Cortical speech-evoked response patterns in multiple auditory fields are correlated with behavioral discrimination ability

T. M. Centanni, C. T. Engineer, and M. P. Kilgard
School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, Texas

Submitted 5 February 2013; accepted in final form 10 April 2013

Centanni TM, Engineer CT, Kilgard MP. Cortical speech-evoked response patterns in multiple auditory fields are correlated with behavioral discrimination ability. J Neurophysiol 110: 177–189, 2013. First published April 17, 2013; doi:10.1152/jn.00092.2013.—Different speech sounds evoke unique patterns of activity in primary auditory cortex (A1). Behavioral discrimination by rats is well correlated with the distinctness of the A1 patterns evoked by individual consonants, but only when precise spike timing is preserved. In this study we recorded the speech-evoked responses in the primary, anterior, and posterior areas of the rat and evaluated whether activity in these fields is better correlated with speech discrimination ability when spike timing information is included or eliminated. Spike timing information improved consonant discrimination in all four of the auditory fields examined. Behavioral discrimination was significantly correlated with neural discrimination in all four auditory fields. The diversity of speech responses across recordings sites was greater in posterior and ventral auditory fields compared with A1 and anterior auditory fields. These results suggest that, while the various auditory fields of the rat process speech sounds differently, neural activity in each field could be used to distinguish between consonant sounds with accuracy that closely parallels behavioral discrimination. Earlier observations in the visual and somatosensory systems that cortical neurons do not rely on spike timing should be reevaluated with more complex natural stimuli to determine whether spike timing contributes to sensory encoding.

auditory cortex; diversity; parallel hierarchy; spike timing; rat

Address for reprint requests and other correspondence: T. M. Centanni, School of Behavioral and Brain Sciences, Univ. of Texas at Dallas, 800 W Campbell Rd., GR 4.1, Richardson, TX, 75080 (e-mail: tmcentanni @gmail.com).

Moxon 2004; Perez et al. 2013; Ranasinghe et al. 2012b; Shetake et al. 2011).

The uniqueness of these patterns is correlated with behavioral ability of rats in a wide range of tasks. Rats are able to discriminate human speech sounds in quiet (Engineer et al. 2008; Perez et al. 2013), background noise (Shetake et al. 2011), and after spectral or temporal degradation (Ranasinghe et al. 2012b). The neural responses in A1 can predict behavioral ability in all of these tasks using a Euclidean distance (ED) classifier. Sounds that create contrasting patterns of activity, such as the opposite patterns evoked by /b/ and /d/, correspond to pairs of sounds that are more easily discriminated by rats (Engineer et al. 2008; Ranasinghe et al. 2012a, 2012b; Shetake et al. 2011). Pairs of sounds that have similar spectrotemporal profiles, such as /t/ and /l/, cause similar neural responses and are more difficult for rats to behaviorally discriminate (Engineer et al. 2008). Degraded stimuli, like those caused by the addition of background noise or a vocoder, can cause delayed and/or weakened neural responses. The more severe the degradation to the neural response patterns, the more impaired the rats were at the corresponding behavior task (Ranasinghe et al. 2012b; Shetake et al. 2011). The preservation of the spike timing information is crucial to these correlations: if spike timing information is removed, no tasks are correlated with the differences in patterns of neural activity (Engineer et al. 2008; Ranasinghe et al. 2012a, 2012b; Shetake et al. 2011).

Other sensory systems appear to rely predominantly on spike rate, rather than spike timing for discrimination, especially in sensory fields higher up in their respective pathways. In the visual system, the primary visual cortex relies on precise spike timing to discriminate between stimuli that evoke similar number of spikes (Montemurro et al. 2008). As the information is passed to higher level regions, neurons begin to use spike rate information rather than spike timing to make behavioral decisions. For example, when monkeys were asked to identify the direction of motion of a visual stimulus after a delay, the rate of firing in the posterior parietal cortex predicted the behavioral decision of the subject (Shadlen and Newsome 1996, 2001). The somatosensory system functions in a similar manner. For example, when monkeys were trained to detect whether the rate of two tactile vibrations were the same, high-level areas of the somatosensory system perform the task by comparing the firing rate of neurons to each stimulus (Romo and Salinas 2003). A well-received theory states that temporal information is consistently transformed into a rate code as the information moves up the appropriate neural network (Ahissar et al. 2000). This is plausible in the auditory system as well (Buonomano and Merzenich 1995; Wang et al. 2008). The
The present study was designed to test the theory that neural responses to consonant speech sounds would also undergo a transformation throughout the auditory pathway. Early auditory cortex requires precise spike timing information to encode consonant sounds (Centanni et al. 2013; Engineer et al. 2008; Ranasinghe et al. 2012a, 2012b; Shetake et al. 2011), while vowel stimuli rely on spike rate, even at this early sensory level (Perez et al. 2013). We therefore used consonant stimuli in the present study to answer the following question: Do auditory fields higher in the auditory pathway than A1 use spike timing or spike rate to encode and discriminate consonant speech sounds?

METHODS

Anesthetized recordings. We acquired 1,253 multiunit recordings from the auditory cortex of 15 experimentally naive rats. The recording procedure is explained in detail elsewhere (Engineer et al. 2008). In brief, animals were anesthetized with pentobarbital (50 mg/kg) and given supplemental dilute pentobarbital (8 mg/ml) as needed to maintain areflexia, along with fluids to prevent dehydration. A tracheotomy was performed to ensure ease of breathing throughout the experiment. A1 and several nearby auditory fields were exposed via craniotomy and durotomy. Four Parylene-coated tungsten microelectrodes (1–2 MΩ) were simultaneously lowered to layer IV/V of either left or right auditory cortex (~600–800 μm). Responses were collected in four auditory fields: A1, anterior auditory field (AAF), posterior auditory field (PAF), and ventral auditory field (VAF). These fields are widely considered core regions in the rat (Doron et al. 2002; Polley et al. 2006; Storace et al. 2010), but recent evidence suggests that these fields receive input from more than one thalamic region, which may suggest a hierarchical organization (Smith et al. 2012; Storace et al. 2012).

