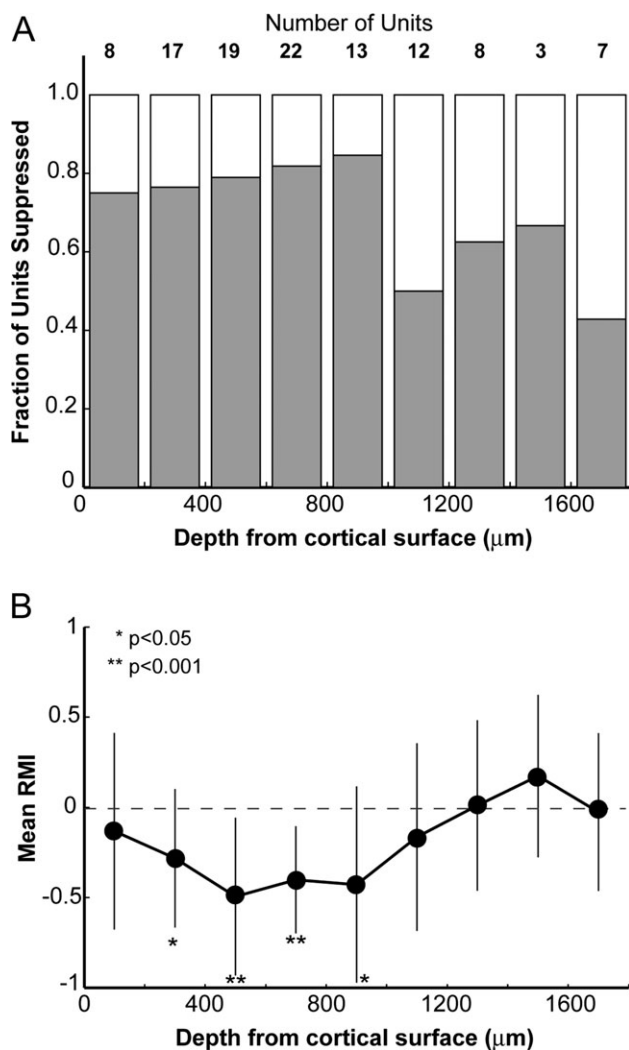


Figure 8. Laminar profile of vocalization responses. (A) The fraction of suppressed units (defined as unit median RMI < 0) is shown as a function of depth from the cortical surface and is differently distributed between superficial and deeper depths. The number of units recorded in each depth bin is indicated at the top. (B) The mean RMI, calculated by averaging individual unit median RMIs in each depth bin, is plotted versus the recording depth. Suppression was strongest at superficial depths. Error bars are the standard deviation calculated from unit median RMIs. Statistical significance of the mean RMI deviation from 0 is indicated in appropriate depth bins (one-way *t*-test, **P* < 0.05, ***P* < 0.001).

correlated to the variations within the acoustics of the produced vocalizations themselves. The diversity of modulations between neurons was observed in adjacent units recorded on the same electrode. A laminar analysis revealed an increased prevalence of suppression in upper cortical layers. Finally, measures of global activity in the auditory cortex during vocalization show an overall weakly excitatory response, in contrast to the dominant suppression recorded from single units. Although vocalization-induced modulation of the primate auditory cortex has been demonstrated (Müller-Preuss and Ploog, 1981; Eliades and Wang, 2003), the correlations found in the variability of responses from single neurons and the diversity of responses between neurons shed new light on possible mechanisms and functions of auditory-vocal interactions in a sensory-motor system for vocal production.

