

# Cortical activity patterns predict speech discrimination ability

Crystal T Engineer, Claudia A Perez, YeTing H Chen, Ryan S Carraway, Amanda C Reed, Jai A Shetake, Vikram Jakkamsetti, Kevin Q Chang & Michael P Kilgard

Neural activity in the cerebral cortex can explain many aspects of sensory perception. Extensive psychophysical and neurophysiological studies of visual motion and vibrotactile processing show that the firing rate of cortical neurons averaged across 50–500 ms is well correlated with discrimination ability. In this study, we tested the hypothesis that primary auditory cortex (A1) neurons use temporal precision on the order of 1–10 ms to represent speech sounds shifted into the rat hearing range. Neural discrimination was highly correlated with behavioral performance on 11 consonant-discrimination tasks when spike timing was preserved and was not correlated when spike timing was eliminated. This result suggests that spike timing contributes to the auditory cortex representation of consonant sounds.

The debate about the importance of spike timing began with the first recordings of neural activity and remains unresolved<sup>1,2</sup>. Although coding strategies based on precise timing have the potential to transmit more information than strategies based on firing rate averaged over long intervals, psychophysical studies of tactile modulation rate and visual movement indicate that rate-based descriptions of sensory events provide the best predictions of behavioral discrimination ability<sup>2–6</sup>. The use of precise temporal information by somatosensory cortex has been rejected because neurometric analysis predicts much better discrimination ability than is observed behaviorally<sup>5</sup>.

The auditory system is sensitive to precise temporal information and is a logical place to study perceptual correlates of neural representations based on precise spike timing<sup>1,7–9</sup>. However, few behavioral studies have examined the relationship between neural activity and auditory discrimination<sup>8,10–12</sup>. Psychophysical studies have demonstrated that newborn and adult humans, as well as rats and chinchillas, can reliably distinguish consonants based on acoustic information found within 40 ms of sound onset<sup>13–19</sup>. Similarly, the onset response of neurons in the central auditory system recorded in awake and anesthetized subjects reliably encodes the rapid acoustic transitions that provide information about consonant identity<sup>20–24</sup>. A1 lesions impair judgments of complex sounds, including speech<sup>25–29</sup>. Here we report that the precise spatio-temporal activity pattern evoked by the onset of consonant sounds is well correlated with the ability of rats to discriminate these sounds.

## RESULTS

### Neural responses to consonant sounds

We recorded neural responses to 20 English consonants (**Fig. 1** and **Supplementary Fig. 1** online) from single neurons and multiunit clusters of A1 neurons in awake and barbiturate-anesthetized rats. To

illustrate the response to each sound, we constructed neurograms from the average onset response of 445 multiunit A1 recording sites ordered by characteristic frequency (**Fig. 2** and **Supplementary Fig. 2** online). As expected, each consonant evoked a distinct spatiotemporal activity pattern in A1 (**Supplementary Video 1** online).

Consonants differing only in their place of articulation resulted in different spatial activity patterns<sup>14,21,22</sup>. For example, the /s/ sound activated high frequency neurons, whereas /sh/ activated mid-frequency neurons (**Fig. 2**, third column and **Supplementary Data** online). Manner of articulation (for example, stop, fricative or glide) substantially altered the temporal profile of the population response (**Fig. 2**, top row). As in earlier studies, stop consonants generated the sharpest onset peaks<sup>20,30,31</sup>. Nasals, glides and liquids resulted in the weakest onset responses; fricatives and affricates resulted in intermediate onset responses (**Supplementary Data**). Whereas the voiced stop consonants (/b/, /d/, /g/) evoked a single burst of activity, unvoiced stop consonants (/p/, /t/, /k/) resulted in a second peak of activity at voicing onset, consistent with previous reports in cats, monkeys and humans (**Supplementary Fig. 3**, **Supplementary Video 2** and **Supplementary Data** online)<sup>20,24,30,31</sup>.

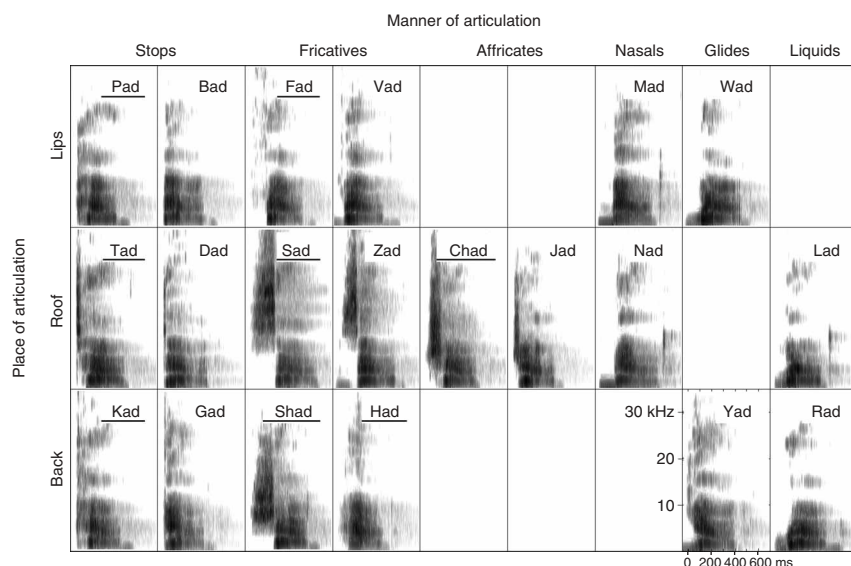
### Relating onset response similarity and behavior

Although it is reasonable to expect that sounds that evoke similar cortical responses will be more difficult to discriminate than sounds that evoke distinct responses, this is the first study to test whether this relationship requires precise temporal information (that is, 1-ms bins) or whether the rate-based strategies observed in visual and somatosensory cortex (that is, 50- to 500-ms bins) predict behavioral performance.

