Published in final edited form as:

Nat Neurosci. 2010 November ; 13(11): 1428–1432. doi:10.1038/nn.2641.

Categorical Speech Representation in Human Superior Temporal Gyrus

Edward F. Chang[1,2,*], **Jochem W. Rieger**[2,4,*], **Keith Johnson**[5], **Mitchel S. Berger**[1], **Nicholas M. Barbaro**[1], and **Robert T. Knight**[1,2,3]

[1]Department of Neurological Surgery, University of California San Francisco

[2]Helen Wills Neuroscience Institute, University of California Berkeley

[3]Department of Psychology, University of California, Berkeley

[4]Department of Neurology, Otto-von-Guericke University, Magdeburg, Germany

[5]Department of Linguistics, University of California Berkeley

Abstract

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. Here we show 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. While speech-sound responses were distributed, spatially discrete cortical loci were found to underlie specific phonetic discrimination. Thus, we demonstrate direct evidence for acoustic-to-higher order phonetic level encoding of speech sounds in human language receptive cortex.

INTRODUCTION

A fundamental property of speech perception is that listeners map continuously variable acoustic speech signals onto discrete phonetic sound categories1–3. This "phonetic" mode of listening4 lays the phonological foundation for speaking new words5 and mapping speech

Correspondence to: Edward F. Chang, MD, W.M. Keck Foundation Center for Integrative Neuroscience, Departments of Neurological Surgery and Physiology, University of California, San Francisco, 505 Parnassus Avenue, M779, San Francisco, CA 94143, (Fax) 415 353-3907, (Phone) 415 385-5280, changed@neurosurg.ucsf.edu, Jochem W. Rieger, PhD, Klinik für Neurology, Otto-von-Guericke Universität, Leipzigerstr. 44, 39120 Magdeburg, Germany, (Fax) +49 39167290224, (Phone) +49 3916117511, jochem.rieger@med.ovgu.de.

*these authors contributed equally to this manuscript

AUTHOR CONTRIBUTIONS

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

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 6, 7.

A convergence of research supports a key role of the posterior superior temporal gyrus (pSTG) in Wernicke's area for higher-order auditory processing of speech sounds8–13. Current noninvasive neurophysiologic methodologies (e.g. fMRI, MEG, PET) have provided important insights into speech localization. However, due to 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 how 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 (i.e. veridical representation), or rather, to a higher-order linguistic extraction of phonetic categories? Furthermore, what neural response features (e.g. place, time, 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 epilepsy14 or brain tumor surgery15. While limited to rare clinical settings, high-density electrocorticographic recordings offer the advantage of simultaneous high spatial (millimeters) with real-time temporal (ms, 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 psycho-metric functions, suggesting that pSTG encoding represents high-order invariant representation for speech sounds.

RESULTS

We employed a classic paradigm first described by Liberman and colleagues6 in 1957 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 the phoneme discrimination boundary corresponding to the 50% point of the labeling curve, are the defining psychophysical properties of categorical perception (Fig. 1b and c). Therefore, one does not hear step-like

changes corresponding to the changes in the acoustic signal, but rather perceives essentially quantal jumps from one perceptual category to another.

While subjects were fully awake in the operating room, a customized high-density 64-electrode microarray (4 mm spacing) was placed 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). Examples of the spatial distribution of responses to /ba/, /da/, and /ga/ are shown in Figure 1f, which demonstrate distributed responses across the pSTG.

Since the functional organization of the pSTG exhibits a distributed representation for speech sounds, in contrast to the well-defined gradient of frequency selectivity in the primary auditory cortex16, we implemented an information-based strategy to determine how distributed neural population activity patterns might encode speech. The specific measure we used was the degree to which a multivariate pattern classifier (L1 norm regularized logistic regression17) 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 sounds18. 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 important results were apparent from the averaged matrices. First, when analyzed over successive time epochs, the overall neural pattern dissimilarity gradually increased (Fig. 2a), 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, while 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, unsupervised multidimensional scaling (MDS) was applied to the confusion matrix to construct a geometric space in which the Euclidean distances between different stimuli markers correspond to similarity of their neural responses19. 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 Figures 1–3 for the entire MDS time series).

