

Research Article

Inferior Frontal Regions Underlie the Perception of Phonetic Category Invariance

Emily B. Myers,¹ Sheila E. Blumstein,¹ Edward Walsh,² and James Eliassen³

¹Department of Cognitive and Linguistic Sciences, Brown University; ²Department of Neuroscience, Brown University; and

³Department of Psychiatry, University of Cincinnati

ABSTRACT—*The problem of mapping differing sensory stimuli onto a common category is fundamental to human cognition. Listeners perceive stable phonetic categories despite many sources of acoustic variability. What are the neural mechanisms that underlie this perceptual stability? In this functional magnetic resonance imaging study, a short-interval habituation paradigm was used to investigate neural sensitivity to acoustic changes within and between phonetic categories. A region in the left inferior frontal sulcus showed a pattern of activation consistent with phonetic invariance: insensitivity to acoustic changes within a phonetic category and sensitivity to changes between phonetic categories. Left superior temporal regions, in contrast, showed graded sensitivity to both within- and between-category changes. These results suggest that perceptual insensitivity to changes within a phonetic category may arise from decision-related mechanisms in the left prefrontal cortex and add to a growing body of literature suggesting that the inferior prefrontal cortex plays a domain-general role in computing category representations.*

Mapping differing sensory stimuli onto a common category is fundamental to human cognition. For instance, multiple views of a given face are mapped to a common identity, visually distinct objects (e.g., cups) are mapped to the same object category, and acoustically distinct speech tokens are resolved to the same phonetic category. Within each of these domains, there is variability in the sensory input. The challenge for the perceiver is to determine which attributes are relevant to category membership

and which are not, in order to arrive at a stable percept of the category. This many-to-one mapping problem has been termed the *invariance problem*. A core issue in cognitive neuroscience is how the neural system solves this problem.

The subject of the current investigation is the invariance problem in the speech domain. The speech signal contains multiple sources of variability. For instance, the acoustics of a given speech sound vary as a function of the vocal characteristics of the speaker (Peterson & Barney, 1952), the speech rate (J.L. Miller, 1981), and coarticulation effects from adjacent speech sounds (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Despite these sources of variability, listeners perceive a stable phonetic percept.

Results from the neuroimaging literature have shown that the neural systems involved in phonetic processing are sensitive to the acoustic variability inherent in phonetic categories. Both anterior (left inferior frontal gyrus, or IFG) and posterior (left superior temporal gyrus, or STG; left superior temporal sulcus, or STS) structures show sensitivity to acoustic variation within a phonetic category (Blumstein, Myers, & Rissman, 2005; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005), and temporal and temporo-parietal structures (left STS; left middle temporal gyrus; left angular gyrus, or AG; and left supramarginal gyrus, or SMG) show sensitivity to acoustic variations that distinguish phonetic categories (Celsis et al., 1999; Joanisse, Zevin, & McCandliss, 2007; Zevin & McCandliss, 2005). Thus, the neural system shows sensitivity to differences within as well as between phonetic categories. In the study reported here, we sought to identify areas of the brain that are implicated in listeners' "invariant" percepts of phonetic categories in the context of variability in the speech signal.

In the speech literature, four hypotheses have been proposed to explain how the perceptual system solves the invariance problem. Each implicates different neural areas as sources of

Address correspondence to Emily Myers, Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912, e-mail: emily_b_myers@brown.edu.

phonetic category invariance. Although all four hypotheses propose that phonetic categories are abstract, the crucial issue, and the one that is the focus of this article, is how different acoustic inputs are treated as functionally equivalent. One hypothesis proposes that phonetic invariance is acoustically based, with the acoustic input transformed to more generalized spectral-temporal patterns shared by the variants within a phonetic category (Blumstein & Stevens, 1981; Stevens & Blumstein, 1978). If this is the case, invariant neural responses should emerge in the STG or STS (or both), as these areas are involved in processing the acoustic properties of speech (Belin, Zatorre, & Ahad, 2002; Liebenthal et al., 2005; Scott & Johnsrude, 2003). A second hypothesis proposes that invariance is based on a motor or gestural code, with the acoustic input mapped onto motor patterns or gestures used in producing speech (Fowler, 1986; Liberman et al., 1967). If this is the case, invariant neural responses would be expected in motor planning areas, including the pars opercularis (Brodmann's area, BA, 44 of the IFG), the supplementary motor area, ventral premotor areas (BA 6), and possibly primary motor cortex (BA4; Pulvermuller et al., 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004). These two hypotheses share the assumption that the basis of invariance resides in higher-order properties of the input.