We quantified the difference between each pair of neurograms using euclidean distance (**Figs. 2** and **3**). When 1-ms windows were used, the

School of Behavioral and Brain Sciences, University of Texas at Dallas, 800 W. Campbell Road, Richardson, Texas 75080, USA. Correspondence should be addressed to M.P.K. (kilgard@utdallas.edu) or C.T.E. (novitski@utdallas.edu).

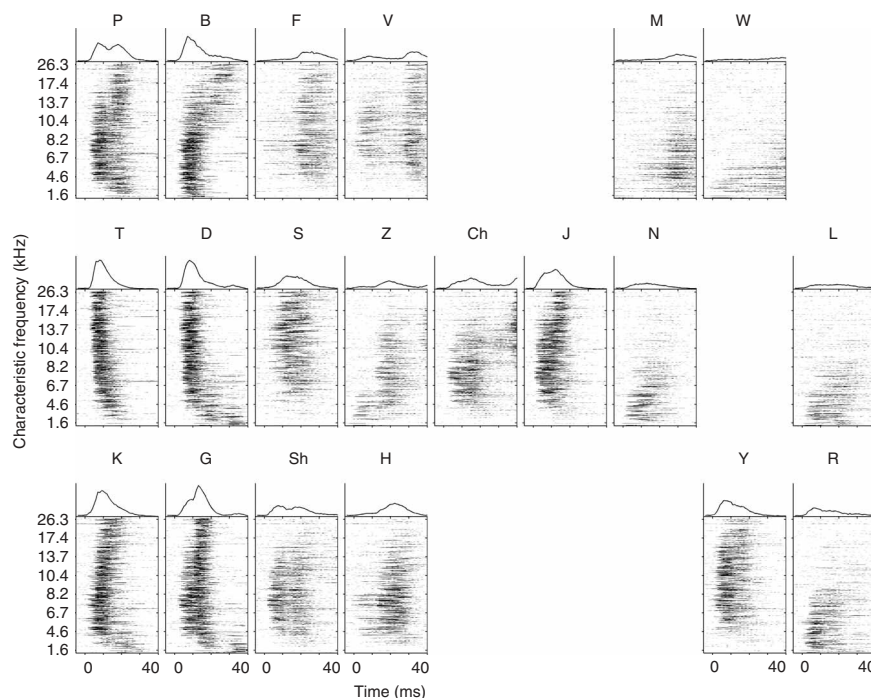
Received 11 January; accepted 17 March; published online 20 April 2008; doi:10.1038/nn.2109



**Figure 1** Spectrograms of each speech sound grouped by manner and place of articulation. Words with unvoiced initial consonants are underlined. Frequency is represented on the y axis (0–35 kHz) and time on the x axis (–50 to 700 ms). Speech sounds were shifted one octave higher to accommodate the rat hearing range. See **Supplementary Figure 1** for a more detailed view of the first 40 ms of each spectrogram.

spatiotemporal patterns evoked by the consonants /d/ and /b/ were much more distinct than the patterns evoked by /m/ and /n/ (**Fig. 3**, part 1), leading to the prediction that /d/ versus /b/ would be one of the easiest consonant pairs to discriminate and /m/ versus /n/ would be one of the hardest. Alternatively, if information about precise timing is not used, /d/ versus /b/ was predicted to be a very difficult discrimination (**Fig. 3**, part 2). To test these contrasting predictions, we evaluated the ability of rats to distinguish between these and nine other consonant pairs using an operant go/no-go procedure wherein rats were rewarded for a lever press after the presentation of a target consonant. The tasks were chosen so that each consonant pair differed by one articulatory feature (place, voicing or manner; **Fig. 1**). Rats were able to reliably discriminate 9 of the 11 pairs tested (**Fig. 4** and **Supplementary Fig. 4** online). These results extend earlier

**Figure 2** Neurograms depicting the onset response of rat A1 neurons to 20 English consonants. Multiunit data was collected from 445 recording sites in 11 anesthetized, experimentally naive adult rats. Average poststimulus time histograms (PSTH) derived from 20 repeats are ordered by the characteristic frequency (kHz) of each recording site (y axis). Time is represented on the x axis (–5 to 40 ms). The firing rate of each site is represented in grayscale, where black indicates 400 Hz. For comparison, the mean population PSTH evoked by each sound is plotted above the corresponding neurogram. For reference, 'gad' evoked the highest population firing rate of 288 Hz. As in **Figure 1**, rows differ in the place of articulation of each consonant, and columns differ in the manner of articulation. See **Supplementary Figure 2** for a direct comparison between stop-consonant spectrograms and neurograms.



observations that monkeys, cats, birds and rodents can discriminate consonant sounds<sup>17,18,31–36</sup>. The wide range of difficulty across the 11 tasks is advantageous for identifying neural correlates.

