Representation of speech categories in the primate auditory cortex

Joji Tsunada,1 Jung Hoon Lee,1 and Yale E. Cohen1,2,3

Department of Otorhinolaryngology: Head and Neck Surgery,2 Department of Neuroscience, and3 Department of Bioengineering, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania

Submitted 12 January 2011; accepted in final form 17 February 2011

Tsunada J, Lee JH, Cohen YE. Representation of speech categories in the primate auditory cortex. J Neurophysiol 105: 2634–2646, 2011. First published February 23, 2011; doi:10.1152/jn.00037.2011.—A “ventral” auditory pathway in nonhuman primates that originates in the core auditory cortex and ends in the prefrontal cortex is thought to be involved in components of nonspatial auditory processing. Previous work from our laboratory has indicated that neurons in the prefrontal cortex reflect monkeys’ decisions during categorical judgments. Here, we tested the role of the superior temporal gyrus (STG), a region of the secondary auditory cortex that is part of this ventral pathway, during similar categorical judgments. While monkeys participated in a match-to-category task and reported whether two consecutive auditory stimuli belonged to the same category or to different categories, we recorded spiking activity from STG neurons. The auditory stimuli were morphs of two human-speech sounds (bad and dad). We found that STG neurons represented auditory categories. However, unlike activity in the prefrontal cortex, STG activity was not modulated by the monkeys’ behavioral reports (choices). This finding is consistent with the anterolateral STG’s role as a part of functional circuit involved in the coding, representation, and perception of the nonspatial features of an auditory stimulus.

AN IMPORTANT CONCEPTUAL MODEL in auditory neuroscience is that spatial and nonspatial sensory information is processed in parallel processing streams (Hackett et al. 1999; Kaas and Hackett 1999; Rauschecker 1998; Rauschecker and Scott 2009; Recanzone and Cohen 2010; Romanski et al. 1999a,b). A “dorsal” pathway processes the spatial attributes of a sound, whereas a “ventral” pathway processes a sound’s nonspatial attributes, such as its identity. The dorsal (spatial) auditory pathway begins in the caudomedial and caudolateral belt regions of the auditory cortex. These regions project to the dorsolateral areas of the prefrontal cortex via the posterior parietal cortex. The ventral (identity) auditory pathway begins in core auditory cortex (in particular, the primary auditory cortex and the rostral field R) and then continues in the anterolateral and mediolateral belt regions of the auditory cortex. These latter areas then project to the ventrolateral prefrontal cortex (vPFC) via the parabelt region of the auditory cortex (Rauschecker 1998; Rauschecker and Tian 2000; Romanski et al. 1999b).

What type of information processing occurs along the ventral (identity) pathway? One possibility is that information processing along this pathway reflects a transformation from stimulus encoding to categorization and ultimately to the decision-making computations that underlie categorical perceptual judgments (e.g., the stimulus is this and not that) (Binder et al. 2004; Gold and Shadlen 2007; Rauschecker and Scott 2009). Several sets of data are consistent with this hypothesis. For example, neurons in regions of the frontal lobe are involved in both auditory categorization and decision making (Binder et al. 2004; Burton and Small 2006; Gifford et al. 2005; Hickok and Poeppel 2007; Lee et al. 2009; Lemus et al. 2009b; Locasto et al. 2004; Russ et al. 2008b). In the core auditory cortex, acoustic processing dominates, which helps shape auditory-category representations (e.g., voice-onset time); however, only later in the belt and parabelt auditory fields are categorical-like responses to phonemes found (Binder et al. 2004; Chang et al. 2010; Husain et al. 2006; Lemus et al. 2009a; Liebenthal et al. 2010; Selezneva et al. 2006; Steinschneider et al. 2005, 2003, 1999, 1995, 1994). However, it remains unclear whether these regions of auditory cortex also represent decision-related activity: whereas some primate studies support a role for the primary auditory cortex in decision making (Selezneva et al. 2006), other studies suggest that the primate auditory cortex is limited to sensory-encoding/categorization (Binder et al. 2004; Lemus et al. 2009a).

To explicitly test whether the nonhuman primate secondary auditory cortex plays a role in categorization vs. decision making, we trained monkeys to discriminate between the human spoken words bad and dad as well as morphs of these words. While the monkeys participated in this task, we recorded from neurons in the superior temporal gyrus (STG). We targeted the region of the STG that, based on MRI coordinates, coincides with the anterolateral belt of the auditory cortex; the anterolateral belt is part of the ventral auditory stream and consequently processes information about auditory identity (Rauschecker and Scott 2009; Rauschecker and Tian 2000). Since this task design is analogous to our previous study that demonstrated a role for the vPFC in decision making (Lee et al. 2009; Russ et al. 2008b), we can compare the current findings with those from our vPFC study to infer the functional and hierarchical computations that occur along the ventral pathway.

