Published in final edited form as:

Hear Res. 2014 June ; 312: 128–142. doi:10.1016/j.heares.2014.03.007.

Natural variability in species-specific vocalizations constrains behavior and neural activity

Kate L. Christison-Lagay^a, Sharath Bennur^b, Jennifer Blackwell^a, Jung H. Lee^b, Tim Schroeder^b, and Yale E. Cohen^{b,c,d,*}

Yale E. Cohen: yaleecohen@gmail.com

^aNeuroscience Graduate Group, University of Pennsylvania, Philadelphia, PA 19104, USA

^bDept. Otorhinolaryngology, U. Pennsylvania, Philadelphia, PA 19104, USA

^cNeuroscience, U. Pennsylvania, Philadelphia, PA 19104, USA

^dBioengineering, U. Pennsylvania, Philadelphia, PA 19104, USA

Abstract

A listener's capacity to discriminate between sounds is related to the amount of acoustic variability that exists between these sounds. However, a full understanding of how this natural variability impacts neural activity and behavior is lacking. Here, we tested monkeys' ability to discriminate between different utterances of vocalizations from the same acoustic class (i.e., *coos* and *grunts*), while neural activity was simultaneously recorded in the anterolateral belt region (AL) of the auditory cortex, a brain region that is a part of a pathway that mediates auditory perception. Monkeys could discriminate between *coos* better than they could discriminate between *grunts*. We also found AL activity was more informative about different *coos* than different *grunts*. This difference could be attributed, in part, to our finding that *coos* had more acoustic variability than *grunts*. Thus, intrinsic acoustic variability constrained the discriminability of AL spike trains and the ability of rhesus monkeys to discriminate between vocalizations.

1. Introduction

The ability of humans and other animals to discriminate between different sound exemplars from the same category (e.g., a particular species-specific vocalization that is elicited by different callers) is limited by the amount and type of acoustic variability that exists between these different exemplars. For example, subtle changes in a vocalization's frequency modulation and pitch enable a monkey to discriminate between certain vocalizations like *coos* (Le Prell and Moody, 1997; Le Prell and Moody, 2000; Zoloth et al., 1979) or to identify a caller (Ghazanfar et al., 2007; Hauser, 1991; Owren and Rendall, 2003; Owren et al., 1992; Rendall et al., 1998). In contrast, the small amount of acoustic variability between *scream* vocalizations limits the ability of rhesus monkeys to discriminate between callers (Le Prell et al., 2002; Owren and Rendall, 2003; Rendall et al., 1998).

Despite the clear importance in understanding the relationship between acoustic variability to behavior, relatively little research (Bao et al., 2013; Bathellier et al., 2012; Liu and Schreiner, 2007), particularly in non-human primates, has investigated how this natural acoustic variability modulates and constrains both neural activity and behavior. In the current study, we quantified the acoustic variability of rhesus vocalizations and simultaneously examined how this variability affected behavioral and neural discriminability (Bennur et al., 2013). Monkeys participated in a task requiring them to discriminate between utterances of vocalizations from the same acoustic class; specifically, *coos* or *grunts*. We found that monkeys were better able to discriminate between *coos* than *grunts*.

While monkeys participated in this task, we recorded spiking activity from neurons in the anterolateral belt (AL) region of the auditory cortex. AL is part of the ventral auditory stream, which is thought to mediate auditory perception, and is known to be modulated strongly by vocalizations (Bizley and Cohen, 2013; Kusmirek et al., 2012; Rauschecker, 2012; Rauschecker and Tian, 2000; Rauschecker and Scott, 2009; Romanski and Averbeck, 2009; Russ et al., 2008b; Tsunada et al., 2011, 2012). Complementary to our behavioral findings, we found that the spiking activity of AL neurons discriminated between *coos* better than it discriminated between *grunts*; that is, AL activity was “more informative” about different *coos* than different *grunts*. This difference between neural discriminability could be attributed, in part, to our finding that *coos* had more acoustic variability than *grunts*. Thus, intrinsic acoustic variability constrained the discriminability of AL spike trains and the ability of rhesus monkeys to discriminate between vocalizations.

2. Materials and methods

The University of Pennsylvania Institutional Animal Care and Use Committee approved the experimental protocols.

2.1. Experimental chamber

Behavioral training and recording sessions were conducted in a darkened room with sound-absorbing foam and with acoustically shielded walls. During these sessions, a monkey (*Macaca mulatta*; Monkey H or Monkey B; both monkeys were adult males) was seated in a primate chair in the center of the room. A speaker (model MSP7, Yamaha) was placed in front of the monkey at a distance of 1.5 m and at eye level. A touch-sensitive lever was attached to the primate chair.

2.2. Auditory stimuli

The stimuli for this experiment were different utterances of vocalizations from two different acoustic classes: *coos* and *grunts*. 42 *grunts* and 31 *coos* were culled from a library of vocalizations (Hauser, 1997). Each vocalization originated from different callers, all of whom were unfamiliar to Monkeys H and B. Although, it is unlikely that the vocalization library encompassed the entire range of the naturally occurring acoustic variability (Egnor and Hauser, 2004), because each vocalization was recorded from a different caller, it is reasonable to assume that the library encompassed a substantial amount of this natural variability.

Prior to the beginning of behavioral training, one *coo* vocalization (i.e., “C1”) and one *grunt* vocalization (i.e., “G1”) was selected to be the “reference” *coo* and reference *grunt*, respectively. A multidimensional-scaling analysis (data not shown), which examined the (dis)similarity between the modulation spectra (Cohen et al., 2007) of each vocalization class, indicated that both reference vocalizations were well within the 95% confidence intervals of their respective distributions. Also, because the reference vocalizations were not at the center of their respective distributions, any potential response bias due to a “perceptual-magnet” effect—if it exists in rhesus (Kuhl, 1991)—was not a likely confounding issue.

For both monkeys H and B, *coo* C1 and *grunt* G1 were the same vocalizations for all training, behavioral, and experimental sessions that are described below. The remaining 30 “non-reference” *coos* (i.e., C2 – C31) and 41 “non-reference” *grunts* (i.e., G2 – G42) are generically referred to as Cx or Gx, respectively.

The digitized vocalizations were presented through a Tucker–Davis Technologies acoustic system that included a D/A converter (sampling rate = 50 kHz), an anti-aliasing filter, and amplifier. The Yamaha speaker was calibrated so that it was flat (± 2 dB) over the frequency range of the vocalizations. Each vocalization was presented at a sound level of 65 dB SPL.

2.3. Behavioral task

Initially, the monkeys participated in a task that required them to discriminate between the reference *coo* and the reference *grunt* (i.e., *coo* C1 and *grunt* G1, respectively). The final version of the task tested the behavioral and neural correlates underlying the monkey’s capacity to discriminate between reference and non-reference vocalizations that came from the same vocalization class. This version of the task was used during all of the neural-recording sessions that are presented in this study. For both tasks, the monkeys released a lever following offset of a vocalization to indicate their choice.

2.3.1. Initial task—As schematized in Fig. 1, each monkey participated in a variant of a two-interval discrimination task (Russ et al., 2008a; Tsunada et al., 2011). The task began with the monkey grasping the touch-sensitive lever that was attached to the primate chair (Cohen et al., 2009). Next, a “sample” stimulus was presented: this stimulus was either *coo* C1 or *grunt* G1 (i.e., a reference vocalization). Following a delay of 775–800 ms, a “test” stimulus was presented; the test stimulus was also one of the reference vocalizations. When the sample and test stimuli were different (e.g., the sample stimulus was *coo* C1 and the test stimulus was *grunt* G1), the monkey was required to wait for offset of the test stimulus and then could release the lever (within 500 ms) to receive a juice reward. However, when the two stimuli were the same (e.g., both were *coo* C1), the monkey maintained his grip on the lever. After another delay of 775–800 ms, a third stimulus, which was the other reference vocalization (i.e., *grunt* G1), was presented. Following offset of this vocalization, the monkey could release the lever to receive a reward. This third stimulus was presented so that the monkey needed to perform an action (i.e., release the lever) on every trial to get a reward. A trial was aborted if the monkey released the lever during the presentation of any vocalization.

2.3.2. Final task—Next, this task was modified so that we could test the monkeys' ability to discriminate between the non-reference *coo* and *grunts*. As noted above, this version of the task was used during all of the behavioral and neural-recording sessions that are presented in this study. For ~70% of the trials, the task structure was identical to that described above in Initial task (see Fig. 1B lower row and compare with Fig. 1A).

For the remaining ~30% of trials, a different task structure was used (see Fig. 1B upper and middle rows). During these trials, the sample stimulus was still a reference vocalization, but the test stimulus was a non-reference vocalization. The non-reference vocalization was constrained to be from the same vocalization class as the reference vocalization.

