Neural Representation of Vocalizations in the Primate Ventrolateral Prefrontal Cortex

Lizabeth M. Romanski, Bruno B. Averbeck, and Mark Diltz

INTRODUCTION

We have recently identified an auditory prefrontal region in a non-human primate that has responses to complex auditory stimuli, including human and monkey vocalizations (Romanski and Goldman-Rakic 2002). The frontal lobe region responsive to vocalizations lies in the projection field of the anterior lateral belt auditory area (AL) below the principal sulcus and may be part of an auditory object stream, specialized for the processing of nonspatial auditory information (Rauschecker and Tian 2000; Romanski et al. 1999b; Tian et al. 2001). Although ventral prefrontal cortex (vIPFC) neurons were shown to respond robustly to familiar conspecific vocalizations and human speech sounds, the salient features of species-specific vocalizations that account for prefrontal auditory responses are largely unknown. Species-typical communication stimuli can encode many types of information including vocalization category, caller identity, body size, and reproductive status (Bradbury and Vehrencamp 1998; Hauser 1996; Hauser and Marler 1993; Owings and Morton 1998). Understanding which, if any, of these features is processed by prefrontal neurons is a pertinent question in characterizing the function of vIPFC.

Sensory information reaching the frontal lobe is highly processed. Thus neurons in vIPFC may be selective for specific features within complex stimuli. Previous studies have asked whether species-specific vocalizations are processed locally by “call detectors” or, in a more distributed fashion, by “cell ensembles” (Newman and Lindsay 1976; Pelleg-Toiba and Wollberg 1991; Wang 2000). A recent examination of marmoset auditory cortex revealed a heterogeneous population of cells responding to vocalizations with a small percentage of cells selectively responsive to particular call types (“call detectors”) and a larger percentage of cells responding more generally to particular acoustic features (Wang et al. 1995). In the macaque auditory cortex, it has been shown that cells in the lateral belt demonstrate greater call selectivity than do cells in the primary auditory cortex, and lateral belt neurons respond similarly to monkey vocalizations with similar acoustic features rather than responding similarly to vocalizations with similar behavioral or functional referents (Rauschecker 1998; Tian et al. 2001). Furthermore, neurons in AL show even greater call selectivity than cells in the caudal belt (area CL) (Tian et al. 2001; Rauschecker and Tian 2000). Because the vIPFC receives direct projections from AL and the rostral parabelt auditory association cortices (Hackett et al. 1999; Romanski et al. 1999a,b), it is possible that prefrontal neurons, like those in the lateral belt, also respond to vocalizations with particular acoustic features. However, there remains the possibility that prefrontal neurons respond to species-specific communication sounds based on their semantic or functional referents. This is supported by human imaging studies, which have suggested that the human inferior frontal gyrus plays a role in semantic processing (Demb et al. 1995; Poldrack et al. 1999). Furthermore, in playback experiments using rhesus macaque vocalizations, monkeys respond behaviorally in a similar manner to vocalizations with similar functional referents regardless of acoustic morphology (Gifford et al. 2003; Hauser 1998). The frontal lobe, with its widespread connections to sensory and motor systems is a likely candidate to encode some of the behavioral features or functional referents of communication sounds.

In the present study, we tested vIPFC auditory cells in awake, behaving macaque monkeys with a library of rhesus macaque calls recorded from a group of unfamiliar callers. The behavioral and social context under which these calls were produced, as well as their acoustic features, have been well characterized (Ghazanfar and Hauser 1999; Gouzoules et al. 1984; Hauser 1998; Hauser and Marler 1993; Owren and...
Rendall 2001). We asked whether vlPFC cells would respond to these communication sounds from unfamiliar callers and what type of selectivity prefrontal auditory neurons have with regard to communication relevant sounds typical of the rhesus macaque repertoire. Further, we asked if prefrontal auditory cells would respond similarly to calls with similar functional referents or if other features of the vocalizations could account for neuronal response specificity.

**METHODS**

**Subjects and surgical methods**

We made extracellular recordings in two rhesus monkeys (*Macaca mulatta*), 6 and 7 yr old, weighing 8.0 and 10.5 kg. All methods were in accordance with National Institutes of Health standards and were approved by the University of Rochester Care and Use of Research Animals committee. Prior to recordings a titanium head post was surgically implanted in the skull of both animals to allow fixation of the head during recordings. When training was complete, animals were implanted with a 20-mm recording cylinder (Narishige, Japan) placed over the ventrolateral prefrontal cortex. The recording cylinders were placed to maximize recordings in area 12 and 45 as defined anatomically by Preuss and Goldman-Rakic (1991), and physiologically by Romanski and Goldman-Rakic (2002). In addition, a scleral search coil was implanted in one eye to monitor eye movements via an electromagnetic coil.

**Localization of prefrontal auditory neurons**

We relied on a standard protocol for finding the macaque frontal auditory area. The auditory area was initially targeted using known stereotaxic coordinates, specifically aimed at 32 mm anterior to the interaural line according to the atlas of Paxinos et al. (2000). Once recordings began, we explored the cylinder to determine the location of neurons responsive to face and other visual stimuli at short latencies (60–100 ms). On finding these cells, we moved anterior and lateral to find the auditory prefrontal neurons. After completion of the experiments we performed a MRI in one monkey with a marker placed at the middle of the recording cylinder. Because the nonmagnetic recording cylinder cast an artifact over the recorded hemisphere, we reflected the coordinates to the opposite hemisphere and have depicted the approximate location of recording tracks through the auditory area (Fig. 1), which is located at the level of the inferior prefrontal dimple (Paxinos et al. 2000) when this small sulcus is present. Although it was not possible to image the second animal, the anterior-posterior stereotaxic coordinates of the recording cylinders for both animals are within 3 mm of each other. Thus we assume that the recordings in both monkeys are from the same ventral prefrontal region although there may be slight differences in the specific part of the auditory area recorded from in each animal.

**Apparatus and stimuli**

All training and recording was performed in a sound-attenuated room lined with Sonex (Acoustical Solutions). Auditory stimuli were presented to the monkeys by either a pair of Audix PHS-vs speakers (frequency response ± 3 dB, 75–20,000 Hz) located on either side of a center monitor, or a centrally located Yamaha MSP5 monitor speaker (frequency response: 50 Hz to 40 kHz), located 30 in from the monkey’s head. The auditory stimuli ranged from 65 to 80 dB SPL measured at the level of the monkey’s ear with a B&K sound level meter, and a Realistic audio monitor.

