

# Categorical speech representation in human superior temporal gyrus

Edward F Chang<sup>1,2,6</sup>, Jochem W Rieger<sup>2,3,6</sup>, Keith Johnson<sup>4</sup>, Mitchel S Berger<sup>1</sup>, Nicholas M Barbaro<sup>1</sup> & Robert T Knight<sup>1,2,5</sup>

Speech perception requires the rapid and effortless extraction of meaningful phonetic information from a highly variable acoustic signal. A powerful example of this phenomenon is categorical speech perception, in which a continuum of acoustically varying sounds is transformed into perceptually distinct phoneme categories. We found that the neural representation of speech sounds is categorically organized in the human posterior superior temporal gyrus. Using intracranial high-density cortical surface arrays, we found that listening to synthesized speech stimuli varying in small and acoustically equal steps evoked distinct and invariant cortical population response patterns that were organized by their sensitivities to critical acoustic features. Phonetic category boundaries were similar between neurometric and psychometric functions. Although speech-sound responses were distributed, spatially discrete cortical loci were found to underlie specific phonetic discrimination. Our results provide direct evidence for acoustic-to-higher order phonetic level encoding of speech sounds in human language receptive cortex.

A fundamental property of speech perception is that listeners map continuously variable acoustic speech signals onto discrete phonetic sound categories<sup>1–3</sup>. This ‘phonetic’ mode of listening<sup>4</sup> lays the phonological foundation for speaking new words<sup>5</sup> and mapping speech into writing. In categorical speech perception, a continuum that gradually morphs from one syllable to another is transformed into perceptually discrete categories whose members closely resemble each other<sup>6,7</sup>.

A number of studies support the notion that the posterior superior temporal gyrus (pSTG) in Wernicke’s area is important for higher order auditory processing of speech sounds<sup>8–13</sup>. Current noninvasive neurophysiologic methodologies (for example, functional magnetic resonance imaging (fMRI), magnetoencephalography and positron emission tomography) have provided important insights into speech localization. However, because of limitations in simultaneous spatial and temporal resolution, these approaches have been unable to offer a mechanistic account for speech representation in humans. As a result, fundamental questions remain unresolved regarding the manner in which the functional organization of pSTG supports the perceptual features of aural speech. In particular, do pSTG neural activity patterns correspond to precise spectrotemporal changes in the external acoustic signal (that is, veridical representation) or to a higher order linguistic extraction of phonetic categories? Furthermore, what neural response features (for example, place, time and amplitude) are critical for representing the discriminability of different phonemes as fundamental contrastive linguistic units?

To answer these questions, we recorded cortical local field potentials from the pSTG in four human subjects undergoing awake craniotomy with speech mapping as part of their epilepsy<sup>14</sup> or brain tumor

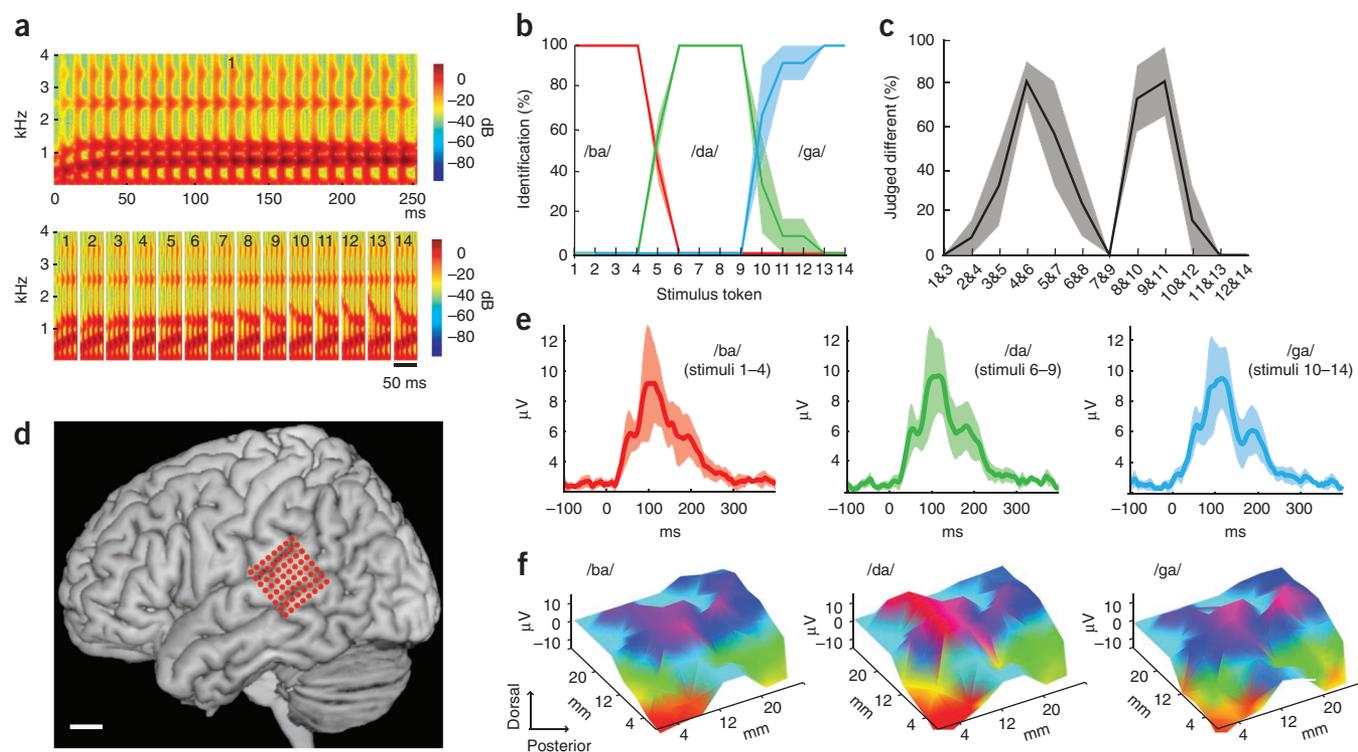
surgery<sup>15</sup>. Although limited to rare clinical settings, high-density electrocorticographic recordings offer the advantage of simultaneous high spatial (millimeters) with real-time temporal (millisecond) resolution, in addition to excellent signal-to-noise properties. We found that listening to speech sounds that differed by small acoustic steps evoked highly distributed cortical activation in the pSTG. Multivariate analyses revealed, however, that the neural response patterns were strongly organized along phonetic categories and did not demonstrate sensitivity for gradual acoustic variation. We found a high level of concordance between neuro- and psychometric functions, suggesting that pSTG encoding represents higher order invariant representation for speech sounds.