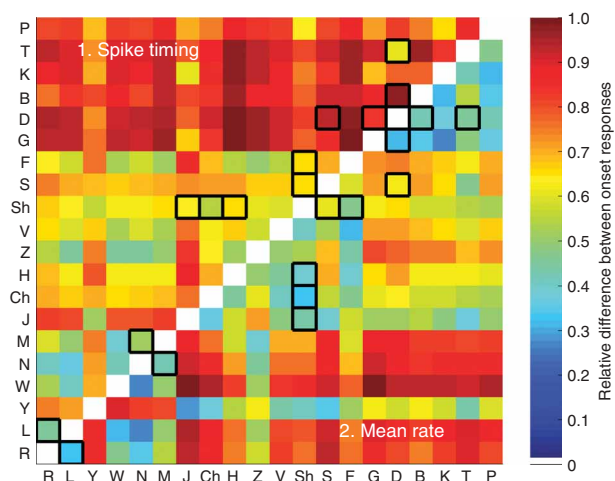
Consistent with our hypothesis that A1 representations make use of precise spike timing, /d/ versus /b/ was one of the easiest tasks (**Fig. 4**), and differences in the A1 onset response patterns were highly correlated with performance on the 11 tasks when 1-ms bins were used ( $R^2 = 0.75$ ,  $P = 0.0006$ , **Fig. 5a**). A1 responses were not correlated with behavior when spike timing information was removed ( $R^2 = 0.046$ ,  $P = 0.5$ ; **Fig. 5b**; **Supplementary Fig. 5a** and **Supplementary Data** online).

### Neural discrimination predicts behavioral discrimination

Although it is interesting that the average neural response to each consonant was related to behavior, in practice, individual speech sounds must be identified during single trials, not based on the average of many trials.

Analysis using a nearest-neighbor classifier makes it possible to document neural discrimination on the basis of single trial data and allows the direct correlation between neural and behavioral discrimination in units of percentage correct. This classifier (which compares the poststimulus time histogram (PSTH) evoked by each stimulus presentation with the average PSTH evoked by each consonant and selects the most similar; see Methods) is effective in identifying tactile patterns and animal vocalizations using cortical activity<sup>8,37</sup>.

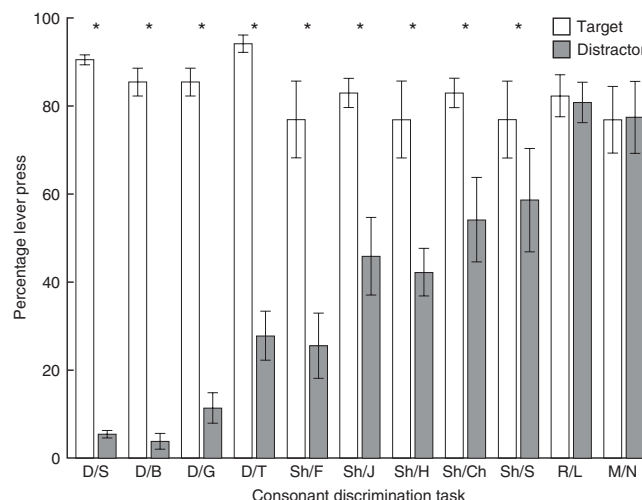
Behavioral performance was well predicted by classifier performance when activity was binned with 1-ms precision. For example, a single



**Figure 3** Predictions of consonant discrimination ability based on onset response similarity. The euclidean distance between the neurograms evoked by each consonant pair (**Fig. 2**) was computed using (1) spike timing with 1-ms bins or (2) average firing rate over a 40-ms bin (see Methods). Distance measures were normalized by the most distinct pair, such that dark red squares indicate pairs that are highly distinct and blue indicates pairs that are similar. Pairs that were tested behaviorally are outlined in black. (1) refers to the upper left half of the figure and (2) refers to the lower right.

sweep of activity from one multiunit cluster was able to discriminate /d/ from /b/  $79.5 \pm 0.8\%$  (mean  $\pm$  s.e.m.) of the time and /m/ from /n/  $60.1 \pm 0.7\%$  of the time; 50% is chance performance. Consistent with previous psychophysical evidence that the first 40 ms contain sufficient information to discriminate consonant sounds<sup>13–16</sup>, the correlation between the behavioral and neural discrimination was highest when the classifier was provided A1 activity patterns during the first 40 ms of the cortical response ( $R^2 = 0.66$ ,  $P = 0.002$ ; **Figs. 5c** and **6a**, part 1, **Supplementary Fig. 6** and **Supplementary Data** online). This correlation was equally strong in awake rats ( $R^2 = 0.63$ ,  $P = 0.004$ ; **Supplementary Fig. 7** online). Neural discrimination correlated well with behavior provided that onset responses were used (5–100 ms) and temporal information was preserved (1–10 ms, **Supplementary Fig. 5b**).

Because of a ceiling effect caused by greatly improved neural discrimination, the correlation between the behavioral and neural

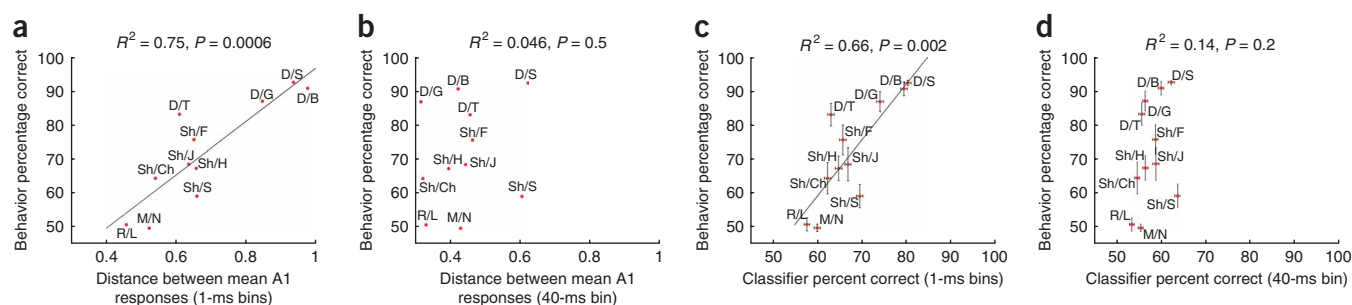


**Figure 4** Behavioral discrimination of consonant sounds. Rats successfully discriminated 9 of 11 consonant pairs evaluated. Open bars, target sound for each go/no-go task; filled bars, nontarget sound. Error bars, s.e.m. across rats. Asterisks, significant discrimination (one-tailed, paired  $t$ -test,  $P < 0.05$ ).