Specifically, when the sample stimulus was *coo* C1, the test stimulus was one of the remaining 30 *coo* non-reference vocalizations (i.e., C2 – C31). When the sample stimulus was *grunt* G1, the test stimulus was one of the remaining 41 *grunts* (i.e., G2 – G42).

As before, these trials began with the monkey grasping the touch-sensitive lever and the presentation of a reference vocalization as the sample stimulus. Following a delay of 775–800 ms, a non-reference vocalization was presented as the test stimulus. If the monkey released the lever following offset of the (test-stimulus) non-reference vocalization, the trial was over, and the monkey received a reward on a randomly selected 50% of trials (Fig. 1B upper row). The decision to give the monkey a reward was independent of his behavioral response. Because the monkeys did not receive consistent reinforcement on these trials, the trials were infrequent (~30% as noted above) in order to maintain the monkeys' motivation. We chose this strategy because we did not want to explicitly train the monkey to report that any particular non-reference vocalization (i.e., a *coo* Cx or a *grunt* Gx) was the same as or different from the reference vocalization (i.e., *coo* C1 or *grunt* G1). That is, the goal of this portion of the behavioral task was to allow monkeys to “spontaneously” (Bennur et al., 2013; Gifford et al., 2005) report whether the two vocalizations were the same or different.

If the monkey maintained his grip on the lever following offset of the test stimulus, a third vocalization was presented (Fig. 1B middle row). This third stimulus was always a reference vocalization but from a different vocalization class than the sample stimulus (e.g., if the sample stimulus was *coo* C1, the third stimulus would be *grunt* G1). Following offset of this stimulus, the monkey could release the lever to receive a reward.

To clarify the type of trial being discussed, we refer to a trial based on the combination of the sample and test stimuli. When both the sample and test stimuli were the reference *coo* C1, we refer to the trial as “C1C1”. However, when the test stimulus was not a reference vocalization (i.e., Cx), we refer to the trial as “C1Cx”. Trials with *grunts* have a corresponding analogous nomenclature.

2.4. Recording sites

We recorded from well-isolated neurons in the anterolateral belt (AL) region of the auditory cortex of monkeys H and B. AL is part of the ventral auditory stream, which is thought to mediate auditory perception, and is known to be modulated strongly by vocalizations (Bizley and Cohen, 2013; Kusmirek et al., 2012; Rauschecker, 2012; Rauschecker and Tian, 2000; Rauschecker and Scott, 2009; Romanski and Averbeck, 2009; Russ et al.,

2008b; Tsunada et al., 2011, 2012). Under isoflurane anesthesia and using aseptic surgical techniques, each monkey was implanted with a head-positioning post and a recording chamber. AL recordings were obtained from Monkey H's left hemisphere and Monkey B's right hemisphere. AL was identified anatomically by pre- and postoperative magnetic-resonance images of each monkey's brain and functionally by their response properties (Russ et al., 2008b; Tsunada et al., 2011).

2.5. Recording Procedure and Strategy

Prior to each recording session, a subset of non-reference vocalizations (typically, 5 from each vocalization class [*grunts* and *coos*]) was randomly selected as test stimuli. This subset was chosen to ensure that, for each recorded neuron, we had sufficient number of trials per test stimulus to perform meaningful statistical analyses on the neural and behavioral data sets. Next, the electrode was advanced into AL. Once spiking activity was found, the monkey participated in trials of the final version of the vocalization-discrimination task (see Final task in Behavioral Task for more details); the vocalization class (i.e., *coo* or *grunt*) of the reference stimulus and the non-reference vocalizations varied on a trial-by-trial basis. Neurons were not selected based on their (a) sensitivity to particular *coos* or *grunts* nor (b) frequency tuning or other response properties. This strategy was justified because AL neurons respond broadly to different types of auditory stimuli (Kusmierek et al., 2012; Rauschecker and Tian, 2000; Russ et al., 2008b; Tsunada et al., 2011, 2012).

Single-unit extracellular recordings were obtained with a tungsten microelectrode (~1.0 M Ω @ 1 kHz; Frederick Haer & Co.) or a 4-core-multifiber microelectrode ("tetrode"; ~0.8 M Ω @ 1 kHz; Thomas Recording GmbH) that was seated inside a stainless-steel guide tube. Extracellular neural signals were sampled at 24 kHz, amplified, and filtered (0.6–6.0 kHz) with a multi-channel recording system (Tucker–Davis Technologies). Spiking activity from individual neurons was extracted offline and sorted with a spike-sorting algorithm (WaveClus) (Quiroga et al., 2004; Tsunada et al., 2011, 2012) that ran on the Matlab (The Mathworks Inc.) programming platform.

2.6. Analyses

Behavioral and neural analyses were limited only to those trials when the monkey correctly followed the task structure: those trials when the monkey released the lever following offset of the test stimulus or offset of the third stimulus (see Fig. 1). Trials were not analyzed if the monkey did not release the lever nor were trials analyzed if the monkey responded early (i.e., prior to offset of the test stimulus or prior to offset of the third vocalization).

2.6.1. Behavioral analyses—The monkeys' performance was calculated independently for each reference vocalization (*coo* C1 or *grunt* G1). Performance is presented as a function of the test stimulus and the probability (p ; i.e., the average percentage of trials) that the monkeys reported that the sample and test stimuli were different. The 95%-confidence interval on each probability value was calculated using the following formula: $1.96 * (p * (1 - p) / n)^{0.5}$; n was the number of trials (Zar, 1996). A proportion test determined whether a particular probability value was significantly different than chance (50%) at a level of $p <$

0.05. A proportion test also determined whether two probability values were significantly different, also at a level of $p < 0.05$.

2.6.2. Neurophysiological analyses—Neural activity was tested during the “baseline”, “sample”, or “test” periods. The baseline period was the 500-ms period that directly preceded the sample stimulus. The sample period was the time period when the sample stimulus was presented. The test stimulus was the time period when the test stimulus was presented. A neuron’s “firing rate” (spikes/second) was the number of action potentials divided by the duration of a given analysis period.

We only quantified those neurons from which we were able to collect data from a minimum of 120 successful trials of different combinations of sample and test stimuli. Further, the neurons reported in this study were those classified as “auditory”. An AL neuron was “auditory” if its firing rate during the baseline period was significantly different ($p < 0.05$, t -test, Bonferroni corrected for multiple comparisons) from either (a) the firing rate during the sample period or (b) the firing rate during the test period. A receiver-operating characteristic (ROC) analysis (Green and Swets, 1966) and a metric-space analysis (Victor and Purpura, 1996, 1997) quantified the degree to which AL activity was informative about different *coos* and different *grunts*; these two analyses are discussed below.

2.6.3. ROC analysis—To test how the mean firing rates of each neuron were modulated by the sample and test stimuli, signal-detection theory was applied to generate an ROC curve (Green and Swets, 1966). For each neuron, two distributions of firing rates were constructed. One distribution contained the trial-by-trial firing rates generated during the sample-stimulus period. The second distribution contained the trial-by-trial firing rates generated during the test-stimulus period. From these two distributions, an ROC curve was generated. The area under this curve represented the probability that an ideal observer could differentiate between the two distributions. An ROC value of 0.5 indicates that the two distributions overlap completely and that an ideal observer can only differentiate between these distributions by chance. An ROC value of 1.0 indicates that the two distributions do not overlap and that an ideal observer can perfectly differentiate between the sample- and test-period firing rates.

2.6.4. Metric-space analysis—A metric-space analysis (Victor and Purpura, 1996, 1997) quantified the difference between a pair of spike trains through a stepwise transformation of one spike train into the second spike train. This analysis was restricted to spike trains elicited during the test-stimulus period and those from the same vocalization class. Each step of this transformation consisted of inserting or deleting a spike that was associated with a cost: the cost of inserting or deleting a spike had a value of 1. Effectively, this meant that the metric-space value was equal to the difference between the number of spikes in the two spike trains; we did not consider the effect of spike timing. For each neuron, we calculated the metric-space value in response to the neural activity that was elicited by each pairwise combination of test-stimulus vocalizations. Metric-space values from individual neurons were averaged as a function of each combination of vocalizations, and we report the average value. Because we chose vocalizations randomly, not all pairwise combinations were tested.

2.6.5. Acoustic analyses—A variety of acoustic analyses were utilized to test for differences between *coos* and between *grunts*.

2.6.6. Dynamic-time-warping analysis—The acoustic difference between two vocalizations was quantified using a dynamic time warping (DTW) algorithm (Ellis, 2003; Turetsky and Ellis, 2003) that was instantiated in the Matlab programming environment. The input to this algorithm was the spectrogram of each vocalization. The algorithm calculated, in an optimal fashion, the minimum number of steps needed to transform one vocalization into the second vocalization; the number of steps was proportional to the degree of dissimilarity between the two vocalizations.