Brief (25 ms) tones with 5-ms ramps were presented at 90 randomly interleaved frequencies (1–48 kHz) at 16 intensities (0–75 dB sound pressure level) to determine the characteristic frequency (CF) of each site. We also presented seven English consonant-vowel-consonant speech sounds (/dad/, /sad/, /tad/, /bad/, /gad/, /shad/, and /chad/) previously tested in our laboratory (Engineer et al. 2008; Floody et al. 2010; Ranasinghe et al. 2012b; Shetake et al. 2011). The speech stimuli were randomly interleaved and presented at 20 repeats per recording site. Sounds were shifted up one octave into the rat hearing range using the STRAIGHT vocoder (Kawahara 1997; Fig. 1). Each sound was calibrated with respect to its length so that the most intense 100 ms of the stimulus length was heard at 60-dB sound pressure level. All sounds were presented ~10 cm from the contralateral ear of the rat. As the effect of speaker location was beyond the scope of our study, the speaker was always located outside of the pinna and aimed directed into the ear canal. With this configuration, the sound level was always greater to the contralateral ear (which corresponded to our recording sites) vs. the ipsilateral ear.

Neural data analysis. To define the borders between auditory fields, recording sites were analyzed to select the CF of each site, as well as to obtain bandwidth, latency, peak firing, and end-of-peak response information. Firing latency is defined as the point in time (ms) that average firing rate (across all repeats) first crosses two standard deviations above the spontaneous firing rate, threshold is defined as the lowest intensity that evoked a response from the multiunit site, and bandwidths were calculated at 10, 20, 30, and 40 dB above threshold and defined as the range of frequencies that evoked responses at the current intensity. Each field was defined as established in the literature, using CF gradients, gradient reversals, and toning curve properties (Doron et al. 2002; Higgins et al. 2010; Pandya et al. 2008; Polley et al. 2006). A1 sites were defined as having sharp tuning, short onset latency (between 10 and 20 ms from tone onset), high firing rate (100 Hz or greater), and organized tonotopically so that CFs ranged from low to high in a posterior-anterior direction. AAF sites were defined using the same parameters as A1, but with a reversed tonotopy, such that CFs ranged from low to high in an anterior-posterior direction. The VAF field was located anatomically between AAF and A1. We first located the tonotopic gradient reversal at the edges of AAF and A1. Next, sites were analyzed for whether or not they exhibited nonmonotonic features. Nonmonotonous sites were defined as sites in which the response bandwidth at 40 dB above threshold was not wider than responses 30 dB quieter. VAF as a field also had a higher average CF compared with the other fields (Polley et al. 2006). PAF sites were defined as having long onset latency (greater than 30 ms) and broad tuning curves and were located immediately posterior to A1. These methods of defining site boundaries are consistent with previous work in rodent models (Doron et al. 2002; Higgins et al. 2010; Pandya et al. 2008;...
Polley et al. 2006). Out of the 1,253 sites we acquired, 1,116 of these were included in the subsequent analyses.

Single-trial response patterns to each of the isolated speech sounds were compared using a nearest neighbor classifier (Engineer et al. 2008; Foffani and Moxon 2004, 2005; Perez et al. 2013; Popescu and Polley 2010; Ranasinghe et al. 2012a, 2012b; Shetake et al. 2011). We used ED to compare single-trial activity from five random sites to the average poststimulus time histogram (PSTH) template evoked by 19 repeats each of two different stimuli. Activity was binned using 1 ms temporal precision over a 40-ms window to encompass the spike timing precision present in the initial consonant (Engineer et al. 2008; Porter et al. 2011; Ranasinghe et al. 2012b). The classifier then compared the response of each single trial with the average activity template (PSTH) evoked by all repeats of each of the speech stimuli presented. The current trial being considered was not included in the PSTH to avoid artifact. The classifier attempted to identify the stimulus that evoked the current single-trial activity pattern by selecting the template that was closest to the single trial in units of ED. ED is calculated using the formula:

$$\text{Euclidean Distance} = \sqrt{\sum_{j=1}^{\#bins} \sum_{i=1}^{\#sites} (X_{ij} - Y_{ij})^2}$$

where $\#_{sites}$ is each recording site, and $\#_{bins}$ is each of forty 1-ms bins being compared against activity evoked by speech sound X vs. speech sound Y. We chose ED as our metric for two reasons. First, it is a well-established metric for this type of neural classification (Engineer et al. 2008; Foffani and Moxon 2004, 2005; Perez et al. 2013; Popescu and Polley 2010; Ranasinghe et al. 2012a, 2012b; Shetake et al. 2011). Second, this metric is inherently resistant to variations in spike rate. Since some auditory fields fire fewer spikes to auditory stimuli than others, we wanted to ensure that this variable was accounted for in evaluating the neural encoding ability of various auditory fields. We ran the classifier so that each site was included in the classifier analysis at least once (i.e., the classifier ran once for each site recorded), and performance was calculated as the average performance across all classifier runs. For example, the classifier ran 399 times to evaluate every A1 site, 303 times to evaluate every AAF site, and so on. We used an ANOVA and t-tests to compare the accuracy of the classifier across all auditory fields. When appropriate, a Bonferroni correction was used to correct for multiple comparisons.

We then compared speech-evoked neural responses between pairs of sites in each field that had CFs within one-fourth octave. This analysis was designed to compare the differences in neural responses between similarly tuned neural sites as a measure of firing redundancy (Chechik et al. 2006; Cohen and Kohn 2011). We counted the number of evoked spikes to each speech sound within the same 40-ms window used for the classifier. We compared the number of spikes evoked by each sound across each pair of sites and quantified the relationship using the correlation coefficient. When neural firing redundancy decreased, we refer to this as an increase in diversity.

