Published in final edited form as:

Hear Res. 2013 November ; 305: 135–143. doi:10.1016/j.heares.2013.07.011.

Coding of Vocalizations by Single Neurons in Ventrolateral Prefrontal Cortex

Bethany Plakke, Mark D. Diltz, and Lizabeth M. Romanski

Dept. Neurobiology & Anatomy, Univ. of Rochester, Box 603, Rochester, NY 14642

Abstract

Neuronal activity in single prefrontal neurons has been correlated with behavioral responses, rules, task variables and stimulus features. In the non-human primate, neurons recorded in ventrolateral prefrontal cortex (VLPFC) have been found to respond to species-specific vocalizations. Previous studies have found multisensory neurons which respond to simultaneously presented faces and vocalizations in this region. Behavioral data suggests that face and vocal information are inextricably linked in animals and humans and therefore may also be tightly linked in the coding of communication calls in prefrontal neurons. In this study we therefore examined the role of VLPFC in encoding vocalization call type information. Specifically, we examined previously recorded single unit responses from the VLPFC in awake, behaving rhesus macaques in response to 3 types of species-specific vocalizations made by 3 individual callers. Analysis of responses by vocalization call type and caller identity showed that ~ 19 % of cells had a main effect of call type with fewer cells encoding caller. Classification performance of VLPFC neurons was ~ 42% averaged across the population. When assessed at discrete time bins, classification performance reached 70 percent for coos in the first 300 ms and remained above chance for the duration of the response period, though performance was lower for other call types. In light of the sub-optimal classification performance of the majority of VLPFC neurons when only vocal information is present, and the recent evidence that most VLPFC neurons are multisensory, the potential enhancement of classification with the addition of accompanying face information is discussed and additional studies recommended. Behavioral and neuronal evidence has shown a considerable benefit in recognition and memory performance when faces and voices are presented simultaneously. In the natural environment both facial and vocalization information is present simultaneously and neural systems no doubt evolved to integrate multisensory stimuli during recognition.

Keywords

Auditory; Communication; Frontal Lobe; Macaque; Monkey; Language

Previous work has identified an auditory region in the ventral lateral prefrontal cortex (VLPFC) that is responsive to complex sounds including species-specific vocalizations (Romanski and Goldman-Rakic, 2002). Species-specific vocalizations are complex sound

© 2013 Elsevier B.V. All rights reserved.

Please address all correspondence to: Lizabeth M. Romanski, Ph.D., University of Rochester Medical Center, Dept. of Neurobiology and Anatomy, Box 603, 601 Elmwood Ave., Rochester, NY 14642, liz_romanski@urmc.rochester.edu, Phone: 585-273-1469, Fax: 585-442-8766.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

stimuli with varying temporal and spectral features which can provide unique information to the listener. The type of call delivered can indicate different behavioral contexts (food call vs. alarm call) which each elicit a unique response. Vocalizations can also provide information about the individual uttering the call, and thus would include information on gender, social status, body size, and reproductive status (Hauser and Marler, 1993; Hauser, 1996; Bradbury and Vehrencamp, 1998; Owings and Morton, 1998). Knowing how the brain encodes species-specific vocalizations can contribute to our understanding of language and communication processing.

Neurophysiological studies have begun to examine the specific acoustic features which neurons in the auditory cortex encode. Work by Wang and colleague's has demonstrated that neurons in marmoset auditory cortex lock to the temporal envelope of natural marmoset twitter calls (Wang et al., 1995). In addition, it has been shown that the auditory cortex in marmosets has a specialized pitch processing region (Bendor and Wang, 2005) and that marmosets use both temporal and spectral cues to discriminate pitch (Bendor et al., 2012). Neurons in primary auditory cortex that are not responsive to pure tones are highly selective to complex features of sounds, features that are found in vocalizations (Sadagopan and Wang 2009).

Other key areas involved in vocalization processing include the belt and parabelt regions of auditory cortex. While neurons in auditory core areas are responsive to simple stimuli, complex sounds including noise bursts and vocalizations activate regions of the belt and parabelt (Rauschecker et al., 1995; Rauschecker, 1998; Rauschecker et al., 1997; Kikuchi et al., 2010). Recently, it has been reported that the left belt and parabelt are more active to complex spectral temporal patterns (Joly et al., 2012). Furthermore, the selectivity of single neurons in the anterolateral belt for vocalizations is similar to that of VLPFC (Romanski et al., 2005; Tian et al., 2001) which are reciprocally connected (Romanski et al., 1999a,b).

This hierarchy of complex sound processing continues along the superior temporal plane (STP) (Kikuchi et al., 2010; Poremba et al., 2003) to the temporal pole (Poremba et al., 2004). An area on the supratemporal plane has also been identified as a vocalization area in macaque monkeys (Petkov et al., 2008) and cells in this region are more selective to individual voices than call type (Perrodin et al., 2011). These auditory regions including the belt, parabelt, and STP all project to VLPFC (Hackett et al., 1999; Romanski et al., 1999), which is presumed to be the apex of complex auditory processing in the brain.

Previous work has shown that VLPFC neurons are preferentially driven by species-specific vocalizations compared to pure tones, noise bursts and other complex sounds (Romanski and Goldman-Rakic, 2002). In terms of vocalization coding, when non-human primates are tested in passive listening conditions without specific discrimination tasks, single unit recordings from VLPFC neurons show similar responses to calls that are similar in acoustic morphology (Romanski et al., 2005). Examination of VLPFC neuronal responses using a small set of call type categories with an oddball-type task, has suggested that auditory cells in VLPFC might encode semantic category information (Gifford et al., 2005), including the discrimination between vocalizations that indicate food vs. non-food (Cohen et al., 2006).