2.6.7. Correlation between modulation spectra—The degree of similarity between a pair of vocalizations was also examined by calculating the correlation between their modulation spectra (Cohen et al., 2007; Singh and Theunissen, 2003). This analysis was appropriate because neurons in the auditory pathway are sensitive to both spectral and temporal modulations (Cohen et al., 2007; Klein et al., 2000; Miller et al., 2002; Singh and Theunissen, 2003; Woolley et al., 2005). To calculate the modulation spectra, we first calculated each vocalization's spectrogram. This spectrogram was generated from a bank of 229 Gaussian-shaped filters (center frequencies between 32 Hz and 10 kHz) whose gain function had a bandwidth of 32 Hz (measured as a standard deviation). The modulation spectrum was the average of the power (amplitude squared) of the two-dimensional Fourier transform of this spectrogram. This analysis was conducted in Matlab.

2.6.8. Analysis of acoustic features—Finally, we extracted a number of acoustic features from the vocalizations and examined how these different features covaried. These features were selected because they were used in previous vocalization studies (Kusmierek et al., 2012; Tchernichovski et al., 2000). The duration and Wiener entropy of each vocalization was derived using the Sound Analysis Pro program (Tchernichovski et al., 2000). Wiener entropy is a measure of the spectrotemporal structure of an auditory stimulus. As the bandwidth of an auditory stimulus increases from narrow band to broadband, its entropy value increases from a large negative number and approaches zero. The standard deviation of intensity (i.e., the variability in a vocalization's intensity), the spectral center of gravity (which measured, on average, the maximum frequency in the vocalization's spectrum), and the harmonic-to-noise ratio (i.e., “harmonicity”, which represents a vocalization's mean degree of acoustic periodicity) were calculated using the Praat software platform (Boersma, 1993, 2001).

3. Results

3.1. Rhesus monkeys discriminate between coos more reliably than between grunts

Monkeys participated in a two-interval discrimination task in which they reported whether a sample and test stimulus were the same or different; see Fig. 1 and Behavioral Task in Material and Methods for more details. The (first) sample stimulus was always a “reference” vocalization (either the *coo* C1 or the *grunt* G1); these reference vocalizations (C1 and G1) were the same in all of the behavioral and recording sessions. The (second) test stimulus could be either a reference vocalization or a “non-reference” vocalization; a non-reference

vocalization is generally referred to as *coo* Cx or *grunt* Gx. The behavioral results were generated from 128 recording sessions. On each recording day, we randomly selected a subset of non-reference vocalizations (typically, 5 *grunts* and 5 *coos*) as test stimuli. This subset was chosen to ensure that, for each recorded neuron, we had sufficient number of trials per test stimulus to perform meaningful statistical analyses on the neural and behavioral data sets. The data from each behavioral session were averaged together and Fig. 2 shows the average behavior of both monkeys; the results were averaged because their behavior was comparable.

When the sample and test stimuli were the same reference vocalizations (C1C1 (Fig. 2A) or G1G1 (Fig. 2B) trials), the monkeys reliably reported that these stimuli were the same: on average, the monkeys' reports of "different" were significantly (proportion test, $p < 0.05$) less than chance (i.e., 50%). In contrast, when the monkeys engaged in C1G1 or G1C1 trials (i.e., trials when the sample and test stimuli were different reference vocalizations; data not shown), their reports of "different" were significantly greater than chance (C1G1: 94% [95% confidence interval = $\pm 3.4\%$]; G1C1: 93% [95% confidence interval = $\pm 6.5\%$]; proportion test, $p < 0.05$).

Our main interest was trials in which the test stimulus was a non-reference vocalization—C1Cx or G1Gx trials. During these trials, for all of the tested non-reference vocalizations, the monkeys reports of "different" were significantly (proportion test, $p < 0.05$) greater than chance. However, the monkeys' reports of "different" were a function of the vocalization stimulus. Specifically, for the vast majority of C1Cx trials, the reports of "different" averaged 80% (Fig. 2A). In contrast, during G1Gx trials, the reports of "different" ranged between 59% and 90% (Fig. 2B). Indeed, on average, the monkeys more often reported (t -test, $p < 0.05$) that test stimuli were different during C1Cx trials than during G1Gx trials. Additionally, the variability in the responses of "different" was smaller (F-test for equality of variances, $p < 0.05$) for C1Cx trials than for G1Gx trials. Thus, the monkeys could discriminate between *coos* better than they could discriminate between *grunts*. Finally, we could not identify any pattern that correlated response times with their behavioral reports; this is not surprising because the monkeys were not trained nor were they encouraged to respond as fast as possible.

3.2. AL spiking activity was more informative about different coos than different grunts

While monkeys were participating in this vocalization-discrimination task, we simultaneously recorded single-neuron spiking AL activity. We recorded from 257 AL neurons (monkey B: 75 neurons, monkey H: 182 neurons). As noted in Material and Methods, we did not (a) select neurons based on their sensitivity to particular *coos* or *grunts*; nor did we (b) tailor the sample and test stimuli to a neuron's particular frequency tuning or other response properties.

Of these 257 neurons, 246 were classified as "auditory". A neuron was "auditory" if its mean firing rate during either the sample-stimulus or the test-stimulus period was significantly (t -test, $p < 0.05$, Bonferroni corrected for multiple comparisons) different from its mean firing rate during the baseline period. Two example auditory neurons are shown in Fig. 3. Fig. 3A shows an example AL neuron that was recorded during C1Cx trials, whereas

Fig. 3B shows an example neuron that was recorded during G1Gx trials. These two neurons illustrate some of the diversity in neural responsivity that we encountered. The AL neuron in Fig. 3A was not responsive to the sample stimulus (C1) but was modulated robustly by all of the test stimuli. However, it appeared have a graded response to each these *coo* test stimuli: its response to *coo* C3 was highest, the response to *coo* 27 was lowest, and *coo* C2 had an intermediate response. The neuron in Fig. 3B had a different response profile. This neuron responded to both sample and test *grunt* stimuli. This neuron seemed to respond in a “binary” fashion. Its response to *grunts* G1, G19, and G24 was similar but lower than the similar responses elicited by *grunts* G17, G30, and G31.

Of these 246 auditory neurons, 137 had significantly (t -test, $p < 0.05$) different firing rates in response to *coos* than to *grunts* during either the sample- or test-stimulus period. A second population of 67 neurons had significantly (t -test, $p < 0.05$) different firing rates in response to *grunts* than to *coos*. A third population of neurons were auditory but were not selectively modulated by either *coos* or *grunts*. We could not identify any other differences between these populations of neurons and, as a consequence, they were grouped together for subsequent analyses. Because we collected neural activity in response to ~5 non-reference vocalizations per recording session, we report data in terms of population analyses; see Recording Procedure and Strategy in Materials and Methods for more details.

3.2.1. ROC analysis—In the first analysis, we calculated ROC values (Green and Swets, 1966) for each neuron. The ROC value quantified the difference between the firing-rate distributions that were elicited during the sample-stimulus and test-stimulus periods. An ROC value of 0.5 indicates that an ideal observer can only differentiate between these distributions by chance. ROC values >0.5 indicates that an ideal observer can differentiate between these distributions.

Fig. 4A shows the distributions when reference *coo* C1 was the sample stimulus, whereas Fig. 4B shows the distribution when reference *grunt* G1 was the sample stimulus. If we first consider the C1C1 and G1G1 trials (black histograms in Figs. 4A and B, respectively), we find that the median ROC values were close to 0.5 (C1C1: 0.58 and G1G1: 0.60) but still significantly (Mann–Whitney test, $p < 0.05$) higher than this chance value of 0.5.

The ROC distributions for C1Cx and G1Gx trials are the grey histograms in Fig. 4A and B, respectively. The median value was 0.81 for C1Cx trials, and the median value for G1Gx trials was 0.73. Each median value was significantly (Mann–Whitney test, $p < 0.05$) higher than 0.5 and significantly higher than the respective C1C1 and G1G1 ROC values. The median C1Cx value was also significantly (Mann–Whitney test, $p < 0.05$) higher than the median G1Gx value.

Finally, we found that ROC values during C1G1 and G1C1 trials (white histograms in Figs. 4A and B, respectively) were significantly (Mann–Whitney test, $p < 0.05$) higher than 0.5. The median C1G1 value was 0.85 was significantly (Mann–Whitney test, $p < 0.05$) higher than the median G1C1 value (0.72). Each of these values was also significantly higher than its respective C1C1 and G1G1 ROC value. Additional analyses could not identify the basis for this difference between the C1G1 and G1C1 ROC values.

Overall, these data indicate that AL neurons responded differentially to the sample and test stimuli. However, this response difference was magnified when either the sample or test stimulus was a *coo*. Thus, there was more neural discriminability for *coo*-elicited AL activity than for *grunt*-elicited activity. In other words, *coo*-elicited activity was more informative than *grunt*-elicited activity.

