Behavioral/Systems/Cognitive

Representation of Con-Specific Vocalizations in the Core and Belt Areas of the Auditory Cortex in the Alert Macaque Monkey

Gregg H. Recanzone

Center for Neuroscience and Department of Neurobiology, Physiology and Behavior, University of California at Davis, Davis, California 95616

Auditory cortical processing in primates has been proposed to be divided into two parallel processing streams, a caudal spatial stream and a rostral nonspatial stream. Previous single neuron studies have indicated that neurons in the rostral lateral belt respond selectively to vocalization stimuli, whereas imaging studies have indicated that selective vocalization processing first occurs in higher order cortical areas. To test the dual stream hypothesis and to find evidence to account for the difference between the electrophysiological and imaging results, we recorded the responses of single neurons in core and belt auditory cortical fields to both forward and reversed vocalizations. We found that there was little difference in the overall firing rate of neurons across different cortical areas or between forward and reversed vocalizations. However, more information was carried in the overall firing rate for forward vocalizations compared with reversed vocalizations in all areas except the rostral field of the core (area R). These results are consistent with the imaging results and are inconsistent with early rostral cortical areas being involved in selectively processing vocalization stimuli based on a firing rate code. They further suggest that a more complex processing scheme is in play in these early auditory cortical areas.

Key words: macaque; auditory cortex; vocalizations; single neuron; physiology; cerebral cortex

Introduction

The ability to identify natural sounds is a fundamental function of the central auditory system. In humans and macaque monkeys, the ability to discriminate speech and con-specific vocalizations is an integral part of normal social interactions. The auditory cortex, particularly in the left hemisphere, is known to be necessary for this function in macaques (Heffner and Heffner, 1989; Harrington et al., 2001). Recent anatomical and physiological studies in macaque auditory cortex have suggested that there are parallel and hierarchical processing streams that selectively process spatial and nonspatial information (Rauschecker and Tian, 2000; Kaas and Hackett 2000). Evidence supporting this hypothesis includes the findings from Tian et al. (2001), where neurons in the caudal regions of the auditory cortex are more responsive to the spatial location of vocalization stimuli, whereas neurons in the rostral regions are more selective for the type of vocalization. Furthermore, the neuronal responses across the population of neurons in the caudal area CM are more closely correlated with sound localization ability compared with neurons in the primary auditory cortex (Recanzone et al., 2000b).

and neurons in caudal auditory cortical areas have sharper spatial tuning than core and more rostral belt areas (Woods et al., 2006).

It is therefore tempting to speculate that vocalizations may be processed selectively along one pathway by the auditory cortex given the importance of this class of stimuli. Electrophysiological studies have shown that neurons have robust responses to vocalization stimuli in the rostral belt areas of auditory cortex (Tian et al., 2001), the ventral prefrontal cortex (Cohen et al., 2004; Romanski et al., 2005; Russ et al., 2008), which is the target of projections from the rostral stream. Imaging studies in macaques have noted vocalization responses all along the superior temporal gyrus, with left hemisphere dominance at the rostral pole (Poremba et al., 2004; Petkov et al., 2008). Further, BOLD activity was greater for vocalization stimuli compared with equally complex nonvocalization stimuli in the more anterior auditory cortical areas, but not in the belt areas that have been studied electrophysiologically (Petkov et al., 2008). Thus, it is unclear how the responses to vocalizations described at the single neuron level in the belt fields (Rauschecker et al., 1995; Tian et al., 2001; Russ et al., 2008) relates to the lack of vocalization specificity as measured by the BOLD response.

In this study the responses of single neurons to four con-specific vocalization exemplars presented in forward and time-reversed directions were recorded in five cortical areas; the primary auditory cortex (A1), rostral field (R), caudo-medial (CM), caudo-lateral (CL) and middle lateral (ML) areas. The dual stream hypothesis predicts that responses of neurons in rostral fields will have a more selective response to forward vocalizations compared with reversed vocalizations, whereas the imaging re-

sults predict that there will be robust activity to both forward and reversed vocalizations in all five cortical areas.

Materials and Methods

Data were collected from three adult male macaque monkeys (monkeys F, G, and L) aged 5–12 years and weighing 7–12 kg that also participated in a study of the spatial processing of auditory cortical neurons (Woods et al., 2006). All animal procedures were approved by the Institutional Animal Care and Use Committee at University of California at Davis and followed Society for Neuroscience guidelines. All stimulus generation, presentation, data acquisition, and behavioral control were controlled by a TDT system run on a PC using customized software. Experiments were conducted in a double-walled sound-attenuating booth measuring 2.4 × 3.0 × 2.0 m (l × w × h; inner dimensions; IAC) lined with 3 inches of echo-attenuating foam. Sounds were presented through a 16 speaker array located 1 m from the center of the interaural axis at 0 degrees elevation. Speaker locations spanned the entire 360 degrees in azimuth at 22.5 degree intervals, but only one location was used for any given neuron (see below). Speakers were 2 inches in diameter with a flat frequency profile between 0.05 and 12.0 kHz, with 6 dB/octave roll-off at higher and lower frequencies.

Stimuli. Stimuli consisted of four monkey calls that were recorded in the vivarium housing the individual animals. Calls were recorded over a period of several hours in the absence of the investigator, thus the caller identity is unknown. These four exemplars were selected as they were the most common types of calls recorded (>97% of all recorded calls), and the specific calls used were the most common of that particular class of call. These vocalizations were therefore very familiar to the studied monkeys. Stimuli were digitized at a sampling rate of 43.5 kHz and "padded" with zero energy just before and after the vocalization. Reversed vocalizations were generated by simply transposing the order of the voltage values for each of the four vocalizations. Stimuli were presented at an average intensity of 65 dB SPL measured at the center of the interaural axis in the absence of the monkey.

Recordings. In a separate surgical procedure under sterile conditions, a restraining head post and recording cylinder were implanted over the left hemisphere [see Recanzone et al. (2000a) for details]. On each experimental day the animal sat in a custom built primate chair specifically designed for auditory experiments to reduce potential reflections of sounds near the head. A tungsten microelectrode (FHC) was inserted into the cerebral cortex using a dorsal approach via the Crist guide tube/ grid system (Crist Instruments). The electrode was advanced with a hydraulic micro drive until neuronal activity was encountered. Search stimuli included noise, tones, clicks, and other complex stimuli. Once a neuron was isolated using a time-amplitude window discriminator (BAK), the spatial tuning profile was determined for broadband noise at four different stimulus intensities (Woods et al., 2006). Vocalization stimuli were then presented from the speaker location that gave the best response (best location), which was usually in the contralateral hemisphere. This maximized the response for each individual neuron, although there is generally little difference in the response to broadband noise across a significant sector of contralateral space (Woods et al., 2006). The monkey was trained to depress a lever to start a trial. Three to seven different stimuli were played from the best location, and then a stimulus was presented from either straight ahead (if that was not the best location) or from directly opposite to the right ear (if that was not the best location). The animal was trained to release the lever to receive a fluid reward when it detected this change in location. The data from this report are restricted to stimuli that were presented from the best location. Monkey G was not able to learn this task and was given fluid rewards after the location change. All animals were continuously monitored via a closed circuit infrared camera system and remained alert throughout the recording session. During each experimental session, 49 other stimuli that were part of a different series of experiments were randomly interleaved with these 8 vocalization stimuli. The results from responses to those nonvocalization stimuli are not reported here. Each vocalization stimulus was presented on 8–12 trials.