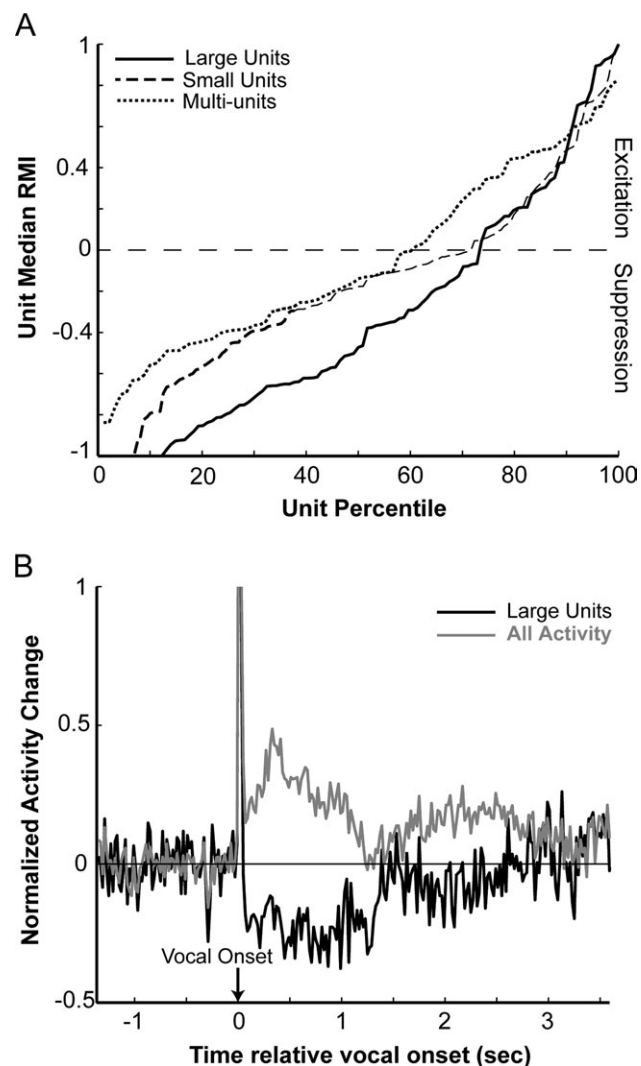


Figure 9. Measures of global activity in the auditory cortex. (A) The distribution of median RMI is shown for units of different isolation qualities. 'Large units': well-isolated single-units with large action potential size (SNR > 10 dB, solid line). 'Small units': sorted single-units with smaller action potential size (dashed line). 'Multi-units': multi-unit background activity (dotted line). Small and multi-units showed less suppression and more excitation than large units. (B) The onset-aligned PSTH, including both suppressed and excited vocalization modulations, comparing responses summed from only large units (black) with summed activity (grey) including all three unit categories (large, small and multi). Overall large unit activity is suppressed, while global activity is weakly increased.

Auditory-Vocal Interaction is Temporally Specific during Vocalizations

We have previously shown that vocalization-induced suppression began prior to, and excitation after, the onset of vocalization (Eliades and Wang, 2003). All neural activities returned to baseline after the end of vocalization. While both of these properties hold for multi-phase as well as single-phase vocalizations, the transient activity during the inter-phase period reveals more complex temporal dynamics. During this period there is a short-lived return of activity following the suppression of the first phrase. This activity is centered just after the end of the first phrase, and actually begins shortly before the vocal offset. We consider this rebound to be temporally specific, despite the slow time scale over which it occurs, because it would be entirely plausible for suppression to be maintained

during the interval in anticipation of the second phrase. Such rebounds between phrases of repetitive sensory-motor events have also been observed during cricket stridulation (Poulet and Hedwig, 2003). The transient activity could represent a neuron's gradual return to spontaneous firing before being suppressed again at the beginning of the second vocal phrase. An alternate explanation is that the firing could represent a sensory response to the wideband breath sound made when the animal inhales between phrases. This alternative is unlikely because most units recorded were from A1 and are generally unresponsive to wideband sounds, and because the responses occurred early in the inter-phrase period, while the inhalation sound did not occur until later in the period. A third possibility is that the transient burst represents a release from inhibition. Neurons suppressed for prolonged periods of time often show a release from inhibition effect, a burst of spikes following the end of suppression (Kuwada and Batra, 1999). This alternative gains strength in light of a similar inter-phrase burst occurring for excitatory vocalization responses, instead of a decrease in firing towards spontaneous that might be expected in the absence of auditory stimulation, which suggests that excitatory responses may also be subject to a degree of inhibitory modulation.

Origins of the Correlation between Vocalization-related Modulation and Vocal Acoustics

Animals in the present study produced multiple successive vocalizations that varied in their acoustic parameters. Recent behavioral data has shown that marmosets increased the intensity and phrase duration of their twitter call (a social call) in the presence of background noise (Brumm *et al.*, 2004). This demonstrates that marmosets, like humans and other animals, exhibit auditory feedback control of their vocalizations. In contrast, the animals studied in our experiments vocalized largely in quiet conditions (with the exception of intermittent external sounds). It is therefore likely that the observed acoustic variations of vocalizations in our experiments were due to animal's voluntary control rather than as a result of altered auditory feedback. We also observed a correlation between vocal intensity and frequency that was not reported during noise masking (Brumm *et al.*, 2004). The discrepancy could be attributable to a number of factors, including different vocalization types (twitter versus phee), the generally higher intensity of phee calls, differences between feedback and voluntary control, or simply that the effects on vocal frequency were not examined during noise masking.