3.2.2. Metric-space analysis—Next, we conducted a metric-space analysis (Victor and Purpura, 1996, 1997) to further quantify the relationship between AL activity and the vocalizations. This analysis quantified the difference between pairs of spike trains by a step-wise transformation of one spike train into the second spike train. The larger the metric-space value, the more dissimilar (different) were the two spike trains; see Material and Methods for more details.

Fig. 5 shows the results of the metric analysis when we considered differences between the spike count that was elicited by two different vocalizations. The results show that *coo*-elicited neural activity had larger metric-space values (i.e., were more dissimilar or “more informative”) than *grunt*-elicited activity. Indeed, the mean *coo* value was 22 (std = 17.3), whereas the mean *grunt* value was 16 (std = 14.4). These values were significantly (*t*-test, $p < 0.05$) different from one another and suggest that *coo*-elicited neural activity was more informative than *grunt*-elicited neural activity.

This metric-space analysis can also take the timing of neurons into account when quantifying spike-train dissimilarity. When we conducted the metric-space analysis and accounted for the relative timing of action potentials (with a precision of 1 ms), a similar pattern emerged to that when using rate-code information alone: the spike trains elicited by *coos* (mean = 42.2; std = 29.7) were significantly (*t*-test, $p < 0.05$) more dissimilar than those for *grunts* (mean = 37.2; std = 23.4).

3.3. Coos had more acoustic variability than grunts

Next, we quantified the acoustic variability that existed between *coos* and between *grunts*. We then correlated this variability with specific acoustic features of the vocalizations.

3.3.1. Dynamic-time-warp analysis (DTW)—The first analysis calculated the degree of acoustic similarity between a reference vocalization and each within-class non-reference vocalizations using the DTW algorithm (Ellis, 2003; Turetsky and Ellis, 2003); see Materials and Methods for more details. As the DTW value increased, it indicated that the reference vocalization and a non-reference vocalization became increasingly dissimilar.

Fig. 6A shows the distribution of DTW values for the 31 *coos* and the spectrograms for three select *coos*. Some of the differences between these vocalizations become apparent when comparing their spectrograms. The durations of all three vocalizations were different: the reference *coo* C1 was shorter than the other two non-reference *coos*. Moreover, we can identify differences between the spectral properties of reference *coo* C1 and the two non-reference *coos*. For example, the two non-reference *coos* had more power in the lower frequencies and less power in higher frequencies than reference *coo* C1.

The distribution of DTW values for the 42 *grunts* and the spectrograms for three select *grunts* is shown in Fig. 6B. Like with the *coos*, differences emerged between these different *grunts*. For example, reference *grunt* G1 and *grunt* G20 were more pulsatile than the *grunt* G15. Also, the power spectrum of *grunt* G15 was more broadband than the other two *grunts*.

Next, we quantified the distributions of DTW values. The median DTW value of the non-reference *coos* was 82.6 (range: 17.2–151.8; coefficient of variation = 0.50), and the median DTW value of the non-reference *grunts* was 14.9 (range: 9.8–31.5; coefficient of variation = 0.29). The non-parametric Leven's test for equality of variances (Levene, 1960) indicated that the *coo* distribution of DTW values had significantly ($p < 0.05$) more variance than the *grunt* distribution, indicating that *coos* had more acoustic variability than *grunts*.

3.3.2. Modulation-spectra correlation—Next, we calculated the correlation coefficient between two vocalizations' modulation spectra. The modulation spectrum characterizes the spectrotemporal modulations present in a vocalization (Singh and Theunissen, 2003; Theunissen et al., 2004).

For illustrative purposes, we computed the average modulation spectrum for the *coos* and for the *grunts*. The average *coo* and *grunt* modulation spectra are shown in the lower left and upper right, respectively, in Fig. 7. For both vocalization classes, most of the modulation-spectra energy occurred at low spectral and temporal frequencies. These spectra had high power at low modulation frequencies that rapidly decreased at higher frequencies. Further, most of the power for the medium to higher spectral modulation frequencies was found at only the very lowest temporal modulations. These patterns are characteristic of all natural sounds (Singh and Theunissen, 2003) including rhesus vocalizations (Cohen et al., 2007).

The differences between the modulation spectra of different *coos* was quantified by computing the correlation between their modulation spectra; a similar analysis was conducted for *grunts*. The results of this analysis are plotted in a matrix format in Fig. 7. Overall, the correlation coefficients between different *coos* and different *grunts* were relatively high (mean ~0.7–0.8), indicating that these vocalizations had a high degree of similarity. However, a more detailed examination indicated that, on average, *coos* had lower correlation coefficients than *grunts* (t -test, $p < 0.05$), indicating more variability between *coos* than between *grunts*. A similar result was obtained when we tested the correlation between the log frequency spectrograms, which was generated as part of the process of modulation-spectra computation (data not shown).

3.3.3. Analysis of acoustic features—The next series of analyses evaluated a number of acoustic features in order to identify how these features contributed to the variability between *coos* and between *grunts*; see Material and Methods for more details. Specifically, we tested a vocalization's duration, Wiener entropy, standard deviation of intensity, spectral center of gravity, and mean harmonicity. Because *grunts* are noisy and have very little harmonic structure, we did not calculate their harmonicity.

The mean values of these acoustic features appear in Table 1. F-tests for equality of variances indicated that the *coo* distributions of duration, entropy, and standard deviation of

the intensity had significantly ($p < 0.05$) more variance than those of their respective *grunt* distributions. Only the *grunt* distribution of spectral center of gravity had significantly ($p < 0.05$) more variance than the *coo* distribution.

As expected, several of these acoustic features were correlated with one another (Fig. 8). Specifically, for the *coos*, the standard deviation of the intensity and the duration positively correlated ($r = 0.54$, $p < 0.05$; Fig. 8A). Additionally, we found significant correlations between the entropy and the spectral center of gravity ($r = 0.51$, $p < 0.05$; Fig. 8B) and between the entropy and the mean harmonicity ($r = -0.5$, $p < 0.05$; Fig. 8C). For the *grunts*, the only significant correlation we could identify was between the entropy and the spectral center of gravity ($r = 0.54$, $p < 0.05$; Fig. 8D).

Because many of the acoustic features were correlated, we conducted a principal-component analysis (PCA) to gain further insight into the relationship between these features and vocalization variability. The PCA transformed these correlated acoustic features into a set of uncorrelated variables that allowed an identification of the largest source of variance in this set of feature values.

The first principal component accounted for 70% of the variance between *coos*. Both the mean harmonicity ($r = 0.71$) and the entropy ($r = -0.63$) of the *coos* correlated strongly with this first principal component. This means that most of the variance between *coos* was explained by differences between their harmonicity, or equivalently, their entropy. The second principal component accounted for 24% of the variance. The duration of the vocalization ($r = -0.61$) and the standard deviation of the intensity ($r = -0.77$) accounted for most of the variance in the second principle component suggesting that *coos* can be well differentiated based on the values of these two acoustic features.

A second PCA analysis indicated that the variance between the *grunts* was more evenly distributed than the variance between the *coos*. Here, the first principal component accounted for only 55% of the variance. The spectral center of gravity correlated best ($r = -0.63$) with the first principal component; the standard deviation of the intensity ($r = -0.49$) and the entropy ($r = 0.57$) were also correlated but to a lesser degree. Thus, *grunts* can be differentiated by their average maximum frequency as well as the variance in their intensity and the degree to which they are tonal. The second principal component accounted for 31% of the variance and the *grunt* duration accounted for much of this variance ($r = -0.72\%$).

Together, these three sets of analyses indicated that, on average, *coos* had more acoustic variability than *grunts*.

3.4. Correlation between behavioral responses, neural responses, and acoustic variability

Finally, we conducted a series of analyses to determine whether we could identify relationships between the acoustic variability and (a) our finding that monkeys discriminated between *coos* better than *grunts* (Fig. 2); and (b) our finding that AL activity was more informative for *coos* than it was for *grunts* (see Figs. 4 and 5).

Fig. 9 plots the behavioral data (shown independently in Fig. 2) but as a function of the DTW value (see Fig. 6). This serves to visualize whether the monkeys' behavioral reports correlated with the degree of the acoustic similarity between the reference and non-reference vocalizations. The monkeys' behavior did not correlate with DTW value. That is, their reports of "different" did not increase as the acoustic dissimilarity between the reference vocalization and a non-reference exemplar increased. Organizing the behavioral responses as a function of the modulation-spectra correlations or different acoustic-feature values did not meaningfully alter this pattern of the behavioral reports.

We also considered the relationship between the variability in the acoustic features of the vocalizations and AL spiking activity. The first analysis undertook a population-level analysis in which we correlated the test-period firing rate with each of the quantified acoustic features. None of these correlations yielded any significant ($p > 0.05$) trends; see Figs. 10 and 11. Moreover, other correlations between neural activity and behavior (e.g., correlations between the metric-space values for a C1Cx [or a G1Gx] trial and behavioral performance) failed to yield significant ($p > 0.05$) relationships.