Recordings from a subset of 4 rats were mapped with a different speaker than the remaining 11 rats (Motorola Optimus Bullet Horn Tweeter vs. TDT FF1 free field speaker). To ensure that the speaker difference did not significantly change the quality of our neural recordings, we compared the classifier’s performance on four consonant tasks (/dad/ vs. /bad/, /gad/, /sad/, or /tad/) using sites from three auditory fields in male rats mapped with each speaker. There were no significant differences in classifier performance across speaker (Motorola Bullet Horn Tweeter speaker; 74.4 ± 1% correct vs. 71.2 ± 1% correct with FF1 speaker; $P = 0.55$), so all data were combined for analysis.

**Behavioral testing.** We trained 10 rats to discriminate speech sounds using an operant go/no-go procedure. Each rat trained for two 1-h sessions per day (5 days/wk). Rats first underwent a shaping period in which they were taught to press a lever to hear the target sound and obtain a 45-mg sugar pellet reward. Once the rat was able to earn 100 pellets independently for two consecutive sessions, the rat was advanced to detection. During detection, the rat learned to withhold pressing the lever until the target sound was presented. Silent periods were randomly interleaved into the session to serve as catch trials. Rats were initially given 8 s to respond to the sound and this window was gradually decreased to 3 s. Once the rat achieved a d’ of 1.5 or greater for 10 sessions, the rat was advanced to discrimination. During discrimination, the rat learned to press the lever for a target speech sound and not press for several distractor speech sounds. A total of six rats were trained to discriminate the sound /dad/ from /bad/, /gad/, /sad/, and /tad/ (data previously published in Engineer et al. 2008), and four rats trained on /dad/ vs. /bad/, /gad/, /sad/, and /tad/ as well as /dad/ vs. /sad/, /dar/, /dad/, and /gard/. Rats were only rewarded for pressing the lever to their respective target sound (presented ~44% of the time). A response to a distractor sound (presented ~44% of the time) or to a silent catch trial (presented ~11% of the time) resulted in a time out in which the cage lights were extinguished and the program paused for 6 s. Each discrimination task lasted for 20 training sessions over 2 wk. Behavior percent correct was evaluated on the last day of training and is reported as the mean ± SE across rats. Training took place in soundproof, double-walled booth which was lined with foam to reduce noise. Inside the booth, a house light, video camera, and speaker were mounted outside an 8 × 8-in. cage which contained the lever and food dish. A pellet dispenser was mounted outside the booth to reduce noise. During the experiment, rats were food deprived to above 85% of their original body weight. Rats were housed individually and maintained on a reverse 12:12-h light-dark cycle. Behavioral percent correct for correlation with classifier performance was calculated using data from the last day of training. All protocols and recording procedures were approved by the University of Texas at Dallas Institutional Animal Care and Use Committee.

**RESULTS**

Topographic organization of tone frequency tuning was used to identify the four auditory fields: A1, AAF, VAF, and PAF. A1 and AAF exhibited clearly ordered frequency maps (correlation between site location and CF $R = -0.77$ and $R = -0.69$, respectively), while VAF and PAF did not ($R = -0.05$ and $R = -0.37$, respectively). CFs in A1 increased in a posterior-to-anterior direction, while CFs in AAF increased in an anterior-to-posterior direction (Fig. 2). Firing latency to tones was significantly different across all four fields [one-way ANOVA, $F(3,1112) = 181.51$, $P < 0.001$; Fig. 3]. VAF had the narrowest bandwidths at 40 dB above threshold (2.4 ± 0.1 octaves), followed by A1 (2.6 ± 0.1 octaves, unpaired t-test, $P < 0.01$), AAF (2.8 ± 0.1 octaves, unpaired t-tests vs. A1 and VAF, $P < 0.01$), and PAF had the longest bandwidths (3.5 ± 0.1 octaves, unpaired t-tests vs. AAF, A1, and VAF, $P < 0.01$; Fig. 3C). Frequency tuning and tonotopic organization in each field was consistent with earlier reports (Carrasco and Lomber 2011; Doron et al. 2002; Jakkamsetti et al. 2012; Pandya et al. 2008; Polley et al. 2007; Puckett et al. 2007; Takahashi et al. 2011, see Methods and Fig. 3), so we are confident that the boundaries between fields were accurately defined. In total, we recorded from 303 multiunit sites in AAF, 399 sites in A1, 206 sites in VAF, and 208 sites in PAF.

**Neural discrimination of consonants is better when spike timing information is preserved.** Neural activity from each field can be used to discriminate between consonant sounds. As in previous studies, A1 responded to each speech sound differently (Engineer et al. 2008; Perez et al. 2013; Ranasinghe et al. 2012a, 2012b; Shetake et al. 2011). We used this classifier to
test the hypothesis that other auditory fields use spike timing information to identify speech sounds. We calculated the ED between the single-trial neural response PSTH and each of two PSTH templates created from the average neural activity evoked by each of two sounds. The classifier then guessed which of the template sounds likely caused the single-trial pattern of activity by choosing the template which was most similar to the single trial (i.e., had the smallest ED; see Fig. 2. Example of an auditory cortex map from one anesthetized, adult rat. Microelectrode recordings were acquired from layer IV/V of 15 experimentally naive rats. We recorded responses from each of four fields: anterior (AAF), primary (A1), ventral (VAF), and posterior auditory fields (PAF). Although there is variability between animals, tonotopic organization and latency were generally consistent within the fields, and these parameters were used to identify boundaries between fields. AAF was organized from low-frequency sites to high-frequency in an anterior-to-posterior direction, while A1 was organized from low to high in a posterior-to-anterior direction. VAF was located anatomically between the two fields, but had no tonotopic gradient. PAF was located posterior to A1 and also had no tonotopic gradient. Sites outlined in black and with black text represent the individual examples shown in Fig. 6.