The modulations of single neurons during these vocalizations were often found to be related to the acoustics of the vocalizations. It is unclear, however, whether the differences in modulation were specifically correlated to the acoustic parameters or if they varied more generally with the type of vocalization (i.e. single- versus multi-phrased). In many neurons, response modulations appeared to vary categorically between single- and multi-phrased vocalizations. It is likely, however, that this was just an artifact of sampling in these neurons. There were many examples where vocalizations of one phrase category had acoustics closely matching those of the other phrase category. In these examples, the neuron's responses were closer to vocalizations of similar acoustic parameters than vocalizations with the same number of phrases, but different acoustics. It is possible, on the other hand, that responses were instead categorical for acoustics. It would be hard to definitively answer the question without more samples of vocalization with

intermediate acoustic parameters. The few units that were sampled with such intermediate parameters tended to show a more continuous, though nonlinear, relationship between modulation and acoustics, and may suggest some form of tuning instead of a categorical separation.

This observed correlation of vocalization-induced modulations with acoustic parameters of vocalization is also interesting given the role of auditory feedback in controlling vocal production. While such correlation is not direct evidence that auditory cortical neurons are actively participating in self-monitoring of feedback, like has been seen in the bat brainstem (Smotherman *et al.*, 2003), it demonstrates that they are at least capable of encoding variations in vocal production, despite often being strongly suppressed. It is unclear, however, whether this correlation observed in the cortical neurons is due to auditory feedback or if it arises from internal signals, originating in vocal control centers, containing a neural representation of the produced vocalization (i.e. a corollary discharge). Both alternatives are consistent with a behavioral role for cortical modulation, though they may differ mechanistically in how they integrate sensory feedback.

The feedback hypothesis would argue that neurons in the auditory cortex are subjected to the same amount of suppression regardless of the type of vocalization being produced, although with diversity from neuron to neuron. The suppression, beginning prior to vocalization, is modified during vocal production as the neuron integrates acoustic feedback, the effects of which would likely be related to the inherent auditory properties of the neuron. The suggestion that all neurons, including those that are excited by vocalization, are subjected to some degree of inhibition is consistent with the transient rebound seen during the inter-phrase period for both suppressed and excited multi-phrase vocalizations. This hypothesis would not seem to accord, however, with the observation that vocal acoustics-modulation correlation is unrelated to passive auditory tuning, as we might expect when integrating feedback auditory responses. One possible functional mechanism for the suppression that could be reconciled with these observations would be if the suppression itself reflected an alteration in auditory receptive fields to maximize the ability of neurons to encode incoming feedback. Such alterations in receptive fields have been observed perisaccadically in parietal visual areas (Kusunoki and Goldberg, 2003). Receptive field changes could conceivably include changes in frequency and amplitude tuning to focus around the range of vocalization acoustics. If this were the case, it might explain how auditory neurons show sensitivity to acoustics well outside their passive tuning range, even to the point of neurons strongly excited during vocalization, but whose receptive fields do not include the spectrum produced during vocalization. This hypothesis would also suggest that any efferent-afferent comparison necessary for feedback monitoring and resulting vocal control would likely have to take place outside the auditory cortex, perhaps directly in vocal control centers, where efferent information would be available.

A second hypothesis is that the corollary discharge contains specific information about the acoustics of the vocalization being produced. The correlation between acoustics and auditory cortical modulation, including both suppression and excitation, would not arise from feedback as much as feed-forward inputs. Specific modulation of auditory neurons would allow efferent-afferent comparisons to take place directly in the

auditory cortex, the output of which could be fed to vocal control centers. Specific modulation would also allow cancellation of expected feedback to maintain sensitivity to external stimuli. It would not, however, explain why suppressed, but not excited, neurons respond poorly to external stimuli (Eliades and Wang, 2003), unless the effects of modulation are broader for suppression and more specific for excitation.

Regardless of the mechanism, the relationship between vocalization-related modulation and vocalization acoustics explains the high degree of response variability seen in a subset of auditory cortical neurons (Eliades and Wang, 2003). The variation in these neurons can now be understood as non-random and as a result of the variation in the vocalizations produced. This also suggests that our earlier attempts to classify neurons as either suppressed or excited may have been overly simplistic. While most neurons can be divided this way, a few switch behaviors depending on the vocalization and may need to be separately classified as bimodal. All neurons, whether suppressed, excited or bimodal, also need to be understood in terms of how their modulation varies with vocal acoustics, at the very least including the direction of correlation (positive or negative).