However, a different finding emerged when the metric-space values (Fig. 5)—which quantified the degree of the spike-train dissimilarity—were correlated with different measures of acoustic variability: in this analysis, we could account for some of the dissimilarity between AL spike trains by the variability in the *coo* and *grunt* distributions of acoustic features. Specifically, for each pair of vocalizations, we calculated the ratio between one of the quantified acoustic features (e.g., the ratio between the center of gravity of one test-stimulus *coo* and the center of gravity of a different test-stimulus *coo*) and correlated this value against the metric-space value that was generated from the neural activity that was elicited by these vocalizations. We also correlated this metric-space value against the corresponding modulation-spectra correlations and the log-frequency-spectra correlations (see Analysis of acoustic features).

The panels in Fig. 12 show these correlations. For the *coos*, we found that the ratio of the mean harmonicity (Fig. 12G) and the ratio of the standard deviation of the intensity (Fig. 12B) had small ($r = 0.13$ and $r = 0.12$, respectively) but significant ($p < 0.05$; Bonferroni corrected) correlations. Similarly, for the *grunts*, the center of gravity (Fig. 12H), the standard deviation of the intensity (Fig. 12I), the entropy (Fig. 12K), and the modulation-spectrum correlation (Fig. 12L) were significantly correlated ($r = -0.18$, $r = 0.15$, $r = 0.17$, $r = -0.15$, respectively; $p < 0.05$; Bonferroni corrected).

4. Discussion

We analyzed the acoustical features of a set of *coos* and *grunts* and found that, across a variety of different forms of analyses, *coos* had more acoustic variability than *grunts*. We also found that monkeys could discriminate between *coos* better (and more consistently) than they could discriminate between *grunts*. We also found that AL activity discriminated between different *coos* better than it discriminated between different *grunts*. These observations paralleled our observation that *coos* had more acoustic variability than *grunts*.

This more informative neural activity could be the neural basis by which monkeys are capable of discriminating between *coos* better than can discriminate between *grunts*.

4.1. Behavior

The monkeys reliably reported that within-class, non-reference vocalizations were different from the reference vocalization (Fig. 2). The probability that monkeys reported that two stimuli were “different” did not vary with the magnitude of the acoustic difference between the stimuli. However, the monkeys reported that non-reference *coos* were “different” more often and with more consistency than they reported that non-reference *grunts* were “different”. This finding is consistent with a previous behavioral study that found that rhesus monkeys have more difficulty in discriminating between callers producing *grunts* than between callers producing *coos* (Rendall et al., 1998).

Our pattern of results cannot be attributed wholly to the particular features of the selected reference vocalization. For example, if the reference *coo* was located near the edge of some measure of *coo* variability, then it would be reasonable to expect that the monkeys would report that the reference and non-reference *coos* were different. However, this not the case: as noted in Materials and Methods, a multi-dimensional-scaling analysis of the modulation spectra (Cohen et al., 2007) of each vocalization class indicated that both reference vocalizations were well within the 95% confidence intervals of their respective distributions. Also, because the reference vocalizations were not at the center of their respective distributions, any potential response bias due to a “perceptual-magnet” effect—if it exists in rhesus monkeys (Kuhl, 1991)—is unlikely.

Both our findings and those of Rendall et al. (1998) point to the idea that because *coos* had more acoustic variance than *grunts*, monkeys were better able to discriminate between *coos* than between *grunts*. Our findings are also consistent with previous work that suggested that the acoustic variability between *coos* is sufficient for a listener to individuate between callers (Hauser, 1991). Indeed, across a number of measures (Figs. 6 and 7), we demonstrated that there was more acoustic variability between *coos* than between *grunts*. For example, the correlation between the modulation spectra of pairs of *coos* was lower than between pairs of *grunts* (Fig. 7). Because *coos* have more acoustic variability than *grunts*, it follows that *coo* discriminations should be easier than *grunt* discriminations.

What was the basis for the monkeys’ behavioral strategy of consistently reporting “different”? The monkeys were initially trained (with consistent reinforcement) to report that the reference *coo* C1 and reference *grunt* G1 were different. In contrast, they did not receive consistent reinforcement when the test vocalization was a non-reference vocalization (C1Cx or G1Gx trials). That is, we wanted the monkeys to “spontaneously” (Bennur et al., 2013; Gifford et al., 2005) report whether the reference and non-reference vocalizations were the same or different. We hypothesized that this approach (Ghazanfar and Santos, 2004) would reveal information that would not be apparent had the monkeys been given consistent reinforcement whenever they reported that specific non-reference vocalizations were the same as or different from the reference vocalization.

Although this behavioral-testing strategy was intentional, it is possible that the monkeys' training history and reward schedule biased them toward always reporting that a non-reference vocalization was different. In other words, they might have learned a general strategy of reporting "different" whenever the reference and test stimuli were different, regardless of whether these two stimuli originated from the same or from different vocalization classes. However, previous work from our group suggests that these possibilities cannot account for all of the current findings: following training on two "reference phonemes" that were presented as sample or test stimuli, the monkeys did not always report that a novel "test-stimulus phoneme" was different (Russ et al., 2008a; Tsunada et al., 2011, 2012). Thus, the monkeys' behavioral reports were a function of the relationship between their training history, the auditory stimuli, and the task and not just the stimuli themselves.

Finally, would a similar behavioral pattern have emerged if, instead of testing the acoustic variability that was elicited by different callers producing the same vocalization, we had tested the acoustic variability elicited by the same caller making different utterances of a *coo* or *grunt*? Unfortunately, our library of vocalizations was limited and such comparisons cannot be drawn. A second, non-exclusive question is how these results would generalize to other vocalization classes. We only evaluated two classes of vocalizations but rhesus monkeys have many other vocalization classes (Hauser, 1998). We posit that the relationships reported here between *coos* and *grunts* are likely to hold for other vocalization classes (Fitch and Fritz, 2006; Ghazanfar et al., 2007; Hauser, 1991; Le Prell and Moody, 1997; Le Prell and Moody, 2000; Le Prell et al., 2002; Owren and Rendall, 2003; Owren et al., 1992; Rendall et al., 1998; Zoloth et al., 1979).

4.2. Neural data

Three important considerations with respect to the neural recordings should be discussed. First, our PCA analyses indicated that the greatest source of acoustic variability for *coos* was harmonicity (or entropy), whereas for *grunts*, it was the spectral central of gravity. A natural question to ask is whether AL activity was differentially sensitive to these features. Although different pools of neurons may be differentially sensitive to specific acoustic features (Averbeck and Romanski, 2004, 2006; Freeman et al., 2013; Kusmirek et al., 2012), our data could not identify any such relationship nor could we identify one in the literature. To address this question, AL activity should be recorded in response to synthetic vocalizations (Chakladar et al., 2008; DiMattina and Wang, 2006) that allow for systematic manipulations of these acoustic features.

Second, could differences in the variability of one low-level acoustic feature explain the differences in the neural data? One obvious candidate feature would be vocalization duration: *coos* have a much longer average duration and more variability (~483 ms; std = 164) than *grunts* (~206 ms; std = 54); see Table 1. Consequently, because *coos* have more acoustic variability than *grunts*, it may very well be that neural activity in response to *coos* may be more variable than *grunts*. Whereas this hypothesis may explain some proportion of our data, it cannot wholly explain our findings. Indeed, duration only explained a small fraction of the acoustic variance between *coos*: harmonicity and entropy accounted for far

more variance. Our neural findings also suggest that there is not a simple relationship between the firing rates of AL neurons and a particular acoustic feature: we could not identify any correlation between the test-period firing rate and any of our quantified acoustic features (Figs. 10 and 11) nor was there a significant correlation between duration and the metric-space values (Fig. 12). Furthermore, even though significant, our ROC analysis, which takes into account differences in mean and variance, found that the mean ROC C1Cx value was only somewhat bigger than the mean ROC G1Gx value (0.81 versus 0.73). Thus, it is unlikely that the variability in any one acoustic feature could wholly explain the differences that we found in firing rates between *coos* and *grunts*.