Importantly, it has been shown that VLPFC cells are multisensory (Sugihara et al., 2006). These multisensory neurons are estimated to represent more than half of the population of the anterolateral VLPFC, including areas 12/47 and 45 (Romanski 2012; Sugihara et al., 2006). Vocalization call type coding by cells may then be more dependent upon an integration of not only acoustic features but also the appropriate facial gesture or mouth movement that accompanies the vocalization. Relying on responses to only the auditory

component of a vocalization might then lead to less accurate recognition and categorization in regards to call type but also especially to caller identity.

We asked how well VLPFC neurons encoded the vocalization call type or caller identity of a species-specific vocalizations using previously recorded data (Romanski et al., 2005) to 3 specific vocalization call types from three individual callers, (one coo, grunt, and girney from each of 3 macaque monkeys). We then analyzed the neural response using linear discriminate analysis to assess how accurately cells classified the different call types (coos, grunts, and girneys).

2. Materials and Methods

This analysis and review relies on neurophysiological recordings of 110 cells from Romanski et al., (2005). Methods are described in detail in Romanski et al., (2005) with modifications to the data analysis described here.

2.1 Neurophysiological recordings

As previously described (Romanski et al., 2005) we made extracellular recordings in two rhesus monkeys (*Macaca mulatta*) during the performance of a fixation task during which auditory stimuli including species-specific vocalizations were presented. All methods were in accordance with NIH standards and were approved by the University of Rochester Care and Use of Research Animals committee. Ventrolateral prefrontal areas 12/47 and 45 were targeted (Preuss and Goldman-Rakic, 1991; Romanski and Goldman-Rakic, 2002; Petrides and Pandya, 2002).

2.2 Apparatus and Stimuli

All training and recording was performed in a sound-attenuated room, lined with Sonex (Acoustical Solutions, Inc). Auditory stimuli were presented to the monkeys by either a pair of Audix PH5-vs speakers (frequency response +/- 3dB, 75-20,000 Hz) located on either side of a center monitor, or a centrally located Yamaha MSP5 monitor speaker (frequency response 50 Hz - 40kHz), located 30 inches from the monkey's head. The auditory stimuli ranged from 65 – 80 dB SPL measured at the level of the monkey's ear with a B & K sound level meter, and a Realistic audio monitor.

In this report we have focused on prefrontal responses to 3 Macaque Vocalizations (Coo, Grunt and Girney) recorded on the island of Cayo Santiago by Dr. Marc Hauser (Hauser and Marler, 1993). The three vocalizations are all normally given during social exchanges and were uttered by 3 adult female rhesus macaques on Cayo Santiago. Thus, in this experiment 3 call types (CO, GT and GY) by three female callers were used yielding a 3×3 call matrix (Figure 1). Coos (CO) are given during social interactions including grooming, upon finding food of low value and when separated from the group. Grunts (GT) are given during social interaction such as an approach to groom, and upon the discovery of a low value food item and girneys (GY) are given during grooming and when females attempt to handle infants. In addition to social context, vocalization types in the macaque vocal repertoire have also been categorized according to the presence (or absence) of particular acoustic features (Figure 1; Hauser and Marler, 1993; Hauser, 1996). Grunts are recognized as Noisy calls along with growls and pant threats while coos are marked by the presence of a harmonic stack, often with a dominant fundamental frequency and some evidence of vocal tract filtering. Girneys are considered tonal calls and have dominant energy at a single or narrow band of frequencies. Brain regions which respond to species-specific vocalizations could do so on the basis of acoustic, behavioral context or emotional features of various calls.

2.3 Experimental Procedure

Each day the monkeys were brought to the experimental chamber and were prepared for extracellular recording as previously published (Romanski et al, 2005). Each isolated unit was tested with a battery of auditory and visual stimuli which have been discussed in additional studies (Romanski et al., 2005; Sugihara et al., 2006). For this analysis cells were tested with the 9 auditory stimuli (3 call types $\times$ 3 callers) which were repeated 8 – 12 times in a randomized design. During each trial a central fixation point appeared and subjects fixated for a 500 ms pretrial fixation period. Then the vocalization was presented and fixation was maintained. After the vocalization terminated a 500 ms post-stimulus fixation period occurred. A juice reward was delivered at the termination of the post-stimulus 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 2-3 second inter-trial interval, after which the fixation point would appear and subjects could voluntarily begin a new trial.

2.4 Data Analysis

The unit activity was read into Matlab® (Mathworks) and SPSS where rasters, histograms and spike density plots of the data could be viewed and printed. For analysis purposes, mean firing rates were measured for a period of 500 ms in the intertrial interval (SA), and for 600 ms during the stimulus presentation epoch. Several time bins during the stimulus presentation period were used for this analysis as defined below.

Significant changes in firing rate during presentation of any of the vocalizations were detected using a t-test which compared the spike rate during the intertrial interval with that of the stimulus period. Any cell with a response that was significantly different from the spontaneous firing rate measured in the intertrial interval ($p < 0.05$) was considered vocalization responsive. The effects of vocalization call type and caller identity were examined with a 2-way MANOVA across multiple bins of the stimulus period using SPSS. Cells which were significant for call type or identity were further analyzed.

Vocalization Selectivity—A selectivity index (SI) was calculated using the absolute value of the averaged responses to each stimulus minus the baseline firing rate. The SI is a measure of the depth of selectivity across all 9 vocalizations stimuli presented and is defined as:

$$SI = (n - \sum_{i=1}^n (\lambda_i / \lambda_{\max})) / (n - 1)$$

where n is the total number of stimuli, λ_i is the firing rate of the neuron to the i th stimulus and $\lambda_{\max}$ is the neuron's maximum firing rate to one of the stimuli (Wirth et al., 2009). Thus, if a neuron is selective and responds to only one stimulus and not to any other stimulus, the SI would be 1. If the neuron responded identically to all stimuli in the list, the SI would be 0.