Stimuli included both auditory and visual stimuli. The visual stimuli, which will be analyzed as part of a separate study, included digitized pictures of familiar and unfamiliar laboratory items, toys, food, human faces, monkey faces, patterns, monochrome color fields, and gratings. The auditory stimuli included human and macaque vocalizations as well as nonvocalization stimuli. Human vocalizations were recorded on a Marantz deck, filtered and digitized at 44.1 kHz. Monkey vocalizations were provided by M. D. Hauser and included a large repertoire of rhesus macaque vocalizations recorded on the island of Cayo Santiago, Puerto Rico. These vocalizations have been behaviorally characterized as to vocalization type, context, and caller identity. Rhesus vocalizations were originally characterized according to behavioral context (Gouzoules et al. 1984; Hauser and Marler 1993). The types of vocalizations presented in the current experiment included aggressive calls (AG; barks and pant threats used in present study), given by the aggressor during agonistic encounters, including chasing and physical attack; coos (CO), given during social interactions including grooming, on finding food of low value, and when separated from the group; copulation screams (CS), given by the male during mating; geckers (GK), given by juveniles when rejected; grunts (GT), given during social interaction such as an approach to groom and on the discovery of a low-value food item; girneys (GY), given during grooming and when females attempt to handle infants; harmonic arches (HA), given on discovering a food item of high value; shrill barks (SB), an alarm call given during threatening situations posed by humans during the trapping season; submissive screams (SC); types of screams used in present study are pulsed, noisy/tonal, noisy/undulating, and arched), given by a subordinate during agonistic encounters including a visual or vocal threat from a dominant, a chase, or physical attack; and warbles (WB), given on discovering a food item of high value. The vocalization types used can be further organized from a contextual or functional perspective under broader call classes that include high-value food calls (HA and WB); low-value food calls (GT and CO); social calls (CO, GT, GK, and GY); agonistic calls (AG, SB, SC); and mating calls (CS). We will refer to the different vocalizations used as belonging to a particular call type or a larger call class.

The nonvocalization stimuli were created using digital signal processing software, (SIGNAL, Kim Beeman, Cambridge, MA) and included pure tones, fm sweeps, noise bursts, and modified vocalizations. The modified vocalizations included reversals of the vocalizations in the time domain and noise-filled temporal envelopes created by convolving the amplitude envelope of the vocalization with a band limited noise burst. The noise envelopes preserved the temporal

**FIG. 1.** Localization of the prefrontal auditory area in the macaque monkey. Right: a schematic lateral view of the left hemisphere of a macaque brain, left: a coronal section through the prefrontal cortex. The location of the recording chamber on 1 of the monkeys from the present study was constructed from a MRI and is shown as a circle (black dashed lines) on the lateral brain schematic. Within the recording chamber the recording area where auditory responsive neurons were found is indicated by a square. Auditory responsive neurons are usually found at the approximate level of the inferior prefrontal dimple (ipd), when this small sulcus is present. A coronal section corresponding to the same anterior-posterior level as that of the prefrontal auditory area is shown on the left. The approximate anterior-posterior location of this section is indicated on the lateral brain schematic by a black line at the top of the diagram. Abbreviations are as in Paxinos et al. (2000). IAR, inferior arcuate sulcus; IPD, inferior prefrontal dimple; LOS, lateral orbital sulcus; PS, principal sulcus; SAR, superior arcuate sulcus.
stimulus, presentation, and a 500-ms poststimulus fixation period. A juice reward was delivered at the termination of the poststimulus fixation period, and the fixation requirement was then released. Losing fixation at any time during the task resulted in an aborted trial. There was a 2-s intertrial interval.

Each day, the subjects were brought to the experimental chamber and were prepared for extracellular recording. The head was fixed in place by means of the chronically implanted head-post and a stereotaxic adaptor was placed on the recording cylinder. A parylene-coated Tungsten electrode (0.8–2.0 MΩ at 1 kHz, FHC) was inserted into the target auditory region by a hydraulic microdrive (Narishige MO-95C), which fit over the recording cylinder. The neuronal activity was amplified (BAK MD-4 amplifier), filtered (Khron-Hite, 3700, Avon, MA), discriminated (BAK Window Discriminator) and displayed on an oscilloscope. Discriminated spikes were digitized and saved online. Simultaneous isolation of up to two units was possible with dual time/amplitude window discriminators. The timing of the behavioral contingencies, acquisition and storage of TTL spike data, presentation of odd stimuli, delivery of reward, and monitoring of eye position were controlled by a PC computer running CORTEX (dual-computer mode).

Initially, the cylinder’s placement was explored with auditory and visual stimuli to delineate the boundaries of the auditory responsive region. Each isolated unit (n = 301) was tested with at least one of the 12 vocalization lists. We did not test every cell with an extensive list of multiple examples of each vocalization type due to the limitations of awake-animal recordings. Instead, we used the 12 lists of 10 stimuli each, with 1 stimulus representing each vocalization type. Each cell was tested with a single list, with 9–12 repetitions of each call in a randomized block design. Those neurons that responded well to a single vocalization type or call class were presented with further exemplars from the same caller within those categories and different callers across categories to determine whether the cell was selective for a specific caller or call type. For the next cell isolated, the next stimulus list was used and thus the population of vlPFC cells was tested in a nonbiased manner with our entire stimulus ensemble spread over the population of cells. Testing stopped after two additional 10-type lists were tested or a minimum of four additional call types, if a clear response to more than one type of vocalization was found, or if the isolation of the cell was lost.

### Data analysis

The unit activity saved in Cortex was read into Matlab (Mathworks) and Excel (Microsoft) where rasters, histograms, and spike density plots of the data could be viewed and printed. For analysis purposes, mean firing rates were measured in the intertrial interval (SA), fixation period (FP), and during the early (S1, 0–250, or 0–500 ms) and late (S2, 500–1,000 ms) periods of stimulus presentation. In addition, the mean firing rate was measured during the variable auditory stimulus period by dividing the spike count during the stimulus presentation period by the duration of the auditory stimulus (SRATE). The stimulus duration varied between 100 and 1,200 ms.