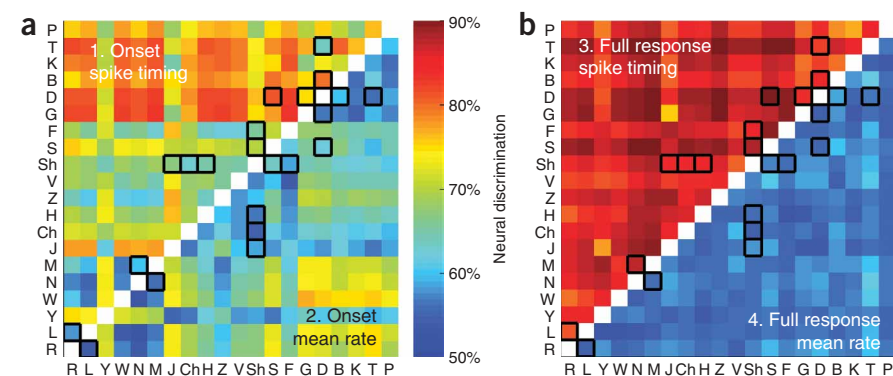
discrimination was not significant ( $R^2 = 0.02$ ,  $P = 0.6$ ) when the classifier was given all 700 ms of activity (**Fig. 6b**, part 3). Neural discrimination was greatly reduced when temporal information was eliminated (that is, mean firing rate over 700 ms) and no relationship with behavior was observed ( $R^2 = 0.06$ ,  $P = 0.5$ ). For example, on the easiest task (/d/ versus /s/), rats were correct on  $92.5 \pm 0.8\%$  of trials, whereas the classifier was correct only  $55.4 \pm 0.6\%$  of the time when spike timing was removed (**Fig. 6b**, part 4). The correlation between classifier and behavior was also not significant when the mean onset response rate was used (40-ms bin,  $R^2 = 0.14$ ,  $P = 0.2$ ; **Figs. 5d** and **6a**, part 2). These results show that the distinctness of the precise temporal activity patterns evoked by consonant onsets is highly correlated with discrimination ability in rats.

### Influence of population size on neural discrimination

To determine the neural population size that best correlates with behavior, we compared behavioral discrimination with neural



**Figure 5** Both average A1 responses and trial-by-trial neural discrimination predicted consonant discrimination ability when temporal information was maintained. **(a)** The normalized euclidean distance between neurogram onset patterns (**Fig. 3**, part 1) correlated with behavioral performance when action potential times were binned with 1-ms precision. **(b)** Distance did not correlate with behavior when a single 40-ms bin was used (**Fig. 3**, part 2). **(c,d)** The average neural discrimination correlated with behavioral discrimination when 1-ms bins were used **(c)**, but not when a 40-ms bin was used **(d)**. Neural discrimination was performed by a nearest-neighbor classifier (see Methods) using a single sweep of neural activity. The 11 consonant pairs are printed next to each data point. Error bars, s.e.m. for behavioral and neural discrimination performance. Solid lines, best linear fits when statistically significant. Analysis using other distance measures is shown in **Supplementary Table 1** online. Classifier performance on each task as a function of tone characteristic frequency is shown in **Supplementary Figure 6**.



**Figure 6** Predictions of consonant discrimination ability based on nearest-neighbor classifier. **(a)** Part 1, neural discrimination of each consonant pair using the 40-ms onset response with 1-ms bins. Part 2, neural discrimination of each consonant pair using the average firing rate over the first 40 ms of the onset response. **(b)** Part 3, neural discrimination using the entire 700-ms stimulus duration with 1-ms bins. Part 4, neural discrimination using the average rate over the 700-ms stimulus duration. Pairs that were tested behaviorally (**Fig. 4**) are outlined in black.

discrimination using individual single units, 16 single units, individual multiunits and sets of 16 multiunits. Stringent spike-sorting criteria were used to increase our confidence that we were recording from individual neurons. We collected a total of 16 well isolated single units from 16 different recording sites distributed across A1. Consonant discrimination was evaluated for each of the 16 single units individually and for the set of all 16 single units. When the classifier was provided with activity from all 16 sites, each pattern was a matrix of 16 columns and a number of rows determined by the bin size used. We used the same technique to evaluate classifier performance using multiunit activity from sets of 16 recording sites randomly selected from the full set of 445 recording sites. Each population size was evaluated with or without precise temporal information using the onset response (that is, 1-ms or 40-ms bins) or the entire response (that is, 1-ms or 700-ms bins).

Neural discrimination using single units did not correlate with behavior regardless of the coding strategy used in the analysis (**Fig. 7a**). The poor correlation may be related to the poor neural discrimination of single units (**Fig. 7b**), which was probably due to the small number of action potentials in single-unit responses compared to multiunit responses. Although discrimination using all 16 single units was better than individual single units, neural discrimination on the 11 tasks was still not significantly correlated with behavior (**Fig. 7**), perhaps because of the anatomical distance between the 16 recording sites.