To confirm these results, another method, unsupervised K-means clustering analysis, was used 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 intra-cluster distances and maximizes inter-cluster differences. The neural responses were organized into three discrete and independent clusters, representing /ba/ (red), /da/ (green), and /ga/ (blue) syllables respectively in Fig. 2c (stimulus color corresponds to each cluster). No errors in cluster membership were found at the peak of discriminability (110 & 120 msec interval start), The neuronal stimulus responses clustered in exactly the same way as found in perception (/ba/ 1–4, /da/ 5–9, and /ga/ 10–13), whereas earlier and later epochs yielded error-prone cluster estimates (see Supplemental Materials for entire cluster error time series). Importantly, the separate organization of response clusters matches 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, neurometric identification functions for each phonetic category were plotted 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. 3a, Supplementary Figure 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; $p<0.05$; Fig. 3b). More importantly, we observed good correspondence between the two neurometric functions: the peaks of the discrimination occur 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. While the evoked potentials showed overlapping representation for speech sounds, discrete differences in cortical activations (<4mm) were observed to underlie phonemic discrimination. Those spatially contrastive differences between various categories are shown in Figure 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 language20. 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 used in this study, we observed a complex distributed pattern of evoked neural activity recorded by a cortical

microarray. The discriminability of these response patterns, however, relies upon transient temporal and local, non-overlapping spatial neural representations.

Without a priori knowledge on functional organization of the pSTG, the use of a multivariate pattern classifier and MDS were useful methods to reveal the critical acoustic features underlying stimulus discriminability. The first MDS dimension correlated linearly with the F2 onset frequency, which in natural speech this parameter cues the feature of place of articulation across /b/ to /d/ to /g/ (i.e. 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, i.e. 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. Therefore, these results support a 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 sounds21–25, and its direct anatomic connections to cortical areas supporting lexical and semantic extraction26–28. In a recent fMRI study, Desai et al. found overall increased activation of the left pSTG after engaging in categorical perception tasks on phonetic and non-phonetic sine-wave syllable tokens29. Our results extend these findings by providing new information about the timing and topography mechanisms intrinsic to stimulus encoding in the pSTG.

While our microarray recordings focused on auditory processing in the pSTG, fMRI has implicated other areas during active phonetic discrimination. Raizada et al. observed selective amplification of left supramarginal gyrus activity in response to the contrastive features of stimulus pairs spanning a /ba/-/da/ category boundary30. Blumstein et al. found invariant neural activation of the left inferior frontal gyrus for sounds morphed along a different acoustic continuum for voice onset time31. These findings suggest that there are several other cortical areas likely involved in the behavioral processes of phonetic detection, working memory, and/or decision-making.

Our results demonstrate that the pSTG implements rapid categorical phonetic analysis, integrating spectro-temporal features to create invariant higher-order linguistic structure32. This pattern is consistent with the pragmatic demands of spoken English: there is a meaning distinction between /b/ and /d/ (e.g. ‘bad’ versus ‘dad’), while 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

The experimental protocol was approved by the University of California, San Francisco and Berkeley institutional review boards and Committees on Human Research and subjects gave their informed consent prior to 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 steps6, (100 Hz step increases ranging from 800 to 2100 Hz) spanning the perceptual phonetic continuum from /ba/ to /da/ to /ga/.

Before surgery, subjects first performed a two-step AX (“same”/”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. Due to 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 prior to the day of surgery. Table 1 shows the patient characteristics included in this study. All subjects underwent neuropsychological language testing, and were found to be normal. Boston naming test, and verbal fluency test were used for preoperative language testing. The Wada-test was used for language dominance assessment.

Before surgery, patients received midazolam (2 mg) and fentanyl (50 to 100 μg). At the start of surgery, propofol (at a dose of 50 to 100 μg per kilogram of body weight per minute) and remifentanil (0.05 to 0.2 μg per kilogram per minute) were given for sedation during scalp incision and craniotomy. After the bone flap was removed, the dura was infiltrated with lidocaine and all anesthetics were discontinued. No anesthesia was administered during routine electrocortical stimulation mapping while the patients were fully awake. Once stimulation mapping was completed, the stimuli were aurally presented via free-field loudspeakers at approximately 80 dB. Patients were instructed to keep their eyes open while passively listening to the stimuli.