A related issue is the apparent lack of correlation between vocalization-related modulations and a neuron's frequency and amplitude tuning. A priori, one might expect that vocal modulation might depend on the relationship between vocalization acoustics and a neuron's receptive field. The feedback model we introduced to explain the correlation between acoustics and modulation is based upon such an assumption. However, this was not the case based on the observations from the present study. Neither a neuron's median RMI nor its vocal-acoustic versus RMI correlation coefficient is related to either its CF or RL tuning. Furthermore, some neurons were found to be strongly excited during vocalization despite the fact that the energy in a vocalization's spectrum fell outside a neuron's receptive field (for example, a neuron whose CF was >30 kHz). One possible explanation for this dissociation could be the presence of inputs outside the classic receptive field (i.e. center excitation surrounded by sideband inhibitions), such as multiple frequency peaks, which have been observed in many neurons in marmoset auditory cortex (Kadia and Wang, 2003). These inputs could account for some of the response modulation during vocalization. The extent of inputs outside the classic receptive field was not characterized for the neurons reported in this study because of experimental limitations. Another explanation could be that vocalization-related modulations, including acoustic-modulation correlations, could be heavily, or entirely, determined by the motor command signal. As we have suggested above, receptive fields could be altered by the modulatory signal to maximize the ability to encode the vocalization being produced. It is also possible that vocalization-related modulation is completely independent of a neuron's receptive field.

Possible Biological Mechanisms Underlying Sensory-Motor Suppression

While it is more than likely that the suppression of auditory cortical neurons during vocalization arises from signals from vocal production centers, the details of how such pathways interact with cortical neurons is unclear. The similarity of observed suppression with the electric fish system (Bell, 1989) suggests that auditory suppression is also a result of GABA-mediated inhibition. It is not known, however, whether the

auditory cortex directly receives such inhibition, or whether the suppression is a reflection of inhibition in lower auditory areas. Previous works have shown reduced brainstem activity in the bat during echolocation (Metzner, 1989), although these may reflect a specialized adaptation for bat echolocation, as studies in primates have failed to find suppression in similar nuclei (Kirzinger and Jurgens, 1991). Sensory-motor suppression has been observed in the primate inferior colliculus during vocalization, but only in non-central areas (Tammer *et al.*, 2004). These areas are not common inputs to lemniscal auditory thalamus (Calford and Aitkin, 1983) and, as a result, are not primary inputs to A1 (Morel and Kaas, 1992). It is unlikely, therefore, that they alone can manifest the strong suppression observed in both A1 and non-primary auditory cortex. Although the non-lemniscal auditory thalamus does project to layers I and VI in A1 (Huang and Winer, 2000), our analysis shows vocalization-induced suppression is strongest in the upper layers (layers II-III and part of layer IV), but not in depths corresponding to layer I. It is possible that layer I neurons could act to suppress layers II/III (but it would require excitatory not suppressed inputs from MGB to A1). This would still be consistent with the auditory cortex as the site of the vocalization-related inhibition. Quantitative analysis of the magnitude of cortical suppression also suggests that the auditory cortex is subjected to additional inhibition beyond what may be inherited from the brainstem (Eliades and Wang, 2003).

Further evidence for cortical mechanisms is provided by multiple neurons recorded simultaneously from single electrodes. Vocalization-induced responses of such pairs were uncorrelated, despite their spatial proximity. Because of the dissipation of electric potentials in the extracellular space, units recorded from a single electrode are limited to a small volume of cortex (Gray *et al.*, 1995), likely from the same or adjacent cortical columns. Stimulus responses of closely spaced neurons are usually similar because of columnar structure of auditory cortex (Abeles and Goldstein, 1972; Merzenich and Brugge, 1973; Merzenich *et al.*, 1975). This was not the case for vocalization-related modulations, where completely suppressed neurons were often adjacent to unresponsive or even excited neurons. The absence of correlation between simultaneously recorded neuron pairs in a column supports the hypothesis that vocalization-induced inhibition has a cortical contribution, as each column generally receives similar subcortical inputs. This architectural complexity also suggests the involvement of local cortical microcircuits rather than some broader, indiscriminate top-down process that would likely modulate neurons in a column similarly. The involvement of a pyramidal cell-interneuron interaction in this circuit, each with different vocalization-related modulations, could explain uncorrelated responses between some neuron pairs and remains an intriguing possibility that might underlie cortical mechanism of vocalization-induced inhibition.

