



Published in final edited form as:

J Neurophysiol. 2008 January ; 99(1): 87–95. doi:10.1152/jn.01069.2007.

Coding of Auditory-Stimulus Identity in the Auditory Non-Spatial Processing Stream

Brian E. Russ, Ashlee L. Ackelson, Allison E. Baker, and Yale E. Cohen

Department of Psychological and Brain Sciences, Center for Cognitive Neuroscience, Dartmouth College, Hanover, New Hampshire

Abstract

The neural computations that underlie the processing of auditory-stimulus identity are not well understood, especially how information is transformed across different cortical areas. Here, we compared the capacity of neurons in the superior temporal gyrus (STG) and the ventrolateral prefrontal cortex (vPFC) to code the identity of an auditory stimulus; these two areas are part of a ventral processing stream for auditory-stimulus identity. Whereas the responses of neurons in both areas are reliably modulated by different vocalizations, STG responses code significantly more vocalizations than those in the vPFC. Together, these data indicate that the STG and vPFC differentially code auditory identity, which suggests that substantial information processing takes place between these two areas. These findings are consistent with the hypothesis that the STG and the vPFC are part of a functional circuit for auditory-identity analysis.

INTRODUCTION

Despite massive parallel and divergent connections between the thalamus and the primary sensory cortices (Felleman and Van Essen 1991; Winer and Lee 2007), the processing of the spatial and non-spatial (i.e., stimulus identity) attributes of a sensory stimulus can be conceptualized as occurring in a hierarchical manner in distinct processing streams (Griffiths and Warren 2004; Poremba and Mishkin 2007; Poremba et al. 2003; Rauschecker 1998; Romanski et al. 1999; Ungerleider and Mishkin 1982). More specifically, it is thought that a “dorsal” pathway processes the spatial attributes of a stimulus. A “ventral” pathway, on the other hand, processes the non-spatial attributes.

In the visual system, the hierarchical processing of stimulus identity has been well characterized. The color, location, and orientation of a visual object are coded in the primary and secondary visual cortices. Higher-order representations of objects are found in subsequent areas such as the inferior temporal cortex where neurons code specific object categories like faces (Fujita et al. 1992; Gross and Sergent 1992; Perrett et al. 1984, 1985) or learned categories like “cat” and “dog” (Freedman et al. 2003). More abstract representations can also be found in the inferior temporal cortex: neurons code a specific

facial expression (emotion), independent of the specific identity of the face (Sugase et al. 1999). However, even neurons in these higher sensory areas can still be modulated somewhat by the low-level sensory features of a stimulus (Freedman et al. 2003). At the level of the prefrontal cortex, neurons appear to be mostly sensitive to the category membership of an object, independent of its specific sensory features (Freedman et al. 2001, 2002).

In contrast, in the ventral auditory pathway, the computational mechanisms that lead from the coding of the sensory features of an auditory stimulus to higher-order representations are relatively unknown. In particular, it is not known how (or even whether) information is transformed between areas of this ventral stimulus-identity pathway. Although a few previous auditory studies have compared response patterns in different primate cortical processing streams (Cohen et al. 2004b; Tian et al. 2001), there are no studies that have examined how information is transformed between different cortical areas that belong to the same processing stream. This type of analysis is necessary to understand how different brain areas interact and how these interactions relate to sensation, perception, and action at the level of brain networks.

Here, we tested and compared how well the responses of neurons in the superior temporal gyrus (STG; which includes the anterolateral belt region of the auditory cortex, a secondary auditory field) and the ventrolateral prefrontal cortex (vPFC) differentiate between different auditory-stimulus identities, specifically species-specific vocalizations. Both anatomical and physiological data suggest that these cortical regions are part of the ventral auditory pathway; this pathway originates in the primary cortex and includes a series of projections through the STG, the parabelt region, and ultimately the vPFC (Rauschecker 1998; Rauschecker and Tian 2000; Romanski and Goldman-Rakic 2002; Romanski et al. 1999; Russ et al. 2007b).

We found that STG and vPFC neurons code the identity of auditory stimuli: they respond differently to different vocalizations. However, between these two areas, there are substantial differences in the information processing. That is, the spike trains of STG neurons code more vocalizations, across a variety of timescales both at the level of individual neurons and at the population level, than do vPFC neurons. These results are consistent with the hypothesis that the STG and the vPFC are part of a functional circuit for auditory-identity analysis.

METHODS

We recorded from neurons in the STG ($n = 64$) from one male and one female rhesus monkey. Two different monkeys (both female) were used to record vPFC neurons ($n = 39$); the data from the vPFC neurons were collected as part of a recent study from our laboratory (Cohen et al. 2007).

Two important points about these monkeys' training history are worth noting. First, none of the monkeys had been operantly trained to associate an auditory stimulus with an action and a subsequent reward. Second, all of the monkeys had similar experience with the

vocalizations used in this experiment. Thus, differences between responses in the vPFC and the STG cannot wholly be attributed to differences in the monkeys' behavioral, recording, or training history.

Under isoflurane anesthesia, the monkeys were implanted with a scleral search coil, head-positioning cylinder, and a recording chamber (Crist Instruments). Recordings in the STG were done in the left hemisphere of the male rhesus and the right hemisphere of the female rhesus. The vPFC recordings were obtained from the left hemisphere of both monkeys. All recordings were guided by pre- and postoperative magnetic resonance images of each monkey's brain at the Dartmouth Brain Imaging Center using a GE 1.5T scanner or a Phillips 3T scanner (3-D T1-weighted gradient echo pulse sequence with a 5-in. receive-only surface coil) (Cohen et al. 2004a; Gifford 3rd and Cohen 2004; Groh et al. 2001). The Dartmouth Institutional Animal Care and Use Committee approved all of the experimental protocols.

Behavioral task and auditory stimuli

The monkeys participated in the "passive-listening" task. A vocalization was presented from the speaker 1,000–1,500 ms after fixating a central red light-emitting diode (LED). To minimize any potential changes in neural activity due to changes in eye position (Werner-Reiss et al. 2003), the monkeys maintained their gaze at the central LED during auditory-stimulus presentation and for an additional 1,000–1,500 ms after auditory-stimulus offset to receive a juice reward.

The vocalizations were recorded and digitized as part of an earlier set of studies (Hauser 1998). We presented a single exemplar from each of the 10 major acoustic classes of rhesus vocalizations; the spectrograms of these stimuli can be seen in Figs. 2 and 3. Each exemplar was filtered to compensate for the transfer-function properties of the speaker and the acoustics of the room. A vocalization was presented through a D/A converter [DA1, Tucker Davis Technologies (TDT)], an anti-aliasing filter (FT6-92, TDT), an amplifier (SA1, TDT; and MPA-250, Radio Shack), transduced by the speaker (Pyle, PLX32), and was presented at an average sound level of 65 dB SPL.

Recording procedures

Single-unit extracellular recordings were obtained with a tungsten microelectrode (1 M Ω) at 1 kHz; FHC) seated inside a stainless steel guide tube. The electrode signal was amplified (MDA-4I, Bak Electronics) and band-pass filtered (model 3362, Krohn-Hite) between 0.6 and 6.0 kHz. Single-unit activity was isolated using a two-window, time-voltage discriminator (Model DDIS-1, Bak Electronics). Neural events that passed through both windows were classified as originating from a single neuron. The amplitudes of the single units typically match and often exceed threefold the noise floor; this criterion was used for recordings in both brain structures. The time of occurrence of each action potential was stored for on- and off-line analyses.

The STG was identified based on its anatomical location, which was verified through anatomical magnetic resonance images of each monkey's brain. We targeted the middle of the anteroposterior extent of the lateral sulcus, at its most lateral position (Ghazanfar et al.