Data acquisition and preprocessing

The electrocorticogram (ECoG) 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, Alachua FL USA).

The ECoG 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 classifier17 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 (e.g. 1&2; 2&3 etc.), doubling number of trials used per dissimilarity estimate. Note that labels in the figures of the main paper 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 done 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 measures 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

Metric multidimensional scaling (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)33. The objective in MDS is to minimize the reconstruction error measured by Kruskall Stress34. 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 analysis35 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 responses7, 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We are grateful to the four patients who participated in this experiment. Also, thanks to A Flinker for help with data acquisition. This research was supported by NIH grants NS21135 (RTK), PO4813 (RTK), F32NS061552 (EC), and K99NS065120 (EC), and FKZ-MK48-2009/003 and RI1511/1-3 (JWR).

REFERENCES

1. Perkell, J.; Klatt, DH., editors. Invariance and variability in speech processes. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986.
2. Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychol Rev. 1967; 74:431–461. [PubMed: 4170865]
3. Diehl RL, Lotto AJ, Holt LL. Speech perception. Annu Rev Psychol. 2004; 55:149–179. [PubMed: 14744213]
4. Liberman AM, Mattingly IG. A specialization for speech perception. Science. 1989; 243:489–494. [PubMed: 2643163]
5. Vihman, M. Phonological Development: The Origins of Language in the Child. Cambridge: Wiley-Blackwell; 1996.
6. Liberman AM, Harris KS, Hoffman HS, Griffith BC. The discrimination of speech sounds within and across phoneme boundaries. J Exp Psychol. 1957; 54:358–368. [PubMed: 13481283]
7. Harnad, SR. Categorical Perception: The Groundwork of Cognition. Cambridge: Cambridge University Press; 1987.
8. Edwards E, et al. Spatiotemporal imaging of cortical activation during verb generation and picture naming. Neuroimage. 2010; 50:291–301. [PubMed: 20026224]
9. Creutzfeldt O, Ojemann G, Lettich E. Neuronal activity in the human lateral temporal lobe. I. Responses to speech. Exp Brain Res. 1989; 77:451–475. [PubMed: 2806441]
10. Boatman D, Lesser RP, Gordon B. Auditory speech processing in the left temporal lobe: an electrical interference study. Brain Lang. 1995; 51:269–290. [PubMed: 8564472]
11. Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. Cereb Cortex. 2005; 15:1621–1631. [PubMed: 15703256]
12. Crone NE, Boatman D, Gordon B, Hao L. Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article, 2001. Clin Neurophysiol. 2001; 112:565–582. [PubMed: 11275528]