## RESULTS

We employed a classic procedure first described in 1957 (ref. 6) to investigate the perceptual and neural organization of stop consonant phonemes. Consonant-vowel syllables were synthesized with 14 equal and parametric changes in the starting frequency of the F2 transition (second vocal tract resonance) that ranged perceptually across three initial consonants /ba/ to /da/ to /ga/ (Fig. 1a). When subjects ascribed one of the three phoneme labels to the stimuli, the psychophysical identification functions demonstrated clear perceptual category borders between /ba/ and /da/ percepts near stimuli 4 and 5 and between /da/ and /ga/ percepts near stimuli 8 and 9 (Fig. 1b). In a psychophysical two-step discrimination task, accuracy was highest for those stimulus pairs that straddled the identification boundary (Fig. 1c). The steep labeling identification functions and peaked discrimination functions shown here, with the peak at

<sup>1</sup>Department of Neurological Surgery, University of California San Francisco, San Francisco, California, USA. <sup>2</sup>Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, California, USA. <sup>3</sup>Department of Neurology, Otto-von-Guericke University, Magdeburg, Germany. <sup>4</sup>Department of Linguistics, University of California Berkeley, Berkeley, California, USA. <sup>5</sup>Department of Psychology, University of California, Berkeley, Berkeley, California, USA. <sup>6</sup>These authors contributed equally to this work. Correspondence should be addressed to E.F.C. (changed@neurosurg.ucsf.edu) or J.W.R. (jochem.rieger@med.ovgu.de).

Received 12 March; accepted 12 August; published online 3 October 2010; doi:10.1038/nn.2641



**Figure 1** Psychophysics of categorical speech perception and speech-evoked responses during intraoperative human cortical recordings. **(a)** Wide-band spectrograms of the stimulus token continuum, synthesized with equal parametric changes in the F2 starting frequency (from 800–2,100 Hz). Top, full spectrogram of a single token with an 800-Hz starting frequency (stimulus 1, duration = 250 ms). Bottom, first 50 ms for each of the 14 stimulus tokens. **(b)** Psychometric identification function with percentage reporting */ba/*, */da/* or */ga/*. **(c)** Psychometric discrimination function (two step). The percentages of responses judged as different versus same are shown. The category boundaries located at peak discrimination are at stimuli 4 and 5 and at 9 and 10. **(d)** Three-dimensional surface reconstruction of representative brain magnetic resonance imaging with superimposed electrode positions over pSTG. **(e)** Grand average rooted mean square (RMS) evoked potentials recorded over pSTG for sound stimuli reliably categorized as */ba/* (tokens 1–4), */da/* (tokens 6–9) and */ga/* (tokens 10–14). The average evoked potentials (RMS, solid line) and standard errors of evoked potential amplitudes (shaded) are shown. Potentials peaked at approximately 110 ms after stimulus onset. **(f)** Topographic plots of evoked potentials at 110 ms for each prototype sound stimulus revealed distributed cortical activation pattern, with some sharply localized differences between stimuli.

the phoneme discrimination boundary corresponding to the 50% point of the labeling curve, are the defining psychophysical properties of categorical perception (Fig. 1b,c). Thus, one does not hear step-like changes corresponding to the changes in the acoustic signal, but instead perceives essentially quantal jumps from one perceptual category to another.