Fig. 3. Tone response properties in AAF, A1, VAF, and PAF mimic previous studies. A: AAF and A1 responded to tones with the shortest onset latency (14.8 ± 0.6 ms and 16.2 ± 0.2 ms; P < 0.01), followed by VAF (19.8 ± 0.8 ms; t-test vs. A1, P < 0.01). PAF fired with the longest onset latency of any field and was significantly different from every other field (33.1 ± 1.2 ms; t-test vs. VAF, P < 0.01). B: A1, AAF, and VAF responded to tones with the same threshold (14.2 ± 0.5 dB, 14.8 ± 0.6 dB, and 13.2 ± 1 dB, respectively), while PAF sites had a significantly higher threshold than the other three fields (18.1 ± 1.0 dB; t-tests with VAF, P < 0.01). C: VAF had the narrowest bandwidths at 40 dB (BW40) above threshold (2.4 ± 0.1 octaves), followed by A1 (2.6 ± 0.1 octaves, t-test vs. VAF, P < 0.01). AAF had broader bandwidths than VAF and A1 (2.8 ± 0.1 octaves, unpaired t-tests with Bonferroni correction, P = 0.01), and PAF had the broadest bandwidths at this intensity level (3.5 ± 0.1 octaves, unpaired t-tests with Bonferroni correction, P < 0.01). D: A1 and AAF fired the most driven spikes to tones (2.8 ± 0.1 spikes and 2.7 ± 0.1 spikes, respectively; t-test, P = 0.14). VAF fired significantly fewer spikes than AAF (2.4 ± 0.1 spikes, P < 0.01), and PAF fired the least amount of driven spikes of any field (1.9 ± 0.1 spikes; t-tests with Bonferroni correction, P < 0.01). *Significant difference.
spike timing information, respectively; Fig. 5). This result is 1.3% vs. 68.2 P Bonferroni correction; and was not significantly different from either field (VAF performance was intermediate between AAF and PAF).

The two-alternative forced choice classifier reached ceiling performance in all fields when greater than 20 sites are used, while performance was close to floor when single sites are used. For the analyses in this report, we used 5 sites per classifier run (marked by the vertical line) to achieve performance well above chance level while avoiding ceiling performance. Classifier was run in all instances using spike timing information: 1-ms temporal bins across a 40-ms analysis window.

Methods). Each run of the classifier used multiunit activity from 1 to 200 recording sites. Classifier performance most closely matched previously published behavioral performance of ~80% correct when the classifier was given data from five randomly selected recording sites (Engineer et al. 2008; Fig. 4). As a result, groups of five sites were used for all classifier analyses in this paper.

We tested the ability of neural activity in each field to discriminate all possible pairs of the seven consonant sounds evaluated in this study and compared classifier performance when spike timing information was preserved or removed. We preserved spike timing information by analyzing the data with 1-ms temporal bins (over a 40-ms analysis window), and we removed timing information by analyzing the data in a single 40-ms bin. The classifier performance was significantly different when spike timing information was preserved vs. when it was not [two-way ANOVA, F(1,1115) = 2.07, P < 0.001]. Across all of the comparisons we used, PAF sites were significantly worse at the neural discrimination task than A1 and AAF (t-tests with Bonferroni correction, P < 0.01; Fig. 5). VAF performance was intermediate between AAF and PAF and was not significantly different from either field (t-tests with Bonferroni correction; P = 0.07 and P = 0.02, respectively; Fig. 5). The classifier performed significantly better when spike timing information was preserved in A1 responses compared with when spike timing information was removed (1-ms bins, 94.4 ± 1.0% correct vs. a single 40-ms bin, 77.3 ± 1.5% correct; t-test, P < 0.01; Fig. 5), which is consistent with our earlier report (Engineer et al. 2008). The three nonprimary fields also performed significantly better when spike timing information was preserved than when it was removed (AAF: 93.1 ± 1.0% vs. 72.2 ± 1.5% correct; P < 0.01, VAF: 86.2 ± 1.3% vs. 68.2 ± 1.4% correct; P < 0.01, PAF: 75.2 ± 1.4% vs. 66.5 ± 1.0% correct; P < 0.01, t-tests for with and without spike timing information, respectively; Fig. 5). This result is not specific to these bin sizes. As expected, the classifier performed significantly better when spike timing information was preserved (1- to 10-ms bin sizes) than when spike timing information was removed (40- to 100-ms bin sizes). When using AAF activity, the classifier was significantly worse at consonant discrimination when 10-ms bins were used instead of 1-ms bins (89.4 ± 1.4% correct with 10-ms bins vs. 93.1 ± 1.0% correct with 1-ms bins, P = 0.02). For all of the other fields, there was no significant difference in performance when 1- or 10-ms bins were used (A1: 92.2 ± 1.0% correct with 10-ms bins vs. 94.4 ± 1.0% correct with 1-ms bins, P = 0.13; VAF: 84.3 ± 1.5% correct with 10-ms bins vs. 86.2 ± 1.3% correct with 1-ms bins, P = 0.32; PAF: 75.4 ± 1.5% correct with 10-ms bins vs. 75.2 ± 1.4% correct with 1-ms bins, P = 0.94). For each of the four auditory fields tested, these results suggest that spike timing can provide information about consonant identity.

Activity from nonprimary fields is nearly as effective as A1 in discriminating speech sounds. As expected from earlier studies (Carrasco and Lomber 2011; Doron et al. 2002; Jakkamsetti et al. 2012; Pandya et al. 2008; Polley et al. 2007; Puckett et al. 2007; Takahashi et al. 2011), the speech-evoked responses slightly differed across the four fields. Differences in the response properties to speech sounds were similar to the differences in the response properties to tones (Fig. 3). AAF and A1 fired quickly after the onset of the speech sound, /d/ and /t/, in a single burst of activity (average latency of 14.2 ± 0.7 ms in AAF and 15.2 ± 0.7 ms in A1; P = 0.32, Fig. 6, B and C). VAF sites responded just as quickly as AAF and A1 to the onset of each consonant sound (14.3 ± 0.8 ms; t-tests with...
Bonferroni correction, \( P = 0.39 \), Fig. 6). The onset of the response in PAF sites was the latest of any of the fields (average latency of 18.1 ± 0.7 ms across all speech sounds; \( t \)-test vs. A1, \( P < 0.01 \)). The result that VAF and PAF sites responded more quickly to speech sounds than to tones may be due to the broader bandwidths in speech stimuli (Barbour and Wang 2003; Petkov et al. 2006; Rauschecker and Tian 2004; Rauschecker et al. 1995). In addition, the peak amplitude of response to speech sounds was highest in AAF (601.8 ± 73.2 Hz), followed by A1 (556.1 ± 70.5 Hz, \( t \)-test vs. AAF, \( P = 0.01 \)), VAF (414.9 ± 54.0 Hz, \( t \)-test vs. AAF and A1, \( P < 0.01 \)), and PAF (314.8 ± 38.2 Hz, \( t \)-tests vs. the other 3 fields, \( P < 0.01 \)). The representative examples shown in black in Fig. 6 (example sites used are outlined in black in Fig. 2) are sites tuned to 10 kHz, but the general timing and strength of the response was consistent across the range of CFs in each field (average responses shown in gray; Fig. 6).