Third, how do our neural findings inform us on the role of AL in vocalization processing and auditory perception? Before addressing this question, it is important to put AL into context of auditory processing. AL is part of the ventral auditory pathway; a pathway that is thought to mediate non-spatial sound perception (i.e., “what is the sound?”) (Bizley and Cohen, 2013; Plakke et al., 2013b; Rauschecker and Tian, 2000; Rauschecker and Scott, 2009; Romanski, 2004; Romanski and Averbeck, 2009). Consequently, neurons in different regions of this pathway have been tested for their sensitivity to vocalizations (Averbeck and Romanski, 2004, 2006; Cohen et al., 2007; Gaucher et al., 2013; Gifford et al., 2005; Kikuchi et al., 2010; Kusmirek and Rauschecker, 2009; Kusmirek et al., 2012; Perrodin et al., 2011; Petkov et al., 2008; Plakke et al., 2013b; Poremba et al., 2013; Romanski and Goldman-Rakic, 2002; Russ et al., 2008b; Tian et al., 2001). In particular, AL neurons respond robustly to specific vocalizations and, relative to other control stimuli, respond preferentially to vocalizations, (Kusmirek et al., 2012; Rauschecker and Tian, 2000; Russ et al., 2008b; Tian et al., 2001). However, AL is not “specialized” for vocalization processing: AL neurons respond to a wide range of stimuli such as human-speech sounds and stochastic sequences of tone bursts (Kusmirek et al., 2012; Tsunada et al., 2011, 2012, 2013).

How does this sensitivity compare with the sensitivity seen in regions of the ventral pathway that are downstream from AL, such as the ventrolateral prefrontal cortex (vPFC)? Several studies have examined vocalization sensitivity in vPFC while monkeys are engaged in passive-listening tasks (i.e., a task in which vocalizations are presented to a monkey but the monkey does not have to use the vocalization to guide an action) (Averbeck and Romanski, 2004, 2006; Cohen et al., 2004; Plakke et al., 2013b; Romanski et al., 2005; Russ et al., 2008b). The results of these studies, in terms of neural sensitivity to different vocalizations, show a wide range of values; these differences may reflect subtle, yet significant, differences in recording locations, differences in neuron-selection strategies, or differences in analysis techniques.

However, given the general role of the PFC in executive function (Miller and Cohen, 2001), we posit that examining vPFC activity during such passive-listening tasks is not an appropriate assay of the computations that occur in this brain region. Instead, when vPFC activity is examined during active behavioral tasks or more complex stimulus-presentation paradigms, it becomes clear that vPFC activity is involved in those functions that utilize vocalization information (and other auditory signals) to help guide goal-directed behaviors, including those behaviors that require non-spatial attention, working memory, decision-

making, and abstract categorization (Gifford et al., 2005; Plakke et al., 2013a,b, 2012; Poremba et al., 2013; Russ et al., 2008a).

In the current study, we found that AL neurons were modulated by the acoustic differences that are inherent in *coos* and *grunts* (see Figs. 4 and 5). In other words, individual AL neurons do not code an invariant representation of a vocalization. These findings, however, do not preclude the possibility that, at the population level, AL contains a categorical (Engineer et al., 2013; Ohl et al., 2001; Selezneva et al., 2006; Tsunada et al., 2011) representation of a vocalization that is invariant to these spectrotemporal variations (DiCarlo et al., 2012). Such an invariant representation would, for example, code *coos* independent of differences in their spectrotemporal structures and independent of other “identity-preserving” transformations such as differences in the level of background noise, rate of vocal production, etc. (DiCarlo et al., 2012). Further study is needed to determine whether such a representation exists either in AL or in an area downstream of AL. Determining this is an important question for auditory neuroscience because it will shed light on how and where the brain constructs perceptual representations of auditory stimuli (i.e., “acoustic objects” or sounds) (Bizley and Cohen, 2013; Griffiths and Warren, 2004; Griffiths et al., 2004; King and Nelken, 2009).

Although representations of vocalizations that are invariant to some spectrotemporal variation may be important in some contexts, it may not always be appropriate or advantageous for listeners to disregard neural representations of variability, as this variability could be used to discriminate between callers (Ghazanfar et al., 2007; Hauser, 1991; Owren and Rendall, 2003; Owren et al., 1992; Rendall et al., 1998). Neural representations that represent the variations between callers are found in more anterior regions of the superior temporal gyrus that process vocalizer identity (Perrodin et al., 2011; Petkov et al., 2008; Poremba et al., 2004). Although the exact auditory-tuning properties of neurons in this anterior brain region are just beginning to be elucidated (Kikuchi et al., 2010), it is feasible that they are tuned to the specific acoustic features of an individual’s voice. Thus, the neural responses reported in this study may contribute to the computations that generate speaker-specific representations in this anterior region.

Acknowledgments

We thank Robert Seyfarth, Joji Tsunada, Marc Schmidt, James Saunders, Maria Geffen, and Heather Hersh for helpful comments on the preparation of this manuscript. We also thank Harry Shirley for outstanding veterinary support. KLCL, SB, and YEC were supported by grants from NIDCD-NIH (R01-DC009224) and from the Boucai Hearing Restoration Fund.

References

- 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]
- Averbeck BB, Romanski LM. Probabilistic encoding of vocalizations in macaque ventral lateral prefrontal cortex. *J Neurosci.* 2006; 26:11023–11033. [PubMed: 17065444]
- Bao S, Chang EF, Teng CL, Heiser MA, Merzenich MM. Emergent categorical representation of natural, complex sounds resulting from the early post-natal sound environment. *Neuroscience.* 2013; 248C:30–42. [PubMed: 23747304]

- Bathellier B, Ushakova L, Rumpel S. Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron*. 2012; 76:435–449. [PubMed: 23083744]
- Bennur S, Tsunada J, Cohen YE, Liu RC. Understanding the neurophysiological basis of auditory abilities for social communication: a perspective on the value of ethological paradigms. *Hear Res*. 2013; 305:3–9. [PubMed: 23994815]
- Bizley JK, Cohen YE. The what, where, and how of auditory-object perception. *Nat Rev Neurosci*. 2013; 14:693–707. [PubMed: 24052177]
- Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *IFA Proc*. 1993; 17:97–110.
- Boersma P. Praat, a system for doing phonetics by computer. *Glott Int*. 2001; 4:341–345.
- Chakladar S, Logothetis NK, Petkov CI. Morphing rhesus monkey vocalizations. *J Neurosci Methods*. 2008; 170:45–55. [PubMed: 18289695]
- Cohen YE, Theunissen F, Russ BE, Gill P. Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *J Neurophysiol*. 2007; 97:1470–1484. [PubMed: 17135477]
- Cohen YE, Russ BE, Gifford GW 3rd, Kiringoda R, MacLean KA. Selectivity for the spatial and nonspatial attributes of auditory stimuli in the ventrolateral prefrontal cortex. *J Neurosci*. 2004; 24:11307–11316. [PubMed: 15601937]
- Cohen YE, Russ BE, Davis SJ, Baker AE, Ackelson AL, Nitecki R. A functional role for the ventrolateral prefrontal cortex in non-spatial auditory cognition. *Proc Natl Acad Sci U S A*. 2009; 106:20045–20050. [PubMed: 19897723]
- DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? *Neuron*. 2012; 73:415–434. [PubMed: 22325196]
- DiMattina C, Wang X. Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations. *J Neurophysiol*. 2006; 95:1244–1262. [PubMed: 16207780]
- Egnor SE, Hauser MD. A paradox in the evolution of primate vocal learning. *Trends Neurosci*. 2004; 27:649–654. [PubMed: 15474164]
- Ellis, D. Dynamic Time Warp (DTW) in Matlab [Online]. 2003. <http://www.ee.columbia.edu/~dpwe/resources/matlab/dtw/>
- Engineer CT, Perez CA, Carraway RS, Chang KQ, Roland JL, Sloan AM, Kilgard MP. Similarity of cortical activity patterns predicts generalization behavior. *PLoS One*. 2013; 8:e78607. [PubMed: 24147140]
- Fitch WT, Fritz JB. Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *J Acoust Soc Am*. 2006; 120:2132–2141. [PubMed: 17069311]
- Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA. A functional and perceptual signature of the second visual area in primates. *Nat Neurosci*. 2013; 16:974–981. [PubMed: 23685719]
- Gaucher Q, Huetz C, Gourevitch B, Laudanski J, Occelli F, Edeline JM. How do auditory cortex neurons represent communication sounds? *Hear Res*. 2013; 305:102–112. [PubMed: 23603138]
- Ghazanfar AA, Santos LR. Primate brains in the wild: the sensory bases for social interactions. *Nat Rev Neurosci*. 2004; 5:603–616. [PubMed: 15263891]
- Ghazanfar AA, Turesson HK, Maier JX, van Dinther R, Patterson RD, Logothetis NK. Vocal-tract resonances as indexical cues in rhesus monkeys. *Curr Biol: CB*. 2007; 17:425–430.
- Gifford GW 3rd, 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 Cognitive Neurosci*. 2005; 17:1471–1482.
- Green, DM.; Swets, JA. *Signal Detection Theory and Psychophysics*. John Wiley and Sons, Inc; New York: 1966.
- Griffiths TD, Warren JD. What is an auditory object? *Nat Rev Neurosci*. 2004; 5:887–892. [PubMed: 15496866]
- Griffiths TD, Warren JD, Scott SK, Nelken I, King AJ. Cortical processing of complex sound: a way forward? *Trends Neurosci*. 2004; 27:181–185. [PubMed: 15046876]