MATERIALS AND METHODS

We recorded from STG neurons of two male rhesus monkeys. While the animals were under isofluorane anesthesia, both monkeys were implanted with a head-positioning cylinder and a recording chamber. Monkey H additionally was implanted with a scleral search coil (Judge et al. 1980). STG recordings were obtained from the left hemisphere of one rhesus monkey (monkey...
and from the right hemisphere of the other rhesus (monkey T). All recordings were guided by pre- and postoperative MRIs of each monkey’s brain. In an earlier study, we recorded from the prefrontal cortex of monkey H while the monkey participated in an auditory categorization similar to the one reported in this study (Russ et al. 2008b). Dartmouth College’s and the University of Pennsylvania’s Institutional Animal Care and Use Committees approved all of the experimental protocols.

Auditory Stimuli

The prototype stimuli were the spoken words bad and dad. Perceptually, these stimuli differ in their place of articulation of each word’s initial consonant (i.e., the place of articulation for /b/ is the lips, and for /d/, it is the roof of the mouth). The prototype stimuli were digitized recordings of an American female adult and were provided by Dr. Michael Kilgard. The prototype bad had the following acoustic properties: the fundamental frequency started at 284 Hz (range: 170–284 Hz); the first formant started at 579 Hz with a steady-state center frequency of 951 Hz; the second formant started at 2,050 Hz with a steady-state center frequency of 2,201 Hz. The prototype dad had the following acoustic properties: the fundamental frequency started at 251 Hz (range: 180–251 Hz); the first formant started at 508 Hz with a steady-state center frequency of 882 Hz; the second formant started at 2,436 Hz with a steady-state center frequency of 2,199 Hz. Neither prototype began with frication.

Morphed versions of the prototypes were created using the STRAIGHT (Kawahara et al. 1999) software package, which is run in the Matlab (The Mathworks) programming environment. Morphing was accomplished by calculating the shortest trajectory between the fundamental and formant frequencies of the two prototypes (Liberman et al. 1957); the shortest trajectory is based on a computed distance metric. Morphed versions of the two prototypes were created at 20, 40, 50, 60, and 80% of the distance along this trajectory. Example spectrograms of these stimuli are shown in Fig. 1. Operationally, the bad prototype was defined as the 0% morph, and the dad prototype was defined as the 100% morph.

In this stimulus design, we tested categorization using two prototypes (bad and dad). This stimulus design is similar to several other successful categorization studies (Ferrera et al. 2009; Freedman and Assad 2006; Koida and Komatsu 2007), especially those auditory studies using speech stimuli like those in this study (Chang et al. 2010; Kuhl and Miller 1975). We have previously tested categorization using multiple prototypes by varying the pitch of the bad-dad stimuli

Fig. 1. Match-to-category task and behavioral performance. A: following the presentation of a reference stimulus, a test stimulus was presented. Auditory stimuli consisted of 2 prototype spoken words (bad or dad) and linear morphs between these 2 prototypes. Bad was operationally defined as the 0% morph, whereas dad was defined as the 100% morph. A reference stimulus could be either of the prototypes or any of the morphs except for the 50% morph. A test stimulus could be any of the auditory stimuli. When the reference and test stimuli belonged to the same category, the monkeys made a saccade to the left light-emitting diode (LED); the monkeys' eye position is indicated by the dotted lines. When the reference and test stimuli belonged to different categories, the monkeys made a saccade to the right LED. B: spectrographic representations of the prototype spoken words and 2 of the morphs; see MATERIALS AND METHODS for more details. Relative power is indicated (dB).
and found that the monkeys’ behavioral responses were independent of the stimulus pitch (Russ et al. 2006).

**Match-to-Category Task**

As schematized in Fig. 1, the task began with a presentation of a “reference” stimulus that was followed by the presentation of a “test” stimulus. The reference stimulus and the test stimulus could be either of the prototypes or any of the morphs with one exception: the 50% morph was not allowed to be a reference stimulus. The reference and test stimuli were 500 ms in duration, and the interstimulus interval was between 1,100 and 1,300 ms. The stimuli were presented from a speaker (Pyle, PLX32) that was placed in front of the monkeys at their eye level. The stimuli were presented at 70 dB SPL.

The monkeys reported whether the reference and test stimuli belonged to the same category or to different categories. They reported this decision by making a saccade to one of two light-emitting diodes (LEDs) that were illuminated 1,100–1,300 ms after test-stimulus offset. The eye position of *monkey H* was monitored with a scleral search coil (Judge et al. 1980). The eye position of *monkey T* was monitored noninvasively with an infrared eye tracker (Eye-Trac6 RS6-HS; Applied Science Laboratories).