2005); see Fig. 1. This region of the STG overlaps with the anterolateral belt region of the auditory cortex, a cortical field thought to be involved in processing the identity of an auditory stimulus (Rauschecker and Tian 2000).

The vPFC was identified by its anatomical location and its neurophysiological properties (Cohen et al. 2004b; Gifford 3rd et al. 2005; Romanski and Goldman-Rakic 2002). The vPFC is located anterior to the arcuate sulcus and area 8a and lies below the principal sulcus. vPFC neurons were further characterized by their strong responses to auditory stimuli.

Recording strategy

An electrode was lowered into the STG or the vPFC. To minimize sampling bias, any neuron that was isolated was tested. Since neurons in both the STG and vPFC were spontaneously active, we did not present any stimuli while we were lowering the electrode and isolating a neuron. Once a neuron was isolated, the monkeys participated in trials of the passive-listening task. In each block, the vocalization exemplars were presented in a balanced pseudorandom order. We report those neurons in which we were able to collect data from >200 successful trials. The intertrial interval was 1–2 s.

Data analysis

In the following text, we discuss the four independent analysis methods used to quantify the capacity of STG and vPFC neurons to code auditory identity. We generated one index value, instantiated two different decoding schemes, and calculated mutual-information values.

VOCALIZATION-PREFERENCE INDEX—On a neuron-by-neuron basis, we determined which vocalization exemplar elicited the highest mean firing rate (the “preferred firing rate”); the firing rate is the number of action potentials divided by the length of a specified time interval. The vocalization-preference index of a neuron is the number of vocalization exemplars that elicit a mean firing rate >50% of the preferred firing rate (Cohen et al. 2004b; Tian et al. 2001). This index was calculated as a function of the time intervals that ranged from 50 to 750 ms following onset of the vocalization.

LINEAR-PATTERN DISCRIMINATOR—A linear-pattern discriminator tested how well the responses of a neuron differentiate between different vocalizations (Schnupp et al. 2006). On a neuron-by-neuron basis and trial-by-trial basis, peristimulus time histograms (PSTHs) were created for each spike train. The PSTHs were calculated as a function of the time interval following vocalization onset (50–750 ms) and the bin size (2–750 ms). Beyond binning the data, we did not further smooth the histograms.

For each neuron, a single spike train was chosen and removed from the data and its PSTH was generated for a specific time interval and bin size. This constituted our “test” data. For the remaining trials, we calculated each of their PSTHs using the same interval and bin size. These “training” data were grouped as a function of each of the 10 vocalization exemplars.

Next, we tested whether we could determine which vocalization elicited the test data by comparing how similar the test data were to the 10 sets of training data. As a similarity metric, we calculated the Euclidian distance between the test data and the mean of each set

of training data. If the test and training data were similar, then the distance should be small. If the test and training data were different, then the distance should be large. The test data were assigned to be the vocalization that had the most similar training set. This procedure was repeated until each trial of a neuron was considered the test data and until each combination of time interval and bin size was considered.

MAXIMUM-LIKELIHOOD ESTIMATOR—A Bayesian classifier calculated how well the firing rates of populations of neurons differentiate between different vocalizations (Yu et al. 2004). Neural data were fit to a multivariate Gaussian distribution

$$f(\mathbf{r}|v) = \frac{1}{(2\pi)^{n/2} |\Sigma_v|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{r} - \boldsymbol{\mu}_v)^T \Sigma_v^{-1} (\mathbf{r} - \boldsymbol{\mu}_v) \right]$$

where \mathbf{r} is a test vector of spike rates for a single trial, $\boldsymbol{\mu}_v$ is the mean vector of the training data for vocalization exemplar v , and Σ_v is the covariance matrix of the training data for test vocalization v . The dimensions of these vectors and the matrix are dependent on the number of “available” neurons (n). Even though we recorded the neurons serially and from different monkeys, we assumed a small amount of covariance between pairs of neurons ($r = 0.1$) (Shadlen and Newsome 1998; Zohary et al. 1994).

A maximum-likelihood estimator decoded the vocalization from the neural responses

$$\hat{v} = \arg \max_v P(v|\mathbf{r})$$

where v is the test vocalization and where \hat{v} is the estimated vocalization. Using Bayes’ rule, we can expand this probability

$$\hat{v} = \arg \max_v \frac{f(\mathbf{r}|v)P(v)}{P(\mathbf{r})}$$

Since each vocalization was equally likely to be present [i.e., $P(v)$ is a constant] and since $P(\mathbf{r})$ was not dependent on the vocalization, it follows that the estimator was maximized when $f(\mathbf{r} | v)$ was maximized

$$\hat{v} = \arg \max_v f(\mathbf{r}|v)$$

The estimator was computed as a function of $n = 1:N$, where N is the number of available neurons. For each value of n , we randomly selected 100 sets of n neurons with replacement. For each of the 100 sets of n neurons, a single trial was removed, its firing rate was calculated, and the test vector \mathbf{r} was created. $f(\mathbf{r} | v)$ was computed with test vector \mathbf{r} and the training vector ($\boldsymbol{\mu}_v$) and the corresponding covariance matrix (Σ_v); these vectors and the covariance matrix were calculated for each vocalization v . The v that maximized $f(\mathbf{r} | v)$ was \hat{v} , the estimated vocalization. This process was repeated for each of the trials in each of the 100 sets of n neurons.

MUTUAL INFORMATION—We quantified the amount of mutual information (Cover and Thomas 1991; Panzeri and Treves 1996b; Shannon 1948a,b) carried in the responses of STG or vPFC neurons to different vocalizations (i.e., different stimulus identities). The advantage of this analysis is that it is a nonparametric index of a neuron's sensitivity to different stimuli. Information (I) was defined as

$$I = \sum_v \sum_r P(v, r) \log_2 \frac{P(v, r)}{P(v)P(r)}$$

where v is the index of each vocalization, r is the index of the neural response, $P(v, r)$ is the joint probability, and $P(v)$ and $P(r)$ are the marginal probabilities. For each recorded neuron, we independently calculated the amount of information as a function of the time interval following vocalization onset (50–750 ms) and the bin size (2–750 ms) using a formulation described by Nelken and colleagues (Chechik et al. 2006; Nelken and Chechik 2007).

To facilitate comparisons across monkeys and to eliminate biases inherent in small sample sizes, we calculated the amount of bias in our information values (Chechik et al. 2006; Grunewald et al. 1999; Nelken and Chechik 2007; Panzeri and Treves 1996b). This correction procedure corrects for erroneously large bit-rate values due to large variances inherent in small sample sizes. We calculated bias in two different ways. First, we calculated the amount of information from the original data and from bootstrapped trials. In bootstrapped trials, the relationship between a neuron's firing rate and the vocalization was randomized and then the amount of information was calculated. This process was repeated 500 times and the mean value from this distribution of values was determined. The amount of bias-corrected information was calculated by subtracting the mean amount of information obtained from bootstrapped trials from the amount obtained from the original data. Second, we calculated the bias using the following formula: $\#bins / (2N \log_2(2))$, where $\#bins$ is the number of nonzero bins and N is the number of samples (Chechik et al. 2006; Nelken and Chechik 2007; Panzeri and Treves 1996a; Treves and Panzeri 1995). The value of this calculation was then subtracted from the original information value. Each of these two calculations was conducted on a neuron-by-neuron basis and as a function of the time interval and bin size (see earlier text). Since the results of both bias calculations were similar, we show only the results from the second bias procedure.

RESULTS

Both STG and vPFC neurons respond to species-specific vocalizations. A response profile from one STG neuron is shown in Fig. 2. The temporal features of this neuron's response were varied and dependent on the specific vocalization exemplar. For instance, when we presented the aggressive vocalization (Fig. 2A), the neuron had a primary-like response to the vocalization. On the other hand, the neuron responded robustly to both the onset and offset of the coo (Fig. 2D) and had periodic bouts of firing in response to the acoustic pulses of the gecker (Fig. 2E). In contrast, the response patterns of vPFC neurons did not show this degree of variability: for each of the 10 different vocalization exemplars, the vPFC neuron in Fig. 3 had mostly strong onset responses with less pronounced sustained responses. These

two response profiles, which are typical of both areas, suggest that vocalizations (stimulus identities) are coded differently in the STG and the vPFC.