13. Howard MA, et al. Auditory cortex on the human posterior superior temporal gyrus. J Comp Neurol. 2000; 416:79–92. [PubMed: 10578103]
14. Penfield, W.; Jasper, H. Epilepsy and the functional anatomy of the human brain. Boston: LIttle, Brown and Company; 1954.
15. Haglund MM, Berger MS, Shamseldin M, Lettich E, Ojemann GA. Cortical localization of temporal lobe language sites in patients with gliomas. Neurosurgery. 1994; 34:567–576. discussion 576. [PubMed: 7516498]
16. Merzenich MM, Brugge JF. Representation of the cochlear partition of the superior temporal plane of the macaque monkey. Brain Res. 1973; 50:275–296. [PubMed: 4196192]
17. Koh K, Kim SJ, Boyd S. An interior-point method for large-scale l1-regularized least squares. Journal of Machine Learning Research. 2007; 8:1519–1555.
18. Miller GA, Nicely PE. An analysis of perceptual confusions among some English consonants. J Acoust Soc Am. 1955; 27:338–352.
19. Iverson P, Kuhl PK. Perceptual magnet and phoneme boundary effects in speech perception: do they arise from a common mechanism? Percept Psychophys. 2000; 62:874–886. [PubMed: 10883591]
20. Liberman AM, Whalen DH. On the relation of speech to language. Trends Cogn Sci. 2000; 4:187–196. [PubMed: 10782105]
21. Binder JR, et al. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex. 2000; 10:512–528. [PubMed: 10847601]
22. Benson RR, Richardson M, Whalen DH, Lai S. Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. Neuroimage. 2006; 31:342–353. [PubMed: 16530428]
23. Uppenkamp S, Johnsrude IS, Norris D, Marslen-Wilson W, Patterson RD. Locating the initial stages of speech-sound processing in human temporal cortex. Neuroimage. 2006; 31:1284–1296. [PubMed: 16504540]
24. Vouloumanos A, Kiehl KA, Werker JF, Liddle PF. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. J Cogn Neurosci. 2001; 13:994–1005. [PubMed: 11595101]
25. Jancke L, Wustenberg T, Scheich H, Heinze HJ. Phonetic perception and the temporal cortex. Neuroimage. 2002; 15:733–746. [PubMed: 11906217]
26. Scott SK, Wise RJ. The functional neuroanatomy of prelexical processing in speech perception. Cognition. 2004; 92:13–45. [PubMed: 15037125]
27. Hickok G, Poeppel D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition. 2004; 92:67–99. [PubMed: 15037127]
28. Whalen DH, et al. Differentiation of speech and nonspeech processing within primary auditory cortex. J Acoust Soc Am. 2006; 119:575–581. [PubMed: 16454311]
29. Desai R, Liebenthal E, Waldron E, Binder JR. Left posterior temporal regions are sensitive to auditory categorization. J Cogn Neurosci. 2008; 20:1174–1188. [PubMed: 18284339]
30. Raizada RD, Poldrack RA. Selective amplification of stimulus differences during categorical processing of speech. Neuron. 2007; 56:726–740. [PubMed: 18031688]
31. Blumstein SE, Myers EB, Rissman J. The perception of voice onset time: an fMRI investigation of phonetic category structure. J Cogn Neurosci. 2005; 17:1353–1366. [PubMed: 16197689]
32. Blumstein SE, Stevens KN. Perceptual invariance and onset spectra for stop consonants in different vowel environments. J Acoust Soc Am. 1980; 67:648–662. [PubMed: 7358906]
33. Iverson P, Kuhl PK. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. J Acoust Soc Am. 1995; 97:553–562. [PubMed: 7860832]
34. Kruskal, JB.; Wish, M. Multidimensional Scaling. Newbury Park: Sage Publications; 1978.
35. Shepard RN. Multidimensional Scaling, Tree-Fitting, and Clustering. Science. 1980; 210:390–398. [PubMed: 17837406]