Since the 50% morph is at the phonemic-category boundary (Russ et al. 2008b) (see results), the monkeys’ behavioral reports were based on whether the reference and test stimuli were on the same side or on different sides of this 50%–morph boundary. That is, when the reference and test stimuli were on the same side of the 50%-morph boundary, they belonged to the same category. Examples of reference and test stimuli that were in the same category are 1) the reference stimulus was a 40% morph and the test stimulus was a 0% morph, or 2) the reference stimulus was a 100% morph and the test stimulus was a 0% morph. In contrast, when the reference and test stimuli were on different sides of the 50%-morph boundary, they belonged to different categories. Examples of reference and test stimuli that were in different categories are 1) the reference stimulus was a 40% morph and the test stimulus was a 100% morph, or 2) the reference stimulus was a 20% morph and the test stimulus was an 80% morph.

The monkeys indicated their behavioral reports with an eye movement. If the reference and test stimuli belonged to the same category, the monkeys were rewarded if they made a saccadic eye movement to the LED that was 20° to the left of the speaker. In contrast, if the reference stimulus and the test stimulus belonged to different categories, the monkeys were rewarded if they made a saccadic eye movement to the LED that was 20° to the right of the speaker. When the test stimulus was a 50% morph (i.e., at the categorical boundary) (Kuhl and Padden 1983; 1982; Russ et al. 2008b) (see results), there was not a “correct” answer. Consequently, the monkeys were rewarded on a random schedule based on their overall performance level (Grunewald et al. 2002).

**Psychophysical Data Analyses**

The monkeys’ performance on the match-to-category task was analyzed as a psychometric function: the independent variable of this fit was the relative morph value of the test stimulus and the dependent variable was the proportion of trials in which the monkey reported that the reference and test stimuli belonged to the same category. The psychometric data were fit with a cumulative Gaussian function. Psychometric data were calculated independently as a function of the distance between the reference stimulus and the prototype.

**Neurophysiological Data Analyses**

The neurons reported in this study were those classified as “auditory.” A STG neuron was classified as auditory, if the firing rate (spikes/second) of a neuron was reliably different (*P < 0.05*) between during the 100-ms period that began with reference- and test-stimulus onset and the baseline period (i.e., the 500-ms period that preceded reference-stimulus onset). This 100-ms period was chosen since many of the neurons had substantial phasic responses to stimulus onset; longer time windows did not substantially modify the population of auditory neurons (Kikuchi et al. 2010). This analysis was conducted independently of the morph value of the stimulus.

**Category Index.** Following from Miller and colleagues (Freedman et al. 2001; 2002), the selectivity of a neuron to a stimulus’ category was tested with the “category index.” Based on the current (see Fig. 2) and previous behavioral data (Russ et al. 2008b), we treated the 0 (i.e., the *bad* prototype), 20, and 40% morphs as one category and the 60, 80, and 100% (i.e., the *dad* prototype) morphs as a second category. The advantage of this index is that it quantifies whether STG neurons responded differentially to stimuli that have the same relative acoustic differences but can either exist 1) within the same category (e.g., a 20
On a neuron-by-neuron basis, we first calculated the “within-category difference” (WCD). The WCD was the average of the absolute difference in test-stimulus-period firing rate between morph pairs that were on the same side of the category boundary: for example, the 0 and the 20% morphs or the 60 and the 100% morphs. Second, we calculated the “between-category difference” (BCD), which was the average of the absolute difference in test-stimulus-period firing rate between morph pairs that were on the different sides of the category boundary: for example, the 40 and 80% morphs. The category index was the difference between the BCD and the WCD, divided by their sum. If the index value is >0, it indicated that a neuron’s response to pairs of between-category stimuli was larger than its response to pairs of within-category stimuli. This result is consistent with a neuron coding for category differences. If the index value is <0, it indicated that a neuron had a larger responses to pairs of within-category stimuli, a result inconsistent with category sensitivity.

To test the temporal dynamics of the category index, distributions of category-index values were calculated from data in consecutive 5-ms bins, relative to reference-stimulus or test-stimulus onset. Five-millisecond intervals were chosen to ensure that the time-varying nature of neural activity was captured in sufficient detail. Confidence intervals on the category-index values were created with a bootstrap analysis. We tested whether the mean value of the category index (±95% confidence-interval) did not include the value of 0 as a function of time.

Regression analysis. A regression analysis modeled the linear relationship between a neuron’s firing rate and the reference and test stimuli (Hernández et al. 2010; Lemus et al. 2009a). The specific form of the regression model was as follows: 

\[
\text{Firing Rate} = a_1(t) * \text{Morph value Reference stimulus} + a_2(t) * \text{Morph value Test stimulus} + a_3(t) .
\]

The \( a_1, a_2, \) and \( a_3 \) are regression coefficients, as a function of time. The \( a_1 \) and \( a_2 \) represent the effects of the reference and test stimulus, respectively, on a neuron’s firing rate. The \( a_3 \) represents any global bias in firing rate. Each morph stimulus (0–100%) was mapped to a Morph value between 0 and 1. With the use of the “regress” function in the Matlab programming environment, regression coefficients were calculated from data in consecutive 5-ms bins. A regression coefficient was reliable (\( P < 0.05 \)) if its 95% confidence interval, as determined by “regress,” did not include 0. The regression analysis was conducted on a neuron-by-neuron basis.