Here, we examined responses to consonant speech sounds in A1 and three additional nonprimary cortical fields. For example, the sound /d/ caused neurons tuned to high frequencies to fire first, followed by lower frequency neurons after a brief delay (Fig. 7A). The consonant sound /g/ caused midfrequency neurons to fire first, followed by firing from high and low neurons milliseconds later (Fig. 7A). The spatiotemporal response patterns to speech were similar in AAF to those in A1 (Fig. 7B). The apparent “blurring” of response in PAF may be caused by broader bandwidths in this field, as sites with broad bandwidths likely responded to multiple aspects of the speech signal. In AAF and A1, some narrowly tuned sites fired to the consonant burst, while other sites fired to the vowel, while PAF neurons fired to both portions of the stimulus signal. We hypothesized that these differences in response patterns may cause longer latency fields to be worse at encoding speech sounds with a short voice onset time, such as /d/ and /b/.

We calculated the similarity (using ED) between patterns of activity evoked by all possible pairs of consonant sounds and compared these differences across auditory fields (Fig. 8). In every field, the patterns of activity evoked by /dad/ and /bad/ were the least similar (i.e., had the largest ED value), and /sad/ and /chad/ were the most similar (i.e., had the smallest ED value; Fig. 8). The difference in neural response patterns between unvoiced consonants (/t/, /s/, /sh/, and /ch/) was higher in VAF and PAF than in A1 (Fig. 8, B and C). The similarity between pairs of neural responses in VAF and PAF were lower than in A1, which suggests that these two fields may be better able to discriminate between these sounds. In spite of this
apparent advantage, classifier performance on pairs of unvoiced consonants was better in A1 than in VAF (90.1 ± 1.8% correct in A1 vs. 82.4 ± 2.6% correct in VAF, \( P < 0.02 \)) and in PAF (69.4 ± 2.2% correct, \( P < 0.01 \)). As expected from previous studies, these results did not significantly change using larger analysis windows (from stimulus onset until between 50–300 ms later), as long as spike timing information was preserved (e.g., 1- to 10-ms temporal bins; Engineer et al. 2008; Perez et al. 2013; Ranasinghe et al. 2012a, 2012b; Shetake et al. 2011). These results suggest that, while VAF and PAF encode unvoiced consonant stimuli differently than A1, the information needed to discriminate between these patterns of activity is present in all three fields. Although the auditory fields fired to speech sounds with different latencies and bandwidths, the similarity between pairs of speech-evoked neural responses was strongly correlated across fields (Fig. 8, not all comparisons shown). 

Neural responses are correlated with behavioral discrimination of consonants. The similarity between speech-evoked patterns in A1 is strongly correlated with behavioral discrimination ability in rats (Engineer et al. 2008; Perez et al. 2013; Ranasinghe et al. 2012b; Shetake et al. 2011). We hypothesized that the similarity between patterns of speech-evoked neural responses in the three nonprimary fields would also be correlated with behavior. Neural activity from all four auditory fields was correlated with behavioral discrimination ability of rats trained to discriminate several sets of consonant speech sounds. The highest performance was achieved by rats trained on tasks in which /dad/ was the target (88.3 ± 2.3% correct; see Table 1 for performance on each pair) and tasks in which /bad/ was the target (86.6 ± 2.7% correct; Table 1), followed by tasks in which /shad/ was the target (79.1 ± 8.5% correct; Table 1). Using the five-site classifier described above in each field (neural data recorded in untrained rats), every field was
strongly correlated with behavior ($R^2 = 0.41$, $P = 0.02$ in AAF, $R^2 = 0.59$, $P < 0.01$ in A1, $R^2 = 0.39$, $P = 0.03$ in VAF, and $R^2 = 0.48$, $P = 0.01$ in PAF). Classifier performance in three of the four fields was correlated when spike timing information was preserved, but not when it was removed (without spike timing in AAF, $R^2 = 0.18$, $P = 0.16$; in A1 $R^2 = 0.28$, $P = 0.07$; and in VAF $R^2 = 0.10$, $P = 0.32$). Classifier performance in PAF was correlated with or without spike timing (without spike timing; $R^2 = 0.55$, $P < 0.01$). PAF neural discrimination ability was able to correlate to rat behavioral ability without spike timing information, even though this field’s classifier performance was significantly worse than the other fields.

Our earlier studies reported that the correlation between speech discrimination performance and neural responses does not depend on using a classifier to perform discrimination using neural activity (Engineer et al. 2008; Perez et al. 2013; Ranasinghe et al. 2012b; Shetake et al. 2011). The EDs between pairs of speech-evoked responses (using all neural sites from each field) was correlated to behavioral discrimination ($R^2 = 0.47$, $P = 0.01$ in AAF, $R^2 = 0.54$, $P = 0.01$ in A1, $R^2 = 0.29$, $P = 0.07$ in VAF, and $R^2 = 0.68$, $P < 0.01$ in PAF), as expected from our earlier studies. The patterns of evoked activity to speech in all four auditory areas were strongly correlated with each other and with behavioral discrimination. We hypothesized that the slight differences in response patterns across fields may contribute differently to the animal’s performance. We used all sites from all four fields and compared the similarity between neural responses to pairs of consonants and behavioral ability. This meta-ensemble was strongly correlated to behavioral performance ($R^2 = 0.46$, $P = 0.02$), but did not seem to perform better or worse than the individual fields. Although we cannot be sure how information in the multiple auditory fields is used by the animal during the task, these results suggest that each auditory field contains comparable information about consonant stimuli.