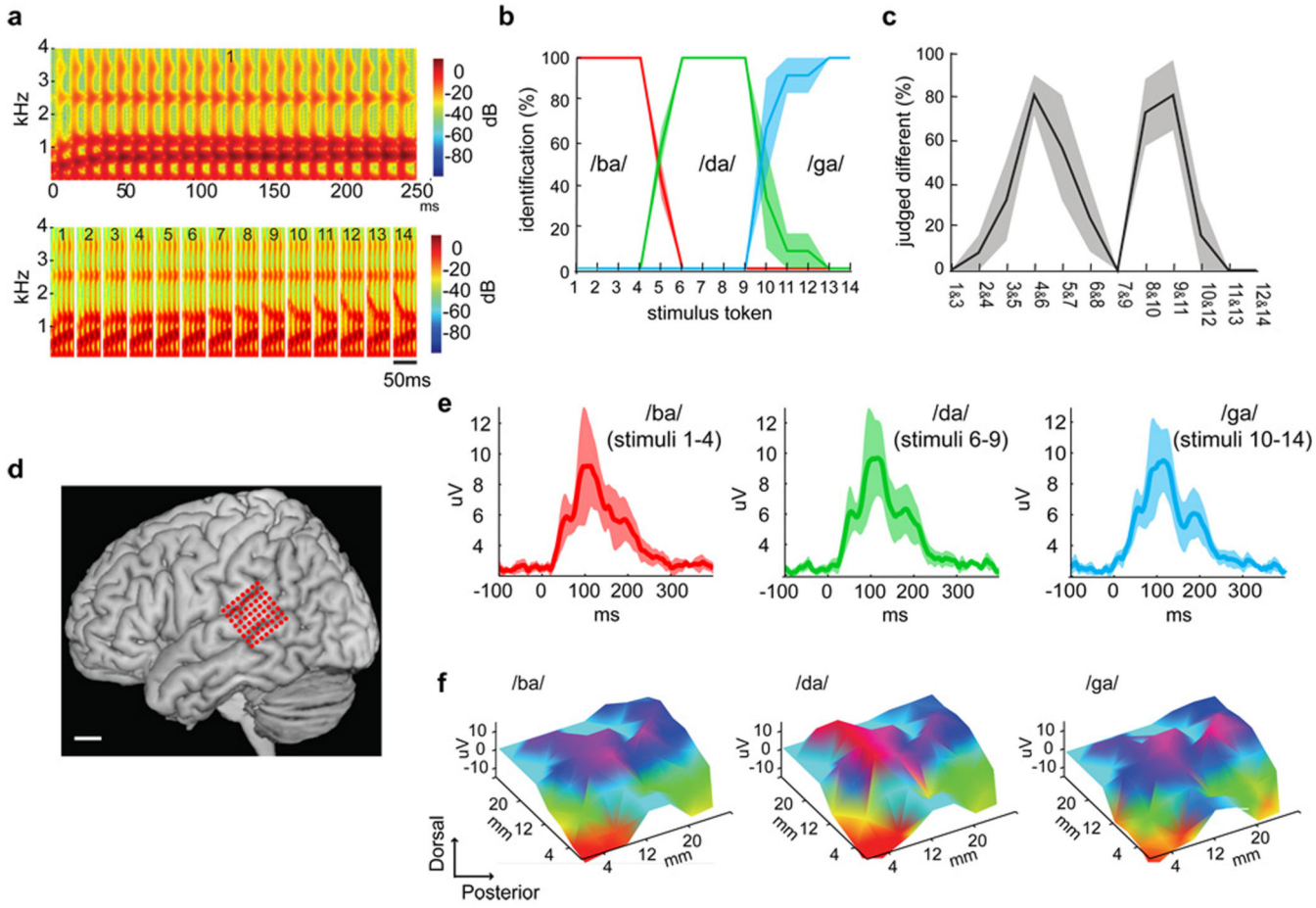

Figure 1. Psychophysics of categorical speech perception and speech-evoked responses during intraoperative human cortical recordings

A. Wide-band spectrograms of the stimulus token continuum, synthesized with equal parametric changes in the F2 starting frequency (from 800 to 2100Hz). Top shows the full spectrogram of a single token with an 800 Hz starting frequency (Stimulus #1, duration=250ms). Bottom shows the 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). Percentage of responses judged as "different" versus "same". The category boundaries located at peak discrimination are at stimuli 4 & 5, and 9 & 10. D. Three-dimensional surface reconstruction of representative brain MRI with superimposed electrode positions over pSTG. E. Grand average rooted mean square (RMS) evoked potentials (EP) recorded over pSTG for sound stimuli reliably categorized as /ba/ (tokens 1–4), /da/ (tokens 6–9), and /ga/ (tokens 10–14). Average EP (root mean square (RMS); solid line) and standard error of EP amplitudes (shaded). Potentials peak at approximately 110 ms after stimulus onset. F. Topographic plots of EPs at 110 ms for each prototype sound stimulus revealed distributed cortical activation pattern, with some sharply localized differences between stimuli. (uV=microvolts, ms=milliseconds, mm=millimeters).