To quantify this observation, we used four independent analysis techniques to quantify STG and vPFC responsivity: we generated one index value, instantiated two different decoding schemes, and calculated mutual-information values.

First, to facilitate comparisons with previous studies in the STG and the vPFC, we calculated a vocalization-preference index (Cohen et al. 2004b; Tian et al. 2001). If the index has a value of one, it indicates that a neuron responded selectively to only one vocalization. Higher values indicate that the neuron responded to a broad number of vocalizations. As seen in Fig. 4, as the time interval increases from 50 to 750 ms, the mean index value for our population of STG neurons increased from a value of about 5.5 to a value of about 8. That is, as we integrated information over longer time intervals, the selectivity got worse. In contrast, the mean index value of vPFC neurons was relatively insensitive to the duration of the time interval; for each of the tested time intervals, the index value was about 8. An ANOVA with post hoc Scheffé tests indicated that STG neurons had reliably ($P < 0.05$) lower (i.e., more selective) vocalization-index values than those in the vPFC except at the longest tested time interval (750 ms).

Next, we used a linear-pattern discriminator (Schnupp et al. 2006) and a maximum-likelihood estimator (Yu et al. 2004) to explicitly quantify the degree to which the responses of STG and vPFC neurons varied with each of the 10 vocalization exemplars. These methods allowed us to test whether we could successfully “decode” the vocalization’s identity from a neuron’s response.

The results from the linear-pattern discriminator are shown in Fig. 5. The decoding capacity of the pattern discriminator was tested as a function of the time interval following the vocalization’s onset and as a function of the temporal resolution (bin size) of the spike trains from the test and training data. Since we presented 10 different vocalizations, the chance level of the discriminator was 10%.

Similar to that seen in other studies (Schnupp et al. 2006), the discriminator’s performance was a function of the duration of the tested time interval and the temporal resolution of the data in an interval. At the shortest time interval (50 ms) and the finest bin resolution (2 ms), the discriminator decoded about 25% of the vocalizations correctly from the spike trains of STG neurons (Fig. 5A). As the temporal resolution became coarser (i.e., larger time bins), the performance of the discriminator decreased in an exponential fashion to about 13%. However, as the duration of the test time interval increased from 50 to 750 ms, the performance of the discriminator improved dramatically. Indeed, when the time interval was 750 ms, the discriminator decoded about 90% of the vocalizations at the finest bin resolution. These longer time intervals were not resistant to bin size: with coarser bin sizes, the performance of the discriminator fell off and reached an asymptotic level of about 18%.

How well does the discriminator perform when it decodes vocalizations from the vPFC data? As shown in Fig. 5B, the general pattern of the discriminator’s performance was the same for the vPFC neurons as it was for the STG neurons: the percentage of correctly

decoded vocalizations increased as the time interval increased and decreased as the temporal resolution of the data became coarser. However, the percentage of correctly decoded vocalizations was worse when the discriminator analyzed the vPFC responses than when it analyzed the STG responses. This can be seen more clearly in Fig. 5C where we overlay the discriminator's performance from the STG and vPFC responses. For a given time interval, as the bin resolution became coarser, the percentage of correctly decoded vocalizations decreased faster for the vPFC responses than it did for the STG responses. Indeed, for time intervals >50 ms, the discriminator could decode reliably ($P < 0.05$) more correct vocalizations from the STG responses than it could from the vPFC responses.

Results of the maximum-likelihood estimator are shown in Fig. 6. The advantage of using this estimator is that since we recorded from different numbers of STG and vPFC neurons, we could use the estimator to directly compare the decoding capacity of similarly sized populations of neurons.

When we focus on the decoding capacity of the STG responses (red data in Fig. 6), we see that a single STG neuron can perform reliably above chance. Whereas the percentage of correctly decoded vocalizations is small, this percentage is comparable to the performance of the linear-pattern discriminator (Fig. 5A) when the discriminator decoded STG responses over a comparable time interval (500 ms) and when a very coarse bin size (i.e., a rate code) was applied to the test and training data; this 500-ms interval was chosen since it approximated the mean duration of our digitized collection of rhesus vocalizations (Cohen et al. 2007; Hauser 1998). As the number of available neurons increased, the percentage of correctly decoded vocalizations increased to about 40% and reached an asymptotic value when the number of simultaneously available neurons reached about 40 neurons. By "available," we do not refer to a number of simultaneously recorded neurons but instead to the number of neurons that were available to be decoded by the likelihood estimator.

The estimator's performance is clearly different for the vPFC neurons (blue data) than that for the STG neurons. The most apparent difference is that the percentage of correctly decoded vocalizations is lower, independent of the number of simultaneously available neurons. Another interesting difference is that, as the number of available vPFC neurons increases, the percentage of correctly decoded vocalizations does not increase at nearly the same rate as it does when more STG neurons become available. Indeed, an ANOVA with post hoc Scheffé tests indicated that STG neurons could reliably ($P < 0.05$) decode more vocalizations than vPFC neurons, at any number of simultaneously available neurons.

Finally, the data shown in Fig. 7 illustrate the mutual-information values calculated as a function of the time interval following the vocalization's onset and as a function of the temporal resolution (bin size) of the spike trains; the time intervals and bin size are the same as those used in the linear-pattern discriminator (Fig. 5). Analogous to the results of the linear-pattern discriminator neurons, the amount of information carried in the responses increased as the time interval increased and decreased as the temporal resolution of the data became coarser for both STG (Fig. 7A) and vPFC (Fig. 7B) neurons; information values >0 are considered to be reliable values (Nelken and Chechik 2007). Like our previous analyses, the amount of information that relates to different vocalization (stimulus) identities was, in

general, greater for STG activity than that for vPFC activity. This can be seen more clearly in Fig. 7C where we overlay the results of the information analyses from the STG and vPFC responses. Indeed, for time intervals >200 ms, the amount of information carried in STG responses was reliably ($P < 0.05$) more than the amount carried in vPFC responses.

One potential confound for this information analysis, as well as the decoding algorithms, is that the apparent differences between STG and vPFC neurons may reflect low-level differences between their response properties such as firing rate: lower firing rates can reduce information about stimulus identity, whereas higher firing rates can falsely increase information (Chechik et al. 2006). To test for this possibility, we calculated and compared the firing rates of STG and vPFC neurons. The average firing rate, which was calculated as a function of each vocalization's duration, averaged between 12 and 17 spikes/s in both the STG and the vPFC. Importantly, we found that the average firing rate elicited by each vocalization exemplar was not reliably different ($P > 0.05$) between STG and vPFC neurons. Thus, the differences between the information capacity of STG and vPFC neurons cannot wholly be attributed to differences between the firing rates of these two populations.

DISCUSSION

Both STG and vPFC responses reliably differentiate between different vocalizations (Figs. 4–7). However, the spike trains of STG and vPFC neurons do not code vocalization identity equivalently. Indeed, we found that STG neurons discriminate between different vocalizations better than vPFC neurons: whereas the discrimination capacity differed as a function of the analysis, across all four independent analyses, we found that the spike trains of STG neurons carry more information about different auditory-stimulus identities than those in vPFC neurons.