Cluster Analysis of the Neuronal Response: To determine whether prefrontal neurons responded to sounds that were similar because of call type or because they were from the same identity caller, (i.e. on the basis of membership in a functional category), we performed a cluster analysis as was done in previous studies (Romanski et al., 2005). For each cell, we computed the mean firing rate during the stimulus period for the responses to the 9 vocalizations tested. We then computed a 9×9 dissimilarity matrix of the mean differences in the responses to each vocalization, which was analyzed in Matlab® using the 'Linkage (average)' and 'Dendrogram' commands to carry out a cluster analysis. A

consensus tree was generated to detect commonalities of association across several groupings of responsive cells. This was accomplished by reading the individual dendrograms into the program CONSENSE® (available at <http://cmgm.stanford.edu/phyliip/consense.html>). CONSENSE® reads a file of dendrogram trees and prints out a consensus tree based on strict consensus and majority rule consensus (Margush and McMorris, 1981).

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. Classification performance was estimated using 4-fold cross validation. This analysis resulted in a stimulus-response matrix, where the stimulus was the vocalization which 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 8-12, the sum of the off-diagonal elements in a particular row).

3. Results

We analyzed neurons recorded in macaque VLPFC with species-specific vocalizations grouped by call type and caller identity. 123 cells (110 from previous Romanski et al., 2005 plus 13 additional from the same animals) were tested with a balanced list of 3 vocalization call types vocalized by 3 female callers (one coo, grunt, and girney from each caller). Recorded neurons, from awake behaving macaques, were analyzed off line to first determine the response to the 9 vocalization stimuli. 81/123 cells were defined as vocalization responsive since the firing rate during early and late portions of the response period were significantly different from baseline with a t-test ($p < 0.5$), for any of the 9 vocalizations.

Examination of the responses of VLPFC cells indicated a varying degree of selectivity such that some cells responded to only a single stimulus in the 9 vocalization set while other cells responded to most stimuli (Figure 2 and Figure 3). To quantify the number of stimuli a given VLPFC cell responded to, we computed a selectivity index for all of the vocalization responsive cells ($n=81$). In Figure 2, the average neuronal response of several neurons to each of the 9 vocalizations is shown with their respective selectivity index. Panel 2A displays an auditory cell with a high selectivity index score (0.74034) which responded best to a Coo from caller 1 and a grunt from caller 3. Panel 2B had a response to all of the grunt exemplars from all three callers ($SI = 0.52327$), suggesting a call type cell. In contrast the responses shown in panels 2C and 2D are increased for more than 4 vocalizations giving a much lower selectivity index than the responses illustrated in Figure 2A and 2B. The cell in Figure 2C responded best to a coo and grunt from the same caller while the response in 2D was best for 2 of the grunt vocalizations. Thus examination of the overall response hinted at selectivity to call type or to caller identity in various cells, which we therefore analyzed in more detail.

Call type and Caller identity

We examined the effect of call type and caller identity of the species-specific vocalizations in single VLPFC neurons. We have previously shown that VLPFC neuron responses to vocalizations are heterogeneous and vary over time, with some cells demonstrating short-lived phasic responses and others having more sustained responses (Figure 3; Romanski et al., 2005). Furthermore, latency and peak response varies across stimuli (Romanski and

Hwang, 2012) indicating that VLPFC responses are differentially affected by features present within the vocalizations with latencies ranging from 29 – 330 ms. To capture the time-varying nature of prefrontal neuronal responses to vocalizations we divided the 600 ms response period into smaller bins and performed a 2-way MANOVA on vocalization responsive cells ($n=81$) with 4×150 ms time bins (Call type $\times$ Caller identity), which would allow us to capture the onset latency of most VLPFC auditory responses (Romanski and Hwang, 2012). In the 2-way MANOVA there were 15 cells with a main effect of vocalization call type (15 cells $p < 0.05$; 6 cells $p < 0.01$; 8 cells $p < 0.001$) and 11 cells with a main effect of caller identity [11 cells $p < 0.05$; 1 cell $p < 0.01$; 4 cells $p < 0.001$]. 14 cells had an interaction of call type and caller identity. Several cells are depicted in Figure 3 that were significant in the 2-way MANOVA. The call type cell depicted in 3A had a similar response to each of the grunts and to each of the coos which are dissimilar from each other. For cells 3A and 3B there was a significant interaction of call type and caller, and a main effect of call type and of caller ($p < 0.01$). The cell depicted in 3C had a similar increase in firing rate to all of the grunts but not to the other vocalization call types (significant main effect of call type $p < 0.01$). Responses varied over time and more vocalization call type cells showed a significant change at the early onset time (0-150 ms time bin) or during the offset of the auditory stimulus (451 – 601 ms time bin). In contrast caller identity cells were more frequently responsive during bin 3, at 301-451 ms after vocalization onset. Thus, call type responsive cells responded earlier suggesting that features defining vocalization type might be discriminated earlier in the response period than caller identity which might involve features that evolve on a slower time scale.

Hierarchical Cluster Analysis

We used a hierarchical cluster analysis to quantify each cell's response to the 9 vocalizations and then computed a consensus tree from these dendrograms in order to determine the common clustered responses across the population (Romanski et al., 2005). We hypothesized that cells would cluster the exemplars from similar call types together as suggested by the call type analysis but some exemplars may be judged as similar more than others. The cluster analysis is one way in which to examine this. We computed the response for each cell subtracted from the cell's average response (mean difference) and performed a hierarchical cluster analysis using the CLUSTER function in MATLAB. The hierarchical cluster analysis yields a dendrogram of each cell's response, for the 9 vocalizations, grouping together stimuli that evoked similar responses. We also computed a cophenetic coefficient (measure of fit) for each responsive cell.