Neurometric analysis and comparison with psychometric data. To quantify the relationship between neural activity and the test stimulus on a neuron-by-neuron basis, signal-detection theory was used to compute a “neurometric” curve (Britten et al. 1992; Green and Swets 1966). The neurometric curve relates the probability that an ideal observer can use a neuron’s test-stimulus firing rate to discriminate between pairs of different test stimuli. By comparing the neurometric curve with the monkey’s psychometric curve, we can gain insight into how well single neurons can guide the monkeys’ choices (Parker and Newsome 1998).

Each neurometric curve was calculated on a neuron-by-neuron basis. We made the simplifying assumption that the neurometric curve can be calculated by comparing the test-stimulus-period activity that was elicited from a particular test-stimulus morph with that elicited from a morph equidistant from the 50% morph (e.g., 0 vs. 100%, 20 vs. 80%, 40 vs. 60%); this latter activity can be thought of as a hypothetical “anti-neuron” (Britten et al. 1996). To
calculate the curve for each of these pairs of test-stimulus morphs, we pooled the firing rates from all correct trials, into two distributions. Next, from these two distributions, a receiver-operating characteristic (ROC) curve was generated. The area under the curve represents the probability that an ideal observer could differentiate between these two test stimuli. This process was repeated for all pairs of test stimuli. The probability values were then plotted as a function of the test-stimulus morph to form a neurometric curve (Barlow et al. 1971; Cohn et al. 1971; Shadlen et al. 1996; Tolhurst et al. 1983). The neurometric curve was fit using a cumulative Gaussian function. Neurometric curves were calculated independently as a function of the distance between the reference stimulus and the prototype.

The relationship between a neuron’s neurometric curve and the corresponding psychometric curve was quantified by correlating the neurometric threshold with the psychometric threshold. The neurometric and psychometric thresholds are the SDs of their respective fitted cumulative Gaussian function (Gu et al. 2007). The threshold represents the amount of change needed to distinguish between two stimuli and indexes the neural sensitivity and behavioral sensitivity of the monkey, respectively.

ROC analyses to test for choice-related activity. A second set of ROC analyses further tested the relationship between neural activity and the monkeys’ behavioral reports (i.e., choices) (Britten et al. 1996, 1992; Freedman et al. 2003). In the first ROC analysis, we tested whether neural activity was modulated as a function of whether the monkey reported that the reference and test stimuli belonged to the same category (i.e., the stimuli were on the same side of the 50%–morph boundary) or to different categories (i.e., the stimuli were on different sides of the 50%–morph boundary). For this analysis, we restricted data only to correct trials and excluded trials in which the test stimulus was the 50% morph. On a neuron-by-neuron basis and trial-by-trial basis, we calculated the test-stimulus firing rates and placed these firing-rate values into one of two distributions based on the monkeys’ behavioral reports. From these two distributions, a ROC curve was generated and the area under this curve was calculated for each test stimulus. Finally, the test-stimulus-dependent ROC values were summed and averaged. These average ROC values from all of the recorded neurons were then pooled to form a distribution of ROC values.

In the second ROC analysis, we calculated “choice probability” (Britten et al. 1996; Gu et al. 2007; Purushothaman and Bradley 2005). Choice probability tested whether neural activity was modulated by the monkeys’ behavioral reports during those trials in which the test stimulus was the 50% morph. On a neuron-by-neuron basis and trial-by-trial basis, we calculated the test-stimulus firing rates and placed these firing-rate values into one of two distributions based on the monkeys’ behavioral reports. From these two distributions, a ROC curve was generated and the area under this curve was calculated for each test stimulus. Finally, the test-stimulus-dependent ROC values were summed and averaged. These average ROC values from all of the recorded neurons were then pooled to form a distribution of ROC values.

In the final analysis, we calculated a different form of choice probability that is called “grand choice probability” (Gu et al. 2007). On a neuron-by-neuron basis, we used data from both correct and incorrect trials to test whether neural activity was modulated as a function of the monkeys’ behavioral reports. For each pair of reference and test stimuli, the test-stimulus firing rate was placed into one of two distributions based on the monkeys’ behavioral report. From these two distributions, a ROC curve was generated and the area under this curve was calculated. The area under the ROC curve was calculated for all pairs of stimuli and averaged. These average ROC values from all of the recorded neurons were then pooled to form a distribution of ROC values.

RESULTS

Psychophysical Performance

Two monkeys participated in the match-to-category task. In this task, a “reference” and a “test” stimulus were presented sequentially. The stimuli were the prototype spoken words dad and dad or morphed versions of these two prototypes; the prototype dad was operationally defined as the 0% morph and the prototype dad was defined as the 100% morph; see MATERIALS AND METHODS for more details. Following stimulus presentation, the monkeys reported whether the two stimuli belonged to the same category or to different categories.