Significant changes in firing rate during the task were detected using a one-way ANOVA (for each of the 3 stimulus periods: early, late, and duration of stimulus) and a repeated-measures ANOVA of vocalization type (10 levels), by time window (2 levels). The repeated-measure time window consisted of the mean firing rate during the intertrial interval (SA) and mean firing rate during the stimulus duration (SRATE). The significance level adopted for these analyses was $P \leq 0.05$. Only neurons with a significant main effect of vocalization type or a significant interaction between vocalization type and time window and a significant one-way ANOVA for one of the stimulus time periods at a level of $P \leq 0.05$ were considered responsive.

VOCALIZATION SELECTIVITY. For those neurons significantly responsive as defined in the preceding text ($n = 124, P \leq 0.05$), we

### Table 1. Subset of vocalization lists used

<table>
<thead>
<tr>
<th>Call Type</th>
<th>List 1</th>
<th>List 2</th>
<th>List 3</th>
<th>List 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG</td>
<td>AH</td>
<td>AM</td>
<td>AL</td>
<td>AX</td>
</tr>
<tr>
<td>CO</td>
<td>AJ</td>
<td>BF</td>
<td>AK</td>
<td>BB</td>
</tr>
<tr>
<td>CS</td>
<td>AI</td>
<td>BX</td>
<td>BL</td>
<td>BN</td>
</tr>
<tr>
<td>GK</td>
<td>BC</td>
<td>CU</td>
<td>BE</td>
<td>CT</td>
</tr>
<tr>
<td>GT</td>
<td>AR</td>
<td>BH</td>
<td>AR</td>
<td>BP</td>
</tr>
<tr>
<td>GY</td>
<td>AA</td>
<td>AT</td>
<td>AA</td>
<td>AQ</td>
</tr>
<tr>
<td>HA</td>
<td>DZ</td>
<td>CP</td>
<td>DZ</td>
<td>CF</td>
</tr>
<tr>
<td>SB</td>
<td>AP</td>
<td>AS</td>
<td>AP</td>
<td>AS</td>
</tr>
<tr>
<td>SC</td>
<td>AL</td>
<td>BW</td>
<td>CR</td>
<td>BJ</td>
</tr>
<tr>
<td>WB</td>
<td>BV</td>
<td>BV</td>
<td>AD</td>
<td>AD</td>
</tr>
</tbody>
</table>

AG, aggressive; CO, coo; CS, copulation scream; GK, gecker; GT, grunt; GY, girney; HA, harmonic arch; SB, shrill bark; SC, submissive scream; WB, warble.
determined the significant source of variance with a post hoc Tukey test. Using the Tukey pairwise comparison \( P \) values, any vocalization type that was significantly different \((P \leq 0.05)\) from at least two other vocalization types in the set of 10, was considered a "preferred stimulus" for that cell. In a less restrictive analysis, we computed a monkey call-preference index (MCPI) for each cell (Tian et al. 2001). The MCPI was the number of calls for which the response (SRATE) of the neuron was \( \geq 50\% \) of the response to the call that elicited the maximum response. In this manner, we could quantify the population for their preferences to individual vocalizations and groups of vocalizations.

**DECODING ANALYSIS.** Linear discriminant analysis (Johnson and Wichern 1998) was used to classify single trial responses of individual neurons with respect to the stimuli which generated them. We also used quadratic discriminant analysis, which takes into account differences in response variability for different stimuli, but the results were similar. Classification performance was estimated using twofold cross-validation. In general cross-validation minimizes overfitting. In this case, because our model was quite simple, it was used to remain consistent with other studies. This analysis resulted in a stimulus-response matrix, where the stimulus was the vocalization that was presented on an individual trial, and the response was the vocalization to which each single trial neural response was classified. Each cell of the matrix contained the count of the number of times that a particular vocalization (the stimulus) was classified as a particular response by the algorithm. Thus the diagonal elements of this matrix contained counts of correct classifications, and the off-diagonal elements of the matrix contained counts of the incorrect classifications. Percent correct performance for each stimulus class was calculated by dividing the number of correctly classified trials for a particular stimulus (the diagonal element of a particular row) by the total number of times a particular stimulus was presented (usually 9–12, the sum of the off-diagonal elements in a particular row).

**INFORMATION THEORETIC ANALYSIS.** We also calculated the partial information contained in the neural responses of a single cell about each stimulus. Partial information was calculated as

---

**FIG. 2.** The spectrograms and waveforms for 10 example vocalizations are shown, 1 for each vocalization type used in the present experiment. The calls have been previously characterized according to acoustic features and functional referents (Hauser and Marler 1993) (see METHODS). While each call type has a number of acoustic features that differentiate them from the other call types, there are some salient features that are common to several vocalization types as shown here. One manner in which to group the calls is by noisy vs. harmonic/tonal calls. The noisy calls include aggressive calls (barks and pant threats were used, call shown is a pant threat), shrill barks, and grunts. Coos, warbles, and harmonic arches have in common the presence of a harmonic stack. Copulation screams and screams (pulsed scream shown) vary in individual examples and may have harmonic or tonal elements present. The girneys and the geckers are quite variable in their acoustic structure and constitute their own groups.
where the sum is taken over all possible responses. The stimulus-response probability distributions were estimated using the classification matrix. The average of the partial information across stimuli gives the total information in the neural response about all of the stimuli. The partial information about a particular stimulus is a measure of how well the response can be predicted when a given stimulus is shown. This is related to the percent correct classification for each stimulus. However, it is possible for a given stimulus to elicit a reliable, consistent response, which leads to high partial information, when the response is incorrect, and thus the percent correct classification is low. Because there may be more information in the raw neural responses than there is in the decoded neural responses, due to mismatches between the model and the real response distribution, these information estimates are lower bounds. However, given the small number of trials, this regularization was necessary. We also corrected for any residual bias, using empirical estimates of the linear and quadratic bias terms (Treves and Panzeri 1995). The quadratic terms were always negligible compared with the linear terms, so we felt that the bias was being effectively controlled. Furthermore, we were primarily interested in the relative information values, as an estimate of the number of stimuli about which cells carry information on average, rather than the absolute amount of information carried about individual stimuli.