A consensus tree of the common clusters was computed for several groupings of the cells including those that had a main effect of call type ($n=15$) or an effect of caller identity (12) at $p < 0.05$. The consensus tree indicates the groups that occur most often in the dendrograms across the group of VLPFC cells. Examination of the consensus tree indicates that small numbers of cells showed similar groupings of the vocalizations (Figure 4). Nonetheless, similar call types were not grouped together by a high percentage of cells even in this small call type responsive group. From the consensus tree it is apparent that coos were grouped together by $\sim 20\%$ of cell type responsive cells. 2 of the grunt exemplars were also clustered by 20% of the cells (Figure 4; Grunt1 and grunt 3). There were 4 cells (27%) that responded in a similar manner to a grunt from speaker 2 and girney from speaker 3. Overall, there were a small number of cells which showed grouping by a similar call type. Cluster analysis and consensus across the significant caller cells resulted in only a few cells showing a grouping by caller (not shown).

Linear Discriminate Analysis

As call type appeared to be a significant factor in the responses of a larger proportion of cells in our sample that caller identity, we further examined cells with a significant effect of call type across any time bin ($n=37$). We used the responses of single VLPFC neurons to determine which call type had been presented using linear discriminant analysis (Romanski et al., 2005). We used a fixed window of 600 ms and divided this into smaller bins and decoded call type using all of the bins simultaneously so that with the 100 ms bin, 6 individual bins were used simultaneously. Analysis of 60, 100, 150 and 300 ms bins in our 600 ms response period showed that classification performance did not differ significantly except at the longest, single 600 ms bin where it declined (Figure 5). In addition, we performed a classification analysis over time using a population vector with four discrete 150 ms bins in a 600 ms response period for the three call types. In this analysis we combined the results across all 37 cells with an effect of call type across any time bin. We found that, on average, cells are more accurate in their classification of coos and this was evident early in the response period during the first 300 ms. While the performance for grunts was above chance at 450 ms, across the population, girneys were not discriminated above chance (Figure 6). On average VLPFC cells decoded a single best call type at 71.4% correct, when ranked with the most accurate bin for that stimulus. In previous studies using a shorter response period of 300 ms, classification performance peaked at 60 ms but did not differ significantly at 100 ms. Our inclusion of several exemplars which lasted longer than 500 ms prompted the need for a longer response period in our analysis. While performance, on average, was above chance and even reached above 70% for coos, classification of even this small number of vocalization categories remained below chance for many cells. Furthermore, while 66% of neurons recorded responded to vocalizations (81/123 cells) only 20% of these were call type responsive. Studies have suggested that most VLPFC vocalization responsive cells are multisensory when tested with appropriate face stimuli (Sugihara et al., 2006). Addition of corresponding face stimuli may increase the accuracy of call type classification by prefrontal neurons.

Discussion

In the present study we found that a small percentage of VLPFC cells are selective for vocalization type and but that less than half of the vocalization responsive cells classified calls above chance. In previous studies when VLPFC cells were tested with a large repertoire of species-specific vocalizations, the neuronal responses were similar for sounds with similar acoustic features (Romanski et al., 2005). When the neuronal response to 2 vocalization call types which both signal high value food, i.e. a harmonic arch and a warble, were evaluated, there was a significant difference in the neuronal response. In contrast, coos and warbles which have similar acoustic structure but different semantic meaning, elicit similar neuronal responses in $\sim 20\%$ of prefrontal neurons (Romanski et al., 2005). In contrast, Cohen and colleagues suggest that VLPFC cells may carry more information about semantic categories than acoustic categories of call type (Gifford et al., 2005) and that prefrontal neurons do not encode simple acoustic features (Cohen et al., 2007). However, more refined analyses suggest that there are complex acoustic features present in vocalizations that prefrontal auditory neurons encode and that prefrontal auditory neurons show responses that are related to specific categories of vocalization call types (Averbeck and Romanski, 2004; 2006). The present study adds to this evidence in demonstrating that some VLPFC neurons show vocalization call type responses, though it does not address their segregation by acoustics or semantics.

In the present study we evaluated the response of neurons to several exemplars from 3 different call types. The exemplars were chosen at random from a library of unfamiliar callers who had recordings of all three call types (coos, grunts and girneys). These three call

types were chosen for this study because they are acoustically distinct, yet coos and grunts are often uttered under similar contexts. We assumed that if prefrontal neurons routinely grouped utterances according to call type that the 3 exemplars from different callers of the same vocalization call type would elicit a similar neuronal response owing to the similarity of dominant acoustic features that define call type. In confirmation of this hypothesis, there was a small population of neurons that were call type responsive in our analysis and clearly showed a similar response across exemplars from different callers to each call type (Figure 2c and 3a,b). However, in our sample of 81 cells, ~ 20% had a main effect of vocalization call type, 14 % with an effect of caller and 17% cells had an interaction of call type and caller. Thus the percent of cells that appear to encode vocalization call type is somewhat small and those encoding the identity of the caller that issued the call are even smaller, considering the robust response of these neurons to vocalizations.