The two other alternatives suggest that phonetic constancy emerges at a level abstracted from the input. One hypothesis is that invariant percepts arise from decision mechanisms acting on multiple sources of information (Magnuson & Nusbaum, 2007). In this case, invariant neural responses should emerge in frontal areas involved in executive processing, such as the IFG (Badre & Wagner, 2004; E.K. Miller & Cohen, 2001; Petrides, 2005). Alternatively, invariant percepts might arise from mapping the speech input to higher-order abstract phonological representations. Under this view, invariant neural responses would emerge in parietal areas, such as the left AG or SMG, which have been implicated in phonological processes (Caplan, Gow, & Makris, 1995) and phonological working memory (Paulesu, Frith, & Frackowiak, 1993).

We report a functional magnetic resonance imaging (fMRI) experiment in which we used a short-interval habituation paradigm to investigate neural sensitivity to acoustic changes within and between phonetic categories (Celsis et al., 1999; Joanisse et al., 2007; Zevin & McCandliss, 2005). Stimuli fell along a voice-onset-time (VOT) acoustic-phonetic series ranging from [da] to [ta]. In this paradigm, repeated presentation of a stimulus (e.g., [ta]) results in a reduction, or "adaptation," of the neural response (Grill-Spector & Malach, 2001). The subsequent presentation of a token at a different point along the acoustic series should cause a release from adaptation (and a concomitant increase in the blood-oxygenation-level-dependent, or BOLD, signal) in a neural area if that area is sensitive to the acoustic difference between the repeated token and the change token. The question was whether there are neural areas that fail to show a release from adaptation for within-category

changes while showing a release from adaptation for between-category changes.

METHOD

Participants

Eighteen healthy subjects from the Brown University community (13 females, 5 males) received modest monetary compensation for their participation. Subjects ranged in age from 19 to 29 (mean age = 23.13 ± 3.46). All were strongly right-handed (Oldfield, 1971; mean score = 16 ± 2.42) and underwent a magnetic resonance safety screening. Subjects gave informed consent in accordance with the human-subjects policies of Brown University.

Stimuli

Stimuli were tokens varying in VOT of the initial consonant. They varied in 25-ms VOT steps: -15 and 10 ms VOT (both "da") and 35 and 60 ms VOT (both "ta"). These stimuli were from a larger series created using a parallel synthesizer at Haskins Laboratory. Each 230-ms stimulus contained five formants. The first three formants started at 200 , 1350 , and 3100 Hz and ended at steady-state frequencies of 720 , 1250 , and 2500 Hz, respectively, and transitioned to steady state over the course of 40 ms. The fourth and fifth formants remained steady throughout at 3600 and 4500 Hz. Stimuli had an average fundamental frequency of 119 Hz. High-pitched versions of the target tokens were created by raising the pitch contour of the test stimuli by 100 Hz (Boersma, 2001).

Each trial consisted of five speech tokens separated by 50-ms interstimulus intervals. In *repeat* trials, all five stimuli were identical. In *within-category* trials, four repeated stimuli were followed by a different stimulus from the same phonetic category. In *between-category* trials, four repeated stimuli were followed by a stimulus from a different phonetic category. In the latter two trial types, the VOT difference between the repeated stimuli and the final stimulus was always 25 ms. Repeat, within-category, and between-category trials were distributed equally across four runs (32 trials per condition). In 36 of the trials, one of the stimuli was replaced by its high-pitched version (the *target*). Targets occurred in every stimulus position within a trial and in trials of each of the three experimental types.

Behavioral Procedure and Apparatus

A discrimination pretest was conducted to confirm that subjects were sensitive to the 25-ms VOT steps. Twenty-two subjects performed a discrimination task on pairs of stimuli that were identical, from the same phonetic category, or from two phonetic categories. Participants were asked to listen to each pair of stimuli and press the "same" button if the stimuli sounded identical and the "different" button if they sounded at all different. Subjects were included in the experiment if they ex-