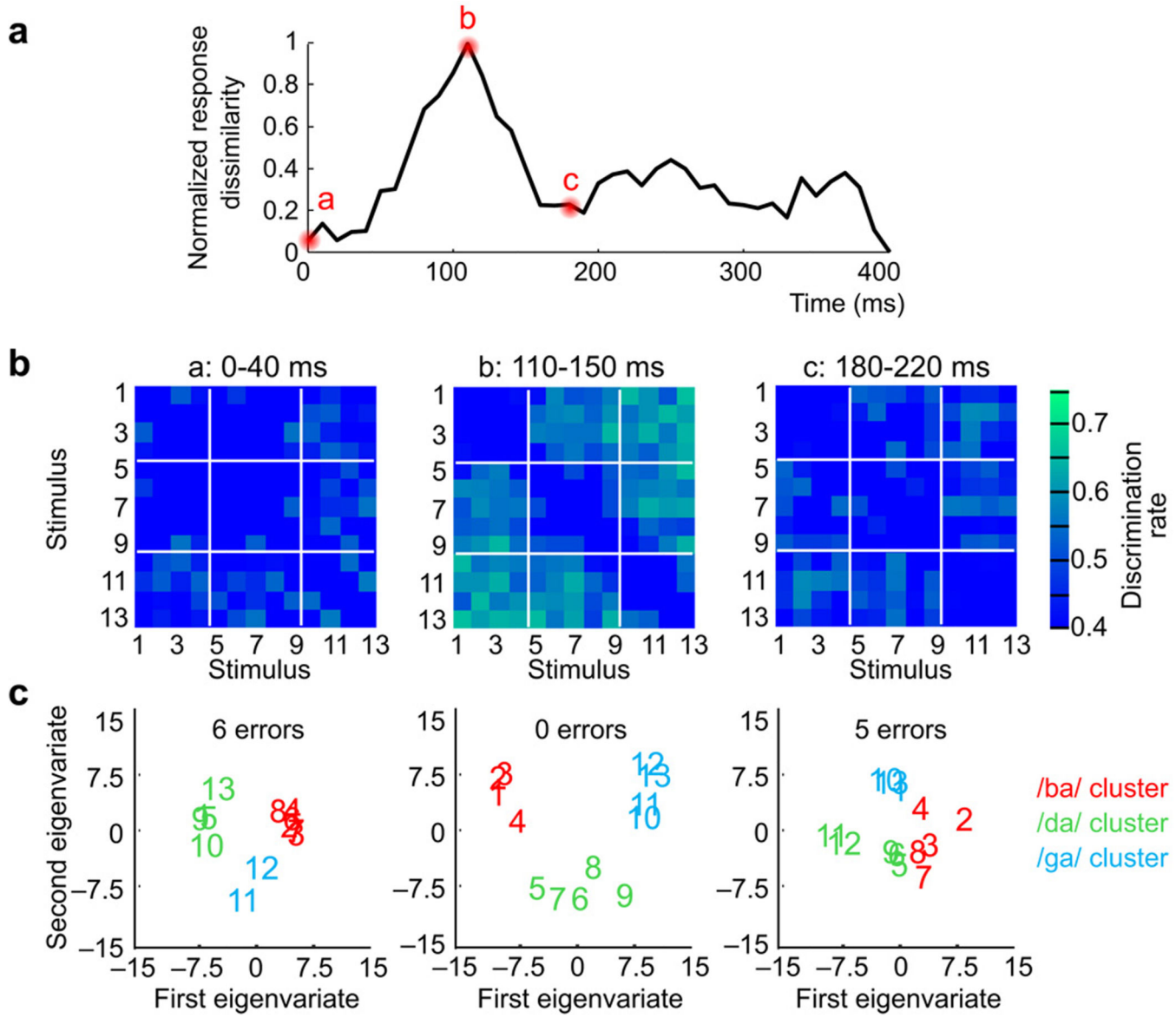

Figure 2. Categorical organization of neural response patterns to a speech-stimulus continuum

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. Peak dissimilarity occurs at the same time as peak of evoked potential magnitude in Figure 1e. B. Structured neural dissimilarity. Neural confusion matrices for three time intervals at 0–40ms (a), 110–150ms (b), and 180–220ms (c) (group average data). Colorbar scaling corresponds to the classifier performance for each pairwise stimulus comparison shown in individual matrix pixels. In the 110–150ms interval, responses to some stimulus pairs, for example, 1 vs 4, 8 vs 5, or 10 vs 13, are nearly indiscriminable while other stimulus pairs elicited responses that were much easier to discriminate, for example 7 vs 11, or 3 vs 9. C. Relational organization of neural pattern response dissimilarity using multidimensional scaling. Neural pattern dissimilarity is proportional to the Euclidean distance (i.e. similar response patterns are

grouped closely together, whereas dissimilar patterns are positioned far apart). K-means clustering results for group membership 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 (i.e. same clustering as in psychophysical results), but 6 errors at 0–40ms, and 5 errors at 180–220 ms.