Our STG findings confirm and extend the seminal work of Rauschecker and colleagues (Rauschecker and Tian 2004; Rauschecker et al. 1995; Tian and Rauschecker 2004; Tian et al. 2001). Similar to the results of our vocalization-preference index (see Fig. 4), they found that the firing rate of STG neurons responds selectively to a subset of vocalizations. Additionally, Rauschecker et al. (1995) demonstrated that STG neurons were sensitive to the temporal structure of certain vocalizations, and their responses to vocalizations were greater than the sum of the responses to filtered versions of vocalizations—two results consistent with a role for the STG in processing auditory-stimulus identity.

How do our vPFC results compare with previous vPFC studies? An important study to consider is one by Romanski and colleagues (2005). In that study, vPFC activity was recorded while monkeys listened passively to species-specific vocalizations and attended to a central LED—a behavioral task that is quite similar to ours. Using a monkey call-preference index (Tian et al. 2001), Romanski et al. reported that roughly 55% of their population responded preferentially to only one or two vocalizations (out of a possible 10). In contrast, our index values (see Fig. 4) indicate that most vPFC neurons responded preferentially to about eight different vocalizations. However, a comparison of their information analysis with our information and decoding analyses show more agreement: both sets of analyses indicate that vPFC neurons transmit between approximately 1 and 2

bits of information, which is slightly above chance (see Figs. 5–7), when comparable analyses periods are examined. Presently, we cannot readily reconcile all of our results with those reported by Romanski et al. These differences may be attributable to subtle, yet significant, differences between recording locations, differences between neuron-selection strategies, or differences between analysis techniques (e.g., they used an analysis period based solely on the duration of a stimulus, whereas we used periods independent of stimulus duration).

As noted earlier, both the linear-pattern discriminator and the information analysis performed better when fine temporal information (i.e., small bin size) was used than when only rate information (i.e., large bin size) was contained in the data sets. The high performance with the temporal information is surprising since it is commonly thought that the temporal precision of cortical neurons is not sufficient to enable reliable temporal codes and that cortical neurons provide information primarily in the form of rate codes (Shadlen and Newsome 1998). There are, however, biophysical and psychophysical data that are consistent with the notion that the auditory system may be capable of using relatively high-resolution temporal information; at least as high as about 50 Hz, which is equivalent to a 20-ms bin size (Fig. 5) (Eggermont 2002; Hirsch 1959; Lu et al. 2001; Saberi and Perrott 1999; Schnupp et al. 2006; Schreiner et al. 1997). It is important to note that our algorithms can only address the amount of information that is *potentially* contained in a neuron's response. They do not make any claims on what kind of code (i.e., temporal, rate, or otherwise) is actually being used by these cortical neurons.

The most compelling result from this study is that vPFC neurons do not discriminate between different auditory-stimulus identities as well as STG neurons (Figs. 4–7). Given that the STG is the main source of auditory input to the vPFC (Rauschecker and Tian 2000; Romanski and Goldman-Rakic 2002; Romanski et al. 1999), we hypothesize that this difference is due to a hierarchical relationship between the STG and the vPFC. In other words, we hypothesize that the vPFC is at a more advanced state of auditory processing than that seen in the STG. Consistent with this hypothesis, we have shown that vPFC neurons are involved in referential categorization (Cohen et al. 2006; Gifford 3rd et al. 2005). In contrast, we propose that the STG may be involved in perceptual categorization or other computations that are needed to create such referential categories (Ashby and Maddox 2005; Rosch 1978).

Obviously, further experiments are needed to confirm this hypothesis and to rule out other potential confounds and interpretations. For example, in our study, the monkeys listened passively to the vocalizations. It may very well be that if the monkeys are required to actively attend to the vocalizations, STG and vPFC neurons may respond similarly. Indeed, the response properties of neurons in the auditory cortex are highly dependent on behavioral state (Fritz et al. 2003). However, in more central areas, the relationship between behavioral state and response properties is not so straightforward: the selectivity of parietal neurons to the shape of different visual objects is similar regardless of whether monkeys are passively viewing them or actively discriminating between them (Serenio and Maunsell 1998), whereas the sensitivity of PFC neurons is highly dependent on the nature of the behavioral task (Rao et al. 1997). Another possibility is that vPFC neurons respond broadly to any type

of auditory stimulus, regardless of whether it is a tone pip or a vocalization, and thus may not code “identity” per se. Further analyses of the auditory-response properties of vPFC neurons are needed to test the degree to which vPFC neurons reflect stimulus identity. Finally, our results may reflect differences in the sensitivity (e.g., differences in integration time, spectrotemporal modulation sensitivity, etc.) of vPFC and STG neurons to the acoustic features of vocalizations. This possibility seems less likely since 1) STG and vPFC neurons are both relatively insensitive to the acoustic features of vocalizations (Cohen et al. 2007; Russ et al. 2007a) and since 2) the coding of the acoustic features is thought to primarily occur in the primary auditory cortex or midbrain (Griffiths and Warren 2004; Nelken et al. 2003).

Nevertheless, our data are consistent with the hypothesis that the pathway leading from the primary auditory cortex to the vPFC forms a functional circuit involved in the coding and representation of the identity of an auditory stimulus. Finally, it may be that these two areas of the auditory pathway (the STG and the vPFC) are analogous to the visual areas in the infratemporal cortex and the prefrontal cortex that process categorical visual information in a hierarchical manner (Freedman et al. 2002, 2003).

Acknowledgments

We thank P. Olszynski and F. Chowdhury for help with data collection and C. Miller, F. Theuinsen, Y.-S. Lee, K. Shenoy, and H. Hersh for constructive comments. M. Hauser generously provided recordings of the rhesus vocalizations.

GRANTS

This work was supported by National Institutes of Health grants and a Burke Award to Y. E. Cohen and a National Research Service Award fellowship to B. E. Russ.

REFERENCES

- Ashby FG, Maddox WT. Human category learning. *Annu Rev Psychol.* 2005; 56:149–178. [PubMed: 15709932]
- Chechik G, Anderson MJ, Bar-Yosef O, Young ED, Tishby N, Nelken I. Reduction of information redundancy in the ascending auditory pathway. *Neuron.* 2006; 51:359–368. [PubMed: 16880130]
- Cohen YE, Cohen IS, Gifford GW 3rd. Modulation of LIP activity by predictive auditory and visual cues. *Cereb Cortex.* 2004a; 14:1287–1301. [PubMed: 15166102]
- Cohen YE, Hauser MD, Russ BE. Spontaneous processing of abstract categorical information in the ventrolateral prefrontal cortex. *Biol Lett.* 2006; 2:261–265. [PubMed: 17148378]
- Cohen YE, Russ BE, Gifford GW 3rd, Kiringoda R, MacLean KA. Selectivity for the spatial and non-spatial attributes of auditory stimuli in the ventrolateral prefrontal cortex. *J Neurosci.* 2004b; 24:11307–11316. [PubMed: 15601937]
- Cohen YE, Theunissen FE, Russ BE, Gill P. The acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. *J Neurophysiol.* 2007; 97:1470–1184. [PubMed: 17135477]
- Cover, TM.; Thomas, JA. *Elements of Information Theory.* New York: Wiley; 1991.
- Eggermont JJ. Temporal modulation transfer functions in cat primary auditory cortex: separating stimulus effects from neural mechanisms. *J Neurophysiol.* 2002; 87:305–321. [PubMed: 11784752]
- Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex.* 1991; 1:1–47. [PubMed: 1822724]
- 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]

- Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. *J Neurophysiol.* 2002; 88:929–941. [PubMed: 12163542]
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J Neurosci.* 2003; 23:5235–5246. [PubMed: 12832548]
- Fritz J, Shamma S, Elhilali M, Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci.* 2003; 6:1216–1223. [PubMed: 14583754]
- Fujita I, Tanaka K, Ito M, Cheng K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature.* 1992; 360:343–346. [PubMed: 1448150]
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci.* 2005; 25:5004–5012. [PubMed: 15901781]
- Gifford GW 3rd, Cohen YE. The effect of a central fixation light on auditory spatial responses in area LIP. *J Neurophysiol.* 2004; 91:2929–2933. [PubMed: 14749308]
- 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 Cogn Neurosci.* 2005; 17:1471–1482. [PubMed: 16197700]
- Griffiths TD, Warren JD. What is an auditory object? *Nat Rev Neurosci.* 2004; 5:887–892. [PubMed: 15496866]
- Groh JM, Trause AS, Underhill AM, Clark KR, Inati S. Eye position influences auditory responses in primate inferior colliculus. *Neuron.* 2001; 29:509–518. [PubMed: 11239439]
- Gross CG, Sergent J. Face recognition. *Curr Opin Neurobiol.* 1992; 2:156–161. [PubMed: 1638146]
- Grunewald A, Linden JF, Andersen RA. Responses to auditory stimuli in macaque lateral intraparietal area. I. Effects of training. *J Neurophysiol.* 1999; 82:330–342. [PubMed: 10400962]
- Hauser MD. Functional referents and acoustic similarity: field playback experiments with rhesus monkeys. *Anim Behav.* 1998; 55:1647–1658. [PubMed: 9642008]
- Hirsch IJ. Auditory perception of temporal order. *J Acoust Soc Am.* 1959; 31:759–767.
- Lu T, Liang L, Wang X. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci.* 2001; 4:1131–1138. [PubMed: 11593234]
- Nelken I, Chechik G. Information theory in auditory research. *Hear Res.* 2007; 229:94–105. [PubMed: 17300891]
- Nelken I, Fishbach A, Las L, Ulanovsky N, Farkas D. Primary auditory cortex of cats: feature detection or something else? *Biol Cybern.* 2003; 89:397–406. [PubMed: 14669020]
- Panzeri S, Treves A. Analytical estimates of limited sampling biases in different information measures. *Network Comput Neural Syst.* 1996a; 7:87–101.
- Panzeri S, Treves A. Analytical estimates of limited sampling biases in different information measures. *Network Comput Neural Syst.* 1996b; 7:87–107.
- Perrett DI, Smith PA, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA. Neurones responsive to faces in the temporal cortex: studies of functional organization, sensitivity to identity and relation to perception. *Hum Neurobiol.* 1984; 3:197–209. [PubMed: 6526706]
- Perrett DI, Smith PA, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA. Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc R Soc Lond B Biol Sci.* 1985; 223:293–317. [PubMed: 2858100]
- Poremba A, Mishkin M. Exploring the extent and function of higher-order auditory cortex in rhesus monkeys. *Hear Res.* 2007; 229:14–23. [PubMed: 17321703]
- 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]
- Rao SC, Rainer G, Miller EK. Integration of what and where in the primate prefrontal cortex. *Science.* 1997; 276:821–824. [PubMed: 9115211]
- Rauschecker JP. Parallel processing in the auditory cortex of primates. *Audiol Neurootol.* 1998; 3:86–103. [PubMed: 9575379]

- Rauschecker JP, Tian B. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci USA*. 2000; 97:11800–11806. [PubMed: 11050212]
- Rauschecker JP, Tian B. Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol*. 2004; 91:2578–2589. [PubMed: 15136602]
- Rauschecker JP, Tian B, Hauser MD. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*. 1995; 268:111–114. [PubMed: 7701330]
- Romanski LM, Averbek BB, Diltz M. Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J Neurophysiol*. 2005; 93:734–747. [PubMed: 15371495]
- Romanski LM, Goldman-Rakic PS. An auditory domain in primate prefrontal cortex. *Nat Neurosci*. 2002; 5:15–16. [PubMed: 11753413]
- 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; 2:1131–1136. [PubMed: 10570492]
- Rosch, E. Principles of categorization. In: Rosch, E.; Lloyd, BB., editors. *Cognition and Categorization*. Hillsdale, NJ: Erlbaum; 1978. p. 27-48.
- Russ BE, Ackelson AL, Baker AE, Theunissen FE, Cohen YE. Auditory spectrotemporal receptive fields in the superior temporal gyrus of rhesus macaques. *Assoc Res Otolaryngol Abstr*. 2007a; 689
- Russ BE, Lee Y-S, Cohen YE. Neural and behavioral correlates of auditory categorization. *Hear Res*. 2007b; 229:204–212. [PubMed: 17208397]
- Saberi K, Perrott DR. Cognitive restoration of reversed speech (Abstract). *Nature*. 1999; 398:760. [PubMed: 10235257]
- Schnupp JW, Hall TM, Kokelaar RF, Ahmed B. Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J Neurosci*. 2006; 26:4785–4795. [PubMed: 16672651]
- Schreiner CE, Mendelson J, Raggio MW, Brosch M, Krueger K. Temporal processing in cat primary auditory cortex. *Acta Otolaryngol Suppl*. 1997; 532:54–60. [PubMed: 9442845]
- Sereno AB, Maunsell JH. Shape selectivity in primate lateral intraparietal cortex. *Nature*. 1998; 395:500–503. [PubMed: 9774105]
- Shadlen MN, Newsome WT. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci*. 1998; 18:3870–3896. [PubMed: 9570816]
- Shannon CE. A mathematical theory of communication (Part 1). *Bell Syst Tech J*. 1948a; 27:379–423.
- Shannon CE. A mathematical theory of communication (Part 2). *Bell Syst Tech J*. 1948b; 27:623–656.
- Sugase Y, Yamane S, Ueno S, Kawano K. Global and fine information coded by single neurons in the temporal visual cortex. *Nature*. 1999; 400:869–873. [PubMed: 10476965]
- Tian B, Rauschecker JP. Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol*. 2004; 92:2993–3013. [PubMed: 15486426]
- Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. Functional specialization in rhesus monkey auditory cortex. *Science*. 2001; 292:290–293. [PubMed: 11303104]
- Treves A, Panzeri S. The upward bias in measures of information derived from limited data samples. *Neural Comput*. 1995; 7:399–407.
- Ungerleider, LG.; Mishkin, M. Two cortical visual systems. In: In: Ingle, DJ.; Goodale, MA.; Mansfield, RJW., editors. *Analysis of Visual Behavior*. Cambridge, MA: MIT Press; 1982.
- Werner-Reiss U, Kelly KA, Trause AS, Underhill AM, Groh JM. Eye position affects activity in primary auditory cortex of primates. *Curr Biol*. 2003; 13:554–562. [PubMed: 12676085]
- Winer JA, Lee CC. The distributed auditory cortex. *Hear Res*. 2007; 229:3–13. [PubMed: 17329049]
- Yu, BM.; Ryu, SI.; Santhanam, G.; Churchland, MM.; Shenoy, KV. Improving neural prosthetic system performance by combining plan and peri-movement activity. *Proc 26th Annu Int Conf IEEE EMBS; San Francisco, CA*. 2004.
- Zohary E, Shadlen MN, Newsome WT. Correlated neuronal discharge rate and its implications for psychophysical performance [published erratum appears in *Nature* 1994 Sep 22;371(6495):358]. *Nature*. 1994; 370:140–143. [PubMed: 8022482]