Figure 2 shows the psychophysical performance of monkey H (Fig. 2A) and monkey T (Fig. 2B) from all of the experimental sessions in which we recorded auditory neurons. For presentation, data were combined from trials in which the reference stimulus was equidistant from (or equivalently different than) one of the two prototypes. The data were organized in this manner to illustrate the fact that the monkeys’ behavioral reports were comparable, independent of the reference stimulus’ distance from the prototype exemplar and independent of the specific prototype. As a result, the data in the top row were generated from those trials when the reference stimulus was 0% different than the two prototypes (i.e., the 0 and the 100% morphs). The data in the middle row were generated from those trials when the reference stimulus was 20% different than the two prototypes (i.e., the 20 morph and the 80% morphs). The data in the bottom row were generated from those trials when the reference stimulus was 40% different than the two prototypes (i.e., the 40 morph and the 60% morphs).

To plot the data from these pairs of reference-stimulus morphs (i.e., the 0 and 100% morphs; the 20 and 80% morphs; and the 40 and the 60% morphs) on the same x-axis, the morph value of the test stimulus needed to be normalized. In this normalization, we calculated the absolute difference between the test-stimulus morph value and the prototype that was closest to the reference stimulus. That is, reference-stimulus morph values between 0 and 40% were mapped to a value of 0%, whereas reference-stimulus morph values between 60 and 100% were mapped to a value of 100%. For example, if the reference stimulus was the 80% morph and the test stimulus was the 60% morph, the x-axis value was 40% (100 – 60%).

The psychometric performance of monkey H is shown in Fig. 2A. When the reference stimulus was 0% different than the prototypes (top row), monkey H rarely reported that the reference stimulus and the test stimulus belonged to the same category when the test stimulus was 60–100% different. In contrast, monkey H regularly reported that the reference stimulus and the test stimulus belonged to the same category when the test stimulus was 0–40% different. Similar results are seen when the reference stimulus was 20% different (middle row) and 40% different (bottom row).

The performance of monkey T is seen in Fig. 2B. Whereas his performance is poorer than that of monkey H, it is clear that the monkey was successfully completing the task. Finally, the behavior of both monkeys is comparable to that seen in other nonhuman animal studies (Kluender et al. 1987; Kuhl and Miller 1975; Kuhl and Padden 1982), including previous work.
from our laboratory (Russ et al. 2008b), of phonetic categorization. Overall, these data indicate that the monkeys perceived the morphs of bad and dad in a categorical manner: the monkeys’ treated the 0–40% morphs as one category and the 60–100% morphs as a second category with the 50% morph serving as the category boundary. Moreover, their behavior did not depend on the morph value of the reference stimulus (compare the Fig. 2, top, middle, and bottom).

**Neurophysiological Data**

One-hundred and ten neurons from the STG were classified as “auditory”; see Materials and Methods. Since the responses from the two monkeys were comparable, data were pooled for analysis.

The response profiles from two STG neurons are shown in Fig. 3. The neuron in Fig. 3A had a categorical response to the stimuli. That is, during presentations of the reference or test stimuli, the neuron responded strongly to the 0 (the prototype bad), 20, and 40% morphs but had a relatively weaker response to the 60, 80, and 100% (the prototype dad). In addition to differences in firing rate, the temporal response profile of this neuron was also categorical. When the 0 (the prototype bad), 20, or 40% morphs were presented, the neuron had a biphasic firing pattern in response to the early components of the stimulus (i.e., the first ~125 ms following stimulus onset); this can be seen most clearly in the raster plot. In contrast, when the 60, 80, or 100% (the prototype dad) morphs were presented, the neuron did not have this early biphasic component but had a biphasic component later in the trial.

**Fig. 3.** Examples of superior temporal gyrus (STG) activity during the match-to-category task. For both A and B, the rasters and peristimulus-time histograms at left are aligned relative to onset of the reference stimulus. Rasters and peristimulus time histograms at right in A and B are aligned relative to onset of the test stimulus. Neural activity is color-coded relative to the morph value of the stimulus; this color code is shown in the legend. Only successful trials are shown. Peristimulus time histograms were generated by calculating the average firing-rate in a 5-ms time bins and then smoothed with a 50-ms kernel.

A neuron with a different response profile is shown in Fig. 3B. This neuron’s response was not categorical but instead was a graded function of the morph value. The neuron’s firing rate was highest in response to the 100% morph (the prototype bad) and lowest in response to the 0% morph (the prototype dad). At morph values between 0 and 100%, the firing rate of the neuron scaled between the firing-rate values elicited by the two prototypes. The latency of this neuron, unlike the example in Fig. 3A, was also modulated by the morph value in a somewhat linear fashion: the 0% morph had the longest response latency, whereas the 100% morph had the shortest response latency.