At the conclusion of all recording experiments, each monkey was given an overdose of barbiturate and perfused through the heart with normal

Table 1. Number of neurons recorded

Area	Monkey F	Monkey G	Monkey L	Total
A1	48	104	79	231
CL	48	0	83	131
CM	45	51	53	149
ML	8	41	34	83
R	57	39	0	96
Total	206	235	249	690

saline followed by 4% paraformaldehyde in phosphate buffer. The brains were removed from the cranium, postfixed and cryo-protected. Coronal sections were cut at 25 μm on a freezing microtome and alternate sections were stained with thionin to reconstruct electrode tracks.

Data analysis. Neurons were classified into one of five different cortical areas based on their physiological response properties of frequency selectivity and latency, by the location within the recording cylinder and by the cytoarchitectonic appearance (Woods et al., 2006). Two different indexes were used to define the selectivity of the neuronal response. The monkey-call index at 50% maximum (MCI_{50}) was calculated as described previously (Tian et al., 2001). The stimulus that elicited the greatest response was considered the "best" stimulus, and the MCI_{50} was the number of stimuli that had a response within 50% of the response to the best stimulus. Forward and reversed calls were analyzed independently so this value is an integer from 1 to 4. A statistical evaluation was also conducted similar to that of Romanski et al. (2005), where the best stimulus was compared against the other three calls (t test with Bonferroni correction) and referred to as the MCI_t. In this case the MCI_t was the number of stimuli that had a response that was not significantly different from the response to the best stimulus ($p > 0.05$).

The linear pattern discriminator model was identical to that used by Russ et al. (2008). For each neuron, the first step was to select one spike train from one trial, referred to as the test trial. A stimulus PSTH was then constructed using all 12 trials for the other 7 stimuli and the remaining 11 trials for that particular stimulus. The next step was to calculate the Euclidian distance between a PSTH constructed from the test trial and each of the other 8 stimulus PSTHs. In the final step the stimulus PSTH with the smallest Euclidian distance was considered the discriminator's selection, and this choice was evaluated as either being correct (the PSTH for the stimulus that was presented) or not. This was repeated for each trial for each of the 8 stimuli, as well as independently for the four forward and four reversed calls. The average percentage correct for each neuron was calculated as the percentage of the total number of trials in which the discriminator selected the correct stimulus. This average was then pooled across neurons from a given area. The binwidth of the PSTHs reflects the amount of temporal information that is available to the discriminator, with small binwidths providing more temporal information than larger binwidths. All neurons within a given cortical area were tested using binwidths of 2, 5, 10, 25, 50, 100, 200, 400, 800, and 1500 ms. The 2 ms binwidth retains the most temporal information, whereas the 1500 ms bin is equivalent to the overall spike rate for that stimulus.

Results

Summary of responses

The results are based on recordings from 690 neurons in the left hemisphere of three monkeys (Table 1). The first step in the analysis was to determine the percentage of neurons that showed a statistically significant change in their response to any of the vocalizations tested as these cells were usually isolated based on their responses to much simpler stimuli (see Materials and Methods). The vast majority of neurons showed a statistically significant change in their response (stimulus response vs spontaneous activity; two-tailed t test with Bonferroni correction; $p < 0.05$), with only 8/690 neurons across all areas and monkeys without a significant response to any of the 8 stimuli tested. These responses were almost entirely excitatory (Tables 2, 3). Across the different

cortical areas, the forward and reversed bark calls elicited excitatory responses most often (97.6–99.0% of neurons) and inhibitory responses least often (0.0–1.2%) of the stimuli tested. The forward and reversed submissive call elicited excitatory responses least often (87.9–95.2%) and inhibitory responses most often (3.5–10.8%) of the calls tested. The coo and grunt calls were within these extremes. The percentage of excitatory and inhibitory responses across the eight stimulus types was not statistically significantly different between cortical areas (paired t test; $p > 0.05$ with Bonferroni correction). Thus, the vast majority of auditory cortical neurons were significantly responsive to these forward and reversed vocalizations regardless of the specific cortical area tested.

Selectivity of responses across vocalizations

The next level of analysis was to determine whether neurons in any of these cortical areas selectively processed specific vocalizations. The responses to each of the 8 stimuli from two different A1 neurons are shown in Figures 1 and 2. The top panel shows the sonogram of the stimulus, with frequencies from 0 to 20 kHz along the y-axis (bottom to top) with the warmer colors showing increasing amounts of power. Below each sonogram are the dot rasters, showing the response of the neuron to each presentation of that stimulus. Each dot represents a single action potential and each row shows a single trial. The poststimulus time histogram is shown below the rasters (bin width 5 ms). In the figure, the time axes were all scaled to be approximately the same size although the vocalizations themselves spanned a temporal range of ~100–1500 ms (note the time scale on the x-axis). The neuron in Figure 1 was typical of the sample in that there was a good excitatory response to each of the different stimuli, and relatively little difference between forward and reversed vocalizations (compare A–D, E–H). In this example, the greatest difference in the responses between vocalizations was for the submissive call, where the overall firing rate was equivalent but the temporal structure of the response clearly differed. This was more than a simple time reversal of the responses, as there are three distinct periods of the response to the forward call, and a broader response to the reversed call. The neuron in Figure 2 showed a different response, with a clearly better response to the coo stimuli (B and F) compared with the other stimuli. There also appears to be some differences between forward and reversed calls, although this difference was again largely in the temporal envelope of the response and not in total activity.

Previous studies have used a vocalization index to measure the selectivity of the neurons to the different calls (Tian et al., 2001; Russ et al., 2008). This metric (MCI_{50}; see Materials and Methods) is the number of vocalizations that elicit a response at least 50% of the maximum response. Analysis of the forward and reversed calls separately showed that, on average, the number of vocalizations that elicited at least 50% of the maximum firing rate was fairly equivalent across cortical areas (Fig. 3A). Area R had the lowest percentage of neurons that responded to all four forward vocalizations within 50% of the peak response, whereas area CM had the most (was least selective) (Fig. 3A). The other three cortical areas were equivalent with respect to the percentage that

had a vocalization index of 4. When the reversed calls were tested, area R again had the lowest percentage with a vocalization index of 4, but it was higher than with the forward vocalizations (Fig. 3B).