Response diversity to speech stimuli is higher in nonprimary fields. Diversity was quantified as the strength of the correlation between the responses evoked in similarly tuned neurons to a variety of speech sounds. Since VAF and PAF are not as clearly organized by CF as AAF and A1, we could not simply analyze pairs of sites that were anatomically close to each other. For this study, we restricted the analysis to pairs of sites that were within one-fourth octaves of each other. Similarly tuned sites in A1 and AAF encode speech stimuli more similarly than similarly tuned sites in VAF or PAF. For example, in a pair of similarly tuned AAF sites, the speech

---

Table 1. Behavioral speech sound discrimination performance of rats

<table>
<thead>
<tr>
<th>Task</th>
<th>/d/ vs. /b/</th>
<th>/d/ vs. /g/</th>
<th>/s/ vs. /ch/</th>
<th>/d/ vs. /t/</th>
</tr>
</thead>
<tbody>
<tr>
<td>%Correct</td>
<td>90.7 ± 2.0</td>
<td>86.9 ± 2.9</td>
<td>92.5 ± 0.8</td>
<td>83.1 ± 3.3</td>
</tr>
<tr>
<td>Task</td>
<td>/b/ vs. /d/</td>
<td>/b/ vs. /g/</td>
<td>/s/ vs. /ch/</td>
<td>/d/ vs. /g/</td>
</tr>
<tr>
<td>%Correct</td>
<td>89.4 ± 5.4</td>
<td>80.6 ± 6.1</td>
<td>85.5 ± 4.8</td>
<td>91.1 ± 5.9</td>
</tr>
<tr>
<td>Task</td>
<td>/d/ vs. /s/</td>
<td>/d/ vs. /ch/</td>
<td>/g/ vs. /s/</td>
<td></td>
</tr>
<tr>
<td>%Correct</td>
<td>90.7 ± 2.7</td>
<td>85.3 ± 4.5</td>
<td>57.6 ± 0.1</td>
<td>82.9 ± 5.4</td>
</tr>
</tbody>
</table>
| Mean performance across rats on the last day of training is reported ± SE. Rats were trained to discriminate speech sounds in one of three tasks: /bad/ vs. /bad/, /gad/ vs. /sad/, /dad/ (n = 10 rats, data previously published in Engineer et al. 2008); /bad/ vs. /dad/, /gad/ vs. /sad/, /dad/ (n = 4 rats); or /shad/ vs. /dad/, /sad/, /chad/, /gad/ (the same 4 rats trained on the /bad/ task).
sound /dad/ evoked more spikes in site #2 than in site #1, but the relationship was still significantly correlated ($R^2/ H_{11005} = 0.77, P/ H_{11021} < 0.01$; Fig. 9). In A1, two sites with a CF of 5 kHz fired the most spikes to the sounds /bad/ and /dad/ and fewest to the sound /sad/, and the relationship between the firing strength of these two sites to all speech sounds was significantly correlated ($R^2/ H_{11005} = 0.86, P/ H_{11021} < 0.01$; Fig. 9B). These results suggest that similarly tuned sites in both A1 and AAF encode speech stimuli with a significant level of redundancy. In PAF and VAF, spike rate among pairs of sites is also significantly correlated ($R^2/ H_{11005} = 0.61, P/ H_{11005} < 0.01$ and $R^2/ H_{11005} = 0.63, P/ H_{11005} < 0.01$ in PAF; Fig. 9, C and D), but several consonant sounds evoke different responses within each pair. For example, in PAF, site #1 fired almost no spikes to the sounds /sad/, /gad/, and /bad/, while site #2 fired strongly to each of these sounds (Fig. 9D). The trends in these single-pair examples are representative of the population of pairs in each of the four auditory fields we tested.

The highest average correlation across pairs of sites with the same CF (±1/4 octave) was observed in A1, followed by AAF ($R = 0.36$ and $R = 0.33$, respectively; $P < 0.01$). In these two fields, 10% of site pairs were in the 75th percentile of their respective distributions ($R$ value above 0.6; Fig. 10, A and B). VAF and PAF were less correlated than AAF or A1 ($t$-tests with Bonferroni correction, $P < 0.01$), but were not significantly different from each other (in VAF; $R = 0.17$ and in PAF; $R = 0.18$, $P = 0.04$). In VAF and PAF, the distribution shifted so that only 8% of pairs were in the 75th percentile ($R$ value above 0.5; Fig. 10, C and D, respectively). These results suggest that pairs of sites in VAF and PAF are encoding the same speech sound with less redundancy than in AAF or A1. The encoding redundancy in A1

Fig. 9. Example correlations from one pair of sites in each field. We counted the number of evoked spikes to each speech sound fired from each of two sites that were tuned within one-fourth octave of each other and quantified these pairs using the correlation coefficient. These examples represent pairs in the 75th percentile in each field. A: AAF sites fire with a similar number of spikes per sound. For example, both of these sites fired the most spikes to /b/ and /g/, and the least number of spikes to /s/. This example had an $R^2$ of 0.77 ($P < 0.01$). B: A1 pairs had the highest correlation, suggesting the largest amount of redundant information. This pair had an $R^2$ of 0.86 ($P < 0.01$). VAF ($R^2$ of 0.61, $P = 0.04$; C) and PAF ($R^2$ of 0.63, $P = 0.03$; D) pairs had weak correlations, suggesting that these fields had less redundancy in information encoding. For example, in both C and D, one site in the pair fired more spikes to /g/ than the other site in the pair.