While subjects were fully awake in the operating room, we placed a customized high-density 64-electrode microarray (4 mm spacing) using stereotactic guidance on the surface of the posterior temporal cortex (defined here as cortical area caudal to the point where the central sulcus intersects the Sylvian fissure; Fig. 1d). Subjects listened passively to a randomized sequence of stimulus tokens. The averaged evoked potential peaked at approximately 110 ms after the stimulus onset (Fig. 1e). The spatial topography of responses to */ba/*, */da/* and */ga/* tokens revealed highly distributed responses across the pSTG (Fig. 1f).

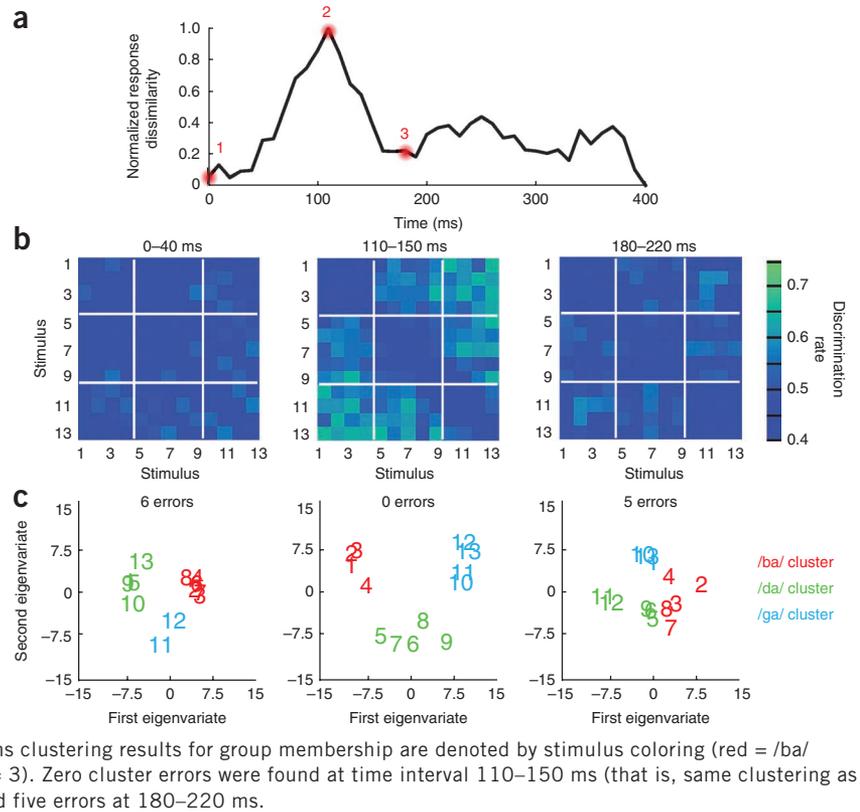
As the functional organization of the pSTG exhibits a distributed representation for speech sounds, in contrast with the well-defined gradient of frequency selectivity in the primary auditory cortex<sup>16</sup>, we used an information-based strategy to determine how distributed neural population activity patterns might encode speech. The specific measure that we used was the degree to which a multivariate pattern classifier (L1 norm regularized logistic regression<sup>17</sup>) was able to distinguish single-trial response patterns of the evoked cortical potentials.

In linguistics, confusion matrices are commonly used to explore the perceptual organization and distinctiveness of speech sounds<sup>18</sup>. We assembled the performance results from pattern classification into neural confusion matrices to organize the neural response dissimilarity across each pair-wise stimulus comparison (Fig. 2). The confusion matrices were calculated for each subject and then averaged for the group using data binned in 40-ms time intervals and advanced by 10-ms steps. Classification performance varied between stimulus pairs, with peak discrimination at 78–79% for each subject.

Two things were apparent from the averaged matrices. First, when analyzed over successive time epochs, the overall neural pattern dissimilarity gradually increased (Fig. 2a and Supplementary Results) and peaked transiently around 110 ms. Thus, the greatest overall neural pattern dissimilarity occurred at the peak response of physiologic evoked potentials, as opposed to early- or longer-latency responses. Second, although the overall discriminability among responses was highest during that interval, specific comparisons in the confusion matrices also showed poor discriminability, suggesting structured organization of response patterns. For example, neural responses to stimuli 1–4 were indiscriminable, whereas those responses to stimuli 7 and 11 were highly discriminable (Fig. 2b).