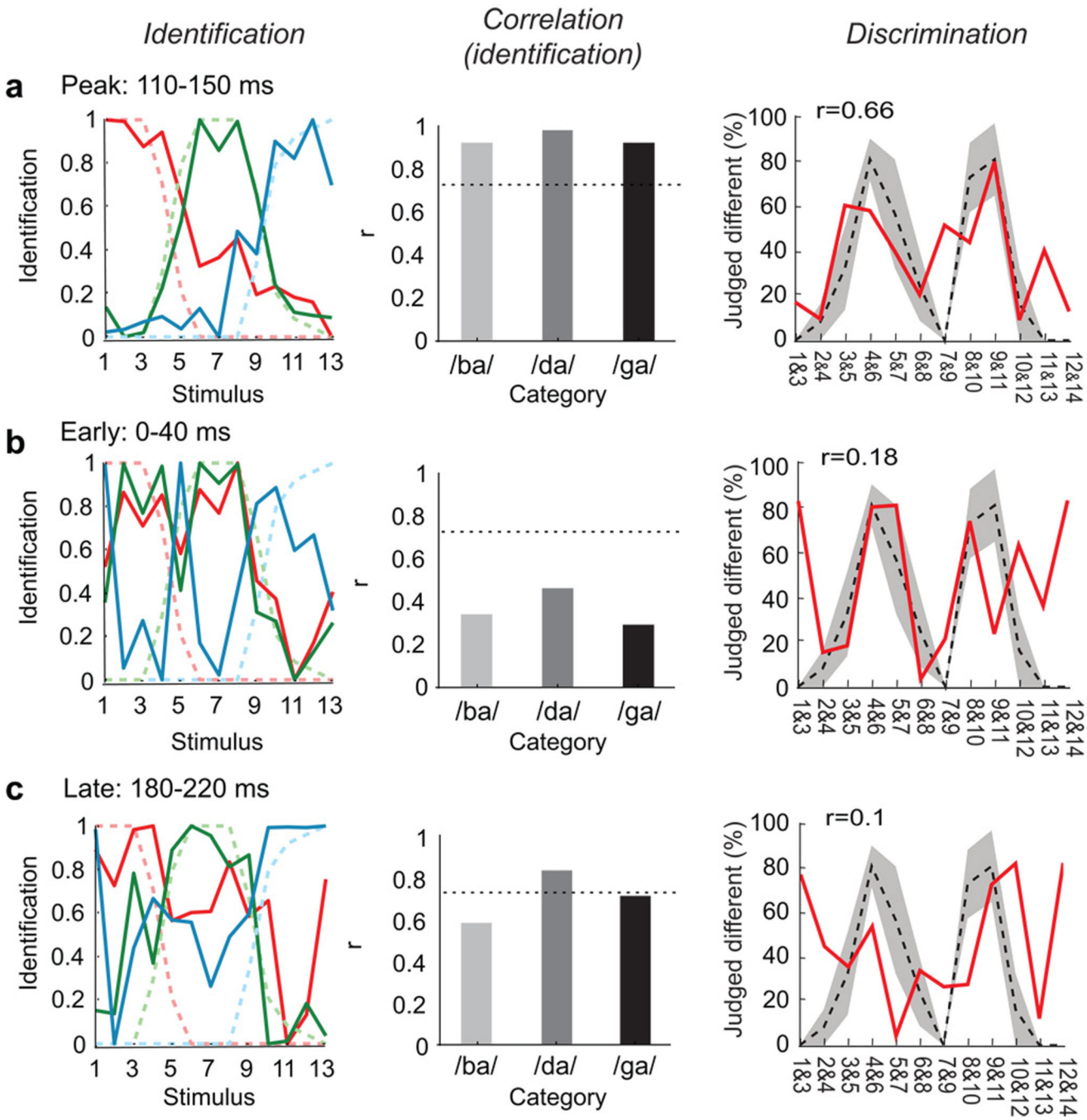

Figure 3. Correlation of neurometric and psychometric category boundaries

Peak encoding at 110–150ms. A. Left, Comparison of neuronal (dark) and psychophysical (light/dashed) -derived identification functions. Neurometric identification functions were determined by 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 correlate well with the psychometric function. ($r=0.66$, $p<0.05$). Early (0–40ms, B) and late (180–220ms, C) epoch field potentials demonstrate poor correlation between neural and psychophysical results (see insets).