Cells which were responsive to the vocalizations but did not have an effect of call type or caller often were selective and responded to 1-3 vocalizations (Figure 2). VLPFC cells have been previously shown to be selective in their responses not only to vocalizations (Romanski and Goldman-Rakic, 2002; Romanski et al., 2005) but also to face stimuli (O'Scalaidhe et al., 1997 O'Scalaidhe et al., 1999). It is not clear whether the selectivity of VLPFC neurons is based on some combination of acoustic features in a particular vocalization which evoke a response or some other aspect of the sound such as its meaning or association (Cohen et al., 2006). However, there were no cells which responded with a similar response to all coo and grunt vocalizations which represent a similar functional category but which are acoustically dissimilar. Responses across the VLPFC population are heterogeneous and correlation of the prefrontal neuronal response with complex features of the vocalizations may be necessary to provide evidence of vocalization call type coding (Averbeck and Romanski, 2006).

We further examined vocalization call type classification using the responses of single VLPFC neurons on individual trials using linear discriminant analysis. Analysis across different bins in the 600 ms response period showed that classification performance did not differ significantly from 60 – 300 ms (Figure 6). Examination of the temporal aspects of call type categorization across the population yielded interesting results. We found that, on average cells were more accurate in their classification of coos and this was evident early in the response period during the first 300 m. Classification performance of coos reached 70 percent in the first 300 ms and remained above chance for the duration of the response period. In previous studies using a shorter response period of 300 ms, classification performance peaked at 60 ms but did not differ significantly at 100 ms (Averbeck and Romanski, 2006). Our inclusion of several exemplars which lasted longer than 500 ms prompted our need for a longer response period in our analysis. Interestingly, Rendall et al., (1998) examined coos, grunts, and screams for spectral features and using a discriminate analysis found that coos were discriminated more accurately than grunts or screams (Rendall et al., 1998). In addition, it was found that coos had a spectral peak which was stable and distinctive (Rendall et al., 1998). Ceugniet and Izumi (2004) demonstrated that macaques could discriminate coos from different individuals and that call duration, pitch, and harmonics were used to make discriminations between calls. Thus, it is possible that these features (call duration, pitch, harmonics, and spectral cues) led to the greater accuracy of classification of coos found here for VLPFC neurons. In fact, even humans are better at discriminating coos from different individual monkeys than screams from those same individuals (Owren and Rendall, 2003). If coos themselves have more cues that make discrimination easier, then this information may be parsed out in earlier auditory regions (belt, parabelt, STG) and routed to VLPFC. Perhaps having multiple coos from the same individual would have increased our discrimination accuracy for identity, or increased the number of identity responsive cells.

The overall classification performance of VLPFC neurons in our analysis of a small number of vocalization categories was $\sim 42\%$ (Figure 5). There are several possibilities as to why the percent correct was not higher. First, our data is based only on a small population of neurons recorded singly which may not be adequate to support the complex function of vocalization call type processing. Classification across simultaneously recorded cells would most likely yield superior accuracy (Georgopoulos and Massey, 1988; Averbach et al., 2003). Second, studies have shown higher classification performance when assessing responses to categories with clear boundaries such as vocalization and non-vocalization stimuli. For example, population decoding in early auditory processing area R (rostral field), found classification accuracy of natural sounds (monkey calls and environmental sounds) to exceed 90% correct 80 msec after stimulus onset (Kusmirek et al., 2012). Finally, our sample of neurons included cells from both areas 12/47 and 45 which may be specialized for different functions with regard to vocalizations. Area 12/47 receives more input from rostral belt and parabelt regions while area 45 receives a denser projection from caudal parabelt, the STS and from inferotemporal cortex (Romanski et al., 1999; Romanski and Goldman-Rakic, 2002; Barbas, 1988; Webster et al., 1994). In addition, area 12/47 has more unimodal auditory responsive cells. Both regions receive input from the STG and STS and have multisensory responsive neurons. Thus, vocalizations may be processed differently in the more auditory responsive area of 12/47 compared to area 45, which may be more specialized for face and object processing (Romanski, 2012; Romanski, 2007). Focusing our analysis on a larger sample of cells which have similarly tuned responses to specific complex features may increase the categorization performance of cells in VLPFC.

Possibly the most important reason for the modest classification performance of VLPFC cells to vocalization stimuli is the fact that most of these cells are multisensory and may typically process vocalizations in a multisensory manner. In the current study we have examined responses only to vocalizations which may be only half of the stimulus that elicits the best response for many prefrontal neurons. In previous work where VLPFC neurons were tested with separate and simultaneous combinations of vocalizations and the accompanying dynamic facial gesture it has been shown that more than half the recorded population was multisensory (Sugihara et al., 2006). VLPFC cells show multisensory enhancement or suppression to face-vocalization combinations. Moreover, neuronal responses demonstrate more information for combined face-vocalization stimuli than for the unimodal components alone (Sugihara et al., 2006). These strong responses to face-vocalization pairs suggest that this region of VLPFC may be specialized for face-vocalization integration (Romanski 2012). It is likely that VLPFC neurons would show significantly enhanced classification of vocalization call types when both face and vocalization information is present. Even neural responses in auditory cortex are enhanced when appropriate visual information is present (Kayser et al., 2010). Importantly, it has been shown that discrimination of vocalizations improves when a congruent face stimulus accompanies the vocalization (Chandrasekaran et al., 2011). In this study, behavioral performance of rhesus macaques discriminating vocalizations improved when congruent avatar face stimuli accompanied the vocalizations.

It is well known that combining auditory and visual stimuli can enhance accuracy and decrease reaction time (Stein and Meredith, 1993). This has been demonstrated in many brain regions including the inferior frontal gyrus in the human brain during neuroimaging where the PFC accumulates information about auditory and visual stimuli to be used in recognition (Noppeney et al., 2010). Moreover, training of face-voice pairs can enhance activity in VLPFC for voice recognition (von Kriegstein and Giraud, 2006) and the interaction between face and voice processing involves connections between face areas, voice areas, and VLPFC (von Kriegstein and Giraud, 2006).