The following analyses quantify the degree to which these neurons represent the stimulus categories and whether they are modulated by the monkeys’ behavioral reports.

**Category index.** We first tested whether neurons coded the stimulus category through a category index. This index quantified whether STG neurons responded differentially to pairs of stimuli that had the *same* acoustic differences but can either exist 1) within the same category (e.g., a 20 vs. a 40% morph) or 2) between different categories (e.g., a 40 vs. a 60% morph). If the index value is >0, it indicated that neurons had larger responses to pairs of between-category stimuli than to pairs of within-category stimuli. This result is consistent with the hypothesis that the neurons coded for category differences. If the index value is <0, it indicated that neuron had larger responses to within-category stimuli, a result inconsistent with category sensitivity.

To test the temporal dynamics of this category index, we computed category-index values in 5-ms consecutive bins starting at reference- or test-stimulus onset. The results of this analysis can be seen in the Fig. 4. Before reference (Fig. 4A)- or test-stimulus onset (Fig. 4B), the value of the category index was not reliably different from 0. However, following reference- or test-stimulus onset, category-index values rapidly increased to values that were reliably (*P* < 0.05) >0 (the peak value was ~0.13–0.18). Thus, on average, the firing rates of STG neurons reflected the category membership of the auditory stimuli.

**Regression analysis.** We hypothesized that if STG neurons contained information about the monkey’s behavioral reports (choice), the test-stimulus firing rate should be strongly modulated by *both* the test stimulus and the preceding reference stimulus. To test this hypothesis, we modeled the test-stimulus firing rates of STG neurons as a linear function of the reference stimulus and the test stimulus (Lemus et al. 2009a; Lemus et al. 2009b; Romo et al. 2004). Firing_Rate = a1(t) * Morph_valueReference stimulus + a2(t) * Morph_valueTest stimulus + a3(t). The *a*1, *a*2, and *a*3 are regression coefficients, as a function of time. The *a*1 and *a*2 represent the effects of the reference and test stimulus, respectively, on a neuron’s firing rate; *a*3 represents any global bias in firing rate (see MATERIALS AND METHODS for more details). An example of this analysis from a single neuron is shown in Fig. 5A; the data in this figure were derived from the neural data in Fig. 3A. In contrast to our hypothesis, the neuron’s firing rate was reliably modulated by only the test stimuli (i.e., *a*3 ≠ 0) and was essentially independent of the reference stimuli (i.e., *a*1 = 0).

Population data are shown in Fig. 5B and are consistent with the aforementioned single-neuron example. For each neuron, we found those time points (with 5-ms resolution) where the value of the *a*1 coefficient or the *a*2 coefficient was reliably (*P* < 0.05) different than 0. Next, using this pair of coefficient values, we calculated an index: the difference between the absolute value of the *a*2 coefficient and the absolute value of the *a*1 coefficient, divided by their sum. If the index value was >0, it indicated that the absolute value of the *a*2 coefficient was greater than that of the *a*1 coefficient. If the index value was <0, it indicates that absolute value of the *a*1 coefficient was greater than that of the *a*2 coefficient. Since this index normalizes for differences in overall firing rate between different neurons, data could be pooled for a population analysis. Similar to the
single-neuron example (Fig. 5A), the distribution of index values indicated that, on average, the \( a_2 \)-coefficient values were reliably greater (\( P < 0.05 \)) than the \( a_1 \)-coefficient values. That is, the firing rates of STG neurons were strongly modulated by the test stimuli but were less dependent on the reference stimulus. Since STG activity was only strongly dependent on the test stimulus, it suggests that STG neurons did not convey information about the monkey’s behavioral reports (choice) (Lemus et al. 2009a; Lemus et al. 2009b; Romo et al. 2004). In subsequent analyses, we further flesh out this observation.

**Neurometric analysis.** Using signal-detection theory, we computed a neurometric curve (Britten et al. 1992; Green and Swets 1966) (see MATERIALS AND METHODS) to test the relationship between neural activity and the monkeys’ behavior. A neurometric curve relates the probability that an ideal observer can discriminate between pairs of different test stimuli from a neuron’s test-stimulus firing rate. By comparing the neurometric curve with a psychometric curve, we can gain insight into how well single neurons guide the monkeys’ perceptual choices (Parker and Newsome 1998).

Figure 6 shows examples of neurometric and psychometric curves. The neural data in this figure were generated from the example neuron shown in Fig. 3A. In Fig. 6, the data are collapsed as a function of their distance from the prototype. Consequently, the top represents data when the reference stimulus was one of the two prototypes; the middle represents data when the reference stimulus was 20% different from either of the two prototypes; and the bottom represents data when the reference stimulus was 40% different from either of the two prototypes. The psychometric curves in Fig. 6, top, middle, and bottom, show the monkey’s psychophysical behavior during those trials of the match-to-category task in which we collected the neural data that went into the neurometric curves. As can be seen, the neurometric and psychometric curves in Fig. 6 do not overlap, a result that is consistent with the hypothesis that this STG neuron is not predictive of the monkey’s behavior.