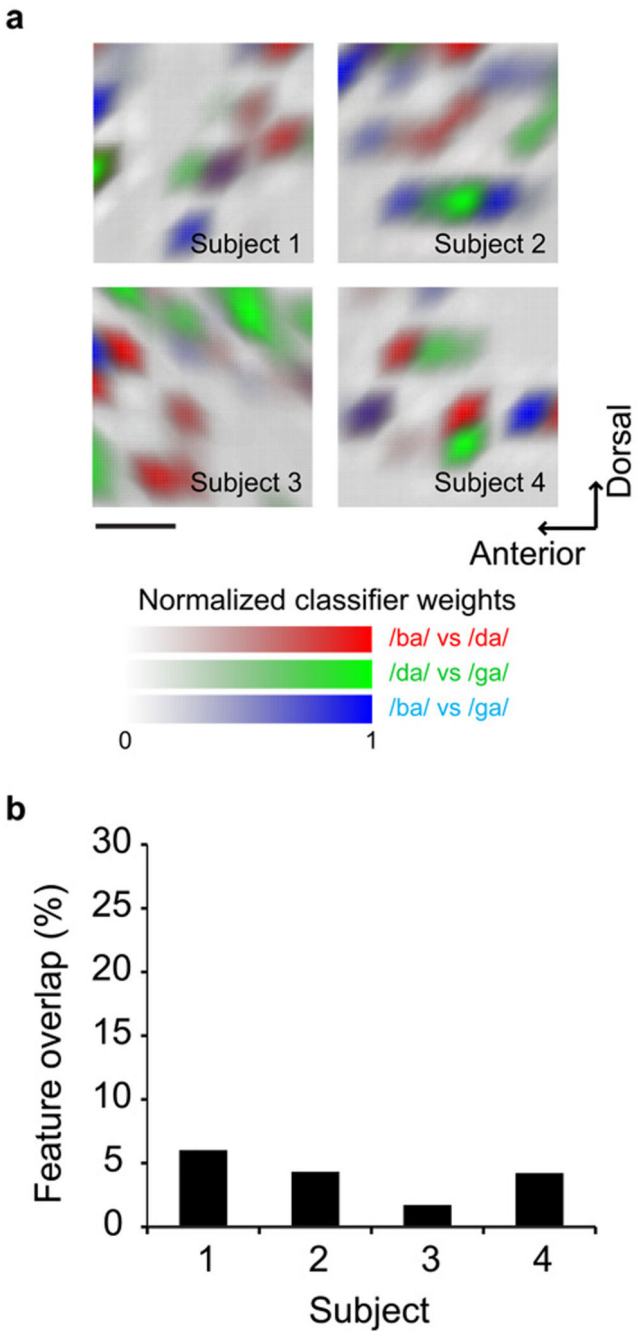

Figure 4. Topography of discriminative cortical sites in the pSTG underlying categorical speech perception

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 is not distributed over the pSTG but instead concentrated in few cortical sites. B. The informative loci overlap very little between comparisons of the features (on average 3.9 +/−0.88%), (indicated by mixed colors such as magenta, cyan, or orange in panel A) suggesting that the neuronal categorization is not accomplished by simply scaling the

responses in the same network but rather is a function of spatially discrete and local selectivity.

Table 1

Patient characteristics

Subject	Age	Gender	Diagnosis	Preoperative language testing	Language dominance
One	47	Male	Left mesial temporal lobe epilepsy	Normal	Left
Two	50	Male	High grade glioma-left frontal cortex	Normal	Left
Three	59	Male	Low grade glioma-right Frontal cortex	Normal	Right
Four	54	Male	Low grade glioma-left frontal cortex	Normal	Left