Neural discrimination using 16 randomly selected multiunit sites correlated with behavior but did so only when temporally precise onset responses were used (**Fig. 7a**). Although the dependence on temporally precise onset responses was similar to results based on single multiunit sites, the average neural performance using 16 multiunit

sites significantly exceeded actual behavioral performance (**Fig. 7b**). This excessive accuracy resulted in a ceiling effect, which probably explains why the correlation with behavior was lower when large populations were used. After exploring a large set of neural readouts using various time windows and population sizes, we found that discrimination using onset activity patterns from individual multiunit sites correlated best with behavioral discrimination.

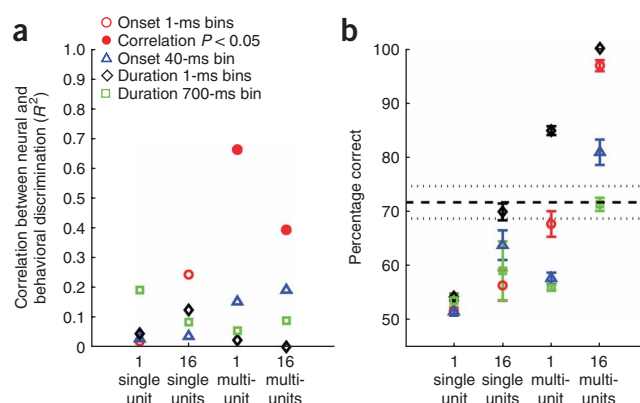
Our observation that multiunit responses were highly correlated with behavioral performance is consistent with earlier reports that multiunit responses are superior to single-unit responses for identifying complex stimuli. For example, V1 single units provide an unreliable estimate of the local contrast in natural images, whereas multiunit responses encode this information efficiently<sup>38</sup>. Similarly, multiunit clusters in the bird homolog of A1 are better than single units at discriminating song from simpler sounds, including tones, ripples and noise<sup>39</sup>.

## DISCUSSION

Although theoretical studies have suggested that precise spike timing can provide a rapid and accurate code for stimulus recognition and categorization<sup>40,41</sup>, studies in visual and somatosensory cortex have indicated that firing rates averaged across 50–500 ms are best correlated with behavior<sup>2–6</sup>. Our results suggest that the representation of consonant sounds in A1 is based on time windows that are approximately 50 times more precise.

The greater temporal precision observed in this study could be specific to the auditory system<sup>1,7–9</sup>. However, it is also possible that spike timing is important in all modalities when transient stimuli are involved<sup>38,39,42</sup>. The latter hypothesis is supported by observations of a rate-based code for steady-state vowels<sup>23,43–45</sup> and by computational studies showing that cortical neurons can efficiently extract temporal patterns from populations of neurons in a manner that promotes

**Figure 7** Neural discrimination using the onset activity pattern from individual multiunit sites was best correlated with behavior. **(a)** Percent of variance across the 11 behavioral tasks that was explained using individual single units, 16 single units, individual multiunits and sets of 16 multiunits. Each population size was evaluated with or without precise temporal information using the onset response or the entire response. Filled symbols, statistically significant correlations between neural and behavioral discrimination (Pearson's correlation coefficient,  $P < 0.05$ ). Regardless of population size, neural discrimination based on mean firing rate (that is, 40- or 700-ms bin) did not correlate with behavioral performance on the 11 tasks. **(b)** Mean neural discrimination on the 11 consonant-discrimination tasks using individual single units, 16 single units, individual multiunits and sets of 16 multiunits. Dashed line, average behavioral discrimination; dotted lines, s.e.m. for behavioral discrimination. Error bars, s.e.m. across the 11 tasks. The complete distributions of neural discrimination by single and multiunit activity for each task are provided in **Supplementary Figure 8** online.





accurate consonant categorization<sup>46</sup>. It will be important to test whether neural correlates of transient visual and tactile stimuli make use of spike timing.

Error-trial analysis in an awake behaving preparation, as well as lesion and microstimulation experiments, are needed to evaluate our hypothesis that consonant processing depends upon precise spike timing in A1. Recordings in higher cortical areas will be needed to establish whether temporal patterns or mean firing rates are better correlated with behavioral discrimination.

## METHODS

**Speech stimuli.** We recorded 20 English words ending in /ad/ (as in 'sad') in a double-walled, soundproof booth. The initial consonants differed in voicing (voiced /d/ versus voiceless /t/), place of articulation (lips /b/ versus back of mouth /g/) or manner of articulation (fricative /sh/ versus nasal /n/) (Fig. 1). The fundamental frequency and spectrum envelope of each word was shifted up in frequency by a factor of two using the STRAIGHT vocoder<sup>47</sup> in order to better match the rat hearing range. The intensity of each sound was adjusted so that the intensity during the most intense 100 ms was 60 dB SPL.

**Operant training procedure and analysis.** Eleven rats were trained using an operant go/no-go procedure to discriminate words differing in their initial consonant sound. Each rat trained for two 1-h sessions each day (5 d/week). Rats first underwent a shaping period during which they were taught to press the lever. Each time the rat was in close proximity to the lever, the rat heard the target sound and received a pellet (45-mg sugar pellet). Eventually, the rat began to press the lever without assistance. After each lever press, the rat heard the target sound and received a pellet. The shaping period lasted until the rat was able to obtain at least 100 pellets per session for two consecutive sessions. This stage lasted on average 3.5 d. After the shaping period, rats began a detection task in which they learned to press the lever each time the target sound was presented. Silent periods were randomly interleaved with the target sounds during each training session. Sounds were initially presented every 10 s, and the rat was given an 8-s window to press the lever. The sound interval was gradually decreased to 6 s, and the lever-press window was decreased to 3 s. Once rats reached the performance criteria of a  $d' \geq 1.5$  for ten sessions, they advanced to a consonant discrimination task. The quantity  $d'$  is a measure of discriminability of two sets of samples based on signal detection theory.