The relationship between a neuron’s neurometric curve and the monkey’s psychometric curve was quantified by correlating the neurometric threshold with the psychometric threshold. The threshold represents the amount of change needed to distinguish between two stimuli and indexes the neural sensitivity and behavioral sensitivity of the monkey, respectively (Gu et al. 2007). Figure 7, A (monkey H) and B (monkey T), summarizes the relationship between the neurometric and psychometric thresholds for our population of auditory STG neurons. As is readily apparent, there is considerable scatter in both graphs. Some neurons had neurometric thresholds that were comparable to the psychometric thresholds. Other neurons had neurometric thresholds that were better than the psychometric thresholds, indicating neurons that were more sensitive than the monkey’s behavior. However, a large population of neurons did not appear to be predictive of behavior: these neurons had larger neurometric thresholds than psychometric thresholds. Indeed, on average, we could not identify a reliable (\( P > 0.05 \)) correlation between the neurometric and psychometric thresholds for any of the reference stimuli.

**ROC analyses to test for choice-related activity.** Next, we conducted a series of ROC analyses to test further whether STG activity is modulated by the monkeys’ behavioral reports. An example of the relationship between STG activity and behavioral report is shown in Fig. 8; the data in this Fig. were derived from that shown in Fig. 3A. The neuron’s response was independent of the monkey’s report. That is, the neuron’s response during trials when the monkey reported that the
stimuli belonged to the same category (black data) was comparable to its response during trials when he reported that the stimuli belonged to different categories (grey data). Hence, like the results of the regression analysis (see Fig. 5), this neuron does not appear to be predictive of the monkey’s behavioral reports.

This observation was quantified with three ROC analyses (Britten et al. 1996, 1992; Freedman et al. 2003; Purushothaman and Bradley 2005); see MATERIALS AND METHODS for more details. If a neuron’s ROC value equals 0.5, it indicates that there was not any relationship between a neuron’s firing rate and the monkeys’ behavioral reports. If the ROC value is >0.5 (i.e., more activity when the monkey reported that the reference and test stimuli belonged to the same category) or <0.5 (i.e., more activity when the monkey reported the reference and test stimuli belonged to the different categories), it indicates that firing rates were predictive of the monkeys’ behavioral reports.

In the first ROC analysis, we tested whether neural activity was modulated as a function of the monkeys’ behavioral reports (same category or different categories); this analysis was limited to correct trials and excluded those trials in which the test stimulus was a 50% morph. Figure 9A shows the distribution of ROC values that were generated from the population of auditory STG neurons. The mean value of this distribution was 0.5 (SD = 0.05; 95% confidence interval = 0.49–0.51) a value that was not reliably different from 0.5.
neurons. Across three different analyses, we found that the monkeys’ choice-probability was different from 0.5 (mean = 0.48 – 0.51). For both of these analyses, the mean value of the distributions was not reliably different from 0.5 (P > 0.05). Also, the vast majority of STG neurons (86 and 96%, respectively) had choice-probability values and the grand-choice-probability values that were not reliably different (P > 0.05; permutation test) from 0.5. Thus, across three different analyses, we found that the monkeys’ behavioral reports did not reliably modulate the firing rates of neurons.

DISCUSSION

Neural activity from the STG was recorded while monkeys categorized the words bad, dad, and morphs of these two words. The monkeys reported that stimuli on one side of the 50%-morph phonetic-category boundary were the same, whereas stimuli on different sides of the 50%-morph boundary were different. That is, they perceived bad, dad, and morphs of these two prototypes in a categorical fashion. This behavioral result is similar to that seen in other human and nonhuman animal studies (Chang et al. 2010; Kluender and Alexander 2008; Kluer and Alexander 1987; Kuhl and Miller 1975; Kuhl and Padden 1982; Liebenthal et al. 2010; Russ et al. 2008b; Sinnott and Gilmore 2004). We found that, as a population, neural responses emphasized the categorical nature of the bad-dad stimuli, although the population also contained neurons that appeared to respond to the graded nature of the acoustics of the sounds. Further analyses indicated that STG activity, on average, was not predictive of the monkeys’ choices.

Three previous studies have examined the role of the primate auditory cortex in sensory encoding/categoryization and decision making (Binder et al. 2004; Lemus et al. 2009a; Seleznева et al. 2006). Our results are, in general, consistent with a human-functional MRI study by Binder et al. as well as a recent single-unit study by Lemus et al.: in both of these studies, neural activity in the auditory cortex correlated best with elements of sensory encoding/categoryization and was not modulated by the subjects’ choices. Our data are also consistent with a large body of human-imaging studies that suggest that phonetic analysis and processing occurs in nonprimary cortical regions of the auditory cortex (Chang et al. 2010; Guenther et al. 2004; Hickok and Poeppel 2007; Husain et al. 2006; Kluender and Alexander 2008; Leaver and Rauschecker 2010; Liebenthal et al. 2010; Olesler et al. 2010; Olesler et al. 2007; Poeppel et al. 2004; Raizada and Poladrack 2007; Wolmetz et al. 2010; Zatorre and Binder 2000), whereas the primary auditory cortex codes the spectral and temporal properties of speech sounds (Engineer et al. 2008; Mesgarani et al. 2008; Steinschneider and Fishman 2010).