Voice recognition in the human brain is also a multisensory process and includes voice regions found in the superior temporal sulcus STS (Belin et al., 2000) and VLPFC (Fecteau et al., 2005). In an fMRI study with face and voice recognition, Brodmann's areas 45 and 47 were activated during the delay period in these working memory tasks (Rama & Courtney, 2005). Our current data showing vocalization identity cells in VLPFC, which included areas 45 and 47, could provide the underlying neuronal mechanism for these effects,

Data from behavioral studies in animals and humans shows that recognition and memory are enhanced when both face and vocal information are present. In fact, use of auditory information alone results in inferior performance on memory tasks compared to performance with visual stimuli alone (Fritz et al., 2005). Training non-human primates to perform auditory working memory (WM) paradigms typically takes longer than for visual WM tasks and it has been shown that performance accuracy is greater for visual WM than for auditory WM tasks in non-human primates (Scott et al., 2012). Surprisingly, despite our superior language abilities, human subjects also perform better when given faces than when given voices as memoranda (Yarmey et al., 1994; Hanley et al., 1998). It has been demonstrated that recognition of speaker identity from a single utterance only resulted in 35% accuracy, with females outperforming males (Skuk, and Schweinberger, 2013). Even when face and voice stimuli are familiar and the content and frequency of presentation is carefully controlled in a similar manner as our animal experiments, faces are more easily remembered and more semantic information is retrieved from human faces compared to their voices (Barsics and Brédart, 2012). Thus, the sub-optimal performance of our population of VLPFC responsive neurons in categorizing vocalization call types is perhaps expected when only vocal information is available.

Finally, our present study relies on data collected during stimulus presentation and it has been shown that prefrontal neurons are most active during working memory, categorization and goal directed behavior (Freedman et al., 2001; Russ et al., 2007). In humans it has been found that task demands can activate different frontal regions (Owen et al., 1996) and that modality and task demands can interact within VLPFC (Protzner and McIntosh 2009). In our lab, context during a non-match-to-sample task has been shown to modulate firing rates, with some cells showing an effect of task conditions and others maintaining a preference for a particular stimulus (Hwang and Romanski, 2009). An analysis of neuronal responses when face and vocal information are actively used in a recognition paradigm and carefully compared with performance using only auditory or only visual information is needed to fully understand the processing of communication and the impact of face and vocalization integration on recognition.

Acknowledgments

Grants: This research was funded by the National Institute for Deafness and Communication Disorders (DC 04845, LMR and DC05409, Center for Navigation and Communication Sciences).

References

- Averbeck BB, Crowe DA, Chafee MV, Georgopoulos AP. Neural activity in prefrontal cortex during copying geometrical shapes. II. Decoding shape segments from neural ensembles. *Exp Brain Res.* 2003; 150(2):142–153. [PubMed: 12669171]
- Averbeck BB, Romanski LM. Probabilistic encoding of vocalizations in macaque ventral lateral prefrontal cortex. *J Neurosci.* 2006; 26:11023–11033. [PubMed: 17065444]
- Averbeck BB, Romanski LM. Principal and independent components of macaque vocalizations: constructing stimuli to probe high-level sensory processing. *J Neurophysiol.* 2004; 91:2897–2909. [PubMed: 15136606]

- Barbas H. Anatomic organization of basoventral and mediodorsal visual recipient prefrontal regions in the rhesus monkey. *J Comp Neurol.* 1988; 276:313–342. [PubMed: 3192766]
- Barsics C, Bredart S. Recalling semantic information about newly learned faces and voices. *Memory.* 2012; 20(5):527–534. [PubMed: 22646520]
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. *Nature.* 2000; 403:309–312. [PubMed: 10659849]
- Bendor D, Osmanski MS, Wang X. Dual-pitch processing mechanisms in primate auditory cortex. 2012; 32:16149–16161.
- Bendor D, Wang X. The neuronal representation of pitch in primate auditory cortex. *Nature.* 2005; 436(7054):1161–1165. [PubMed: 16121182]
- Bradbury, JW.; Vehrencamp, SL. *Principles of Animal Communication.* Blackwell; Oxford: 1998.
- Ceugniet M, Izumi A. Vocal individual discrimination in Japanese monkeys. 2004; 45:119–128.
- Chandrasekaran C, Lemus L, Trubanova A, Gondon M, Ghazanfar AA. Monkeys and humans share a common computation for face/voice integration. 2011; 7:e1002165.
- Cohen YE, Hauser MD, Russ BE. Spontaneous processing of abstract categorical information in the ventrolateral prefrontal cortex. *Biol Lett.* 2006; 2(2):261–265. [PubMed: 17148378]
- Cohen YE, Theunissen F, Russ BE, Gill P. Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *J Neurophysiol.* 2007; 97(2):1470–1484. [PubMed: 17135477]
- Fecteau S, Armony JL, Joannette Y, Belin P. Sensitivity to voice in human prefrontal cortex. *J Neurophysiol.* 2005; 94:2251–2254. [PubMed: 15928057]
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science.* 2001; 291:312–316. [PubMed: 11209083]
- Fritz J, Mishkin M, Saunders RC. In search of an auditory engram. *Proc Natl Acad Sci U S A.* 2005; 102(26):9359–9364. [PubMed: 15967995]
- Georgopoulos AP, Massey JT. Cognitive spatial-motor processes. 2. Information transmitted by the direction of two-dimensional arm movements and by neuronal populations in primate motor cortex and area 5. *Exp Brain Res.* 1988; 69(2):315–326. [PubMed: 3126080]
- Gifford GW III, Maclean KA, Hauser MD, Cohen YE. The neurophysiology of functionally meaningful categories: macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations. *J Cogn Neurosci.* 2005; 17:1471–1482. [PubMed: 16197700]
- Hackett TA, Stepniewska I, Kaas JH. Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res.* 1999; 817:45–58. [PubMed: 9889315]
- Hanley JR, Smith ST, Hadfield J. I recognise you but I can't place you: an investigation of familiar-only experiences during tests of voice and face recognition. *Q J Exp Psychol.* 1998; 51(1):179–195.
- Hauser, MD. *The Evolution of Communication.* MIT Press; Cambridge, MA: 1996.
- Hauser MD, Marler P. Food associated calls in rhesus macaques (*Macaca mulatta*) I. Socioecological factors. *Behav Ecol.* 1993; 4:194–205.
- Hwang J, Romanski LM. Comparison of face and non-face stimuli in an audiovisual discrimination task. 2009; 578:5.
- Johnson, RA.; Wichern, DW. *Applied Multivariate Statistical Analysis.* Prentice Hall; Saddle River, NJ: 1998.
- Joly O, Ramus F, Pressnitzer D, Vanduffel W, Orban GA. Interhemispheric differences in auditory processing revealed by fMRI in awake rhesus monkeys. *Cereb Cortex.* 2012; 22(4):838–853. [PubMed: 21709178]
- Kayser C, Logothetis NK, Panzeri S. Visual enhancement of the information representation in auditory cortex. *Curr Biol.* 2010; 20(1):19–24. [PubMed: 20036538]
- Kikuchi Y, Horwitz B, Mishkin M. Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *J Neurosci.* 2010; 30(39):13021–13030. [PubMed: 20881120]
- Kusmierek P, Ortiz M, Rauschecker JP. Sound-identity processing in early areas of the auditory ventral stream in the macaque. 2012; 107:1123–1141.