To examine the similarity relationships across all stimuli, we applied unsupervised multidimensional scaling (MDS) to the confusion matrix to construct a geometric space in which the Euclidean distances between different stimuli markers correspond to the similarity of their neural

**Figure 2** Categorical organization of neural response patterns to a speech-stimulus continuum. **(a)** Rapid and transient neural representation for speech stimulus discriminability. Time series of the total normalized neural pattern dissimilarity derived from classifier performance aggregated across all pair-wise stimulus comparisons are shown. Peak dissimilarity occurred at the same time as peak of evoked potential magnitude in **Figure 1e**. **(b)** Structured neural dissimilarity. Shown are neural confusion matrices for three time intervals: 0–40 ms (1), 110–150 ms (2) and 180–220 ms (3) (group average data). Color bar scaling corresponds to the classifier performance for each pair-wise stimulus comparison shown in individual matrix pixels. In the 110–150-ms interval, responses to some stimulus pairs, for example, 1 versus 4, 8 versus 5 or 10 versus 13, were nearly indistinguishable, whereas other stimulus pairs elicited responses that were much easier to discriminate, such as 7 versus 11 or 3 versus 9. **(c)** Relational organization of neural pattern response dissimilarity using MDS. Neural pattern dissimilarity was proportional to the Euclidean distance (that is, similar response patterns are grouped closely together, whereas dissimilar patterns are positioned far apart). K-means clustering results for group membership are denoted by stimulus coloring (red = /ba/ sounds, green = /da/ sounds, blue = /ga/ sounds,  $k = 3$ ). Zero cluster errors were found at time interval 110–150 ms (that is, same clustering as in psychophysical results), six errors at 0–40 ms and five errors at 180–220 ms.



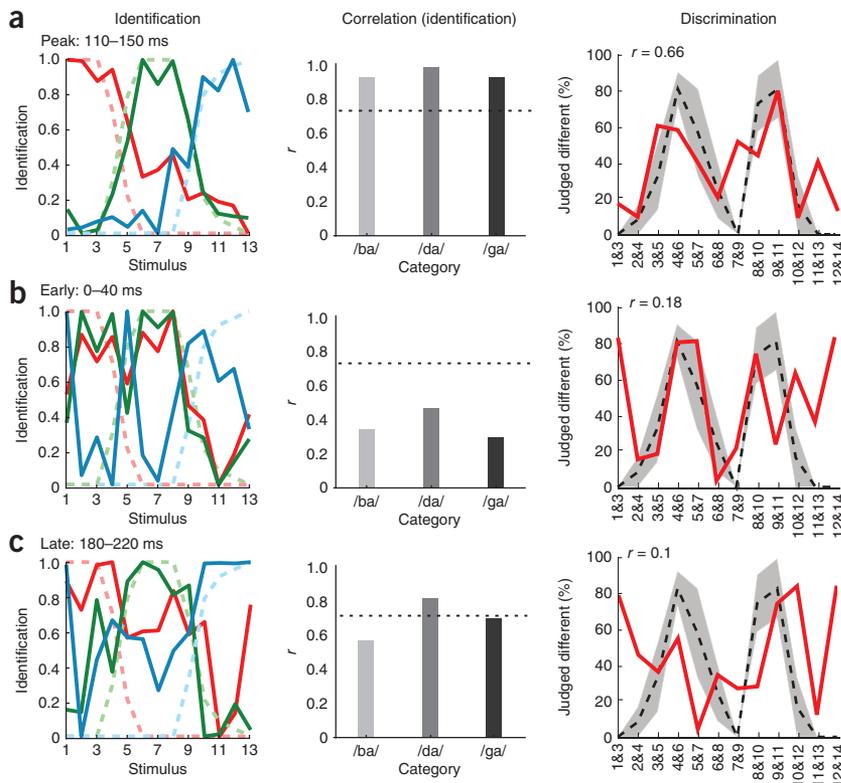
responses<sup>19</sup>. Stimuli placed close together elicited similar neural response patterns, whereas stimuli positioned far apart elicited dissimilar response patterns. Visual inspection of the MDS plots suggested that, during maximal neural response discriminability (110–150 ms), neuronal

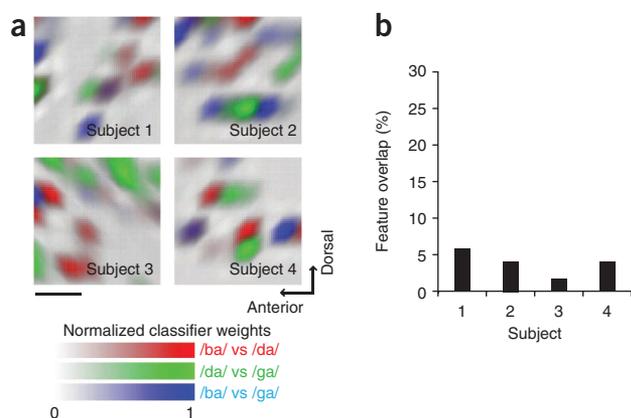
responses to different stimuli organized into three discrete groupings (**Fig. 2c**, see **Supplementary Fig. 1** for MDS residual stress values).