Although the diversity between neural modulations was not explainable based upon a columnar organization, an analysis of the laminar distribution of modulations revealed differences between the distributions of suppression and excitation. Inhibition was unequally concentrated in upper cortical layers, while deeper layers received a more balanced number of inhibited and excited neurons. One possible explanation for the predominance of inhibition in the upper layers is that it could result from cortical connections projecting to the

auditory cortex. The upper layers are the primary target for both long- and short-range inhibitory cortico-cortical connections, connections that can act to inhibit their target through local GABAergic interneurons (Fitzpatrick and Imig, 1980; Hirsch and Gilbert, 1991). If correct, this would suggest that vocalization-related modulatory signals arise from a cortical source and act directly on the auditory cortex itself. One possible source for such signals could be prefrontal areas that are reciprocally connected with the auditory cortex (Hackett *et al.*, 1999; Romanski *et al.*, 1999), areas that are responsive to vocalizations and other sounds (Romanski and Goldman-Rakic, 2002). This hypothesis is tempting given the role of frontal-temporal interactions in human speech (Penfield and Roberts, 1959; Geschwind, 1970).

Comparison between Monkeys and Human Studies

Our recordings in the non-human primate auditory cortex have revealed both excitatory and inhibitory responses of single cortical neurons during vocalization. When we broaden the type of activity recorded to include not just well isolated single units, but also the multi-unit background, the resulting overall pattern is weakly excitatory. A likely reason that this overall pattern is excitatory, despite a large population of suppressed neurons, is the asymmetry of neural excitation and inhibition. Most auditory cortical neurons in awake animals have low spontaneous discharge rates. The maximum amount of discharge rate reduction observable by extracellular recordings is bounded by zero (no firing), while stimulus-driven firing rates can be many times higher than spontaneous discharge rates. This asymmetry allows a much smaller number of excitatory neurons to mask a larger number of inhibited neurons when summed together as a group. The direct implication is that studies of auditory-vocal interaction are best conducted using single-units, or one may miss the inhibitory interaction being hidden in multi-unit recordings.

Human studies have thus far reported only dampened activation in auditory cortical areas during speech (MEG: Numminen *et al.*, 1999; Curio *et al.*, 2000; Gunji *et al.*, 2001; Houde *et al.*, 2002; PET: Paus *et al.*, 1996; Wise *et al.*, 1999; ECoG: Crone *et al.*, 2001). Such signals recorded during speaking are elevated over baseline, but are smaller than signals evoked by the same speech sound when played back through a speaker. These results resemble the weak activation seen in the global summed activity of multi-units in the marmoset (Fig. 9B), and this could suggest that the reduced strength of net activation during speaking, compared with playback of sound, may be attributable to underlying suppression by individual neurons. However, care must be taken when making such a comparison because single and multiple unit recordings are only indirectly linked to signals obtained from imaging techniques used in humans. It is also important to note that our global activity measure does not equate to field potential measurement, which would be a closer comparison for MEG and ECoG methods.

When comparing the results from marmosets and humans, one should also keep in mind potential differences between the species. It is clear that both human speech and the architecture of the human brain are far more complex than marmoset vocalizations and brain structure. Nonetheless, there are certain parallels of auditory cortical anatomy between humans, Old World primates (Hackett *et al.*, 2001) and New World primates (Morel and Kaas, 1992), including the connections between the

auditory cortex and the prefrontal cortex (Morel and Kaas, 1992; Hackett *et al.*, 1999). It remains unknown whether other primates, particularly Old World primates like the macaque, have similar mechanisms of auditory-vocal interaction. This is likely the case, given the similar modulations seen in both humans and marmosets, and it would be interesting to study auditory-vocal interactions in an intermediate species, such as the macaque.

Functional Models for Sensory-Motor Interaction during Vocalization

Sensory-motor interactions that modulate sensory processing have been described in a number of systems. The common proposed mechanism involves a neural signal, termed efference copy (Holst and Mittelstaedt, 1950) or corollary discharge (Sperry, 1950), relaying information from motor control areas that influences the activities of sensory neurons. The precise form of this signal is unclear, since it is rarely measured directly, though it has been suggested to contain a representation of the expected sensory responses produced by a motor action (Bell, 1989; Poulet and Hedwig, 2003). In most cases, these discharges result in the inhibition of sensory neurons, similar to what we have seen during primate vocalization. The weakly electric fish, perhaps the most well characterized model of neuronal sensory-motor interactions, uses corollary discharges from the electric organ to influence central sensory neurons through GABA-mediated inhibition (Bell, 1989). In addition to the auditory cortex, other sensory cortices have also been shown to be inhibited during sensation-generating motor activities, including the primate visual cortex (Judge *et al.*, 1980) and the somatosensory cortex (Rushton *et al.*, 1981; Blakemore *et al.*, 1998). Vocalization-induced modulation of auditory cortical neurons is another example of a behaviorally relevant sensory-motor interaction that, presumably, shares common mechanisms and functions with other systems.