Fig. 10. Distribution of correlation coefficients between speech-evoked responses in pairs of recording sites. In each field, we found pairs of sites tuned within one-fourth octave of each other and compared the number of evoked spikes to each of the seven consonant sounds presented. We quantified these comparisons using the correlation coefficient. A: pairs of sites in AAF were strongly correlated with each other when comparing the number of evoked spikes to speech sounds. AAF site pairs had an average $R$ value of 0.33. B: A1 pairs were the most correlated with each other (an average $R$ value of 0.36, $t$-test vs. AAF, $P < 0.01$). VAF (C) and PAF (D) pairs were the least correlated, with an average $R$ value of 0.17 in VAF and 0.18 in PAF. Both of these fields were less correlated than AAF or A1 ($t$-tests with Bonferroni correction, $P < 0.01$), but were not different from each other ($P = 0.04$).
may increase the efficacy of this field in driving downstream neurons (Eggermont 2007).

These results suggest that similarly tuned neurons encode speech stimuli with various levels of redundancy across the auditory fields from which we recorded. This difference in redundant firing across similarly tuned neurons supports the hypothesis that information is transformed across the synapses of the auditory pathway. In spite of this increased diversity, no auditory field was better correlated to behavioral performance than any other. It is unlikely that neural circuits use this method of calculating similarity between individual neural responses. We report that neural firing patterns in each of four auditory fields can be used to achieve comparable levels of performance on a consonant discrimination task using ED as a metric. This result suggests that the information needed to accomplish such a task is encoded in each of the fields we investigated.

Neural activity from multiple fields can be used to identify consonant sounds. Identifying sounds from a list of many possibilities is different than simply categorizing sounds into two categories (“go” and “no go”). To determine whether neural responses in multiple fields would be able to identify consonant speech sounds, we tested the classifier on a seven-category forced-choice task with and without spike timing information. The classifier compared single-trial activity to the average template of each of seven sounds instead of two sounds used above and in previous studies (Centanni et al. 2013; Engineer et al. 2008; Perez et al. 2013; Ranasinghe et al. 2012a, 2012b; Shetake et al. 2011). The classifier was able to identify seven different sounds at well above chance level using activity from any of the auditory fields tested (chance is 14%; t-tests for all fields’ performance vs. chance, P < 0.001; Fig. 11). Spike timing information improved classifier performance for each field (P < 0.001). The classifier performed best when using A1 activity, followed by AAF (t-test vs. A1, P < 0.01), VAF (t-tests vs. AAF and A1, P < 0.01), and PAF (t-tests vs. all other fields, P < 0.01). Spike timing improved classifier performance more in some fields than others. The classifier benefited most from having spike timing when AAF activity was used (increase of 40.6 ± 0.7%), followed by A1 (36.6 ± 0.7%), VAF (30.8 ± 1.2%), and PAF (13.6 ± 1.2%). These results suggest that neural activity patterns in all four fields can be used to accurately identify different consonant sounds.

**DISCUSSION**

**Summary of results.** This study was designed to test whether spike timing information contributes to speech processing in nonprimary auditory fields. We used an established nearest neighbor classifier to demonstrate that spike timing information improves accuracy on a neural discrimination task in all auditory fields. The classifier performance in each field was correlated with behavioral ability of rats trained to discriminate the same sounds in an operant go/no-go paradigm. The response to speech sounds between recording sites with the same CF is less redundant in long latency fields of the auditory pathway. Our results suggest that, while the various auditory fields process speech sounds differently, each fields’ neural discrimination ability is strongly and independently correlated with the behavioral performance of rats.

Anesthesia may affect neural responses. Our recordings were obtained from the auditory cortex of anesthetized adult rats. Neural recordings in awake animals differ from recordings in anesthetized animals, especially to repetitive stimuli, and may have affected our recordings using speech stimuli. Auditory cortex neurons in awake animals respond strongly to repetitive noise burst stimuli and encode more information about these sounds than the anesthetized cortex (Anderson et al. 2006; Dong et al. 2011). Basic tuning properties in the rat auditory cortex change under anesthesia, including a reduction in the number of active neurons and sharper tuning curves in those active neurons (Gaese and Ostwald 2001). Although response properties of neurons can differ when animals are awake compared with anesthetized, discrimination between similar sounds using cortical activity from awake and anesthetized animals is comparable (Engineer et al. 2008; Hromádka et al. 2008; Huetz et al. 2009). In awake rats and monkeys, response patterns evoked by speech sounds are just as accurate at encoding the stimulus as in anesthetized cortex (Centanni et al. 2013; Engineer et al. 2008; Stein Schneider et al. 1994). In spite of the firing differences to tones and repetitive stimuli caused by anesthesia, speech sound responses in A1 are not qualitatively different in the awake vs. the anesthetized rat. The ability to record speech sound responses in the anesthetized animal ensures a low spontaneous firing rate (Anderson et al. 2006; Rennaker et al. 2007). The reduction of spontaneous firing makes the evoked responses easily visible and reduces the variability in identifying driven recordings. Responses in nonprimary visual cortex to complex stimuli are similar in anesthetized and awake monkeys (Jazayeri et al. 2012; Schmolesky et al. 1998; Stoner and Albright 1992), which suggests that responses in non-AI of awake and anesthetized subjects may also be similar. In human subjects with intracranial electrodes, anesthesia diminishes activity in nonprimary auditory fields, but the general pattern of activity is comparable (Howard et al. 2000). Additional studies are needed to deter-

![Graph](http://jn.physiology.org/Downloadedfrom)
mine how anesthesia affects neural encoding of speech sounds in nonprimary auditory fields during behavioral tasks and in passive listening conditions.