In contrast, Seleznева et al. (2006) demonstrated that neurons in the auditory cortex not only reflected the stimulus category (i.e., the direction of the pitch change of tones in a sequence) but were also modulated by the monkeys’ choices. What could explain the differences between our study and the Seleznева findings? We hypothesize that these differences relate to differences between the task demands of the current study and the Seleznева study. In the Seleznева study, the monkeys learned a complex rule that required them to categorize the pitch direction of a sequence of tones, independent of the absolute frequency of the tones and independent of when the pitch direction occurred within a tone sequence. Indeed, the Seleznева study is consistent with a large literature on human evoked potentials suggesting that the auditory cortex is modulated by the behavioral response of a subject (Celsis et al. 1999; Kayser et al. 1998; Novak et al. 1990; Woods et al. 1994). Nevertheless, it is conceivable that this categorization task might require different or more neural resources than the one used in the current study. Additionally, the means by which monkeys learned the pitch-direction task of Seleznева et al. (2006) and our task may be substantially different. Thus it may be that differences between the task demands and learning requirements differentially engaged the auditory cortex.

Our auditory-categoryization task required that the monkeys report whether the reference and test stimuli belonged...
to the same category or to different categories. Where and how do these decision signals evolve in the cortex? Previous studies from our laboratory tested activity in the vPFC during a comparable auditory-categorization task and found that, unlike STG neurons, vPFC neurons reflected the monkeys’ behavioral reports. This suggests that a hierarchical relationship exists between STG and the vPFC whereby STG provides the “sensory evidence” to form the decision and vPFC activity encodes the output of the decision process. This hierarchical relationship is consistent with the STG’s role as a part of functional circuit involved in the coding, representation, and perception of the nonspatial features of an auditory stimulus (i.e., auditory identity) (Cohen et al. 2009, 2007; Rauschecker 1998; Rauschecker and Scott 2009; Rauschecker and Tian 2000; Romanski and Averbeck 2009; Romanski et al. 1999b; Russ et al. 2008a,b). This relationship may also be somewhat analogous to that observed in visual decision making: neural activity in parts of extrastriate visual cortex is highly sensitive to features of the visual stimulus and only weakly related to choice, whereas neural activity in area LIP of the parietal cortex, parts of the PFC, including the frontal eye fields, encode the input to, dynamics of, and output of the decision process, including an accumulation of evidence into a decision variable that governs the behavioral response (Britten et al. 1992; Celelbrini and Newsome 1995; 1994; Horwitz and Newsome 1999; Kim and Shadlen 1999; Newsome and Pare 1988; Roitman and Shadlen 2002; Salzman et al. 1990, 1992; Shadlen and Newsome 2001).

In addition to reflecting the outputs of a decision process, neurons between the STG and the vPFC may be increasingly sensitive to more abstract categorical information (Griffiths and Warren 2004). Indeed, vPFC neurons code the referential information that a vocalization transmits, independent of the vocalizations’ acoustic properties (Gifford et al. 2005): vPFC activity is not modulated by acoustically distinct vocalizations that transmit the same information (e.g., high-quality food) but is modulated by acoustically-distinct vocalizations that transmit different types of information (i.e., low-quality vs. high-quality food). Consistent with this hypothesis, vPFC neurons have the capacity to code more complex properties (e.g., referential information) of a vocalization than STG neurons (Miller and Cohen 2010). These differences between STG and vPFC auditory-categorical activity are also consistent with studies of visual categorization: the responses of prefrontal neurons tend to vary with the rules mediating a task or the behavioral significance of stimuli during a categorization task, whereas responses in the inferior temporal cortex tend be better correlated with the stimulus’ physical properties than prefrontal neurons (Ashby and Spiering 2004; Freedman et al. 2003).

**ACKNOWLEDGEMENTS**

We thank Kate Christison-Lagay, Subhash Bennur, Maria Geffen, and Heather Hersh for helpful comments on the preparation of this manuscript. We also thank Harry Shirley for outstanding animal support.

**GRANTS**

Y. E. Cohen was supported by grants from the National Institute on Deafness and Other Communication Disorders.

**DISCLOSURES**

No conflicts of interest, financial or otherwise, are declared by the author(s).


J Neurophysiol • VOL 105 • JUNE 2011 • www.jn.org

Downloaded from http://jn.physiology.org/ on October 9, 2016 by 10.220.33.6