An alternative metric is to test whether the greatest firing rate to a particular vocalization is statistically significantly greater than the response to other vocalizations. This was tested using a t test (with Bonferroni correction) between the call that gave the highest firing rate to each of the other three calls, with the forward and reversed calls analyzed separately (MCI_t; see Materials and Methods). These results are shown in Figure 4. This metric showed much more selectivity by the neurons compared with the MCI_{50}. For example, almost half of the neurons in each area had a significantly greater firing rate to the best call compared with the other three calls (Fig. 4, left column), and only ~10% of neurons had the same firing rate to all four calls as they did to the best call (far right column). However, there was no difference across cortical areas in the percentage of neurons that have a statistically significantly similar response to other calls compared with the call that elicited the greatest response. Thus, while neurons within both core (A1 and R) and belt (CM, CL and ML) areas can be quite selective for particular calls using a statistically based metric, the samples of neurons recorded across each of these five areas were not significantly more or less selective.

Forward versus reversed call responses

The next consideration was how the firing rate to the forward calls compared with the firing rate to the same calls time-reversed. Temporally reversing the call preserves the spectral complexity of the stimulus; however, temporally reversed calls do induce different behavioral responses if they are temporally asymmetrical (Ghazanfar et al., 2001) such as the bark, grunt, and submissive calls used here, but not if they are temporally symmetrical (Le Prell and Moody, 2000), similar to the coo call used here. It was reasoned that a neuron that was vocalization-selective would have a different response to the naturally occurring forward vocalizations compared with the same call but presented reversed in time. This was tested by comparing the overall firing rate from the forward vocalization to that from the reversed vocalization (with Bonferroni correction). Table 4 shows the results from this analysis across neurons from each cortical area. The first finding to note was that none of the tested neurons showed a statistically significant difference in the response between the for-
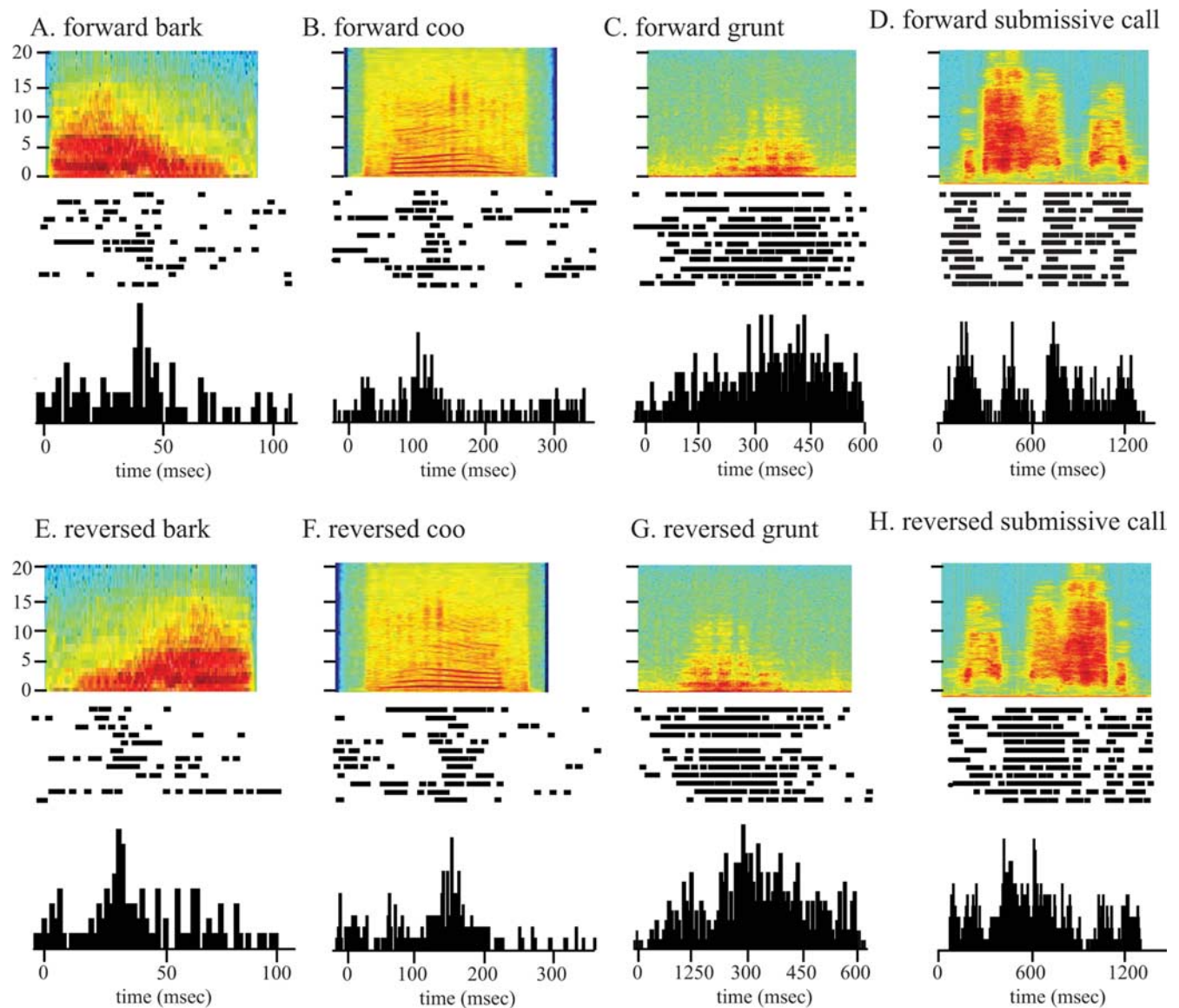
Table 2. Percentage of statistically significantly responsive neurons (excitatory responses)

Area	Forward calls				Reversed calls			
	Bark	Coo	Grunt	Subm.	Bark	Coo	Grunt	Subm.
A1	98.3	96.1	97.0	94.4	98.7	96.5	96.1	95.2
CL	98.5	91.6	96.9	88.5	98.5	94.7	92.4	89.3
CM	98.0	92.6	91.9	88.6	98.0	91.3	92.6	87.9
ML	97.6	86.7	94.0	88.0	97.6	91.6	90.4	90.4
R	99.0	92.7	97.9	90.6	99.0	92.7	97.9	93.8

Subm., Submissive.

Table 3. Percentage of statistically significantly responsive neurons (inhibitory responses)

Area	Forward calls				Reversed calls			
	Bark	Coo	Grunt	Subm.	Bark	Coo	Grunt	Subm.
A1	0.4	2.6	1.8	4.4	0.0	2.2	2.6	3.5
CL	0.0	6.9	1.5	9.9	0.0	3.8	6.1	9.2
CM	0.0	5.4	6.0	9.4	0.0	6.7	5.4	10.1
ML	1.2	12.0	4.8	10.8	1.2	7.2	8.4	8.4
R	0.0	6.3	1.0	8.3	0.0	6.3	1.0	5.2

Subm., Submissive.