To confirm these results, we used another method, unsupervised K-means clustering analysis, to examine the independent grouping

of neural response patterns. This method is well-suited for exploring categorical data organization because it extracts a clustering of the data that minimizes intracluster distances and maximizes intercluster differences.

**Figure 3** Correlation of neurometric and psychometric category boundaries. Peak encoding at 110–150 ms. **(a)** Left, comparison of neuronal-derived (dark) and psychophysical-derived (light/dashed) identification functions. Neurometric identification functions were determined using the MDS distance between each stimulus position and the three cluster means. Middle, correlation between neurometric and psychometric identification functions (Pearson's correlation, 0.92 for /ba/, 0.98 for /da/ and 0.92 for the /ga/ category; dotted line, threshold of corrected  $P$  value at 0.05). Right, comparison of neural (red) and psychophysical (black/dashed) discrimination functions. The neurometric discrimination functions were derived from the distance of the stimulus responses in MDS space. At 110 ms, both the position of the maxima and the general shape of the neurometric function correlated well with the psychometric function ( $r = 0.66$ ,  $P < 0.05$ ). Early (0–40 ms, **b**) and late (180–220 ms, **c**) epoch field potentials indicated poor correlation between neural and psychophysical results (insets).





**Figure 4** Topography of discriminative cortical sites in the pSTG underlying categorical speech perception. **(a)** The degree of separability of the various evoked activations at each electrode position is shown as classifier weights. The spatial patterns indicate that discriminative neuronal activation was not distributed over the pSTG, but was instead concentrated in few cortical sites. **(b)** The informative loci overlapped very little between comparisons of the features (on average  $3.9 \pm 0.88\%$ ; indicated by mixed colors such as magenta, cyan or orange in **a**), suggesting that the neuronal categorization is not accomplished by simply scaling the responses in the same network, but is instead a function of spatially discrete and local selectivity.

The neural responses were organized into three discrete and independent clusters, representing the /ba/, /da/ and /ga/ syllables (**Fig. 2c**). No errors in cluster membership were found at the peak of discriminability (110- and 120 ms interval start). The neuronal stimulus responses clustered in exactly the same way as we observed in perception (/ba/ 1–4, /da/ 5–9 and /ga/ 10–13), whereas earlier and later epochs yielded error-prone cluster estimates (see **Supplementary Fig. 2** for entire cluster error time series). Notably, the separate organization of response clusters matched the robust perception that /ba/, /da/ and /ga/ are perceived as independent and unique phonetic entities, rather than speech sounds occurring along a linear acoustic or even phonetic continuum.

To evaluate how well the neural pattern correlated to the psychophysical behavior, we plotted neurometric identification functions for each phonetic category using the normalized distance in MDS space between each stimulus position and the three cluster means. This revealed a similar appearance to the psychometric identification functions, with steep boundaries occurring between phoneme categories (Pearson's correlation,  $r > 0.9$  for each function at 110-ms intervals start,  $P < 0.05$ ; **Fig. 3**, and **Supplementary Figs. 2** and **3** for entire cluster-error time series and combined MDS and K-means solutions, respectively, and **Supplementary Fig. 4**). A neurometric discrimination function was also derived from distances between individual stimulus positions in MDS space. This also achieved good correlation with the psychometric functions for discrimination (Pearson's correlation,  $r = 0.66$  at 110-ms intervals start,  $P < 0.05$ ). Notably, we observed good correspondence between the two neurometric functions; the peaks of the discrimination occurred for the same stimuli as the steepest parts of the identification, thus fulfilling the criterion for neural categorical organization. This organized representation was transient, spanning the neuronal response from 110–160 ms.

To determine the spatial organization of phonetic representation, we next identified the cortical sites contributing to stimulus discriminability by extracting the most informative electrodes as determined by the classifier. Although the evoked potentials showed

overlapping representation for speech sounds, discrete differences in cortical activations ( $<4\text{mm}$ ) were observed to underlie phonemic discrimination. We plotted these spatially contrastive differences between various categories (**Fig. 4**). The small overlap between these loci suggests that phonetic encoding is not simply a scaling of the response amplitudes in the same neuronal population.