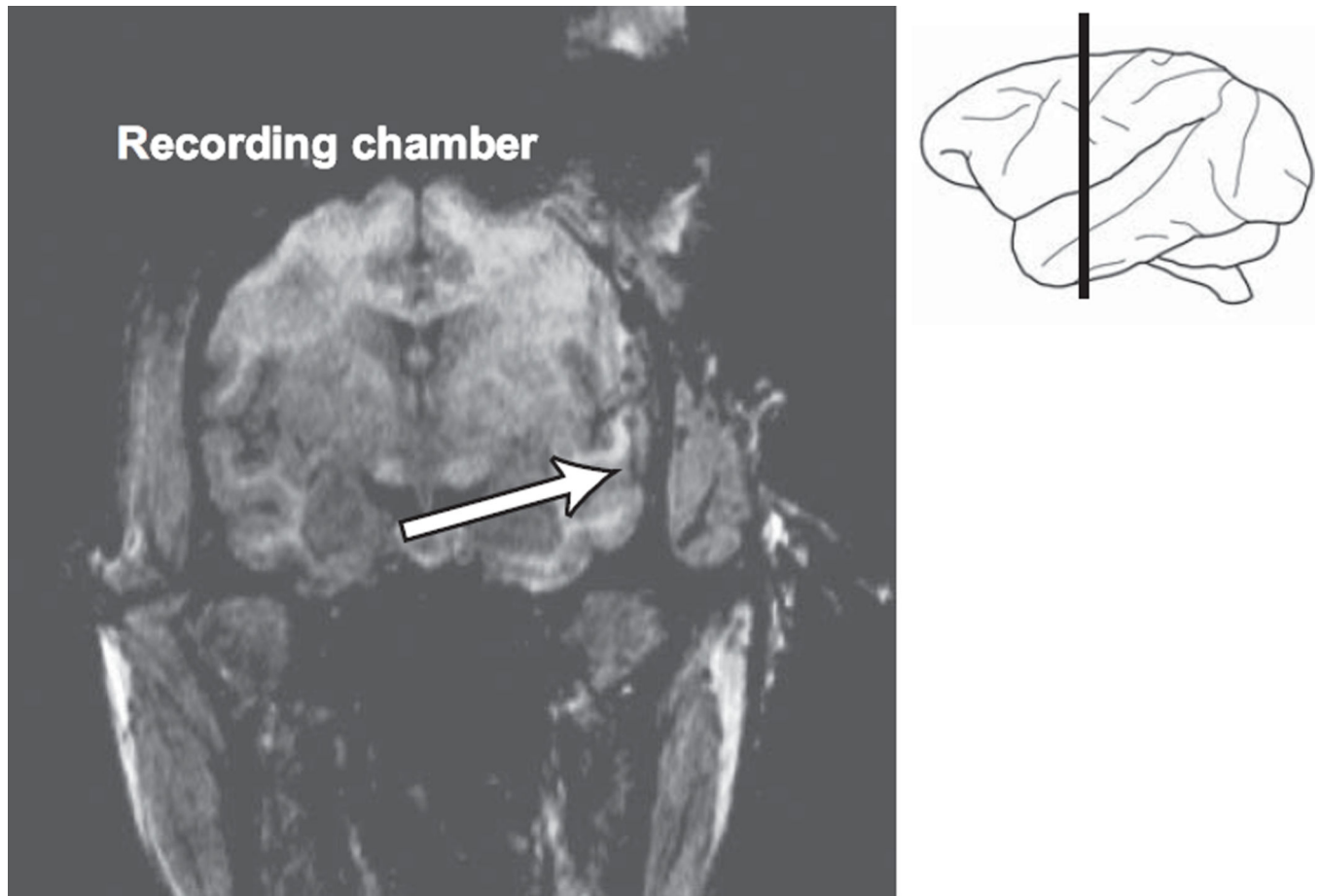


Fig. 1.

A coronal magnetic resonance image illustrating the approach of a recording electrode, as indicated by the white arrow, through the recording chamber and into the superior temporal gyrus (STG). The approximate plane of section of this image is illustrated by the vertical black line through the schematic of the rhesus brain.

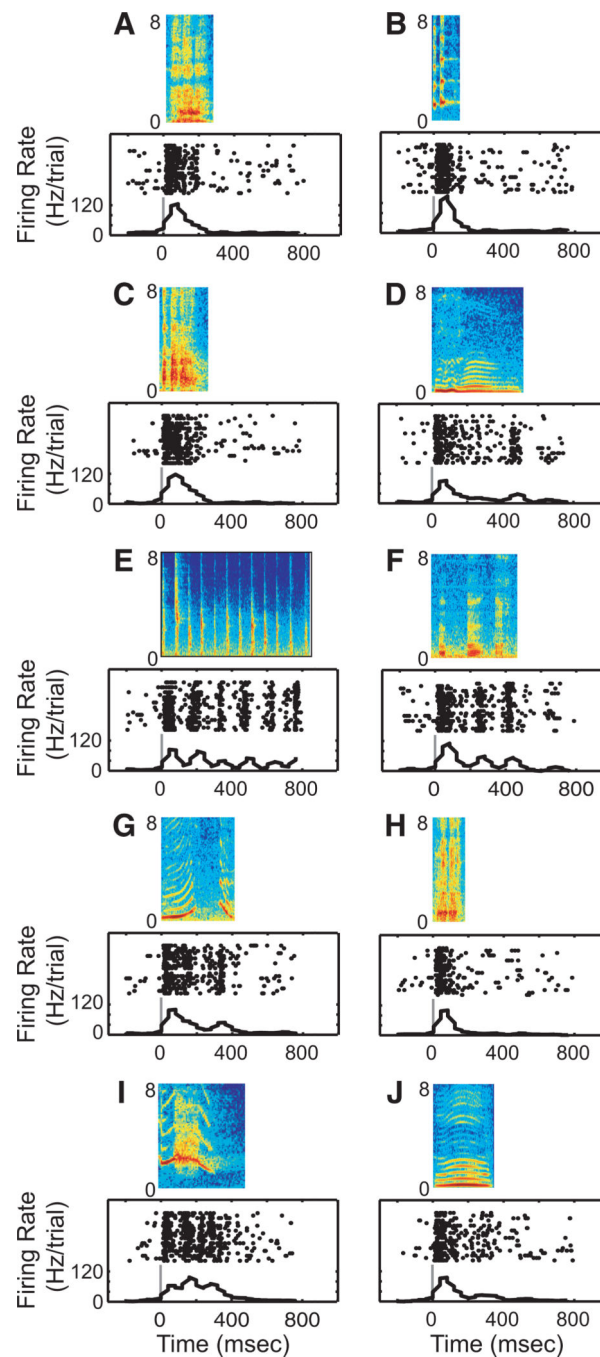


Fig. 2. Response profile of STG neuron to each of the 10 vocalization exemplars: aggressive (*A*), copulation scream (*B*), grunt (*C*), coo (*D*), gecker (*E*), girney (*F*), harmonic arch (*G*), shrill bark (*H*), scream (*I*), and warble (*J*). The rasters and peristimulus time histograms are aligned relative to onset of the vocalization; the solid gray line indicates onset of the vocalization. The histograms were generated by binning spike times into 40-ms bins.