hibited sensitivity to within-category stimuli relative to repeat stimuli. “Sensitivity” was defined as either longer reaction time (RT) for “same” judgments for within-category pairs than for repeat pairs or significantly more “different” judgments for within-category pairs than for repeat pairs. On this basis, 4 subjects were excluded, leaving 18 subjects. Paired t tests ($N = 18$) showed significant differences in error rate (percentage of “different” responses) between within-category pairs ($M = 8.78\%$, $SD = 5.9\%$) and repeat pairs ($M = 3.67\%$, $SD = 4.15\%$), $t(17) = 4.273$, $p_{\text{rep}} = .992$, $d = 1.007$. Despite discriminating within-category pairs below chance, subjects showed sensitivity to within-category contrasts. Between-category pairs were discriminated at near-ceiling rates (mean percentage of “different” responses = 93.8% , $SD = 8.44\%$). The RT difference between within-category pairs ($M = 848.8$ ms, $SD = 134.2$ ms) and repeat pairs ($M = 817.4$ ms, $SD = 127.9$ ms) approached significance, $t(17) = 1.532$, $p_{\text{rep}} = .849$, $d = 0.361$.

During the main experiment, participants listened to each trial while in the scanner and pressed a button when they heard the high-pitched target syllable. Ten practice trials were presented during the anatomical scan.

fMRI Data Acquisition

Functional and anatomical brain images were acquired with a 3-T Siemens Trio scanner. High-resolution three-dimensional T1-weighted anatomical images were acquired for anatomical coregistration (repetition time = 1,900 ms, echo time = 4.15 ms, inversion time = 1,100 ms, 1-mm^3 isotropic voxels, 256×256 matrix). Functional images consisted of fifteen 5-mm-thick echo-planar axial slices with a 3-mm isotropic in-plane resolution, acquired in an ascending, interleaved order. Functional slabs were positioned to image peri-sylvian cortex (repetition time = 1 s, echo time = 30 ms, flip angle = 90° , field of view = 192 mm^3 , 64×64 matrix). Each functional volume was acquired with a 1-s repetition time. Auditory stimuli were pre-

sented during a 2-s silent interval after every 12th echo-planar volume. The first trial began 12 volumes into each run to avoid saturation effects. A single auditory trial occurred in each silent gap (see Fig. 1). A total of 33 trials and 408 functional volumes were collected in each run.

Functional Data Analysis

Image Preprocessing

MR data analysis was performed using AFNI (Cox, 1996). In order to accurately interpret the time course of functional volume acquisition, we inserted two dummy volumes in place of each silent gap. These volumes were censored from further analysis, as were the first two volumes of each block of 12 volumes (because of T1 saturation effects). Functional data sets were corrected for slice acquisition time, and runs were concatenated and motion-corrected using a six-parameter rigid-body transform (Cox & Jesmanowicz, 1999). Functional data sets were resampled to 3 mm^3 , transformed to Talairach and Tournoux space, spatially smoothed with a 6-mm Gaussian kernel, and converted to percentage-signal-change units.

Statistical Analysis

Each subject’s preprocessed functional data were submitted to a regression analysis. A vector was created for each stimulus type (between-category, within-category, repeat, target) by convolving the start time of each stimulus with a gamma-variate function. The target trials were not further analyzed. Each subject’s mean whole-brain time course was included as a nuisance regressor to account for potential effects of signal destabilization due to the clustered acquisition design. The six parameters output by the motion-correction process were also included as nuisance regressors. By-run mean and linear trends were removed from the data. The 3dDeconvolve analysis returned by-voxel fit coefficients for each stimulus condition.

A mixed-factor analysis of variance (ANOVA) was performed on the fit coefficients; subjects was a random factor, and stimulus

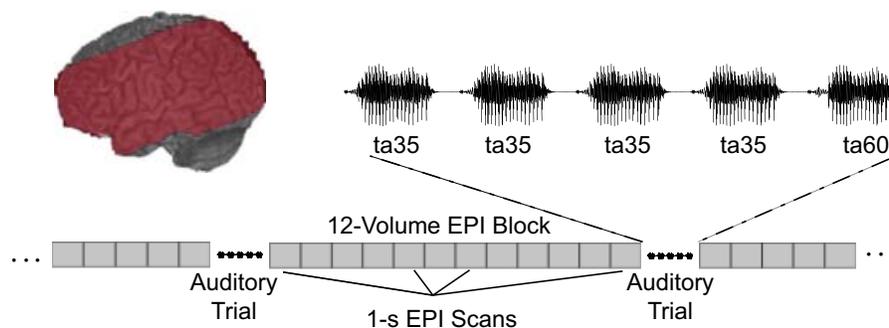


Fig. 1. Diagram of stimulus presentation during functional data acquisition. The red area in the illustration of the brain indicates the extent of echo-planar-imaging (EPI) coverage for a typical subject. Each trial consisted of a five-syllable train presented during the silent gap between successive sets of 12 EPI scans. The trial shown is a within-category trial in which the adapting stimulus is the syllable “ta” with a 35-ms voice-onset time (VOT) for the initial consonant, and the final stimulus is the same syllable with a 60-ms VOT.