- Margush T, McMorris FR. Consensus n-trees. *Bull Math Biol.* 1981; 43:239–244.
- Noppeney U, Ostwald D, Werner S. Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *J Neurosci.* 2010; 30(21):7434–7446. [PubMed: 20505110]
- O'Scalaidhe SP, Wilson FA, Goldman-Rakic PS. Areal segregation of face-processing neurons in prefrontal cortex. *Science.* 1997; 278:1135–1138. [PubMed: 9353197]
- O'Scalaidhe SPO, Wilson FAW, Goldman-Rakic PGR. Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: Evidence for intrinsic specialization of neuronal coding. *Cereb Cortex.* 1999; 9:459–475. [PubMed: 10450891]
- Owen AM, Evans AC, Petrides M. Evidence for a two-stage model of spatial working memory processing with the lateral frontal cortex: A positron emission tomography study. *Cereb Cortex.* 1996; 6:31–38. [PubMed: 8670636]
- Owings, DH.; Morton, ES. *Animal Vocal Communication: A New Approach.* Cambridge University Press; Caombridge: 1998.
- Owren MJ, Rendall D. Salience of caller identity in rhesus monkey (*Macaca mulatta*) coos and screams: perceptual experiments with human (*Homo sapiens*) listeners. *J Comp Psychol.* 2003; 117(4):380–390. [PubMed: 14717639]
- Perrodin C, Kayser C, Logothetis NK, Petkov CI. Voice cells in the primate temporal lobe. *Curr Biol.* 2011; 21(16):1408–1415. [PubMed: 21835625]
- Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK. A voice region in the monkey brain. *Nat Neurosci.* 2008; 11(3):367–74. [PubMed: 18264095]
- Petrides M, Pandya DN. Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur J Neurosci.* 2002; 16(2):291–310. [PubMed: 12169111]
- Poremba A, Malloy M, Saunders RC, Carson RE, Herscovitch P, Mishkin M. Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature.* 2004; 427(6973):448–51. [PubMed: 14749833]
- Poremba A, Saunders RC, Crane AM, Cook M, Sokoloff L, Mishkin M. Functional mapping of the primate auditory system. *Science.* 2003; 299:568–572. [PubMed: 12543977]
- Preuss TM, Goldman-Rakic PS. Ipsilateral cortical connections of granular frontal cortex in the strepsirrhine primate Galago, with comparative comments on anthropoid primates. *J Comp Neurol.* 1991; 310:507–549. [PubMed: 1719039]
- Protzner AB, McIntosh AR. Modulation of ventral prefrontal cortex functional connections reflects the interplay of cognitive processes and stimulus characteristics. 2009; 19:1042–1054.
- Rama P, Courtney SM. Functional topography of working memory for face or voice identity. *Neuroimage.* 2005; 24(1):224–234. [PubMed: 15588614]
- Rauschecker JP. Cortical processing of complex sounds. *Curr Opin Neurobiol.* 1998; 8(4):516–521. [PubMed: 9751652]
- Rauschecker JP, Tian B, Hauser M. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science.* 1995; 268(5207):111–114. [PubMed: 7701330]
- Rauschecker JP, Tian B, Pons T, Mishkin M. Serial and parallel processing in rhesus monkey auditory cortex. *J Comp Neurol.* 1997; 382(1):89–103. [PubMed: 9136813]
- Rendall D, Owren MJ, Rodman PS. The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *J Acoust Soc Am.* 1998; 103(1):602–614. [PubMed: 9440345]
- Romanski LM. Integration of faces and vocalizations in ventral prefrontal cortex: implications for the evolution of audiovisual speech. *Proc Natl Acad Sci U S A.* 2012; 109(Suppl 1):10717–10724. [PubMed: 22723356]
- Romanski LM. Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cereb Cortex.* 2007; 17(Suppl 1):i61–9. [PubMed: 17634387]
- Romanski LM, Averbek BB, Diltz M. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J Neurophysiol.* 2005; 93(2):734–747. [PubMed: 15371495]
- Romanski LM, Bates JF, Goldman-Rakic PS. Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J Comp Neurol.* 1999; 403:141–157. [PubMed: 9886040]