The possible functions of efference copy-mediated inhibition are twofold. First, it may play a role in distinguishing self-generated from external sensory stimuli. Central electrosensory neurons in the fish perform a subtractive comparison of efferent and afferent signals, the output of which reflects environmental stimuli, but not the fish's own electric discharges (Bell, 1989). The cricket cercal system is suppressed during stridulation (rubbing of wings to generate sound) in order to prevent saturation, and the resulting loss of acoustic sensitivity, of auditory neurons due to self-generated sounds (Poulet and Hedwig, 2003). This alteration of sensory activity to remove self-generated inputs appears quite robust, but is not limited to biologically natural sensory events (e.g. fish electric discharge, touch, animal vocalization or human speech). While MEG measurements from the human auditory cortex are reduced during speech, they are also reduced when self-generated sounds are artificially produced, such as a tone played following a button press (Martikainen *et al.*, 2005). While such a role is possible during vocalization, our previous data have suggested that suppression reduces the auditory sensitivity to external sounds (Eliades and Wang, 2003). It is possible that the excitatory neurons, who respond to external sound normally during vocalization, have this function.

The second possible function of efferent-mediated sensory-motor interaction is self-monitoring for the control of motor behavior. Every action generates sensory feedback, and this

feedback is often used in on-line motor control. Visual feedback, for example, is used for both oculomotor (Sommer and Wurtz, 2002) and arm movement (Goodbody and Wolpert, 1999) control. Somatosensory feedback is used to control the amount of force applied when grasping an object (Edin *et al.*, 1992), as well as providing a source of feedback control for the lips during speech (Gracco and Abbs, 1985). The auditory equivalent of this sort of sensory-motor control is the monitoring of vocal/speech feedback in order to maintain desired acoustic production, including the control of vocal frequency, intensity and temporal patterns. It is clear that this feedback plays an important role in vocal production in humans and, perhaps, primates. However, the neural mechanisms, including the possible involvement of efferent signals, remain unclear. The echolocating of bats is one example where specific neural structures have been demonstrated to have such a role. Brainstem neurons in the nuclei surrounding the bat lateral lemniscus are suppressed during the production of bat echolocation sounds (Metzner, 1989), similar to what we have observed in primate cortex. These nuclei, which may represent a specialized adaptation for echolocation, are involved in the control of echolocation frequency production when presented with frequency-shifted feedback, a phenomenon known as Doppler-shift compensation (Smotherman *et al.*, 2003). Application of bicuculine to block GABA-ergic inhibition in these areas resulted in an elimination of feedback compensation behavior. A role for the auditory cortex in vocal production has been demonstrated in humans, where intra-operative electrical stimulation has been shown to disturb, but not interrupt, ongoing speech production (Penfield and Roberts, 1959). The auditory cortex has also been implicated in the neural origins of stuttering during speech (Fox *et al.*, 1996). Whether or not our observations of cortical modulation in primates plays a role in vocal production remains to be seen. The correlation between neural modulation and vocalization acoustics suggests that cortical neurons are at least capable of encoding variations in auditory feedback, a prerequisite for feedback monitoring.

Notes

The authors would like to thank Drs T. Lu and L. Liang for assistance in data collection and A. Pistorio for assistance in animal training and care and for assistance in preparing this manuscript. This work is supported by NIH/NIDCD grant DC005808 (X.W.).

Address correspondence to Dr Xiaojin Wang, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, 720 Rutland Avenue, Ross 419, Baltimore, MD 21205, USA. Email: xwang@bme.jhu.edu.

References

- Abeles M, Goldstein MH Jr (1972) Responses of single units in the primary auditory cortex of the cat to tones and to tone pairs. *Brain Res* 42:337-352.
- Agamaite JA, Wang X (1997) Quantitative classification of the vocal repertoire of the common marmoset (*Callithrix jacchus jacchus*). *Assoc Res Otolaryngol Abstr* 20:573.
- Aitkin L, Park V (1993) Audition and the auditory pathway of a vocal new world primate, the common marmoset. *Prog Neurobiol* 41:345-367.
- Bell CC (1989) Sensory coding and corollary discharge effects in mormyrid electric fish. *J Exp Biol* 146:229-253.
- Blakemore SJ, Wolpert DM, Frith CD (1998) Central cancellation of self-produced tickle sensation. *Nat Neurosci* 1:635-640.
- Brumm H, Voss K, Kollmer I, Todt D (2004) Acoustic communication in noise: regulation of call characteristics in a New World monkey. *J Exp Biol* 207:443-448.