Figure 1. Responses to the eight different stimuli by a representative A1 neuron. In each panel, the top shows the sonogram of the stimulus with the *y*-axis denoting the frequency from 0 Hz (bottom) to 20 kHz (top) on a linear scale. Time is on the *x*-axis, and the color corresponds to the energy of the stimulus with warmer colors indicating greater power. Below the sonograms are the spike rasters. Each line is a different trial (top line is the first trial of that stimulus type) and each tic mark represents a single spike. Below this is the poststimulus time histogram (PSTH). Binwidth is 5 ms. ***A–D*** show forward calls, ***E–H*** are the time-reversed calls. This neuron shows little difference in the spike rate between forward and reversed calls, but some indication of a difference in the temporal response.

ward and reversed presentation of three or all four of the four calls presented. Thus, we did not find any example of a neuron that was selective for forward calls over reversed calls (or reversed over forward) that generalized across either three or all four call types used here. While a fair percentage of neurons did respond differentially to the forward and reversed presentation of at least one call (10.7–14.8%; column 3), a much smaller proportion responded differentially between the forward and reversed version of two of the calls (0.8–6.0%; column 5). The total percentage of neurons that did respond differently between the forward and reversed calls was relatively modest, with area CL showing the smallest percentage of neurons (11.5%) compared with areas A1 (14.3%) and CM (16.1%) with a greater proportion of ML (19.3%) and R neurons (20.5%). These data suggest that there is a slight preference for encoding these spectrally complex stimuli in areas R and ML compared with the other cortical areas tested.

This analysis does not, however, test whether the neurons are specifically tuned to the conspecific vocalizations. Table 5 takes the cases from Table 4 where there was a statistically significantly different response between the forward and reversed calls and differentiates between cases where the forward vocalization elicited a greater response than the reversed vocalization. The forward bark call most reliably elicited the greater response compared with the reversed call (73.1%; χ^2, df = 1, p = 0.019), whereas there was not a clear distinction for the other calls (all p values >0.05). Interestingly, only neurons in area CL had a greater response to forward calls that was statistically significantly different from chance (χ^2; df = 1, p = 0.0124), although it also had the lowest percentage of neurons that showed a significant difference between forward and reversed calls (15 neurons and 16 cases) (Table 4). This is in stark contrast to what would be predicted from the dual stream hypothesis, where neurons in ML or potentially area R should be the most vocalization selective, and areas CL and CM the least selective.
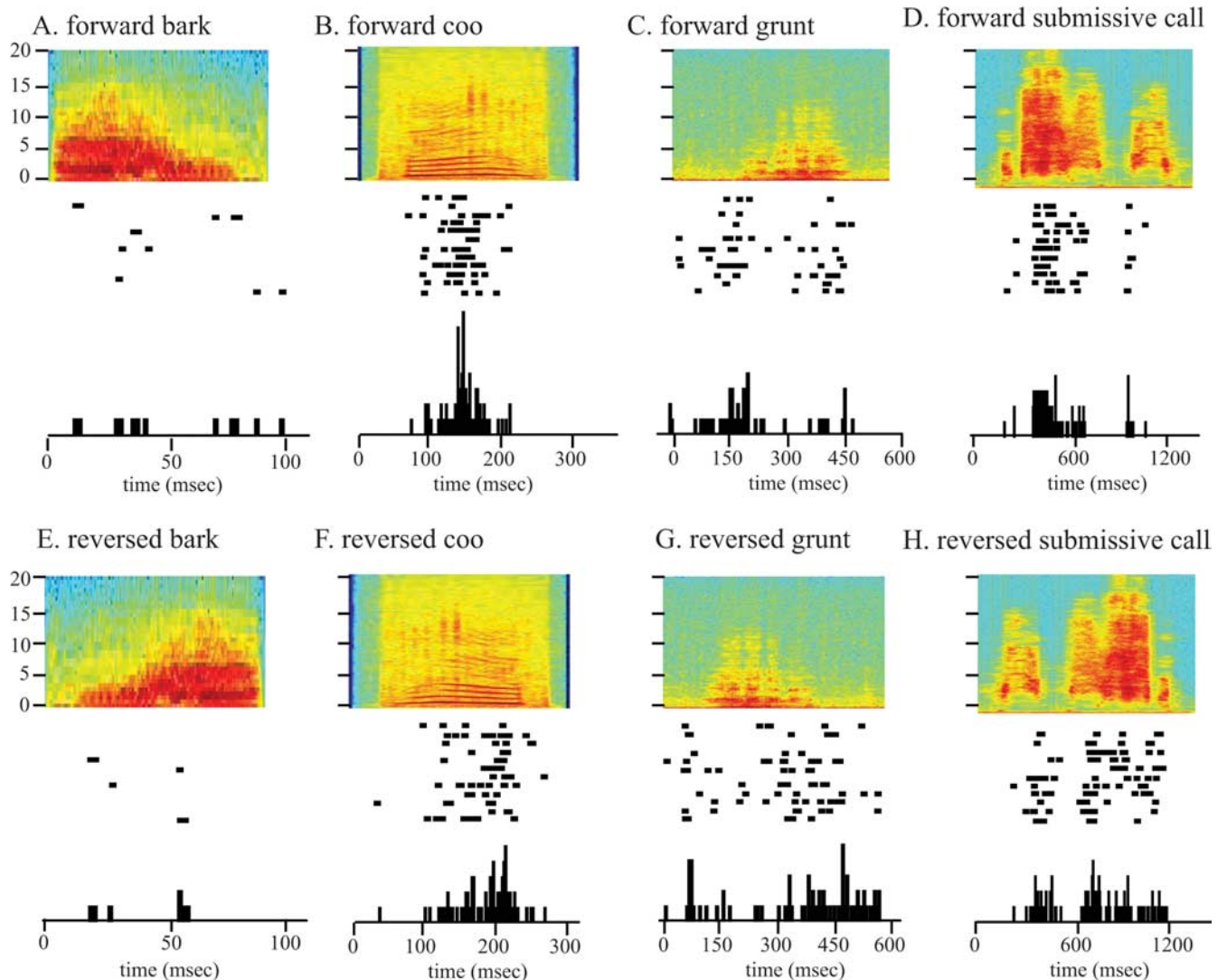
Figure 2. Responses from a second example cell. Conventions as in Figure 1. This neuron responded best to the forward and reversed coo stimuli (**B** and **F**).