During each consonant discrimination task, rats learned to discriminate the target sound from the distractor sounds. Trials began every 6 s, and silent catch trials were randomly interleaved 20–33% of the time. Rats were only rewarded for lever presses to the target (conditioned) stimulus. Pressing the lever on a stimulus other than the target resulted in a time-out during which the house light was extinguished and the training program paused for a period of approximately 6 s. Training took place in a soundproof, double-walled training booth that included a house light, a video camera for monitoring, a speaker (Optimus Bullet Horn Tweeter) and a cage (8 inches length  $\times$  8 inches width  $\times$  8 inches height) that included a lever, lever light and pellet receptacle. A pellet dispenser was mounted outside the double-walled, foam-lined booth to reduce noise. Rats were food deprived to motivate behavior, but were fed on days off to maintain between 80% and 90% ad lib body weight. Rats were housed individually and maintained on a reverse 12-h light-dark cycle.

Each consonant discrimination task lasted for 20 training sessions over 2 weeks. Six rats performed each of four different consonant-discrimination tasks (/d/ versus /s/, /d/ versus /t/, /r/ versus /l/, and /d/ versus /b/ and /g/), and five rats performed each of three different consonant-discrimination tasks (/m/ versus /n/, /sh/ versus /f/, /s/ and /h/; and /sh/ versus /ch/ and /j/). Each group of rats trained on each of the tasks for 2 weeks, in the order given. We subsequently tested them for 2 d on each task to ensure that discrimination ability was not strongly influenced by task order (see **Supplementary Data**). Data shown in **Figure 4** was collected on the seventh and eighth days (that is, four sessions) of training on each task. Over these 2 d, each rat performed  $940 \pm 173$  trials (mean  $\pm$  s.d.). An example learning curve is shown in **Supplementary Figure 4**.

**Recording procedure.** We recorded multiunit ( $n = 445$ ) and single-unit ( $n = 16$ ) responses from right primary auditory cortex (A1) of anesthetized, experimentally naive, female Sprague-Dawley rats in a soundproof recording booth ( $n = 11$  rats). Rats were anesthetized with pentobarbital (50 mg kg<sup>-1</sup>) and received supplemental dilute pentobarbital (8 mg ml<sup>-1</sup>) every 0.5–1 h as needed to maintain areflexia, along with a 1:1 mixture of dextrose (5%) and standard Ringer's lactate<sup>48</sup> to prevent dehydration. Heart rate and body temperature were monitored throughout the experiment. Four Parylene-coated tungsten microelectrodes (1–2 M $\Omega$ , FHC) were simultaneously lowered to 600  $\mu$ m below the surface of the right primary auditory cortex (layer 4/5). Electrode penetrations were marked using blood vessels as landmarks.

We recorded multiunit A1 responses ( $n = 40$ ) in six awake rats using chronically implanted microwire arrays, which have been described in detail in previous publications<sup>49,50</sup>. Briefly, 14-channel microwire electrodes were implanted in the right primary auditory cortex using a custom-built mechanical insertion device to rapidly insert electrodes in layers 4/5 (depth, 550  $\mu$ m)<sup>50</sup>. Restraint jackets were used to minimize movement artifacts during recording sessions, conducted 1–7 d after implantation.

Twenty 60-dB speech stimuli were randomly interleaved and presented every 2,000 ms for 20 repeats per site at each recording site. Brief (25-ms) tones were presented at 81 frequencies (1–32 kHz) at 16 intensities (0–75 dB) to determine the characteristic frequency of each site. All tones were separated by 560 ms and randomly interleaved. Sounds were presented approximately 10 cm from the left ear of the rat. Stimulus generation, data acquisition and spike sorting were performed with Tucker-Davis hardware (RP2.1 and RX5) and software (Brainware). Single units refer to well isolated waveforms likely to have been evoked by a single neuron. Multiunits include action potentials from more than one nearby neuron. The University of Texas at Dallas Institutional Animal Care and Use Committee approved all protocols and recording procedures.

**Statistical analysis.** Neurogram similarity was computed using euclidean distance. The euclidean distance between any two neurograms ( $X$ ,  $Y$ ) is the square root of the sum of the squared differences between the firing rate at each bin ( $j$ ) for each recording site ( $i$ ). For the analysis in **Figure 3**, part 1 and **Figure 5a**, we used activity from 40 1-ms bins from all 445 sites to compute the similarity between neurogram pairs. For the analysis in **Figure 3**, part 2 and **Figure 5b**, we used activity from a single 40-ms bin from each of 445 sites to compute the similarity between neurogram pairs

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{n_{\text{sites}}} \sum_{j=1}^{n_{\text{bins}}} (x_{ij} - y_{ij})^2}$$

where  $n_{\text{sites}}$  and  $n_{\text{bins}}$  are the total numbers of sites and bins, respectively.