**RESULTS**

**Response types and ANOVA results**

We recorded a total of 301 auditory responsive cells from the left hemisphere of one monkey and right hemisphere of a second monkey. We were able to test 245 of these auditory cells with 1 of the 12 vocalization lists. Prefrontal neuron responses to the auditory stimuli included cells with short-lived phasic activity, which was coincident with stimulus onset and/or offset, and cells with a longer lasting, tonic response to the auditory stimuli which lasted the length of, or beyond, the stimulus duration. Some cells exhibited a combination of phasic and tonic activity. The timing of the responses to the different vocalizations varied. This is partly due to individual latency variation among the cells but is also due to the presence or absence of salient features in the vocalizations which appear to differentially evoke responses in the cells. This observation alone suggests that the stimulus-evoked responses of vIPFC neurons are partly determined by features present within the vocalizations.

We analyzed the 245 vocalization-tested cells with a one-way repeated measures ANOVA (repeated measure = 2, SRATE, SA × vocalization type). We also computed a one-way ANOVA of the neuronal firing during the auditory stimulus period (minus the spontaneous rate) by vocalization type. Of the 245 cells tested with all 10 vocalization types, 124 cells (monkey 1, n = 58; monkey 2, n = 66) were significantly responsive (P ≤ 0.05) during the stimulus period or had a significant effect of vocalization type or an interaction of type by time window. To determine the source of significant variance in these cells, we performed a post hoc Tukey HSD test on the response to the 10 vocalizations. This yielded a matrix of pairwise P values across the 10 vocalization types. The pairwise Tukey data are problematic for quantifying the general “selectivity” of all vIPFC audi-
tory cells for macaque vocalizations because some cells which respond equally well to more than five vocalizations would not show a significant effect in the post hoc comparisons. The post hoc analysis would reveal that there was no significant difference among the vocalization responses without indicating that most of the vocalizations evoked a response above baseline firing rate. A simpler, though less rigorous, procedure involves calculation of a MCPI for each vlPFC cell (Tian et al. 2001). We calculated the MCPI for each cell as a general means of establishing how many vocalizations evoked a response in a given cell.

Selectivity

Most of the vocalization responsive cells demonstrated an increase in firing rate to more than one vocalization with the majority responding to between two and five vocalizations (Fig. 3). Because cells were not systematically tested with nonvocalization stimuli, we cannot state that these neurons were vocalization-specific. Hence, our selectivity results are based on vocalization-responsive cells. The average number of vocalizations that a cell responded to was three. Few cells responded to every vocalization type or to a single vocalization type. Thus our population exhibited a gradient in terms of stimulus selectivity. A small percentage of cells were responsive to six or more vocalization types (n = 11; MCPI > 5). An example of this type of cell is shown in Fig. 4A. The raster and spike density plots for this cell indicate an increase in firing to many calls in our test list and the MCPI calculation indicates responses to eight vocalizations (Fig. 4A). We tested a subset (n = 5) of these “nonselective” cells with nonvocalization stimuli, including calls reversed in the time domain, human speech sounds, FM sweeps, and noise stimuli. All five cells showed a significant response to nonvocalization stimuli.

Of the 124 vocalization responsive cells, the MCPI was 1 for 21 cells, indicating that the response to the other nine vocalizations was not >50% of the maximum response. For 14/21 cells, this vocalization was significantly different from the other vocalizations tested with post hoc analysis (Tukey HSD, P = 0.05). We tested 16 of these 21 cells with additional vocalizations and found that they also responded to the new vocalizations. The additional vocalizations included aggressive (AG) vocalizations from other callers, as well as vocalizations from other speakers. CO, coo; CS, copulation scream; GK, gecker; GY, girney; HA, harmonic arch; SB, shrill bark; SC, submissive scream; WB, warble.

FIG. 4. Nonselective and selective cells. Raster (tics) and spike density (gray shaded function) plots are shown of 2 infrequently encountered cell types. Each panel represents the cell’s response to the vocalization type abbreviated on top. The waveform for the vocalization tested is shown below each plot. In each plot, the prestimulus fixation period (−250 to 0) occurs prior to the onset of the vocalization which is at time 0. A juice reward was given at the end of the trial at time 1,500 ms, which is off the plots. The rasters are shown with trial 1 at the bottom. A: raster and spike density plots are shown for 1 cell that responded to >5 vocalizations in the 10-type list. B: selective cell that responded best to grunt (GT) vocalizations in the original 10-type list is shown in the top 2 rows. On further testing (bottom), this cell responded to GT from other speakers (GT-C2, GT-C3) but also responded to acoustically similar aggressive (AG) vocalizations (pant threat) from other callers (AG, pant threat –C4 and AG, pant threat –C5). CO, coo; CS, copulation scream; GK, gecker; GY, girney; HA, harmonic arch; SB, shrill bark; SC, submissive scream; WB, warble.
vocalizations across call type and caller to determine call type and caller selectivity. A caller-selective cell responded best (using post hoc comparisons, Tukey HSD, \( P \leq 0.05 \)) to vocalizations from a single caller across multiple types and not to vocalization by other callers. A vocalization- or call-type-selective cell exhibited a significant increase in firing to a single vocalization type across multiple callers and not to additional vocalization types. We tested cells that appeared responsive to a single call with at least two additional stimulus lists, four additional exemplars of that call type (by different callers), and as many exemplars of the same caller possible given the limitations of the library of calls used. We also tested seven of these cells with nonvocalization stimuli, including vocalization reversal, noise stimuli, and/or FM sweeps. Using this additional testing procedure, few cells were caller selective \((n = 2)\), and one of these was selective for call type and caller. We found that many of the cells that were call-type responsive in the initial testing were responsive to acoustically similar stimuli of a different vocalization type or even to some nonvocalization stimuli. As an example, the cell portrayed in Fig. 4B had a best response to GTs \((\text{MCPI} = 1); \) post hoc Tukey test, \( P \leq 0.001 \) compared with all 9 other stimuli). With further testing, this cell responded to additional GTs from other callers but was also responsive to pant threats from other callers, an AG vocalization type that is acoustically similar but contextually different to GTs (Fig. 4B). With the additional testing of callers and call types 8/16 cells were found to be call-type selective. The cells that responded selectively to particular vocalization types were not confined to a particular hemisphere in our experiments but were observed in the right hemisphere of one monkey and in the left hemisphere of the second monkey. There was an insignificant trend for call-type cells to be located in the most anterior portion of the auditory responsive recording region \((P = 0.09)\).