While the preceding analysis revealed modest differences between the responses to forward and reversed calls by individual neurons, it is possible that one area may have the same percentage of tuned neurons, but with overall lower firing rates. At the population level, therefore, the overall signal from one area could be considerably smaller than that of another area. A second possibility to consider is whether, across the population, there is a consistently greater response to the forward calls that does not reach statistical significance when tested on the single neuron. To test these possibilities, the mean firing rates across all forward calls and all reversed calls was calculated for each neuron. These averages are shown in Figure 5. Statistical analysis showed that the firing rates of neurons in A1 and CL were not statistically significantly different from each other, but were different from neurons in areas CM, ML and R (t test, $p < 0.01$ with Bonferroni correction). In addition, firing rates for neurons in CM, ML and R were not different from each other ($p > 0.01$). However, these differences in firing rates were relatively modest, with means of ~29 spikes/s for A1 and CL compared with ~24 spikes/s for the other three areas, with SDs of 23–26 spikes/s and 18–23 spikes/s, respectively. This indicates that these two areas could

have relatively larger signals to vocalization stimuli compared with the other areas by ~20%. Analysis of the differences between forward and reversed calls within a cortical area indicated that there was a statistically significant difference for neurons in areas CL and ML (paired t test; $p < 0.01$ with Bonferroni correction). This was somewhat surprising given the nearly identical firing rates and the rather large error bars for neurons in these two areas (Fig. 5A). Regression analysis indicated that the slope of the regression line was near 1.0 (1.01 and 1.08 for CL and ML, respectively) and the intercept of the regression line was near zero (0.70 and 1.84 for CL and ML, respectively) (Fig. 5B,C). This indicates that, while there is a statistically significant effect, it is quite small and likely due to the very large number of comparisons within the t test (8 times the number of neurons, or 1048 and 684 for CL and ML, respectively). Together, these analyses of firing rates indicate that there is little difference between forward and reversed calls across the population of neurons in any of these cortical areas tested, consistent with fMRI results (Petkov et al., 2008). The differences that were seen were small and were not consistent with the hypothesis that rostral auditory cortical areas are more selective for forward over reversed vocalizations compared with more caudal cortical areas.
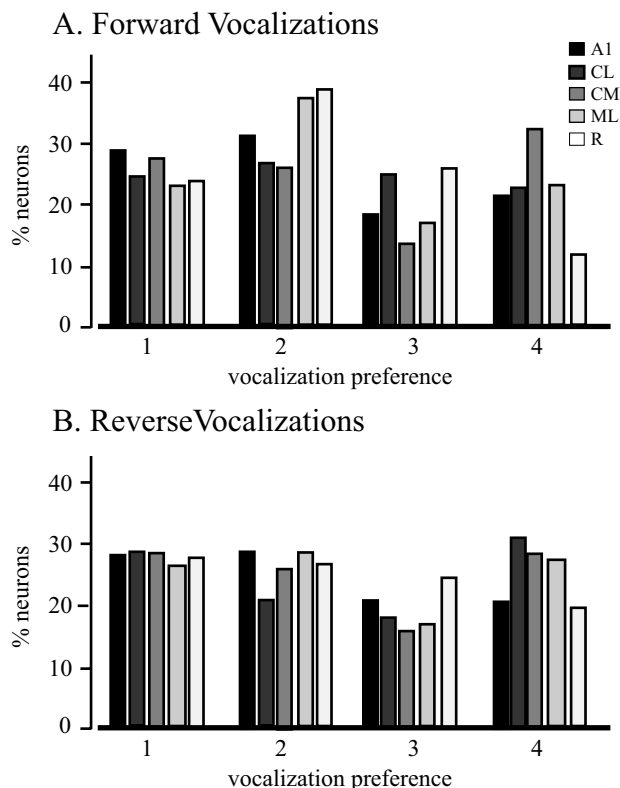
A. Forward Vocalizations

B. ReverseVocalizations

Figure 3. MCI$_{50}$ across cortical areas. **A** shows the monkey call index calculated as the number of forward calls where the response was at least 50% of the best call. Shading corresponds to the different cortical areas (see inset). **B** shows the monkey call index for the reversed calls. There is little difference between forward and reversed calls, and little difference between cortical areas.
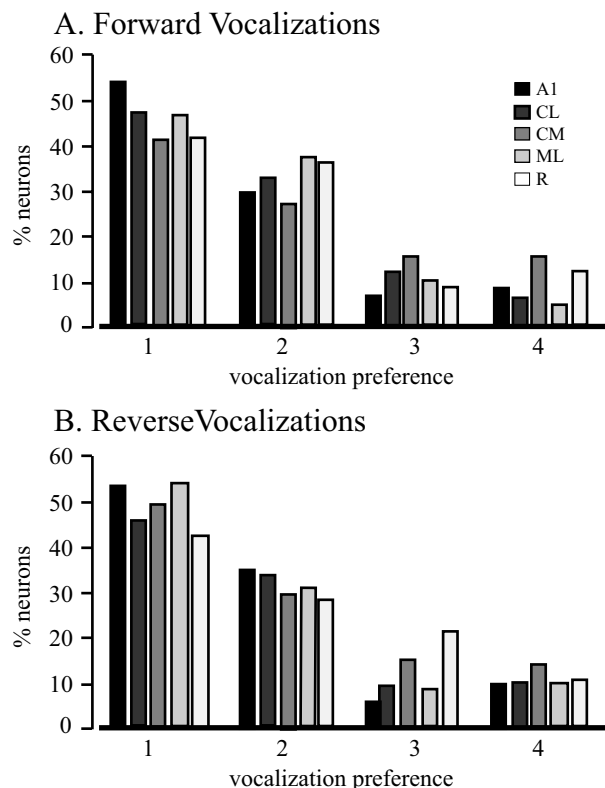
A. Forward Vocalizations

B. ReverseVocalizations

Figure 4. MCI$_t$ across cortical areas. The MCI$_t$ is the number of calls that were not statistically significantly different from the call that gave the best response. Data analysis was restricted to forward (**A**) and reversed (**B**) calls. This metric showed much more selectivity; however, there was little difference between cortical areas or between forward and reversed calls.

Linear pattern discriminator performance

While firing rate comparisons are common in this class of study, they do not test whether differences in the responses could be used to discriminate between the different vocalizations. As can be seen from Figure 1, although approximately the same number of spikes can be elicited by the forward and reversed submissive call stimuli, there can be a clear difference in the temporal pattern of the response. A recent study comparing neurons in the auditory cortex to those in the ventral prefrontal cortex (Russ et al., 2008) used a linear pattern discriminator model to determine how much the temporal features of the response could differentiate between different calls. To determine the extent that neurons in different areas could use the pattern information in discriminating between calls, we applied the same analysis (see Materials and Methods). We used the entire stimulus duration and varied the bin size of the PSTHs generated for each trial and neuron. The percentage of times neurons in each cortical area were able to accurately discriminate these eight different vocalizations, measured by the smallest euclidean distance between the single trial and the PSTH for each of the 8 stimuli, is shown in Figure 6 using a bin width of 2 ms, which was the smallest that we tested. The mean accuracy across all neurons within a particular cortical area was equivalent, and ranged from 89.2 to 92.1% in areas R and ML, respectively. These small differences were not statistically significantly different (unpaired t test; all p values >0.05 without Bonferroni correction).