We used a nearest-neighbor classifier to quantify neural discrimination performance based on single-trial activity patterns<sup>8,37</sup>. The classifier binned activity using 1-ms to 700-ms intervals and then compared the response of each single trial with the average activity pattern (PSTH) evoked by each of the speech stimuli presented. The given trial being considered was not included in the average activity pattern, to prevent artifact. This model assumes that the brain region reading out the information in the spike trials has previously heard each of the sounds 19 times and attempts to identify which of the possible choices was most likely to have generated the trial under consideration. It uses euclidean distance to determine how similar each response was to the average activity evoked by each of the sounds. The classifier guesses that the single-trial pattern was generated by the sound whose average pattern it most closely resembles (that is, minimum euclidean distance). The onset response to each sound is defined as the 40-ms interval beginning when neural activity exceeded the spontaneous firing rate by three s.d. Error estimates are s.e.m. Pearson's correlation coefficient was used to examine the relationship between neural and behavioral discrimination on the 11 tasks ( $n = 11$ ).

*Note: Supplementary information is available on the Nature Neuroscience website.*

## ACKNOWLEDGMENTS

The authors would like to thank J. Roland, R. Jain and D. Lishrop for assistance with microelectrode mappings. We would like to thank R. Rennaker for technical assistance and training and for providing microelectrode arrays and inserter. We would also like to thank M. Perry, C. Heydrick, A. McMenamy, A. Meepe,

C. Dablain, J. Choi, V. Badhiwala, J. Riley, N. Hatate, P. Kan, M. Lazo de la Vega and A. Hudson for help with behavioral training. We would also like to thank S. Blumstein, Y. Cohen, H. Read, S. Denham, L. Miller, S. Edelman, V. Dragoi, H. Abdi, P. Assmann, X. Wang and R. Romo for their suggestions about earlier versions of the manuscript. This work was supported by grants from the US National Institute for Deafness and Other Communicative Disorders and the James S. McDonnell Foundation.

# AUTHOR CONTRIBUTIONS

C.T.E., C.A.P., R.S.C. and A.C.R. collected behavioral training data. C.T.E., C.A.P., Y.H.C., R.S.C., V.J. and K.Q.C. recorded anesthetized cortical responses. J.A.S. recorded awake cortical responses. M.P.K. and C.T.E. wrote the manuscript and performed data analysis. All authors discussed the paper and commented on the manuscript.

Published online at <http://www.nature.com/natureneuroscience>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