The effect of training on neural responses to speech sounds. The speech-evoked responses in untrained rats is correlated with behavioral ability in several different tasks, including consonant and vowel discrimination in quiet (Engineer et al. 2008; Perez et al. 2013), in various levels of background noise (Shetake et al. 2011), and after spectral or temporal degradation (Ranasinghe et al. 2012b). Extensive behavioral training does change neural firing patterns in auditory cortex (Reed et al. 2011; Takahashi et al. 2011) and may therefore affect the ability of the classifier to predict stimulus identity. Training in ferrets increased the amount of information encoded (as measured by bits), but did not seem to affect the fundamental nature of the spatiotemporal response patterns (Schnupp et al. 2006b). Additional studies are needed to determine whether speech sound training alters speech responses in primary and nonprimary cortex of nonhuman animals. Human imaging studies suggest that speech training will enhance neural responses (Kraus et al. 1995; Tremblay et al. 2001).

Diversity increases throughout the auditory pathway. Diversity in the speech-evoked responses in cortical auditory fields may contribute to the ability of an animal to generalize to stimuli or accurately perceive stimuli in adverse listening conditions (Kilgard 2012). The redundancy of encoded information decreases significantly as information is passed from the cochlea to the nonprimary cortical fields. In the inferior colliculus (IC), neural responses in neurons tuned to the same frequency are highly correlated with each other (Chechik et al. 2006). In the thalamus and the A1, the response patterns in similarly tuned neurons are already substantially different, likely due to the transformation of information to reflect different stimulus characteristics (Chechik et al. 2006; Sen et al. 2001; Spitzer and Semple 1998; Winer et al. 2005). Earlier studies have indicated that greater diversity results in a more robust representation of sensory information (Lyons-Warren et al. 2012; Morisset and Ghallab 2008; Schnupp 2006a; Sharpee et al. 2011; Shimizu et al. 2000). Novel natural stimuli, such as songbird vocalizations, evoke highly redundant patterns of activity in the cat IC compared with A1 (Chechik et al. 2002, 2006). Both the spike rate and the spike timing information of a stimulus encoded by IC neurons are strongly correlated across neurons with the same CF, while similarly tuned A1 neurons encode less redundant information about the same stimulus. This result is likely due to the longer integration time in A1 compared with IC (Chen et al. 2012). We show that the nonprimary field, PAF, has significantly longer integration times than A1 and may represent a continuation of this hierarchical organization. Classifier performance using PAF activity was correlated with behavior both with and without spike timing information. With larger numbers of sites, PAF is able to perform the neural discrimination task with higher accuracy (Fig. 4). Since the brain has access to the entire field of neurons, the performance difference we show here may not accurately reflect functional differences across fields. Activity from more PAF neurons (compared with the other fields) may be needed to complete the same tasks. Our results support these earlier findings as we have shown an increase in firing diversity throughout the cortical auditory fields we tested.

The ability of multiple fields to identify complex auditory stimuli may be beneficial in ensuring the processing of important auditory cues. The presence of background noise dramatically alters the neural responses in the A1 in rats. When 60 dB of speech-shaped background noise is added to a speech sound stimulus, A1 sites fire with a delayed latency and lower amplitude compared with speech sounds presented in quiet (Shetake et al. 2011). In spite of the severe degradation to the neural response, rats are still able to behaviorally discriminate speech sounds with this level of background noise significantly above chance levels. If nonprimary auditory fields are encoding different aspects of the speech stimulus, as has been previously suggested (Rauschecker et al. 2009), this may explain the robust speech discrimination ability of rats at this signal-to-noise ratio. Similarly, spectral or temporal degradation with a vocoder causes degraded neural firing patterns in rat A1, while the behavioral performance remains significantly above chance (Ranasinghe et al. 2012b). Other auditory fields may encode the speech stimuli in a way that is more robust to such interference, allowing the rat to accomplish the behavioral task in adverse listening environments. Additional cortical deactivation experiments are needed to evaluate whether each auditory field is capable of compensating for the loss of other auditory field activity.

Evidence for an integrated and parallel hierarchical organization. There are two opposing theories for sensory organization in the brain that are currently being debated. The first suggests that the auditory system is organized into separate streams of information processing. Similar to the visual system, a “what” and a “where” pathway may also exist in the auditory system (Lomber and Malhotra 2008; Rauschecker and Scott 2009; Recanzone 2000). Deactivation of AAF in cats causes selective impairment on pattern discrimination tasks, while deactivation of PAF causes impairment on spatial awareness tasks (Lomber and Malhotra 2008). The second theory suggests that the auditory system functions as an arrangement of integrated but parallel groups of neurons (Recanzone 2000; Sharpee et al. 2011). Our results, as well as previous work in rats, show that, as information moves farther up the auditory pathway, onset latency significantly increases and suggests an order of processing (Jakkamsetti et al. 2012; Pandya et al. 2008; Polley et al. 2007; Puckett et al. 2007; Storace et al. 2012). The encoding of different stimulus features across different levels of the auditory system may help the brain to better encode spatial location (Recanzone 2000; Walker et al. 2011) or help the brain process information in adverse listening environments (Ranasinghe et al. 2012b; Shetake et al. 2011). Our data suggest that, while multiple auditory fields encode speech sounds in a similar but not identical manner, each field is highly correlated with behavioral discrimination ability. The similarity in correlative ability across all fields supports the view that processing of speech sounds involves neural activity that is distributed across multiple auditory cortex fields.

ACKNOWLEDGMENTS

The authors thank K. Im and N. Moreno for help with acquiring microelectrode recordings. We also thank D. Polley, N. Khodaparast, and A. Sloan for comments on earlier versions of this manuscript.

J Neurophysiol • doi:10.1152/jn.00092.2013 • www.jn.org
GRANTS
This work was supported by the National Institute of Deafness and Other Communications Disorders at the National Institutes of Health (RO1DC010433).

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

AUTHOR CONTRIBUTIONS
Author contributions: T.M.C. conception and design of research; T.M.C. and C.T.E. performed experiments; T.M.C. and C.T.E. analyzed data; T.M.C., C.T.E., and M.P.K. interpreted results of experiments; T.M.C. prepared figures; T.M.C. drafted manuscript; T.M.C., C.T.E., and M.P.K. edited and revised manuscript; M.P.K. approved final version of manuscript.

REFERENCES

J Neurophysiol • doi:10.1152/jn.00092.2013 • www.jn.org