## DISCUSSION

A key element of speech perception is the categorization of acoustically variable inputs into a discrete phonetic code. Understanding the neural basis of this process is a central question in the study of the human capacity for language<sup>20</sup>. We found that the pSTG is robustly organized according to its sensitivity to phonetic objects rather than to the linear changes of spectrotemporal acoustic cues. For the stop consonant-vowel sounds that we used, we observed a complex distributed pattern of evoked neural activity recorded by a cortical microarray. The discriminability of these response patterns, however, relies on transient temporal and local, non-overlapping spatial neural representations.

Without a priori knowledge on functional organization of the pSTG, the multivariate pattern classifier and MDS are useful methods for examining the critical acoustic features underlying stimulus discriminability. The first MDS dimension correlated linearly with the F2 onset frequency, which, in natural speech, cues the feature of place of articulation across /b/ to /d/ to /g/ (that is, location of constriction in the vocal tract from lips to teeth to soft palate). The second MDS dimension correlated with the size of F2 transition (absolute value of the difference between the onset F2 frequency and the vowel F2 frequency), which, in these stimuli, cues the linguistic feature (coronal; that is, not produced by tongue tip position), grouping /b/ and /g/ together. Critically, the grouping patterns observed did not arise from one dimension alone, but instead from the specific combination of two different linguistically relevant feature dimensions: the F2 onset frequency and the F2 formant transition. Thus, these results support the notion that phonetic encoding in the pSTG appears to be facilitated by feature detectors that integrate specific spectrotemporal cues relevant to speech.

The pSTG appears to have a specialized role in phonetic processing because of its specific responsiveness to speech over other sounds<sup>21–25</sup> and its direct anatomic connections to cortical areas supporting lexical and semantic extraction<sup>26–28</sup>. A recent fMRI study found activation of the left pSTG increased overall after engaging in categorical perception tasks on phonetic and nonphonetic sine-wave syllable tokens<sup>29</sup>. Our results extend these findings by providing new information about the timing and topography mechanisms intrinsic to stimulus encoding in the pSTG.

Although our microarray recordings focused on auditory processing in the pSTG, fMRI has implicated other areas during active phonetic discrimination. Selective amplification of left supramarginal gyrus activity has been observed in response to the contrastive features of stimulus pairs spanning a /ba/-/da/ category boundary<sup>30</sup>. Invariant neural activation of the left inferior frontal gyrus was found for sounds morphed along a different acoustic continuum for voice onset time<sup>31</sup>. These findings suggest that there are several other cortical areas that are likely involved in the behavioral processes of phonetic detection, working memory and/or decision making.

Our results indicate that the pSTG implements rapid categorical phonetic analysis, integrating spectro-temporal features to create invariant higher order linguistic structure<sup>32</sup>. This pattern is consistent with the pragmatic demands of spoken English; there is a meaning distinction between /b/ and /d/ (for example, 'bad' versus

'dad'), whereas the distinction between the variations of /b/ carries no meaning. Our results provide a mechanistic account whereby the pSTG functions as a critical locus for phonological processing in the neural representation of human language.

## METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/natureneuroscience/>.

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

## ACKNOWLEDGMENTS

We are grateful to the four individuals who participated in this experiment and to A. Flinker for help with data acquisition. This research was supported by US National Institutes of Health grants NS21135 (R.T.K.), PO4813 (R.T.K.), F32NS061552 (E.F.C.), K99NS065120 (E.F.C.), FKZ-MK48-2009/003 (J.W.R.) and RI1511/1-3 (J.W.R.).

## AUTHOR CONTRIBUTIONS

E.F.C. designed the experiments, collected the data and wrote the manuscript. E.F.C. and J.W.R. analyzed the data, evaluated results and edited the manuscript. J.W.R., N.M.B. and M.S.B. helped with data collection. K.J. and R.T.K. reviewed the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