- Romanski LM, Goldman-Rakic PS. An auditory domain in primate prefrontal cortex. *Nat Neurosci.* 2002; 5:15–16. [PubMed: 11753413]
- Romanski LM, Hwang J. Timing of audiovisual inputs to the prefrontal cortex and multisensory integration. *Neuroscience.* 2012; 214:36–48. [PubMed: 22516006]
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci.* 1999; 2(12):1131–1136. [PubMed: 10570492]
- Russ BE, Lee YS, Cohen YE. Neural and behavioral correlates of auditory categorization. *Hear Res.* 2007; 229(1-2):204–212. [PubMed: 17208397]
- Sadagopan S, Wang X. Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex. *J Neurosci.* 2009; 29(36):11192–11202. [PubMed: 19741126]
- Scott BH, Mishkin M, Yin P. Monkeys have a limited form of short-term memory in audition. *Proc Natl Acad Sci U S A.* 2012; 109(30):12237–12241. [PubMed: 22778411]
- Skuk BG, Schweinberger SR. Gender differences in familiar voice identification. *Hear Res.* 2013; 296:131–40. [PubMed: 23168357]
- Stein, BE.; Meredith, MA. *The Merging of the Senses.* MIT Press; Cambridge: 1993.
- Sugihara T, Diltz MD, Averbach BB, Romanski LM. Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J Neuroscience.* 2006; 26:11138–11147.
- Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. Functional specialization in rhesus monkey auditory cortex. *Science.* 2001; 292:290–293. [PubMed: 11303104]
- von Kriegstein K, Giraud AL. Implicit multisensory associations influence voice recognition. *PLoS Biol.* 2006; 4(10):e326. [PubMed: 17002519]
- Wang X, Merzenich MM, Beitel R, Schreiner CE. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol.* 1995; 74(6):2685–2706. [PubMed: 8747224]
- Webster MJ, Bachevalier J, Ungerleider LG. Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cerebral Ctx.* 1994; 4:470–483.
- Wirth S, Avsar E, Chiu CC, Sharma V, Smith AC, Brown E, Suzuki WA. Trial outcome and associative learning signals in the monkey hippocampus. *Neuron.* 2009; 61(6):930–940. [PubMed: 19324001]
- Yarmey AD, Yarmey AL, Yarmey MJ. Face and voice identifications in showups and lineups. *Appl Cog Psychol.* 1994; 8(5):453–464.

Highlights

1. Prefrontal neurons are responsive to species-specific vocalizations
2. Some neurons have similar responses to exemplars with same vocalization call type
3. Approximately 20% of cells encode the vocalization call type
4. Call type categorization is high during specific response times & call types
5. Categorization should be enhanced with accompanying facial gestures

Figure 1.

The spectrograms and waveforms for the 9 vocalizations used in the current study are shown. There are three exemplars for each vocalization type (Coos, Grunts and Girneys) that were given by 3 callers in rows A, B and C. The calls have been previously characterized according to acoustic features, functional referents and the identity of the callers (Hauser and Marler, 1993; see Methods).

Figure 2.

Individual VLPFC neurons demonstrated a range of selectivity to the 9 vocalizations. A) An auditory responsive neuron selective for two vocalizations but not call type or caller specific (coo 1 and grunt 3); B) A neuron that had the best response to all of the grunt exemplars; C) A neuron which was responsive to > 5 vocalizations but had the highest mean response to coo1 and grunt 1 from the same caller. D) An example of a neuron with a low selectivity score which was responsive to all 9 vocalizations but with the highest average response to two of the grunt exemplars. Error bars are standard error of the mean.

Figure 3.

The neural response of three cells to the 9 vocalizations is shown as rasters and spike density functions. For each plot, the pre-stimulus fixation period (-250 to 0) occurs prior to the onset of the vocalization which occurs at time 0 and the response is shown over 800 ms. A) An example of a call type neuron with responses to the 3 coos (black) and 3 grunts (red). The response to the coos is bi-phasic and similar across coos but different from the grunts. In B and C the response to the 9 vocalizations is grouped by the three vocalization types (coos-black, grunts-red, girneys-blue). B) An example of a call type neuron with the highest average firing rate to grunts and the lowest response to coos. C) A call type responsive neuron which responded best to grunts.

Figure 4.

Neuronal Consensus Trees. Consensus tree, based on the dendrograms for the individual call type responsive cells is shown. Dendrograms were derived for each of the 15 call type cells and their response to the 9 vocalizations (3 calls each from 3 callers) and a consensus tree was generated indicating the common groupings. The consensus tree indicated that in this sample 3 main groupings occurred: Grunt1 and Grunt 2 clustered together in 3 cells; Grunt 2 and Girney3 clustered together in 4 cells and Coo2 and Coo3 evoked a similar response in 3 cells.

Figure 5.

Classification performance as a function of time bin length. This line graph shows how well VLPFC vocalization responsive neurons ($n=37$) categorized vocalizations as coo, grunt and girney using time bins of 60, 100, 150, 300 ms or a single 600 ms time bin. Utilization of small to medium bin sizes leads to decoding above chance levels, whereas using one larger bin size of 600 ms is less accurate. Error bars are standard error of the means.

Figure 6.

Classification performance of call types over the course of the stimulus response period. Decoding performance at 4×150 ms time bins of VLPFC vocalization responsive neurons ($n=37$) is shown for each of the three call types. On average, cells are more accurate in their classification of coos and this is apparent in the first 300 ms of the response period. While the performance for grunts was above chance at 450 ms, girneys were not discriminated above chance (33.3%).