- Hauser MD. Sources of acoustic variation in rhesus macaque (*Macaca mulatta*) vocalizations. *Ethology*. 1991; 89:29–46.
- Hauser, MD. *The Evolution of Communication*. MIT Press; Cambridge, MA: 1997.
- Hauser MD. Functional referents and acoustic similarity: field playback experiments with rhesus monkeys. *Anim Behav*. 1998; 55:1647–1658. [PubMed: 9642008]
- Kikuchi Y, Horwitz B, Mishkin M. Hierarchical auditory processing directed rostrally along the monkey's supratemporal plane. *J Neurosci*. 2010; 30:13021–13030. [PubMed: 20881120]
- King AJ, Nelken I. Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nat Neurosci*. 2009; 12:698–701. [PubMed: 19471268]
- Klein DJ, Depireux DA, Simon JZ, Shamma SA. Robust spectro-temporal reverse correlation for the auditory system: optimizing stimulus design. *J Comp Neurosci*. 2000; 9:85–111.
- Kuhl PK. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Percept Psychophys*. 1991; 50:93–107. [PubMed: 1945741]
- Kusmirek P, Rauschecker JP. Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J Neurophysiol*. 2009; 102:1606–1622. [PubMed: 19571201]
- Kusmirek P, Ortiz M, Rauschecker JP. Sound-identity processing in early areas of the auditory ventral stream in the macaque. *J Neurophysiol*. 2012; 107:1123–1141. [PubMed: 22131372]
- Le Prell CG, Moody DB. Perceptual salience of acoustic features of Japanese monkey coo calls. *J Comp Psychol*. 1997; 111:261–274. [PubMed: 9286095]
- Le Prell CG, Moody DB. Factors influencing the salience of temporal cues in the discrimination of synthetic Japanese monkey (*Macaca fuscata*) coo calls. *J Exp Psychol Anim Behav*. 2000:261–273. [PubMed: 10913991]
- Le Prell CG, Moody D, Hauser MD. Discrete or graded variation within monkey screams? Psychophysical experiments on classification. *Anim Behav*. 2002; 63:47–62.
- Levene, H. Robust tests for equality of variances. In: Olkin, I.; Ghurye, SG.; Hoeffding, W.; Madow, WG.; Mann, HB., editors. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford University Press; Palo Alto, CA: 1960. p. 278–292.
- Liu RC, Schreiner CE. Auditory cortical detection and discrimination correlates with communicative significance. *PLoS Biol*. 2007; 5:e173. [PubMed: 17564499]
- Miller E, Cohen JD. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci*. 2001; 24:167–202. [PubMed: 11283309]
- Miller LM, Escabi MA, Read HL, Schreiner CE. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J Neurophysiol*. 2002; 87:516–527. [PubMed: 11784767]
- Ohl FW, Scheich H, Freeman WJ. Change in pattern of ongoing cortical activity with auditory category learning. *Nature*. 2001; 412:733–736. [PubMed: 11507640]
- 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; 117:380–390. [PubMed: 14717639]
- Owren MJ, Dieter JA, Seyfarth RM, Cheney DL. ‘Food’ calls produced by adult female rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques, their normally-raised offspring, and offspring cross-fostered between species. *Behavior*. 1992; 120:218–231.
- Perrodin C, Kayser C, Logothetis NK, Petkov CI. Voice cells in the primate temporal lobe. *Curr Biol*. 2011; 21:1408–1415.
- Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK. A voice region in the monkey brain. *Nat Neurosci*. 2008; 11:367–374. [PubMed: 18264095]
- Plakke B, Ng CW, Poremba A. Neural correlates of auditory recognition memory in primate lateral prefrontal cortex. *Neuroscience*. 2013a; 244:62–76. [PubMed: 23583656]
- Plakke B, Diltz MD, Romanski LM. Coding of vocalizations by single neurons in ventrolateral prefrontal cortex. *Hear Res*. 2013b; 305:135–143. [PubMed: 23895874]
- Plakke, B.; Hwang, J.; Diltz, MD.; Romanski, LM. Inactivation of Ventral Prefrontal Cortex Impairs Audiovisual Working Memory. Program 878.10 2012 Neuroscience Meeting Planner. Society for Neuroscience; New Orleans, LA. 2012.

- Poremba A, Bigelow J, Rossi B. Processing of communication sounds: contributions of learning, memory, and experience. *Hear Res.* 2013; 305:31–44. [PubMed: 23792078]
- 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; 427:448–451. [PubMed: 14749833]
- Quiroga RQ, Nadasy Z, Ben-Shaul M. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput.* 2004; 16:1661–1687. [PubMed: 15228749]
- Rauschecker JP. Ventral and dorsal streams in the evolution of speech and language. *Front Evol Neurosci.* 2012; 4
- Rauschecker JP, Tian B. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci U S A.* 2000; 97:11800–11806. [PubMed: 11050212]
- Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci.* 2009; 12:718–724. [PubMed: 19471271]
- 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; 103:602–614. [PubMed: 9440345]
- Romanski LM. Domain specificity in the primate prefrontal cortex. *Cogn Affect Behav Neurosci.* 2004; 4:421–429. [PubMed: 15849888]
- Romanski LM, Goldman-Rakic PS. An auditory domain in primate pre-frontal cortex. *Nat Neurosci.* 2002; 5:15–16. [PubMed: 11753413]
- Romanski LM, Averbek BB. The primate cortical auditory system and neural representation of conspecific vocalizations. *Annu Rev Neurosci.* 2009; 32:315–346. [PubMed: 19400713]
- Romanski LM, Averbek BB, Diltz M. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J Neurophysiol.* 2005; 93:734–747. [PubMed: 15371495]
- Russ BE, Orr LE, Cohen YE. Prefrontal neurons predict choices during an auditory same-different task. *Curr Biol : CB.* 2008a; 18:1483–1488.
- Russ BE, Ackelson AL, Baker AE, Cohen YE. Coding of auditory-stimulus identity in the auditory non-spatial processing stream. *J Neurophysiol.* 2008b; 99:87–95. [PubMed: 18003874]
- Selezneva E, Scheich H, Brosch M. Dual time scales for categorical decision making in auditory cortex. *Curr Biol: CB.* 2006; 16:2428–2433.
- Singh NC, Theunissen FE. Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am.* 2003; 114:3394–3411. [PubMed: 14714819]
- Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP. A procedure for an automated measurement of song similarity. *Anim Behav.* 2000; 59:1167–1176. [PubMed: 10877896]
- Theunissen FE, Woolley SM, Hsu A, Fremouw T. Methods of analysis of auditory processing in the brain. *Ann N Y Acad Sci.* 2004; 1016:187–207. [PubMed: 15313776]
- Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. Functional specialization in rhesus monkey auditory cortex. *Science.* 2001; 292:290–293. [PubMed: 11303104]
- Tsunada J, Lee JH, Cohen YE. Representation of speech categories in the primate auditory cortex. *J Neurophysiol.* 2011; 105:2634–2646. [PubMed: 21346209]
- Tsunada J, Lee JH, Cohen YE. Differential representation of auditory categories between cell classes in primate auditory cortex. *J Physiol.* 2012; 590:3129–3139. [PubMed: 22570374]
- Tsunada, J.; Liu, AS.; Gold, JI.; Cohen, YE. Causal Role of Primate Auditory Cortex in Auditory Perceptual Decision-making. *Annual Meeting of the Society for Neuroscience; San Diego, CA.* 2013. p. 636.01
- Turetsky, R.; Ellis, D. Ground-truth transcriptions of real music from force-aligned MIDI syntheses. *4th International Symposium on Music information Retrieval ISMIR-03; Baltimore, MD.* 2003. p. 135–141.
- Victor JD, Purpura KP. Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J Neurophysiol.* 1996; 76:1310–1326. [PubMed: 8871238]
- Victor JD, Purpura KP. Metric-space analysis of spike trains: theory, algorithms, and applications. *Network.* 1997; 8:127–164.

- Woolley SM, Fremouw TE, Hsu A, Theunissen FE. Tuning for spectrotemporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci.* 2005; 8:1371–1379. [PubMed: 16136039]
- Zar, JR. *Biostatistical Analysis*. 3. Prentice Hall; Upper Saddle River, NJ: 1996.
- Zoloth SR, Petersen MR, Beecher MD, Green S, Marler P, Moody DB, Stebbins W. Species-specific perceptual processing of vocal sounds by monkeys. *Science.* 1979; 204:870–873. [PubMed: 108805]

Fig. 1.