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

- Perkell, J. & Klatt, D.H. *Invariance and Variability in Speech Processes* (Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1986).
- Lieberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. Perception of the speech code. *Psychol. Rev.* **74**, 431–461 (1967).
- Diehl, R.L., Lotto, A.J. & Holt, L.L. Speech perception. *Annu. Rev. Psychol.* **55**, 149–179 (2004).
- Lieberman, A.M. & Mattingly, I.G. A specialization for speech perception. *Science* **243**, 489–494 (1989).
- Vihman, M. *Phonological Development: The Origins of Language in the Child* (Wiley-Blackwell, Cambridge, 1996).
- Lieberman, A.M., Harris, K.S., Hoffman, H.S. & Griffith, B.C. The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* **54**, 358–368 (1957).
- Harnad, S.R. *Categorical Perception: The Groundwork of Cognition* (Cambridge University Press, Cambridge, 1987).
- Edwards, E. *et al.* Spatiotemporal imaging of cortical activation during verb generation and picture naming. *Neuroimage* **50**, 291–301 (2010).
- Creutzfeldt, O., Ojemann, G. & Lettich, E. Neuronal activity in the human lateral temporal lobe. I. Responses to speech. *Exp. Brain Res.* **77**, 451–475 (1989).
- Boatman, D., Lesser, R.P. & Gordon, B. Auditory speech processing in the left temporal lobe: an electrical interference study. *Brain Lang.* **51**, 269–290 (1995).
- Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T. & Medler, D.A. Neural substrates of phonemic perception. *Cereb. Cortex* **15**, 1621–1631 (2005).
- Crone, N.E., Boatman, D., Gordon, B. & Hao, L. Induced electrocorticographic gamma activity during auditory perception. *Brazier Award-winning article, 2001. Clin. Neurophysiol.* **112**, 565–582 (2001).
- Howard, M.A. *et al.* Auditory cortex on the human posterior superior temporal gyrus. *J. Comp. Neurol.* **416**, 79–92 (2000).
- Penfield, W. & Jasper, H. *Epilepsy and the Functional Anatomy of the Human Brain* (Little, Brown and Company, Boston, 1954).
- Haglund, M.M., Berger, M.S., Shamseldin, M., Lettich, E. & Ojemann, G.A. Cortical localization of temporal lobe language sites in patients with gliomas. *Neurosurgery* **34**, 567–576 discussion 576 (1994).
- Merzenich, M.M. & Brugge, J.F. Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Res.* **50**, 275–296 (1973).
- Koh, K., Kim, S.J. & Boyd, S. An interior-point method for large-scale l1-regularized least squares. *J. Mach. Learn. Res.* **8**, 1519–1555 (2007).
- Miller, G.A. & Nicely, P.E. An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.* **27**, 338–352 (1955).
- Iverson, P. & Kuhl, P.K. Perceptual magnet and phoneme boundary effects in speech perception: do they arise from a common mechanism? *Percept. Psychophys.* **62**, 874–886 (2000).
- Lieberman, A.M. & Whalen, D.H. On the relation of speech to language. *Trends Cogn. Sci.* **4**, 187–196 (2000).
- Binder, J.R. *et al.* Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* **10**, 512–528 (2000).
- Benson, R.R., Richardson, M., Whalen, D.H. & Lai, S. Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. *Neuroimage* **31**, 342–353 (2006).
- Uppenkamp, S., Johnsrude, I.S., Norris, D., Marslen-Wilson, W. & Patterson, R.D. Locating the initial stages of speech-sound processing in human temporal cortex. *Neuroimage* **31**, 1284–1296 (2006).
- Vouloumanos, A., Kiehl, K.A., Werker, J.F. & Liddle, P.F. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J. Cogn. Neurosci.* **13**, 994–1005 (2001).
- Jäncke, L., Wustenberg, T., Scheich, H. & Heinze, H.J. Phonetic perception and the temporal cortex. *Neuroimage* **15**, 733–746 (2002).
- Scott, S.K. & Wise, R.J. The functional neuroanatomy of prelexical processing in speech perception. *Cognition* **92**, 13–45 (2004).
- Hickok, G. & Poeppel, D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* **92**, 67–99 (2004).
- Whalen, D.H. *et al.* Differentiation of speech and nonspeech processing within primary auditory cortex. *J. Acoust. Soc. Am.* **119**, 575–581 (2006).
- Desai, R., Liebenthal, E., Waldron, E. & Binder, J.R. Left posterior temporal regions are sensitive to auditory categorization. *J. Cogn. Neurosci.* **20**, 1174–1188 (2008).
- Raizada, R.D. & Poldrack, R.A. Selective amplification of stimulus differences during categorical processing of speech. *Neuron* **56**, 726–740 (2007).
- Blumstein, S.E., Myers, E.B. & Rissman, J. The perception of voice onset time: an fMRI investigation of phonetic category structure. *J. Cogn. Neurosci.* **17**, 1353–1366 (2005).
- Blumstein, S.E. & Stevens, K.N. Perceptual invariance and onset spectra for stop consonants in different vowel environments. *J. Acoust. Soc. Am.* **67**, 648–662 (1980).

## ONLINE METHODS

The experimental protocol was approved by the University of California, San Francisco and Berkeley institutional review boards and Committees on Human Research, and the subjects gave their informed consent before testing.

**Stimulus synthesis and behavioral testing.** Speech stimuli were synthesized using the Klatt synthesizer. The critical stimulus variation was created by stepwise changes in the F2 onset frequency over 14 equal steps<sup>6</sup> (100-Hz step increases ranging from 800–2,100 Hz) spanning the perceptual phonetic continuum from /ba/ to /da/ to /ga/.