As mentioned, the majority of vlPFC cells \((n = 92)\) had an MCPI between 2 and 5 and were thus responsive to several vocalization types (Fig. 3). Cells in our population responded to a variety of vocalization types. Using the MCPI to determine “best call” for each cell \((i.e., \) highest mean firing rate) GKS were the preferred type in 31 cells (Fig. 5). Figure 6, A and B, shows the raster and spike density plots for two typical vlPFC cells. The cell shown in A had a MCPI of 3 and had a best response to GY (significantly different, \( P < 0.01 \), from 6 other calls). The cell shown in B had a MCPI of 5 and responded best to AG (significantly different \( P < 0.05 \), from 5 other calls). To further investigate the selectivity or tuning of vlPFC neurons to vocalizations, we examined each responsive cell with additional analysis methods.

### Decoding and information theoretic analysis

Although the MCPI analysis shows that most cells are responsive to between two and five vocalizations, it does not indicate how much information single cells convey about individual calls within our stimulus set. To characterize more specifically the number of stimuli about which single cells provide information, we carried out linear discriminant and partial information analyses (see METHODS) on the neural responses. These analyses give us an estimate of how well individual stimuli can be discriminated by the responses of single neurons and allow us to estimate the number of vocalizations about which a neuron provides information. In general, single cells were not highly effective at discriminating among the 10 calls in the testing set. Particular cells, however, performed well and are shown here. As an example, we have plotted the partial information and percent correct (Fig. 6, C and D) across all vocalization types for the cells shown in A and B. The cell shown in Fig. 6A had a significant response to vocalization type in the ANOVA \((P < 0.00001)\), and the post hoc Tukey test revealed that GY and SC differed significantly from more than half of the vocalizations tested \((P < 0.01)\). The mean percent correct for this cell was 25, and the total information was 1.4 bits. Examination of the response to individual stimuli (Fig. 6C) shows that GY and SC were discriminated correctly at 63 and 75%, respectively (Fig. 6C, ●). The cell had partial information of \( \sim 1.6 \) bits for each of these vocalizations. Comparison with the raster response shows that GY evoked the highest mean firing rate, whereas SC evoked almost no response, i.e., the lowest mean firing rate. These extremes in firing rate allowed these two calls to be discriminated better than other calls it was tested with. This is similar to the second example shown in Fig. 6, B and D, which had a mean percent correct of 12% and total information of 0.95 bits. This cell had a significant response to vocalization type in the ANOVA \((P < 0.00001)\) with significantly different responses (post hoc Tukey test, \( P < 0.05 \)) to AG, GT, CS, and SC. The plot in Fig. 6D shows the partial information across the vocalizations from highest (AG) to lowest (GY). The vocalizations that had the highest mean firing rate (AG) and the lowest (CS) had the highest partial information \((AG, 2.3 \text{ bits}; \) CS, 1.44 bits). CS, which evoked the lowest mean firing rate, also had the highest percent correct score (70%) and thus was discriminated well by this cell. The ANOVA analysis showed that this cell had a significant effect of vocalization type, but post hoc analysis did not distinguish among the four significant vocalizations. Hence, from the point of view of the ANOVA analysis, the cell does not appear selective and the MCPI of this cell is 4. The information theoretic analysis suggests that the cell had a high amount of partial information about one vocalization in particular (AG) while the decoding analysis revealed that the CS vocalization was discriminated far better than any other for this cell, and hence, the cell might be considered selective for this particular call.

We sorted the vocalizations for each cell in descending order based on the partial information transmitted and the percent correct for each vocalization using the responses to the 10 vocalizations from the original 10-type list. This resulted in
information and percent correct tuning curves for each cell. Figure 7 shows a plot of the population average tuning functions. It indicates the average information and percent correct for a given cell across vocalizations ranked for the amount of information transmitted, from most information to least information. The graph shows that for a given cell in our population there is information about one or two vocalizations but that this drops off dramatically across the rest of the vocalization set. Considered in this manner, neurons within our population convey information about one or two vocalizations or might be considered “selective” for two vocalizations.

Hierarchical cluster analysis

Because our data indicated that vlPFC cells respond to multiple vocalizations, we wanted to quantify this response and determine which vocalizations cluster together in our prefrontal cells. In this manner, we can ask whether prefrontal cells categorize or group together vocalizations with similar social function. We used a hierarchical cluster analysis to quantify each cell’s response to multiple vocalizations. The hierarchical cluster analysis yielded a dendrogram of each cell’s response to the 10 vocalizations grouping together stimuli that evoked similar responses. For the cluster analysis, we included any cell that had a significant response \( P \leq 0.05 \) to vocalization type, interaction of response bin by vocalization type or to the stimulus period \( n = 124 \). The analysis resulted in a dendrogram and cophenetic coefficient (measure of fit) for each responsive cell. The corresponding dendrograms of the cells from Fig. 6, A and B, are shown in Fig. 8, A and B. For both cells, the results of the cluster analysis can be explained by groupings that reflect the stimulus-evoked responses seen in the raster/spike density plots. For example, the dendrogram in Fig. 8A shows two major clusters, the “best” call GY (as determined by post hoc Tukey analysis) portrayed as an outlier in the first cluster with SB and HA. The second cluster includes low responses to GT, GK, CS, CO, and WB, with AG and SC as outliers. The dendrogram in Fig. 8B corresponding to the cell in Fig. 6B also reveals two major clusters. The first cluster includes moderate to high responses to SB, WB, AG, and GK with GT as an outlier. The second cluster can be subdivided into a moderate to low firing rate group (CO, HA, GY) and a no-response group (CS, SC). For both cells, the clusters present in the dendrograms correspond with the raster/spike density plots as in Fig. 3.
Figure 9 shows two additional examples. In Fig. 9A, the WB and CO vocalizations, which evoked the highest mean firing rate (black bars) cluster together while the reduced responses to the other eight vocalizations form a second major group (dendrogram, Fig. 9B), which can be subdivided into two smaller groupings shown with dashed and dotted lines around the clusters in the dendrogram and mean firing rate bar graph. The cell portrayed in Fig. 9C responded best to SC (a noisy tonal scream) and CS, which had similar mean firing rate and, as shown in the dendrogram (Fig. 9D), are grouped together. The four calls, which all elicited a small response (AG, GK, GY, HA), group together in the dendrogram (D, dashed line) as do the four calls that elicit no response or a low inhibitory response (CO, GT, WB, SB; dotted line).