- Burnett TA, Freedland MB, Larson CR, Hain TC (1998) Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am* 103:3153-3161.
- Calford MB, Aitkin LM (1983) Ascending projections to the medial geniculate body of the cat: evidence for multiple, parallel auditory pathways through thalamus. *J Neurosci* 3:2365-2380.
- Carmel PW, Starr A (1963) Acoustic and nonacoustic factors modifying middle ear muscle activity in waking cats. *J Neurophysiol* 26:598-616.
- Creutzfeldt O, Ojemann G, Lettich E (1989) Neuronal activity in the human lateral temporal lobe. II. Responses to the subjects own voice. *Exp Brain Res* 77:476-489.
- Crone NE, Hao L, Hart J, Boatman D, Lesser RP, Irizarry R, Gordon B (2001) Electrographic gamma activity during word production in spoken and sign language. *Neurology* 57:2045-2053.
- Curio G, Neuloh G, Numminen J, Jousmaki V, Hari R (2000) Speaking modifies voice-evoked activity in the human auditory cortex. *Hum Brain Mapp* 9:183-191.
- Cynx J, Lewis R, Tavel B, Tse H (1998) Amplitude regulation of vocalizations in noise by a songbird, *Taeniopygia guttata*. *Anim Behav* 56:107-113.
- Edin BB, Westling G, Johansson RS (1992) Independent control of human finger-tip forces at individual digits during precision lifting. *J Physiol* 450:547-564.
- Eliades SJ, Wang X (2003) Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J Neurophysiol* 89:2194-2207.
- Epple G (1968) Comparative studies on vocalization in marmoset monkeys (Hapalidae). *Folia Primatol* 8:1-40.
- Fairbanks G (1955) Selective vocal effects of delayed auditory feedback. *J Speech Hear Disord* 20:333-346.
- Fitzpatrick KA, Imig TJ (1980) Auditory cortico-cortical connections in the owl monkey. *J Comp Neurol* 192:589-610.
- Fox PT, Ingham RJ, Ingham JC, Hirsch TB, Downs JH, Martin C, Jerabek P, Glass T, Lancaster JL (1996) A PET study of the neural systems of stuttering. *Nature* 382:158-161.
- Geschwind N (1970) The organization of language and the brain. *Science* 170:194-195.
- Goodbody SJ, Wolpert DM (1999) The effect of visuomotor displacements on arm movement paths. *Exp Brain Res* 127:213-223.
- Gracco VL, Abbs JH (1985) Dynamic control of the perioral system during speech: kinematic analyses of autogenic and nonautogenic sensorimotor processes. *J Neurophysiol* 54:418-432.
- Gray CM, Maldonado PE, Wilson M, McNaughton B (1995) Tetrodes markedly improve the reliability and yield of multiple single unit isolation from multi-unit recordings in cat striate cortex. *J Neurosci Methods* 63:43-54.
- Gunji A, Hoshiyama M, Kakigi R (2001) Auditory response following vocalization: a magnetoencephalographic study. *Clin Neurophysiol* 112:514-520.
- Hackett TA, Stepniewska I, Kaas JH (1999) Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res* 817:45-58.
- Hackett TA, Preuss TM, Kaas JH (2001) Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol* 441:197-222.
- Henson OW Jr (1965) The activity and function of the middle-ear muscles in echo-locating bats. *J Physiol* 180:871-887.
- Hirsch JA, Gilbert CD (1991) Synaptic physiology of horizontal connections in the cat's visual cortex. *J Neurosci* 11:1800-1809.
- Holst E von, Mittelstaedt H (1950) Das Refferenzprinzip: Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften* 37:464-476.
- Houde JF, Jordan MI (1998) Sensorimotor adaptation in speech production. *Science* 279:1213-1216.
- Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002) Modulation of the auditory cortex during speech: an MEG study. *J Cogn Neurosci* 14:1125-1138.
- Huang CL, Winer JA (2000) Auditory thalamocortical projections in the cat: laminar and areal patterns of input. *J Comp Neurol* 427:302-331.