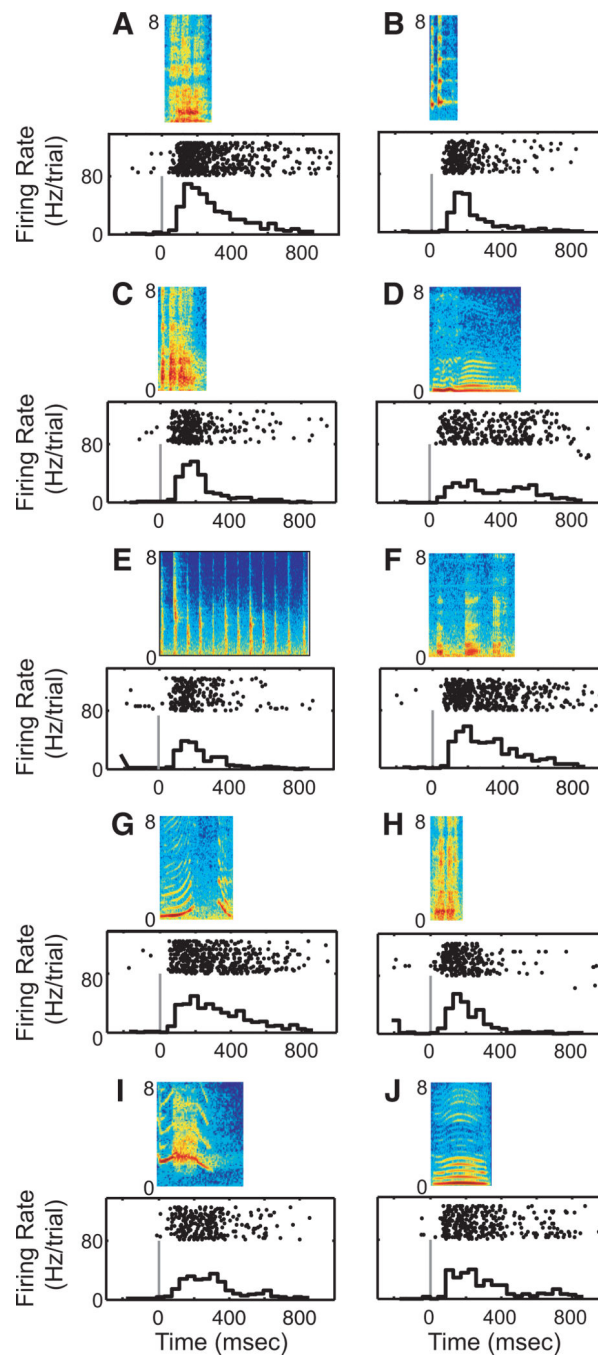


Fig. 3. Response profile of a ventrolateral prefrontal cortex (vPFC) neuron to each of the 10 vocalization exemplars. Same format as Fig. 2. This response profile was modified from Cohen et al. (2007).

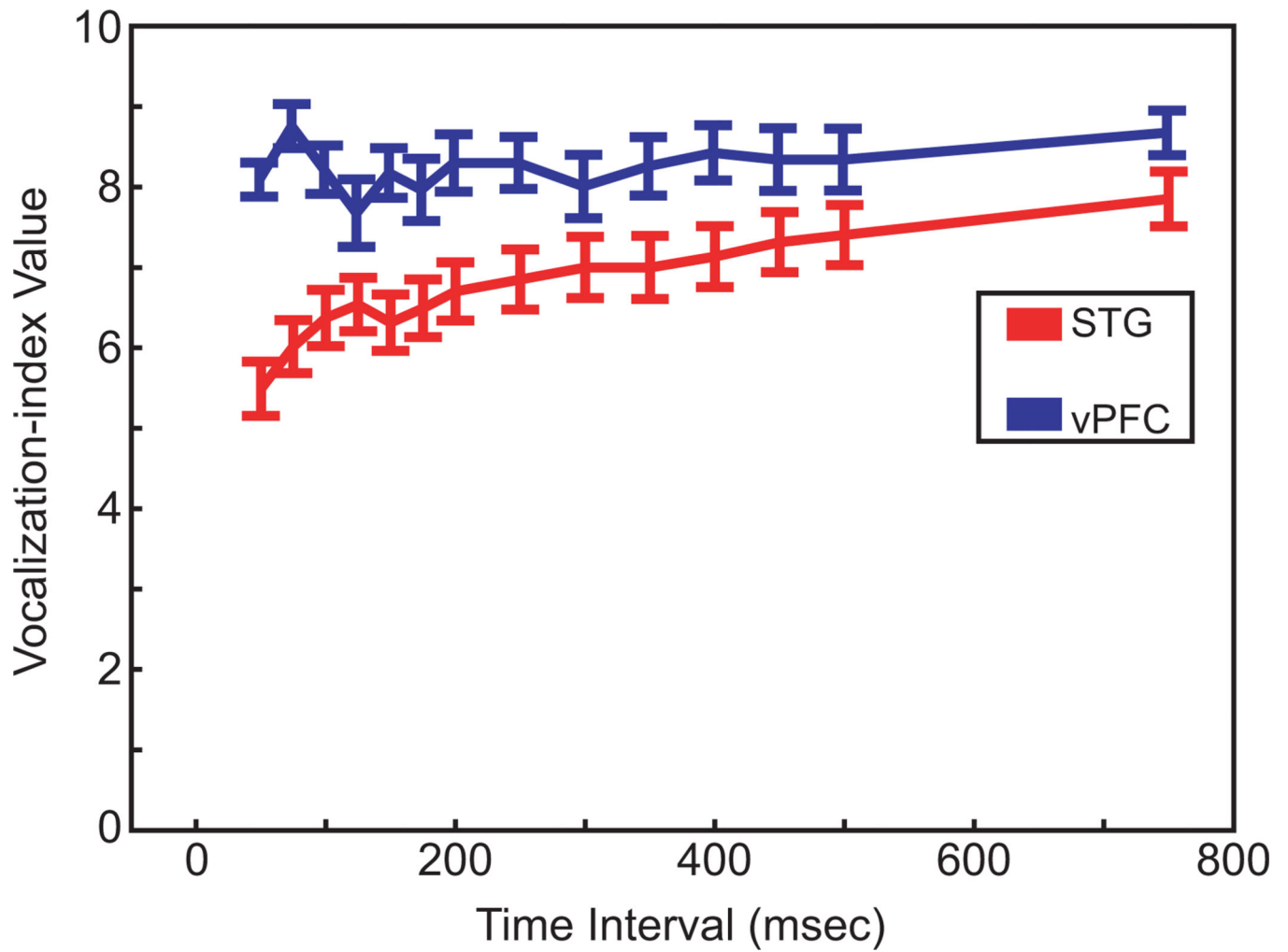


Fig. 4. Vocalization-preference index. The index values are calculated as a function of increasingly longer time intervals following vocalization onset. The STG data are shown in red and the vPFC data are shown in blue. The error bars represent the SE.

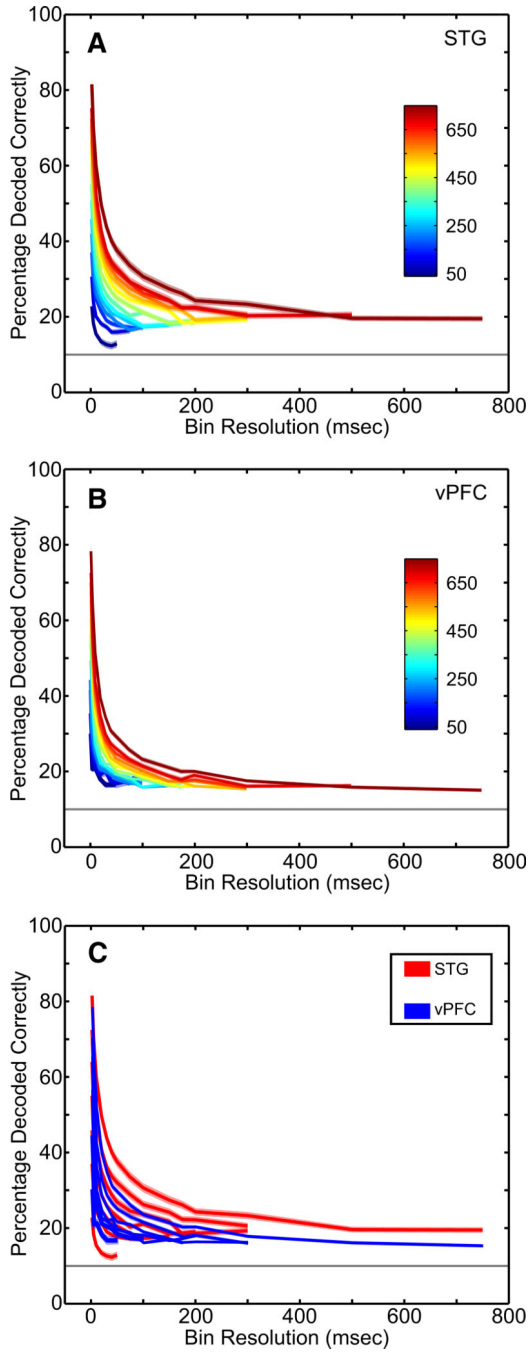


Fig. 5. Decoding capacity of STG and vPFC neurons with a linear-pattern discriminator. A–C: the percentage of correctly decoded vocalizations as calculated from the linear-pattern discriminator. The percentage of correctly decoded vocalizations for the population of (A) STG neurons and (B) vPFC neurons as a function of the time interval following vocalization onset (in ms; see color legend) and the bin size (x -axis). The solid colored lines are the mean value and the shaded areas represent the SE. The solid gray horizontal line indicates chance

performance (10%). C: the STG data (red) and the vPFC data (blue) are overlaid; to facilitate comparison, every other time interval has been removed.