Examining the dendrograms for these two cells, we asked whether the vocalizations that clustered together could be related by acoustic or functional similarity. A similar response by a cell to several vocalizations that have a similar functional referent might suggest that single prefrontal neurons are capable of encoding functional attributes in their responses to communication stimuli. In contrast, cells might respond similarly to calls that have similar acoustic features. To determine the functional classes and behavioral context of the vocalizations, we relied on previously published studies of rhesus macaque vocalizations (Gouzoules et al. 1998; Hauser and Marler 1993; Rendall et al. 1998) (as described in METHODS). To determine the acoustic features common to our 10 vocalization types, we relied both on previously published studies (Hauser 1996; Hauser and Marler 1993) and on our own cluster analysis of the vocalizations themselves (see METHODS). The cell in Fig. 9A had a best response to WB and CO, and these two acoustically similar but functionally different calls (see Fig. 2 and METHODS) clustered together in the dendrogram (Fig. 9B). The cell in Fig. 9C had a significant increase in firing for SC and CS. The dendrogram for this cell (Fig. 9D) shows a clustering of SC and CS for the corresponding cell, both of which are uttered under different behavioral contexts but, for these two tokens used, have similar acoustic features. For each of the 124 responsive cells, we computed a dendrogram for the cell’s response to the 10 vocalization types.

**Consensus tree**

To determine if there were any common clusters of particular vocalizations across the entire population of vocalization-responsive neurons, we computed a consensus tree using the 124 dendrograms. The consensus tree indicates the groups that occur most often in the dendrograms across our population of vIPFC cells. Because our goal was to determine if vIPFC neurons clustered vocalizations with common functional referents or similar acoustic features, we compared the consensus tree derived for the neuronal response with a consensus tree of the vocalizations themselves (see METHODS). The vocalization consensus tree revealed several common clusters across call lists (Fig. 10A). In five of our vocalization lists, CS and SC clustered together. A second cluster was that of WB and CO, which clustered together in five lists and with GT in six lists;
and a third cluster was AG calls and SB. Comparison of this vocalization analysis with previously published data on the acoustic features of rhesus calls (Ghazanfar and Hauser 1999; Hauser 1996; Hauser and Marler 1993; Owren and Rendall 2001) indicates some common findings. For example, AG calls (barks, pant threats) and SB are classified acoustically as noisy

FIG. 9. Mean firing rate (±SE) graphs and corresponding dendrograms for 2 cells. For 2 cells, the vocalization responses are plotted as mean firing rate (± SE) in the bar graphs in C and are accurately reflected in the dendrograms in B and D. The cell in A had a mean firing rate that was highest for WB and CO (black bar) compared with the other vocalization responses. The dendrogram shows clustering of responses with similar firing rate, for example, the high firing rate for WB and CO (dark box in A and B); moderate firing rate for GK GT, and AG (dotted line box in A and B); and low firing rate for GY, HA, CS, and SC (dashed line box in A and B). The cell in C had a mean firing rate that was highest for SC and CS which cluster together (dark box in C and D); moderate firing rate for HA, GY, GK, and AG (dashed line box in C and D); and low firing rate for SB, WB, GT, and CO (dotted line box in C and D), which cluster together.

FIG. 10. Vocalization and neuronal consensus trees. Consensus trees, based on the dendrograms for the vocalizations used in the current study and analyzed according to their mean power at each frequency (A) and for the neuronal response to the vocalizations (B), are shown. A: dendrograms were derived for each of the 12 vocalization lists, and a consensus tree of these was generated indicating the common groupings according to acoustic features of the vocalizations. The consensus tree indicated that in our 12 vocalization lists, 3 groupings occurred in 5 vocalization lists: AG and SB, CO and WB (which clustered with GT in 6 lists), and SC and CS. In B, dendrograms from the neuronal response to these vocalizations were pooled to create this consensus tree of the most common groupings. Although no large groups emerged in our analysis, AG and GT clustered together in 15% of the cells and CO and WB clustered together in 14% of cells. Both of these groups represent clusters of acoustically similar but functionally dissimilar vocalizations.
calls, whereas WB and CO both have a rich harmonic structure (Fig. 2). CS and SC vary considerably but have tonal and harmonic features that are commonly present in many tokens.

The consensus tree of the neuronal responses can be compared with the consensus tree of the vocalizations and the known features of the vocalizations (Figs. 10A and 2). Although there were some common groupings, across our population of 124 cells there were no large clusters that represented >20% of the population. As Fig. 10B indicates, the most common cluster for the neuronal data that occurred in 15% of the cells (19/124) occurred with AG calls (barks, pant threats) and GT. These two vocalization types are noisy calls. The pant threat AG call is so similar to the GT vocalization that they are indistinguishable to humans (Hauser and Marler 1993). However, these acoustically similar calls are given under very different behavioral contexts and with different accompanying facial gestures (Hauser et al. 1993). AG calls are given during agonistic social interactions, whereas GT vocalizations are given under many positive social contexts including approach to groom and finding a food item of low value. A second cluster that occurred in 14% of the cells (17/124 cells) was the cluster of WB and CO. WB are food calls that are given by the monkeys on discovering a food item of high value. CO are given under many circumstances but are common during social interactions, separation from the group, and during the discovery of food items of low value. Although the calls are quite different functionally, they share salient acoustic features because these two vocalization types are both harmonically rich calls, differing primarily in the extent of rapid FM of the fundamental (Fig. 2). Figure 9A illustrates a cell that responded best to WB and CO. The dendrogram shows these two vocalization types clustered together in that individual cell (Fig. 9B). In Fig. 2, the spectrogram for WB and CO is shown indicating the common acoustic features including the presence of a harmonic stack. Furthermore, these two vocalizations are also grouped together in the vocalization consensus tree (Fig. 10A). Ten cells in our population grouped together CS and SB (an alarm call), which differ functionally but have some acoustic features that might be jointly present in these two call types depending on the particular tokens used and the acoustic features focused on. There were several other common groups that are not shown in this neuronal consensus tree because they occurred less often than the clusters included in the tree. For example, 15 cells clustered HA and WB together and 15 cells clustered HA and AG calls together. The clustering of HA and WB together is significant because this represents a grouping of calls with similar context or function and was the only example of functionally related calls that clustered together in our neuronal data. Like WB and CO, HA and WB have some acoustic features in common but can still be differentiated, for instance, by the more dramatic change in the fundamental frequency over time in HA.