Before surgery, subjects performed a two-step AX (same versus different) discrimination task and then an identification task in which they labeled the stimulus as either /ba/, /da/ or /ga/. Subjects then underwent awake craniotomy with speech mapping by electrocortical stimulation as part of their epilepsy or brain tumor surgery. The stimulus tokens were aurally presented in a pseudorandom order via free-field loudspeakers at approximately 80 dB. As a result of time constraints in the operating room, each stimulus token was repeated 25 times, for a total of 350 total trials per subject.

**Subjects and intraoperative testing.** The four subjects in this study underwent awake craniotomy as part of their epilepsy or brain tumor surgery. They gave their written informed consent before the day of surgery. **Supplementary Table 1** shows the patient characteristics included in this study. All subjects underwent neuropsychological language testing and were found to be normal. The Boston naming test and verbal fluency test were used for preoperative language testing. The Wada test was used for language dominance assessment.

**Data acquisition and preprocessing.** The electrocorticogram was recorded using a customized 64-channel subdural cortical electrode microarray, with center-to-center distance of 4 mm. The electrode array was placed on the lateral aspect of the posterior superior temporal gyrus using stereotactic intraoperative neuronavigation. The signal was recorded with a TDT amplifier optically connected to a digital signal processor (Tucker-Davis Technologies).

The electrocorticogram data were digitally low-pass filtered at 50 Hz and resampled at 508.6 Hz. Each channel time series was visually and quantitatively inspected for artifacts or excessive noise. The data was then segmented with a 100-ms stimulus pre-stimulus baseline and a 400-ms post-stimulus interval. The common mode signal was estimated using principal component analysis with channels as repetitions and was removed from each channel time series using vector projection.

**Estimation of neuronal response dissimilarity.** We estimated single-trial pair-wise dissimilarity of the neuronal response patterns evoked by different stimulus tokens using an L1-norm regularized logistic regression classifier<sup>17</sup> applied to the time series data in a leave-one-trial-out cross validation procedure. Dissimilarities were estimated for 40-ms-long data windows, advanced every 10 ms. To increase the ratio of the number of examples to the number of features, we combined responses to adjacent stimuli (for example, 1 and 2, 2 and 3, etc.), doubling the number of trials used per dissimilarity estimate. Note that labels in the figures list only the first stimulus in these combined sets of trials.

Both feature selection and classifier training were performed in the cross-validation loop. Feature selection was carried out by calculating univariate effect sizes for each data sample and discarding samples with small effects from classifier training. L1-norm logistic regression is well suited for classification problems involving high-dimensional feature spaces and relatively few examples for training because it provides good generalization performance even when relatively few training data are available.

Generalization rate expressed as percent correct classifications measured the dissimilarity of the neuronal responses of a stimulus pair. The single trial classification measures of pair-wise neural response dissimilarity were used to construct a confusion matrix for each time interval.

**Derivation of neuronal response classes, neuronal identification and discrimination functions.** MDS was applied to the confusion matrices averaged over all subjects to represent neural response patterns to different phoneme stimuli in a new space in which the distance between neuronal responses represents their relative similarity (and dissimilarity)<sup>33</sup>. The objective in MDS is to minimize the reconstruction error measured by Kruskal Stress<sup>34</sup>. The MDS embedding was calculated in three dimensions, given a priori considerations of how many dimensions would be maximally required. The simultaneous representation of all neuronal responses in on common similarity space allowed us to use K-means cluster analysis<sup>35</sup> to test when, if at all, neuronal responses group in a way that parallels perceptual grouping obtained psychophysically. K-means clustering implements the definition of categorical representation of stimulus responses<sup>7</sup>, hence the obvious choice for  $k$ , the number of expected clusters, was three, the number of perceived phonemes.

To derive the three neuronal identification functions, we calculated three distance functions in MDS similarity space, one between each of the three cluster prototypes and all neuronal responses. These functions can be directly compared to the psychophysical identification functions using a Pearson's correlation analysis. The psychophysical discrimination functions were approximated by calculating the distances of the neuronal responses between consecutive pairs of stimuli in the MDS-representation.

**Reconstruction of spatial informative patterns.** The trained classifier's weight vector quantifies the amount of information each feature provides for classification. Highly informative features receive higher weights and features providing little or no information receive low or zero weights. Features with zero entries in the weight vector do not contribute to the classification results.

The feature weights represent averages over cross validation results and samples per electrode in the analysis interval. The average feature weights represent an estimate of how informative a local neuronal population (per electrode) was judged by the classifier.

33. Iverson, P. & Kuhl, P.K. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *J. Acoust. Soc. Am.* **97**, 553–562 (1995).
34. Kruskal, J.B. & Wish, M. *Multidimensional Scaling* (Sage Publications, Newbury Park, California, 1978).
35. Shepard, R.N. Multidimensional scaling, tree-fitting and clustering. *Science* **210**, 390–398 (1980).