The preceding analysis was focused on a small binwidth (2 ms) and the entire vocalization. The high rate of accuracy and the lack of a difference between neurons in the different cortical areas may

be due to a ceiling effect, as many neurons were perfectly accurate in each cortical area. The Russ et al. (2008) study found that as the bin size increased (temporal information was lost), the accuracy of the discriminator decreased. This indicates that the different stimuli were better discriminated by the temporal pattern of the response as opposed to the absolute firing rate. When the accuracy of the discriminator was tested as a function of the bin size, a similar finding was observed for the neurons in the five cortical areas studied here (Fig. 7A). As with the Russ et al. (2008) study, the greatest accuracy was when the bin size was the smallest (2 ms), and this value decreased and found an asymptote at ∼800 ms. This corresponds to the firing rate alone for all stimuli except for the forward and reversed submissive call. There was no difference in the accuracy for discrimination between cortical areas until a bin size of 50 ms was used, and from that point and all greater bin sizes, the accuracy of area R neurons was significantly better than the other four areas, which were not significantly different from each other (error bars are only shown for one cortical area for clarity). This indicates that the overall firing rate of area R neurons contains more information about the stimulus type than the firing rate of neurons in other cortical areas.

To determine whether this was due to greater accuracy for forward or reversed calls, the same analysis was done but restricted to the four forward and four reversed calls. The results for the forward calls are shown in Figure 7B. Again, there is a decrease in accuracy with an increase in bin size, but this effect was equivalent across all cortical areas. In contrast, when the reversed calls were tested (Fig. 7C), area R neurons were more accurate at discriminating the reversed calls than the forward calls, whereas neurons from the other four areas were less accurate when larger

bin sizes were used, and indeed fell to chance levels. These differences between the accuracy of the discriminator for forward versus reversed calls is highlighted by the difference functions (forward percentage correct minus reverse percentage correct) shown in Figure 7D. For all areas except R, the discriminator was ~20% more accurate for forward vocalizations compared with reversed vocalizations, whereas in area R the accuracy of the discriminator was actually better for reversed calls compared with forward calls. This shows that the difference between area R and the other cortical areas tested with respect to all calls (Fig. 7A) is entirely due to the greater accuracy at discriminating reversed calls. Interestingly, neurons in all other cortical areas fell to chance for the reversed calls with bin sizes greater than ~400 ms, whereas area R neurons remained significantly above chance when only the firing rate was available. This indicates that, for reversed calls in area R, the temporal features of the response accounts for only about one third of the discrimination accuracy. In contrast, the temporal patterns of the response account for virtually all of the accuracy of the discriminator for the reversed calls in the other four cortical areas. For forward calls, the temporal aspect of the response accounts for more than half of the overall discrimination accuracy in all cortical areas. It should be noted, however, that for bin sizes of 25 ms or less, there was no difference in accuracy between cortical areas or between forward and reversed calls.

Discussion

This report details the responses of single neurons in the core and belt areas of auditory cortex in alert macaque monkeys to vocalization stimuli. These results are consistent with early studies in awake squirrel monkeys in that the vast majority of neurons were responsive to these stimuli (Winter and Funkenstein, 1973; Manley and Müller-Preus, 1978; Glass and Wollberg, 1979). We have extended those findings in the alert macaque monkey, as well as tested five different cortical areas along the putative spatial and nonspatial processing streams. The first question addressed was whether vocalization stimuli are selectively processed in different cortical areas. No selectivity was found between the five cortical areas tested using standard metrics based on the overall firing rate (Figs. 3, 4). However, a temporal pattern discriminator model did reveal that neurons in all areas carried enough information to discriminate between these vocalizations above chance levels based on firing rate alone (Fig. 7). A second question that was addressed was whether neurons in any of these cortical areas selectivity respond to vocalization stimuli. This was addressed by comparing responses to forward and the equally acoustically complex reversed vocalizations. Previous studies have indicated that reversed vocalizations are perceived differently than forward vocalizations, and thus they are likely largely behaviorally irrelevant (Ghazanfar et al., 2001). Using traditional spike rate methods, neurons in lateral belt area CL had higher firing rates to forward vocalizations compared with reversed vocalizations (Table 5), although neurons in this area were least likely to have a difference in activity between the two types of stimuli (Table 4). This result is tempered by the population analysis, where there was a small but statistically significant smaller response to forward compared with reversed vocalizations in both CL and ML (Fig. 5). The regression analysis indicated that this difference was very small, and could be biologically irrelevant given that the slope was near 1.0 and intercept was <2 spikes/s. However, again

Table 4. Neurons with statistically significant difference between forward and reverse calls

Area	One call	%	Two calls	%	Total neurons	%
A1	26	11.3	7	3.0	33	14.3
CL	14	10.7	1	0.8	15	11.5
CM	22	14.8	2	1.3	24	16.1
ML	11	13.3	4	4.8	16	19.3
R	12	14.5	5	6.0	17	20.5

the temporal pattern discriminator model showed that forward vocalizations were better discriminated than reversed vocalizations by populations of neurons in all areas except area R. This indicates that, at this level of A1 and belt auditory cortex, vocalizations are not selectively processed based on overall firing rate but could be differentiated based on more complex processing.

These results are not in support of the dual stream hypothesis, in contrast to previous studies investigating spatial response properties where caudal areas have sharper spatial tuning compared with core or rostral belt areas (Recanzone et al., 2000b; Woods et al., 2006). Results from nonspatial studies are mixed, with ventral prefrontal neurons (the targets of the rostral processing stream) showing better responses to different vocalizations compared with rostral auditory belt areas (Romanski et al., 2005). However, ventral prefrontal neurons have also been shown to have slightly worse preference for vocalization compared with anterior belt fields (Cohen et al., 2004; Russ et al., 2008) and also good spatial responses, similar to findings in the lateral intraparietal area (LIP) where both spatial and nonspatial responses were noted (Gifford and Cohen, 2005). The strongest evidence comes from Tian et al. (2001) where neurons in rostral belt areas had smaller vocalization indexes compared with neurons in caudal belt areas. There are several factors that can account for these apparent discrepancies. The first may be due to the areas that were studied, as neurons in area AL have been shown to have a greater response to vocalizations than neurons in ML and CL (Tian et al., 2001; Russ et al., 2008). It was unfortunate that AL was not accessible in these animals. The dual-processing hypothesis predicts that ML neurons have a greater selectivity than CL and CM neurons, as well as core fields A1 and potentially R, and this was not clearly the case. The second is that some previous studies were done in anesthetized animals in acute experiments, thus anesthetic effects may have influenced the neuronal selectivity, although previous studies indicate that such effects are not strongly directional (Benson et al., 1981). The third is that most previous studies used more call exemplars than the four used in this study. These calls were the most often uttered in the vivarium in which these animals were housed and were therefore very familiar and presumably most behaviorally relevant. Nonetheless, presenting more exemplars may have revealed a greater level of selectivity.
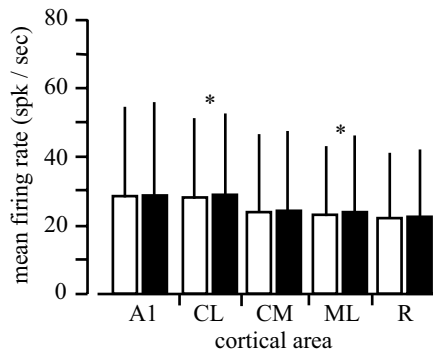
Finally, the difference that most likely explains the discrepancy between studies is that all but one single neuron study relied on forward vocalizations and did not compare the responses with other complex stimuli. The single neuron study that manipulated the calls only did so in a small subset of studied neurons (Tian et al., 2001). They found that the response was greatest to the entire call and was reduced when segments were deleted or replaced by broadband noise. Thus, it seems likely that all of the neurons across these different studies responded well to complex auditory stimuli, but not selectively to vocalizations. This finding is consistent with the imaging data. Poremba et al. (2004) showed that vocalization stimuli activated regions throughout the superior