Because our cells were tested with 12 different vocalization lists that each contained a single exemplar from each of the 10 categories of vocalizations used, our analysis of the entire population is based on responses to different tokens from the same categories. For this reason, we also computed separate consensus trees for the cells tested with each stimulus list. Thus all cells tested with list 1 were analyzed together, all cells tested with list 2, and so on. Because there were ~10 cells tested with each list, the number of cells contributing to each list-consensus tree were quite small, and the overall percentage of cells that had particular groupings of vocalizations was, accordingly, small and few major clusters emerged. Nonetheless, for 3 of the 12 lists, WB and CO, and AG and GT were common clusters just as they were in the overall population consensus tree. Additional clusters of vocalizations were apparent, and in two lists, HA and WB, which are functionally similar calls, were clustered together. The analysis by list showed that similar clusters occurred in the separate lists that are present in the population and that averaging across lists did not appear to obscure any common groupings.

**Balanced analysis of call and caller**

The results in the present study indicated that only two cells were responsive to a particular caller and only eight cells were significantly responsive to call type. One problem in assessing call type and caller specificity in our study is that the vocalization library used to test our cells did not contain examples of every call type for a single caller, and thus different callers had to be used across categories. To more carefully assess the possibility that vlPFC neurons encoded caller or call-type selectively, we used a smaller stimulus set to construct a 3 × 3 matrix of three call types (CO, GT, GY) made by the same three callers (AR, BG, and BQ). We tested 101 cells with this list and analyzed the results in a two-way ANOVA with caller and call type as independent categorical variables and response during the stimulus as the dependent variable. Even when using this limited set of vocalizations, only seven units showed an effect of caller, whereas 16 cells had a significant effect of call type (P < 0.05). There was an interaction effect of call by caller in 10 cells. Examining the data further with a post hoc Tukey test revealed that in the 16 call-type-responsive cells, 10 responded best to GT, 3 to CO, and 3 to GY. These numbers are still small in terms of the population of tested cells (16 call-type cells/101 cells). The fact that even when a balanced call by caller list is used <16% of the population show an effect of call type indicates that the population of cells that respond selectively to call type may, in fact represent a small segment of the total vlPFC auditory population, and is not affected by our testing paradigm.

**Discussion**

The data from the present study confirm that vlPFC auditory neurons respond to a variety of species-specific vocalizations from a library of calls by unfamiliar callers. Furthermore, our data suggest that vlPFC neurons display a gradient of selectivity. We found that using typical parametric statistics the majority of vlPFC neurons recorded responded to two to five vocalizations, whereas smaller percentages of cells responded either selectively to a particular vocalization or category of vocalizations or nonselectively to most auditory stimuli tested. Information theoretic approaches indicated that across the population, vlPFC neurons convey information about one or two vocalizations. Examination of vocalization responses using hierarchical cluster analysis suggests that vlPFC cells do not respond to multiple calls based strictly on functional or referential meaning. Rather these responses may be due to other features including acoustic similarity or some interaction of behavioral context and perceptual features.
Selectivity

In our analysis using 10 vocalization types, the majority of vlPFC cells (n = 103) responded to more than one vocalization type, and on average, vlPFC cells respond to three vocalizations. In contrast, the number of cells in the vlPFC that, by our criteria, responded selectively to a particular call type or caller ID was quite small, 6% for call type in the general population and <1% for caller. Even with a reduced, balanced list of three call types by the same three callers, only 7% of cells showed an effect of caller and 16% call type.

Our selectivity index (MCPI) results (Fig. 3) are remarkably similar to those of the lateral belt auditory area because most cells (82%) responded to more than one vocalization, whereas a smaller number of cells exhibited selectivity for a single vocalization or vocalization category (Tian et al. 2001). Because the lateral belt and parabelt project directly to the vlPFC and the frontal auditory cells recorded from in the present study lie in the projection zone of AL, the similarity between vlPFC neurons and AL neurons is not surprising. However, if the vlPFC is elevated in the processing hierarchy, one might expect increased selectivity in vlPFC. One possible explanation for our results is the fact that no cognitive task was employed by the monkeys in the present study. Performance of a task that requires discrimination or cognitive processing of the vocalizations might reveal additional selectivity by prefrontal neurons. Furthermore, the MCPI data from Tian et al. (2001) used only 7 vocalization types, whereas we have used 10.

When our data were examined using information theoretic and decoding approaches, vlPFC cells do appear more selective as might be expected from prefrontal cortex. Our data indicate that, considered qualitatively, vlPFC cells convey information about one or two vocalizations and that the amount of information rapidly decreases from the best stimulus to the worst (Fig. 7). Although information theoretic analyses have not been used to study macaque lateral belt auditory neurons, it has been used to study face-processing cells in macaque inferotemporal cortex. IT cells do not use a local or “grandmother”-like encoding where single cells show great selectivity for a single object but use a distributed code where each cell responds to a number of face stimuli in a graded fashion (Rolls and Tovee 1995). It is possible that auditory cortical areas such as the lateral belt, parabelt, and superior temporal sulcus, which project to vlPFC auditory cells, may use a distributed code by which to discriminate among species-specific vocalizations. This is ideal for discrimination where fine details must be compared across stimuli (Rolls and Tovee 1995). Once this highly processed information reaches vlPFC, however, our current data suggest that a local code is implemented because on average, vlPFC auditory cells have information about only one or two stimuli. In the marmoset auditory cortex, Wang et al. (1995) suggested the existence of two general populations of auditory cortical neurons, one responding selectively to call types or even callers with the other responding to a wider range of sounds. One notion is that the nonselective cells provide a general and flexible means for sound analysis, whereas call-selective cells may provide a quick and accurate means to detect the occurrence of a commonly heard species-specific vocalization (Wang 2000). Because our ANOVA data suggest a gradient of stimulus selectivity, it is possible that both local (selective) and distributed mechanisms may operate within vlPFC where call-selective neurons may participate in networks that are involved in call-specific actions, whereas less selective neurons would then participate in encoding new sounds and linking up relevant behaviors with sound sequences.