- Judge SJ, Wurtz RH, Richmond BJ (1980) Vision during saccadic eye movements. I. Visual interactions in striate cortex. *J Neurophysiol* 43:1133-1155.
- Kadia SC, Wang X (2003) Spectral integration in A1 of awake primates: neurons with single- and multi-peaked tuning characteristics. *J Neurophysiol* 89:1603-1622.
- Kirzinger A, Jurgens U (1991) Vocalization-correlated single unit activity in the brain stem of the squirrel monkey. *Exp Brain Res* 84:545-560.
- Kusunoki M, Goldberg ME (2003) The time course of perisaccadic receptive field shifts in the lateral intraparietal area of the monkey. *J Neurophysiol* 89:1519-1527.
- Kuwada S, Batra R (1999) Coding of sound envelopes by inhibitory rebound in neurons of the superior olivary complex in the unanesthetized rabbit. *J Neurosci* 19:2273-2287.
- Lane H, Tranel B (1971) The Lombard sign and the role of hearing in speech. *J Speech Hear Res* 14:677-709.
- Lee BS (1950) Effects of delayed speech feedback. *J Acoust Soc Am* 22:824-826.
- Leonardo A, Konishi M (1999) Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature* 399:466-470.
- Lewicki MS (1998) A review of methods for spike sorting: the detection and classification of neural action potentials. *Network* 9: R53-R78.
- Liang L, Lu T, Wang X (2002) Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. *J Neurophysiol* 87:2237-2261.
- Lu T, Liang L, Wang X (2001a) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci* 4:1131-1138.
- Lu T, Liang L, Wang X (2001b) Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. *J Neurophysiol* 85:2364-2380.
- Martikainen MH, Kaneko KI, Hari R (2005) Suppressed responses to self-triggered sounds in the human auditory cortex. *Cereb Cortex* (in press).
- Merzenich MM, Brugge JF (1973) Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Res* 50:275-296.
- Merzenich MM, Knight PL, Roth GL (1975) Representation of cochlea within primary auditory cortex in the cat. *J Neurophysiol* 38: 231-249.
- Metzner W (1989) A possible neuronal basis for Doppler-shift compensation in echo-locating horseshoe bats. *Nature* 341:529-532.
- Morel A, Kaas JH (1992) Subdivisions and connections of auditory cortex in owl monkeys. *J Comp Neurol* 318:27-63.
- Müller-Preuss P, Ploog D (1981) Inhibition of auditory cortical neurons during phonation. *Brain Res* 215:61-76.
- Nonaka S, Takahashi R, Enomoto K, Katada A, Unno T (1997) Lombard reflex during PAG-induced vocalization in decerebrate cats. *Neurosci Res* 29:283-289.
- Numminen J, Salmelin R, Hari R (1999) Subject's own speech reduces reactivity of the human auditory cortex. *Neurosci Lett* 265:119-122.
- Paus T, Perry DW, Zatorre RJ, Worsley KJ, Evans AC (1996) Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur J Neurosci* 8:2236-2246.
- Penfield W, Roberts L (1959) Speech and brain-mechanisms. Princeton, NJ: Princeton University Press.
- Poulet JF, Hedwig B (2003) A corollary discharge mechanism modulates central auditory processing in singing crickets. *J Neurophysiol* 89:1528-1540.
- Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268:111-114.
- Romanski LM, Bates JF, Goldman-Rakic PS (1999) Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J Comp Neurol* 403:141-157.
- Romanski LM, Goldman-Rakic PS (2002) An auditory domain in primate prefrontal cortex. *Nat Neurosci* 5:15-16.
- Rushton DN, Rothwell JC, Craggs MD (1981) Gating of somatosensory evoked potentials during different kinds of movement in man. *Brain* 104:465-491.
- Schuller G, Beuter K, Schnitzler H-U (1974) Response to frequency shifted artificial echoes in the bat *Rhinolophus ferrumequinum*. *J Comp Physiol* 89:275-286.
- Siegel GM, Pick HL Jr (1974) Auditory feedback in the regulation of voice. *J Acoust Soc Am* 56:1618-1624.
- Sinnott JM, Stebbins WC, Moody DB (1975) Regulation of voice amplitude by the monkey. *J Acoust Soc Am* 58:412-414.
- Smotherman M, Zhang S, Metzner W (2003) A neural basis for auditory feedback control of vocal pitch. *J Neurosci* 23:1464-1477.
- Sommer MA, Wurtz RH (2002) A pathway in primate brain for internal monitoring of movements. *Science* 296:1480-1482.
- Sperry RW (1950) Neural basis of the spontaneous optokinetic responses produced by visual inversion. *J Comp Physiol Psychol* 43:482-489.
- Suga N, Jen PH (1975) Peripheral control of acoustic signals in the auditory system of echolocating bats. *J Exp Biol* 62:277-311.
- Suga N, Shimozawa T (1974) Site of neural attenuation of responses to self-vocalized sounds in echolocating bats. *Science* 183:1211-1213.
- Tammer R, Ehrenreich L, Jurgens U (2004) Telemetrically recorded neuronal activity in the inferior colliculus and bordering tegmentum during vocal communication in squirrel monkeys (*Saimiri sciureus*). *Behav Brain Res* 151:331-336.
- Wang X (2000) On cortical coding of vocal communication sounds in primates. *Proc Natl Acad Sci USA* 97:11843-11949.
- Wise RJ, Greene J, Buchel C, Scott SK (1999) Brain regions involved in articulation. *Lancet* 353:1057-1061.