- DeWeese, M.R., Hromadka, T. & Zador, A.M. Reliability and representational bandwidth in the auditory cortex. *Neuron* **48**, 479–488 (2005).
- Parker, A.J. & Newsome, W.T. Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* **21**, 227–277 (1998).
- Liu, J. & Newsome, W.T. Correlation between speed perception and neural activity in the middle temporal visual area. *J. Neurosci.* **25**, 711–722 (2005).
- Pruett, J.R. Jr., Sinclair, R.J. & Burton, H. Neural correlates for roughness choice in monkey second somatosensory cortex (SII). *J. Neurophysiol.* **86**, 2069–2080 (2001).
- Romo, R. & Salinas, E. Flutter discrimination: neural codes, perception, memory and decision making. *Nat. Rev. Neurosci.* **4**, 203–218 (2003).
- Britten, K.H., Shadlen, M.N., Newsome, W.T. & Movshon, J.A. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745–4765 (1992).
- Narayan, R., Grana, G. & Sen, K. Distinct time scales in cortical discrimination of natural sounds in songbirds. *J. Neurophysiol.* **96**, 252–258 (2006).
- Schnupp, J.W., Hall, T.M., Kokelaar, R.F. & Ahmed, B. Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J. Neurosci.* **26**, 4785–4795 (2006).
- Walker, K.M., Ahmed, B. & Schnupp, J.W. Linking cortical spike pattern codes to auditory perception. *J. Cogn. Neurosci.* **20**, 135–152 (2008).
- Ahissar, E. *et al.* Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. USA* **98**, 13367–13372 (2001).
- Orduna, I., Mercado, E. III, Gluck, M.A. & Merzenich, M.M. Cortical responses in rats predict perceptual sensitivities to complex sounds. *Behav. Neurosci.* **119**, 256–264 (2005).
- Wang, L., Narayan, R., Grana, G., Shamir, M. & Sen, K. Cortical discrimination of complex natural stimuli: can single neurons match behavior? *J. Neurosci.* **27**, 582–589 (2007).
- Bertoncini, J., Bijeljac-Babic, R., Blumstein, S.E. & Mehler, J. Discrimination in neonates of very short CVs. *J. Acoust. Soc. Am.* **82**, 31–37 (1987).
- Blumstein, S.E. & Stevens, K.N. Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. *J. Acoust. Soc. Am.* **66**, 1001–1017 (1979).
- Fowler, C.A., Brown, J.M., Sabadini, L. & Weihing, J. Rapid access to speech gestures in perception: evidence from choice and simple response time tasks. *J. Mem. Lang.* **49**, 396–413 (2003).
- Jongman, A. Duration of frication noise required for identification of English fricatives. *J. Acoust. Soc. Am.* **85**, 1718–1725 (1989).
- Kuhl, P.K. & Miller, J.D. Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science* **190**, 69–72 (1975).
- Reed, P., Howell, P., Sackin, S., Pizzimenti, L. & Rosen, S. Speech perception in rats: use of duration and rise time cues in labeling of affricate/fricative sounds. *J. Exp. Anal. Behav.* **80**, 205–215 (2003).
- Miller, G.A. & Nicely, P.E. An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.* **27**, 338–352 (1955).
- Steinschneider, M., Fishman, Y.I. & Arezzo, J.C. Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. *J. Acoust. Soc. Am.* **114**, 307–321 (2003).
- Steinschneider, M., Reser, D., Schroeder, C.E. & Arezzo, J.C. Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. *Brain Res.* **674**, 147–152 (1995).
- Tavabi, K., Obleser, J., Dobel, C. & Pantev, C. Auditory evoked fields differentially encode speech features: an MEG investigation of the P50m and N100m time courses during syllable processing. *Eur. J. Neurosci.* **25**, 3155–3162 (2007).
- Young, E.D. Neural representation of spectral and temporal information in speech. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 923–945 (2008).
- Steinschneider, M. *et al.* Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cereb. Cortex* **15**, 170–186 (2005).
- Cooke, J.E., Zhang, H. & Kelly, J.B. Detection of sinusoidal amplitude modulated sounds: deficits after bilateral lesions of auditory cortex in the rat. *Hear. Res.* **231**, 90–99 (2007).
- Dewson, J.H. III, Pribram, K.H. & Lynch, J.C. Effects of ablations of temporal cortex upon speech sound discrimination in the monkey. *Exp. Neurol.* **24**, 579–591 (1969).
- Heffner, H.E. & Heffner, R.S. Effect of restricted cortical lesions on absolute thresholds and aphasia-like deficits in Japanese macaques. *Behav. Neurosci.* **103**, 158–169 (1989).
- Rybalko, N., Suta, D., Nwabueze-Ogbo, F. & Syka, J. Effect of auditory cortex lesions on the discrimination of frequency-modulated tones in rats. *Eur. J. Neurosci.* **23**, 1614–1622 (2006).
- Wetzel, W., Ohl, F.W., Wagner, T. & Scheich, H. Right auditory cortex lesion in Mongolian gerbils impairs discrimination of rising and falling frequency-modulated tones. *Neurosci. Lett.* **252**, 115–118 (1998).
- Steinschneider, M., Volkov, I.O., Noh, M.D., Garell, P.C. & Howard, M.A. III. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J. Neurophysiol.* **82**, 2346–2357 (1999).
- Wong, S.W. & Schreiner, C.E. Representation of CV-sounds in cat primary auditory cortex: intensity dependence. *Speech Commun.* **41**, 93–106 (2003).
- Kluender, K.R., Diehl, R.L. & Killeen, P.R. Japanese quail can learn phonetic categories. *Science* **237**, 1195–1197 (1987).
- Ramus, F., Hauser, M.D., Miller, C., Morris, D. & Mehler, J. Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* **288**, 349–351 (2000).
- Sinnott, J.M. & Brown, C.H. Perception of the American English liquid /ra-la/ contrast by humans and monkeys. *J. Acoust. Soc. Am.* **102**, 588–602 (1997).
- Toro, J.M., Trobalon, J.B. & Sebastian-Galles, N. Effects of backward speech and speaker variability in language discrimination by rats. *J. Exp. Psychol. Anim. Behav. Process.* **31**, 95–100 (2005).
- Dooling, R.J., Okanoya, K. & Brown, S.D. Speech perception by budgerigars (*Melopsittacus undulatus*): the voiced-voiceless distinction. *Percept. Psychophys.* **46**, 65–71 (1989).
- Foffani, G. & Moxon, K.A. PSTH-based classification of sensory stimuli using ensembles of single neurons. *J. Neurosci. Methods* **135**, 107–120 (2004).
- Weliky, M., Fiser, J., Hunt, R.H. & Wagner, D.N. Coding of natural scenes in primary visual cortex. *Neuron* **37**, 703–718 (2003).
- Grace, J.A., Amin, N., Singh, N.C. & Theunissen, F.E. Selectivity for conspecific song in the zebra finch auditory forebrain. *J. Neurophysiol.* **89**, 472–487 (2003).
- Buonomano, D.V. & Merzenich, M. A neural network model of temporal code generation and position-invariant pattern recognition. *Neural Comput.* **11**, 103–116 (1999).
- VanRullen, R., Guyonnet, R. & Thorpe, S.J. Spike times make sense. *Trends Neurosci.* **28**, 1–4 (2005).
- Richmond, B.J., Optican, L.M. & Spitzer, H. Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. I. Stimulus-response relations. *J. Neurophysiol.* **64**, 351–369 (1990).
- Ohl, F.W. & Scheich, H. Orderly cortical representation of vowels based on formant interaction. *Proc. Natl. Acad. Sci. USA* **94**, 9440–9444 (1997).
- Qin, L., Chimoto, S., Sakai, M. & Sato, Y. Spectral-shape preference of primary auditory cortex neurons in awake cats. *Brain Res.* **1024**, 167–175 (2004).
- Versnel, H. & Shamma, S.A. Spectral-ripple representation of steady-state vowels in primary auditory cortex. *J. Acoust. Soc. Am.* **103**, 2502–2514 (1998).
- Buonomano, D.V. & Merzenich, M.M. Temporal information transformed into a spatial code by a neural network with realistic properties. *Science* **267**, 1028–1030 (1995).
- Kawahara, H. Speech representation and transformation using adaptive interpolation of weighted spectrum: vocoder revisited. *Proc. 1997 IEEE Int. Conf. Acoust., Speech, Signal Process* **2**, 1303–1306 (1997).
- Engineer, N.D. *et al.* Environmental enrichment improves response strength, threshold, selectivity, and latency of auditory cortex neurons. *J. Neurophysiol.* **92**, 73–82 (2004).
- Rennaker, R.L., Ruyle, A.M., Street, S.E. & Sloan, A.M. An economical multi-channel cortical electrode array for extended periods of recording during behavior. *J. Neurosci. Methods* **142**, 97–105 (2005).
- Rennaker, R.L., Street, S., Ruyle, A.M. & Sloan, A.M. A comparison of chronic multi-channel cortical implantation techniques: manual versus mechanical insertion. *J. Neurosci. Methods* **142**, 169–176 (2005).