Perception and categorization of natural stimuli

One issue explored is how vlPFC cells categorize vocalization stimuli. The results of the cluster analysis indicate how individual cells group the vocalizations and the consensus tree reveals the most common clusters within the entire population of vlPFC cells. Our data do not indicate any major clusters representing >20% of the population either acoustically related or functionally related in the overall population. However, there were a number of small groupings (>10% of the population) that occurred where cells responded to several vocalizations with similar acoustic features (e.g., WB and CO: AG and GT) rather than similar functional referents. Figure 9 shows the dendrograms and mean firing rate graphs for two cells indicating this feature. Some of the clusters that emerge from the neuronal data are similar to the groupings that occur in a cluster analysis of the vocalizations themselves and are related by acoustic similarities (Figs. 2 and 10) because the most common cluster within the neuronal population was WB and CO, which are acoustically similar and cluster together in the vocalization analysis. There was a slightly smaller segment of the population (15 cells) that clustered functionally similar calls, WB and HA, together. Overall, the percentage of cells in which groupings occurred that were functionally or acoustically similar was small and therefore may not accurately reflect the essential features common to the stimuli which evoke similar neuronal responses. Further studies using stimuli that emphasize particular acoustic elements may reveal the salient features that evoke prefrontal auditory responses. Tasks that more directly assess the ability of monkeys to categorize vocalization stimuli may demonstrate the ability of prefrontal neurons to classify stimuli along functional or acoustic dimensions. However, the data in the present study using a fixation task do not provide evidence that single vlPFC cells categorize auditory stimuli along functional lines.

Role of vlPFC in call coding: articulatory versus semantic streams

If vlPFC cells respond to vocalizations on the basis of acoustic features, future work can be directed toward identifying the essential and common features contained in different types of species-typical vocalizations that might be encoded by prefrontal neurons. A number of feature reduction (Rauschecker 1998) or extraction techniques (Barbour and Wang 2003; Shamma and Versnel 1995) have been used to examine auditory cortical processing. New techniques have also been developed for extracting elements within communication sounds to determine essential features that drive auditory cortical neurons (Averbeck and Romanski 2004). Use of these techniques can help to delineate features, common to groups of vocalizations, that drive auditory cortical cells and to discern the topography of feature selectivity across auditory cortical regions including the vlPFC.
The results of the present study, which suggest that vlPFC neurons may respond to complex stimuli on the basis of acoustic features, are consistent with its presence in a ventral auditory processing stream that analyzes the features of auditory objects. The notion of an object-based auditory stream is supported by both human (Belin et al. 2000; Binder et al. 2000; Scott et al. 2000; Zatorre et al. 2004) and animal studies (Rauschecker and Tian 2000; Romanski et al. 1999b; Tian et al. 2001). The localization of an auditory object processing stream to the vlPFC (Fig. 1) in the macaque is suggestive of functional similarity between this area and human language processing regions located in the inferior frontal gyrus (Deacon 1992; Romanski and Goldman-Rakic 2002). Separate streams for the processing of spatial versus nonspatial stimuli were originally proposed for the visual system (Ungerleider and Mishkin 1982) and extended to separate domains in the frontal lobe by Goldman-Rakic and colleagues in their analysis of frontal lobe visual working memory (Goldman-Rakic 1996; Wilson et al. 1993). Although neurons of the vlPFC demonstrate robust responses to complex sounds including vocalizations, their spatial selectivity has not been systematically studied and hence their specialization for processing “what versus where” has not yet been confirmed.

Although it is plausible that vlPFC neurons are at the pinnacle of a sensory processing network that analyzes the acoustic features of a sound and matches them up with stored representations of familiar vocalizations, it is also possible that vlPFC is beyond the sensory processing hierarchy and that vlPFC might play a role in phonetic processing or articulatory control. For example, the superior temporal sulcus, which is fairly late in the auditory cortical processing hierarchy, has a robust projection to vlPFC (Romanski et al. 1999a). Rather than analyzing acoustic features for semantic content, the vlPFC might utilize information from STS to modulate motor- and phonetic-based processes and play a role in feedback control of articulation. The feature-based (phonetic) coding of natural sounds by vlPFC suggested in the cluster analysis of the present study is conducive to such a role in feedback control of vocalization motor systems. Furthermore, the location of auditory prefrontal cells in the macaque inferior frontal convexity is similar to the location of Broca’s area in the human inferior frontal gyrus. In the human brain, the posterior aspects of Broca’s area are thought to be involved in the phonetic and motor control of speech, whereas more anterior regions have been shown to be activated during semantic processing, comprehension, and auditory working memory (Buckner et al. 1995; Cohen et al. 1997; Demb et al. 1995; Fiez et al. 1996; Gabrieli et al. 1998; Gelfand and Bookheimer 2003; Pauls et al. 1993; Posner et al. 1999; Price 1998; Stevens et al. 1998; Stromswold 1996; Zatorre et al. 1992). Previous studies using large lesions have suggested a role for the lateral prefrontal cortex in auditory discrimination (Goldman and Rosvold 1970; Gross 1963; Gross and Weiskrantz 1962; Weiskrantz and Mishkin 1958). Further studies focused specifically on the ventrolateral frontal auditory region may clarify its role in the processing of communication relevant stimuli.

ACKNOWLEDGMENTS

The authors thank M. D. Hauser for helpful comments on the manuscript and for generously providing the library of macaque vocalizations. We also thank I. Kanwal, B. Gucu, S. Sally, A. Ghazanfar, and W. E. O’Neill for comments on the manuscript.

GRANTS

This research was funded by the National Institute for Deafness and Communication Disorders Grants DC-04845 to L. M. Romanski and DC-05409, Center for Navigation and Communication Sciences.

REFERENCES


Sources of acoustic variation in rhesus macaque (Macaca mulatta) vocalizations. Gross CG and Weiskrantz L. Gross CG.


The role of articulation in the production of rhesus monkey, Macaca mulatta, vocalizations. LePrell CG, Hauser MD, and Moody DB.

Discrete or graded variation in responses to word stimuli following bilateral parabelt auditory cortex lesions in macaque monkeys. Brain Res 817: 45–58, 1999.