Schematic of the (A) initial and (B) final auditory tasks. The schematics only show the task flow for those trials in which the first stimulus (i.e., the sample stimulus) is the reference *coo* C1; the task flow when the sample stimulus is the reference *grunt* G1 is analogous to that shown in these schematics. The monkeys were first trained to participate in the “initial” task, which utilized only reference vocalizations. As shown in A, this task started when the monkeys grasped a touch-sensitive lever. Next, a sample stimulus (*coo* C1) was presented. After another delay, we presented a test stimulus, which was one of the reference vocalizations (i.e., *coo* C1 or *grunt* G1). If the test stimulus was different from the sample stimulus, the monkey released the lever to receive a juice reward (as shown by the water drop). When the sample and test stimuli were the same, the monkey maintained his grip on the lever and waited for a third stimulus, which was always a reference vocalization different than the sample stimulus. Following offset of this stimulus, the monkey released the lever to get a reward. Recording sessions occurred while the monkeys participated in the final task, which is shown in Panel B. Following the monkey’s grasp of the lever, a sample stimulus was presented; like before, the sample stimulus was always one of the reference vocalizations. After a delay, the test stimulus was presented. On 70% of trials (lower row), the test stimulus could be either of the reference vocalizations (e.g., *grunt* G1). For the remaining 30% of trials (upper and middle rows), the test stimulus was one of the within-class, non-reference vocalizations (i.e., if the sample stimulus was *coo* C1, the test stimulus was a *coo* Cx). If the monkey released the lever following offset of *coo* Cx, the trial was over (upper row). The monkey would receive a reward on randomly selected trials, as indicated by the pair of dice. The decision to reward was independent of the monkey’s behavior. If he maintained his grip on the lever, a third vocalization was presented (middle row). This stimulus was always a reference vocalization that was different than the sample stimulus (i.e., *grunt* G1); following offset of this stimulus, the monkey released the lever to receive a reward.

Fig. 2.

Behavioral performance of the monkeys. In each plot, data are organized as a function of the test-stimulus for (A) *coos* and (B) *grunts*. The y-axis indicates the probability (i.e., the percentage of trials) that the monkeys reported that the sample and test stimuli were different. For data along the x-axis, when the vocalization number equals 1, it indicates that test stimulus was one of the reference vocalizations (i.e., C1C1 or G1G1 trials). All other vocalization numbers indicate data that were generated from trials when the test stimulus was a non-reference vocalization (i.e., either C1Cx or G1Gx trials); the ordering of these trials along the x-axis is arbitrary. Error bars indicate 95% confidence intervals.

Fig. 3.

Example AL neurons. For both A and B, the rasters and peristimulus-time histograms are aligned relative to onset of the stimulus. For each neuron, we illustrate data for either (A) C1Cx or (B) G1Gx trials. Neural activity is color-coded relative to the sample and test stimuli that were presented while the neuron was recorded; this color code is shown in the legend. Only correct trials are shown. Peristimulus-time histograms were generated by calculating the average firing-rate in a 5-msec time bins and then smoothed with a 20-msec kernel. The dotted line indicates stimulus onset.

Fig. 4.

Distribution of neural ROC values. Panel A shows the distribution of ROC values when the sample stimulus was reference *coo* C1. The data in black illustrate C1C1 trials; the data in grey are C1Cx trials; and white are C1G1 trials. Panel B shows the distribution of ROC values when the sample stimulus was reference *grunt* G1. The data in black illustrate G1G1 trials; the data in grey are G1Gx trials; and white are G1C1 trials.

Fig. 5.

Metric-space analysis of the variability in spike count elicited by *coos* and by *grunts*. The matrix in the upper left shows the average metric-space value between spike trains elicited for each pairwise combination of *coos*. An analogous matrix for the *grunt*-elicited spike trains is shown in the bottom right. The ordering of the vocalizations along the matrix axes is arbitrary. The average metric-space value is indicated by color with black showing the smallest value and light purple illustrating those with the most. For matrix elements that are colored white, neural activity was not recorded in response to these vocalizations.

Fig. 6.

Dynamic-time-warp (DTW) quantification of the variability between vocalizations. The DTW algorithm quantified (A) the acoustic similarity between reference *coo* C1 and the other non-reference *coos* or (B) the acoustic similarity between reference *grunt* G1 and the other non-reference *grunts*. In each panel, the vocalizations are organized as a function of increasing DTW value. The numbers on the x-axis indicate the vocalization number. In panel A, vocalization #1 is reference *coo* C1. Similarly, in panel B, vocalization #1 is reference *grunt* G1. Because the reference vocalization was compared with itself, by definition, the DTW value was zero. Above the bar charts in each panel, we display spectrograms of select vocalizations. In A, the leftmost spectrogram is reference *coo* C1, the one in the middle is a *coo* with an intermediate DTW value (C27), and the rightmost one is the *coo* with the largest DTW value (C28). In B, the leftmost spectrogram is reference *grunt* G1, a *grunt* with an intermediate DTW value (G20), and the rightmost one is the *grunt* with the largest DTW value (G15). The power density for each spectrogram is indicated by color using a decibel (dB) scale.

Fig. 7. Modulation-spectra analysis of the variability between *coos* and *grunts*. The matrix in the upper left shows the correlation between the modulation spectra for all pairwise combinations of *coos*, whereas the matrix in the lower right shows the correlation between the modulation spectra for all pairwise combinations of *grunts*. The average correlation value is indicated by color with black showing the smallest value and white illustrating those with the most. The ordering of the vocalizations along the matrix axes is arbitrary. Below the *coo* matrix and above the *grunt* matrix is the average modulation spectrum for our set of *coos* or *grunts*, respectively. For these spectra, power density is indicated by color using a dB scale with red showing the spectrotemporal modulations with the most energy and blue illustrating those with the least. The ordering of the vocalizations is the same as that in Fig. 5.

Fig. 8.

Correlation between the acoustic features of different *coos* and different *grunts*. (A) The correlation between *coo* standard deviation of the intensity and the duration. (B) The correlation between *coo* entropy and the spectral center of gravity. (C) The correlation between *coo* entropy and the mean harmonicity. (D) The correlation between *grunt* spectral center of gravity and the entropy. In each panel, the thick black line is a least-square line fit to the data. This figure shows only those correlations that were significant ($p < 0.05$). au = arbitrary units.

Fig. 9. Behavioral performance of the monkeys as a function of DTW. Panels A and B replot the behavioral data in Fig. 2 as a function of the normalized DTW value; DTW values were normalized independently for *coos* and *grunts* so that, for both, the maximum DTW value equaled one. For ease of presentation, we do not show the error bars displayed in Fig. 2.

Fig. 10.

Correlation between firing rate and the acoustic features of *coo* vocalizations. Population analysis of the correlation between test-stimulus period firing rate and (A) duration, (B) entropy, (C) standard deviation of the intensity, (D) center of gravity, and (E) harmonicity. Each data point represents the mean firing rate; error bars represent the standard deviation. A single neuron can contribute to multiple data points in each panel because neural activity was collected for multiple test stimuli.

Fig. 11.

Correlation between firing rate and the acoustic features of *grunt* vocalizations. Analogous analyses for data generated in response to *grunts*: population analysis of the correlation between test-stimulus period firing rate and (A) duration, (B) entropy, (C) standard deviation of the intensity, and (D) center of gravity. Each data point represents the mean firing rate; error bars represent the standard deviation. A single neuron can contribute to multiple data points in each panel because neural activity was collected for multiple test stimuli.

Fig. 12.

Correlation between metric-space values and the acoustic features of rhesus vocalizations. In the TOP panel, metric-space values, which were calculated from the spike trains elicited by two *coos*, are correlated against the ratio between one of the quantified acoustic features (e.g., the ratio between the center of gravity of one test-stimulus *coo* and the center of gravity of a different test-stimulus *coo*). Panel A illustrates this analysis when the ratio was calculated using the center of gravity; B uses the standard deviation of the intensity; C uses duration; D uses entropy; E uses the modulation spectra; F uses the log frequency spectra; and G uses the mean harmonicity. In the BOTTOM panel, the analogous correlations are shown for *grunts*: (H) the ratio of the center of gravity, (I) the ratio of the standard deviation of the intensity, (J) the ratio of the duration, (K) the ratio of the entropy, (L) the correlation between the modulation spectra, and (M) the correlation between the log frequency spectra. A single neuron can contribute to multiple data points in each panel because neural activity was collected for multiple test stimuli.

Table 1Statistics of the acoustic features of *coos* and *grunts*.

	Coos	Grunts
Duration	$\mu = 482.6$ ms (std = 164) *	$\mu = 205.9$ ms (std = 54)
Wiener entropy	$\mu = -4.28$ (std = 1.0) *	$\mu = -2.15$ (std = 0.59)
Standard deviation of intensity	$\mu = 4.83$ dB (std = 1.9) *	$\mu = 5.97$ dB (std = 1.2)
Spectral center of gravity	$\mu = 861$ Hz (std = 377)	$\mu = 924$ Hz (std = 724) *
Mean harmonicity	$\mu = 14.6$ dB (std = 5.3); CoV = 0.36	

* Significantly ($p < 0.05$) more variance (F-test for equality of variances).