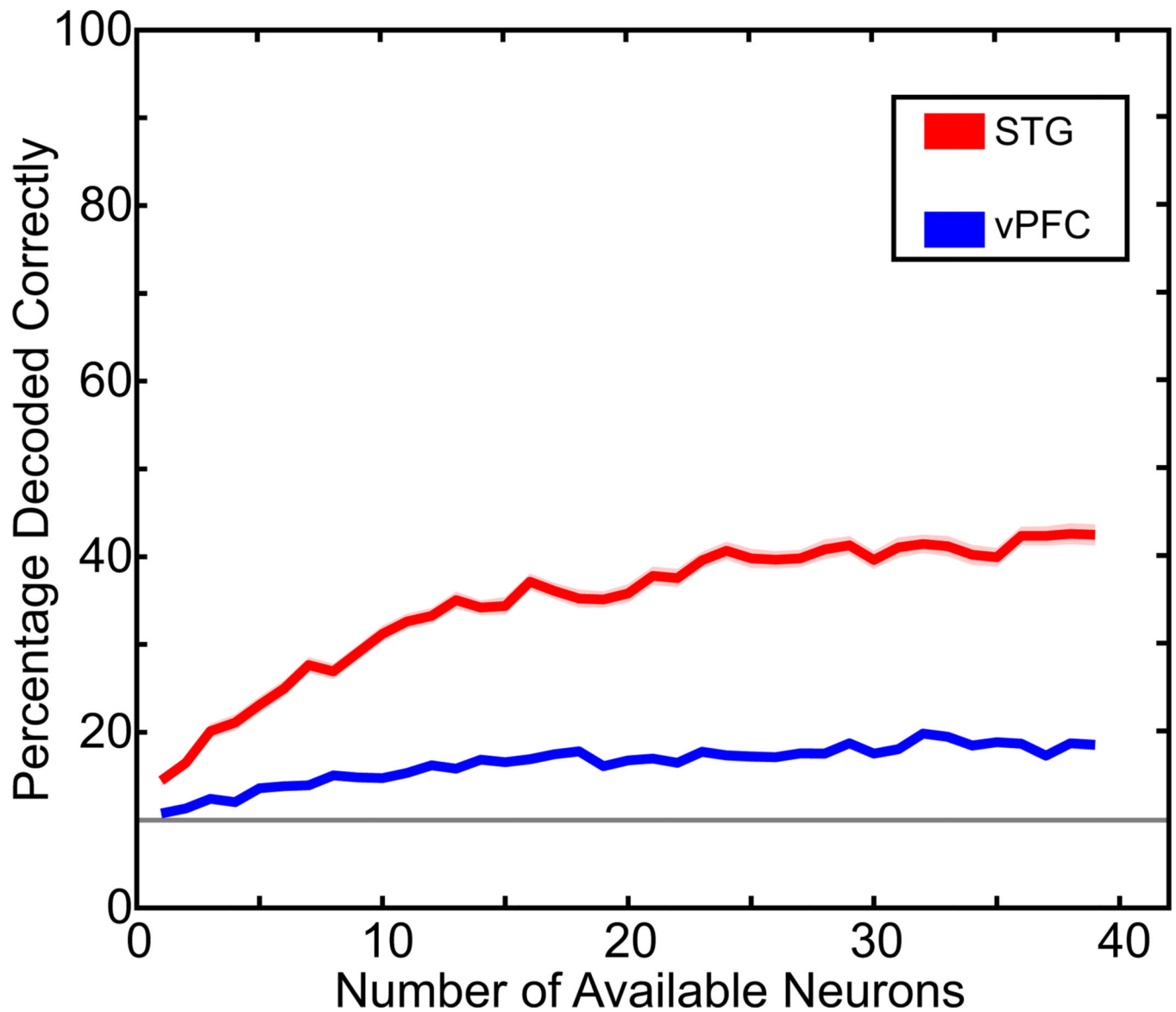


Fig. 6. Decoding capacity of STG and vPFC neurons with a maximum-likelihood estimator. The percentage of correctly decoded vocalizations as calculated from the maximum-likelihood estimator. The STG data are shown in red and the vPFC data are shown in blue. The solid colored lines are the mean value and the shaded areas represent the SE. The solid gray horizontal line indicates chance performance (10%).

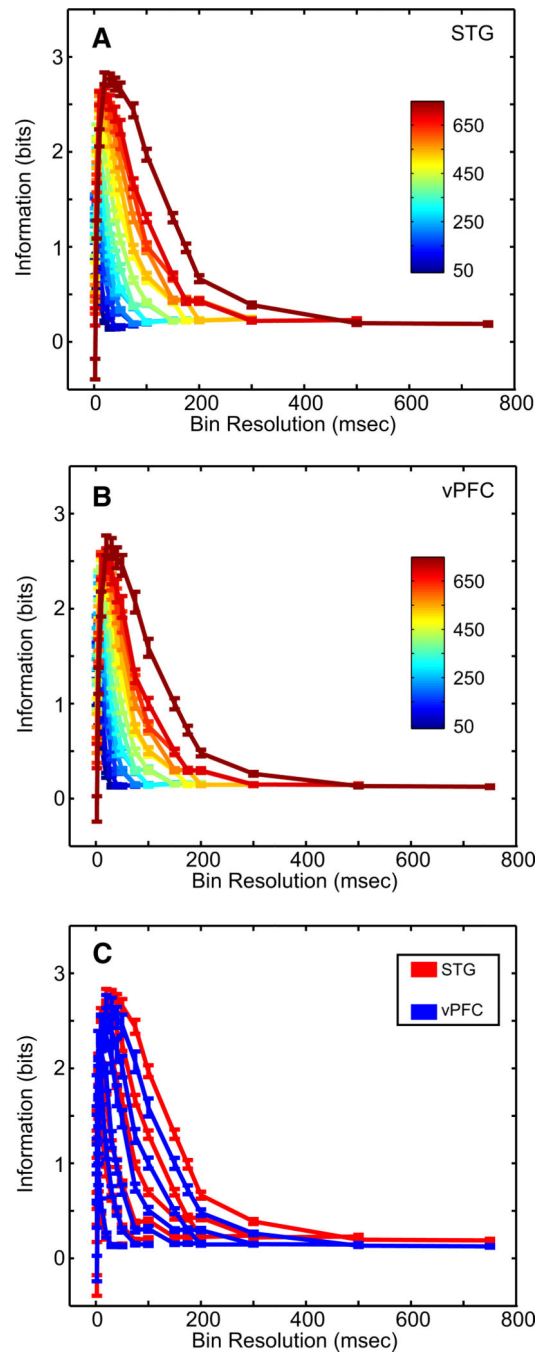


Fig. 7. Mutual information of STG and vPFC responses. A–C: the bias-corrected amount of mutual information (in bits) carried in (A) STG and (B) vPFC activity. Information is plotted as a function of the time interval following vocalization onset (in ms; see color legend) and the bin size (x -axis). The error bars represent SE. C: the STG data (red) and the vPFC data (blue) are overlaid; to facilitate comparison, some of the time intervals have been removed.