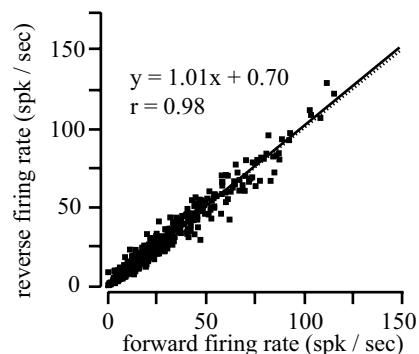**Table 5. Neurons with a greater response to forward versus reversed calls**

Area	Bark	%	Coo	%	Grunt	%	Subm.	%	Total	%
A1	8/10	80.0	2/5	40.0	5/11	45.5	8/14	57.1	23/40	57.5
CL	2/2	100	3/3	100.0	2/2	100.0	6/9	66.7	13/16	81.3
CM	5/5	100	3/6	50.0	2/4	50.0	6/11	54.5	16/26	61.5
ML	1/4	25.0	0/4	0.0	2/5	40.0	3/6	50.0	6/19	31.6
R	3/5	60.0	4/7	57.1	1/4	25.0	6/6	100	14/22	63.6
Tot.	19/26	73.1	12/21	57.2	12/26	46.2	29/46	63.0	72/119	60.5

Tot., Total; Subm., submissive.

A. Population mean firing rates

B. Area CL neurons

$y = 1.01x + 0.70$
$r = 0.98$

C. Area ML neurons

$y = 1.08x + 1.84$
$r = 0.97$

Figure 6. Results from the linear pattern discriminator model. Each bar shows the average percentage correct for neurons in each cortical area using the entire stimulus duration and the smallest PSTH binwidth (2 ms). There is no difference between neurons in the different cortical areas.
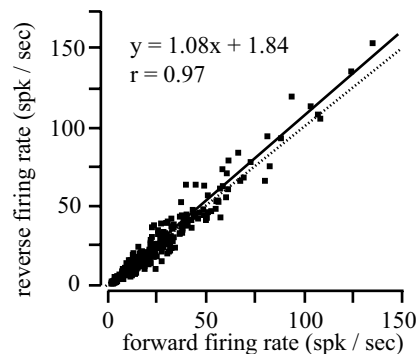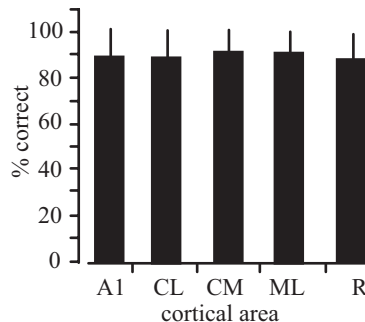
Figure 5. Overall mean firing rate across cortical areas for forward and reversed calls. ***A***, The mean and SD are shown for forward (white bars) and reversed (black bars) calls. Neurons in areas A1 and CL had the highest firing rates, which were not different from each other, but were different from areas CM, ML and R, which were not different from each other. Comparisons between the forward and reversed calls within a cortical area revealed a statistically significant difference for neurons in areas CL and ML (asterisks), where there was a greater firing rate to reversed calls compared with forward calls. ***B***, Regression plot of the firing rate to the forward call (*x*-axis) and reversed call (*y*-axis) for each of the four calls for all CL neurons. The dashed bar shows the line of unity, the solid bar shows the regression line (equation in the inset). ***C***, Regression plot for ML neurons. Conventions as in ***B***.

temporal gyrus as far rostral as the temporal pole. A recent fMRI study also showed an enhanced BOLD response for vocalization stimuli compared with temporally "jumbled" vocalization and other naturalistic sounds in the anterior portion of the superior temporal sulcus (Petkov et al., 2008). They also showed an equivalent BOLD response to natural vocalizations and complex nonvocalization stimuli in core and belt areas, consistent with what was observed in this study at the single neuron level (Fig. 5).

It should be stressed that the dual stream hypothesis cannot be rejected based on these data, however. It may be that vocalization stimuli themselves are inappropriate to differentiate between these two putative processing streams, and stimuli that specifically test some other complex feature such as temporal integration (Bendor and Wang, 2008) could yield results in support of the dual stream model. Vocalizations are spectrally and temporally rich, as well as behaviorally important, and these complex stimulus features may also provide important spatial cues, giving rise to robust responses in caudal belt areas.

This study is also consistent with a previous study in the auditory cortex where a linear pattern discriminator model was tested (Russ et al., 2008). In that study, the average neuron contained sufficient information to accurately predict the call ~80% of the time using the smallest binwidth and largest sample duration. This is consistent with the current study that found ~90% accuracy across the 8 different stimuli. The slight differences could be due to the smaller number of calls used in this study, the variability of the call structures used between the two studies, and to the fact that the Russ et al. (2008) study also used only forward calls. One interesting finding was that forward vocalizations were better discriminated than reversed vocalizations when only spike rate information was used in all areas except R. This indicates that there is information carried in the firing rate alone in neurons in all five cortical areas. Interestingly, for all cortical areas except area R, the linear pattern discriminator model fell to chance when
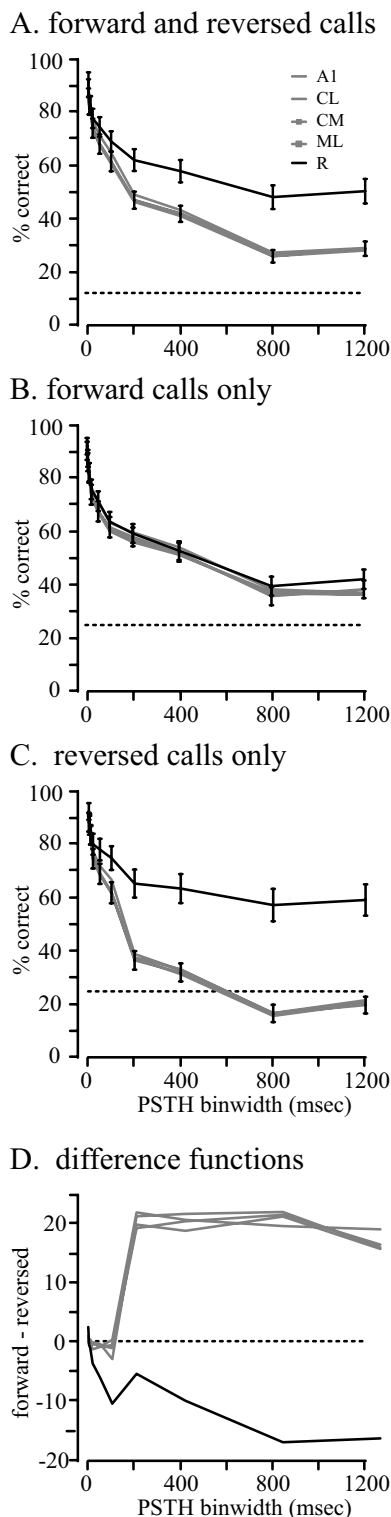
A. forward and reversed calls

B. forward calls only

C. reversed calls only

D. difference functions

Figure 7. Effects of binwidth on the linear discriminator model. *A* shows the average percentage correct as a function of PSTH binwidth using the entire stimulus duration. The 2 ms data are the same shown in Figure 6*A*. The performance of the model decreases as stimulus binwidth increases and temporal information is lost. Neurons in all cortical areas (gray lines) except for area R had nearly identical functions, whereas area R neurons showed better model performance (black line). Chance (12.5%) is shown as the dashed line. *B* shows the results when only forward vocalizations are considered. There is no difference between neurons in the five different cortical areas in this case. *C* shows the results when only the reversed vocalizations are considered. As when the entire stimulus is used, neurons in all cortical areas except for area R dropped dramatically as the binwidth increased, and fell to chance after 400 ms. Neurons in area R remained well above chance, indicating that firing rate alone accounted for much of the accuracy at the smaller binwidth values. *D* shows the difference functions (forward minus reverse) for each of the cortical areas at each binwidth.

using the spike rate alone for the reversed vocalizations. This indicates that these four cortical areas are better at encoding forward vocalizations than reversed vocalizations, but not by selectively responding to a particular vocalization type. The differences in area R neurons from those of other cortical areas is unlikely to reflect selectivity for reversed vocalizations, which never occur naturally, and are likely related to the differences in response properties of area R neurons compared with neurons in the other cortical areas tested. Neurons in area R have been little studied, and rigorous quantification of neurons in this area in comparison to others indicates that they have relatively large spatial receptive fields (Woods et al., 2006), longer minimum latency, more non-monotonic rate/level functions, and sharper frequency tuning compared with A1, CL, ML and CM (Recanzone et al., 2000a; Bendor and Wang, 2008). This feature selectivity could be at the root of the performance of the model, which may be revealed in future experiments.

In summary, this study investigated the response properties of single neurons in five different auditory cortical areas to forward and reversed vocalizations in the alert macaque monkey. Traditional measures of firing rate did not reveal any selectivity for the call type, consistent with recent fMRI studies. This result indicates that, while vocalizations are behaviorally relevant and likely ultimately processed more selectively, it is almost certainly not done in these early (core and belt) auditory cortical areas. Additionally, there is no compelling evidence for selectivity between cortical areas using any of the metric employed in this study. Future studies will be necessary to probe what stimulus features the neurons in these cortical areas are selectively processing.

References

Bendor D, Wang X (2008) Neural response properties of primary, rostral, and trostrotemporal core fields in the auditory cortex of marmoset monkeys. J Neurophysiol 100:888–906.

Benson DA, Hienz RD, Goldstein MH Jr (1981) Single-unit activity in the auditory cortex of monkeys actively localizing sound sources: spatial tuning and behavioral dependency. Brain Res 219:249–267.

Cohen YE, Russ BE, Gifford GW 3rd, Kiringoda R, MacLean KA (2004) Selectivity for the spatial and nonspatial attributes of auditory stimuli in the ventrolateral prefrontal cortex. J Neurosci 24:11307–11316.

Ghazanfar AA, Smith-Rohrberg D, Hauser MD (2001) The role of temporal cues in rhesus monkey vocal recognition: orienting asymmetries to reversed calls. Brain Beh Evol 58:163–172.

Gifford GW 3rd, Cohen YE (2005) Spatial and non-spatial auditory processing in the lateral intraparietal area. Exp Brain Res 162:509–512.

Glass I, Wollberg Z (1979) Lability in the responses of cells in the auditory cortex of squirrel monkeys to species-specific vocalizations. Exp Brain Res 34:489–498.

Harrington IA, Heffner RS, Heffner HE (2001) An investigation of sensory deficits underlying the aphasia-like behavior of macaques with auditory cortex lesions. Neuroreport 12:1217–1221.

Heffner HE, Heffner RS (1989) Effect of restricted cortical lesions on absolute thresholds and aphasia-like deficits in Japanese macaques. Behav Neurosci 103:158–169.

Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. Proc Natl Acad Sci U S A 97:11793–11799.

Le Prell CG, Moody DB (2000) Factors influencing the salience of temporal cues in the discrimination of synthetic Japanese monkey (*Macaca fuscata*) coo calls. J Exp Psychol Anim Behav Proc 26:261–273.

Manley JA, Müller-Preus P (1978) Response variability of auditory cortex cells in the squirrel monkey to constant acoustic stimuli. Exp Brain Res 32:171–180.

Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK (2008) A voice region in the monkey brain. Nat Neurosci 11:367–374.

Poremba A, Malloy M, Saunders RC, Carson RE, Herscovitch P, Mishkin M (2004) Species-specific calls evoke asymmetric activity in the monkey's temporal poles. Nature 427:448–451.

Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci U S A 97:11800 –11806.

Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268:111–114.

Recanzone GH, Guard DC, Phan ML (2000a) Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. J Neurophysiol 83:2315–2331.

Recanzone GH, Guard DC, Phan ML, Su TK (2000b) Correlation between the activity of single auditory cortical neurons and sound localization behavior in the macaque monkey. J Neurophysiol 83:2723–2739.

Romanski LM, Averbeck BB, Diltz M (2005) Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. J Neurophysiol 93:734 –747.

Russ BE, Ackelson AL, Baker AE, Cohen YE (2008) Coding of auditory-stimulus identity in the auditory non-spatial processing stream. J Neurophysiol 99:87–95.

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. Science 292:290 –293.

Winter P, Funkenstein HH (1973) The effect of species-specific vocalization on the discharge of auditory cortical cells in the awake squirrel monkey. Exp Brain Res 18:489 –504.

Woods TM, Lopez SE, Long JH, Rahman JE, Recanzone GH (2006) Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey. J Neurophysiol 96:3323–3337.