condition a fixed factor. First-level analysis involved identifying brain regions that showed significant differences between trial types. To this end, pair-wise comparisons (between-category vs. repeat, between-category vs. within-category, and within-category vs. repeat) were performed. Statistical maps for all comparisons were corrected for multiple comparisons at $p < .05$ (56 contiguous voxels at a voxel-level threshold of $p < .025$). Group statistical maps were displayed on a canonical inflated brain surface in Talairach space (Holmes et al., 1998; *SUMA – AFNI Surface Mapper*, 2006). Within these functionally defined regions of interest, a second-level analysis was performed. Mean percentage signal change for each subject was calculated for each condition, and these data were submitted to paired t tests (Bonferroni-corrected $\alpha = .0167$, critical $t = 2.655$).

RESULTS

Table 1 shows the clusters that emerged in each planned comparison. A significant temporal lobe cluster in the left STG posterior to Heschl’s gyrus and extending into the STS emerged in the between-category/repeat comparison, indicating that this

area showed sensitivity to phonetic-category changes (see Fig. 2 and Table 1). Second-level analysis across the three trial types in this cluster showed significant differences between all conditions (see Fig. 2), with between-category trials resulting in the greatest activation, within-category trials showing less activation, and repeat trials resulting in the least activation—between-category versus repeat: $t(17) = 6.480$, $p_{\text{rep}} = .999$, $d = 1.527$; between-category versus within-category: $t(17) = 2.993$, $p_{\text{rep}} = .908$, $d = 0.706$; within-category versus repeat: $t(17) = 3.053$, $p_{\text{rep}} = .970$, $d = 0.720$. This graded activation pattern suggests that the posterior STG is sensitive not only to between-category differences (Joanisse et al., 2007), but also to within-category differences. No other significant clusters emerged in left temporal areas.

Several frontal clusters emerged in the first-level analysis (see Table 1). The between-category/repeat comparison yielded a large cluster in left inferior frontal regions (Fig. 2). Two separate clusters emerged within this larger inferior frontal region in the between-category/within-category comparison and the within-category/repeat comparison (see Fig. 3 and Table 1). Overlap between these two clusters was minimal (18 voxels). Projection

TABLE 1
Clusters Revealed in Planned Comparisons of Trial Types

Area	No. of voxels	Coordinates of maximum t value			Maximum t	p value
		x	y	z		
Between-category activation > repeat activation						
Left inferior frontal gyrus (pars opercularis, pars triangularis)	465	47	-5	15	6.3410	< .00001
Right inferior frontal gyrus (pars opercularis)	138	-47	-5	6	4.7640	< .0005
Left posterior superior temporal gyrus, middle temporal gyrus	92	62	29	12	5.3400	< .001
Between-category activation < repeat activation						
Right fusiform, right lingual gyrus	62	-29	53	-1	-5.4150	< .025
Between-category activation > within-category activation						
Left inferior frontal gyrus (pars opercularis)	105	41	8	27	4.0090	< .0012
Between-category activation < within-category activation						
Right midorbital gyrus, anterior cingulate cortex	138	-5	-29	-1	-4.2000	< .0005
Within-category activation > repeat activation						
Right inferior frontal gyrus (pars opercularis), right precentral gyrus	149	-47	-5	30	5.1940	< .0003
Right superior temporal gyrus, right supramarginal gyrus	72	-47	32	30	3.7960	< .007
Left insula	60	26	-23	3	3.7750	< .033
Left inferior frontal gyrus (pars opercularis), left precentral gyrus	59	47	-5	21	3.8210	< .036
Within-category activation < repeat activation						
Left fusiform, left inferior occipital cortex	57	38	56	-16	-5.0680	< .046
Right thalamus, putamen	120	-8	20	9	-5.3760	< .0042

Note. All clusters were significant at $p < .05$, corrected (voxel-wise $p < .025$, minimum of 56 contiguous voxels per cluster).

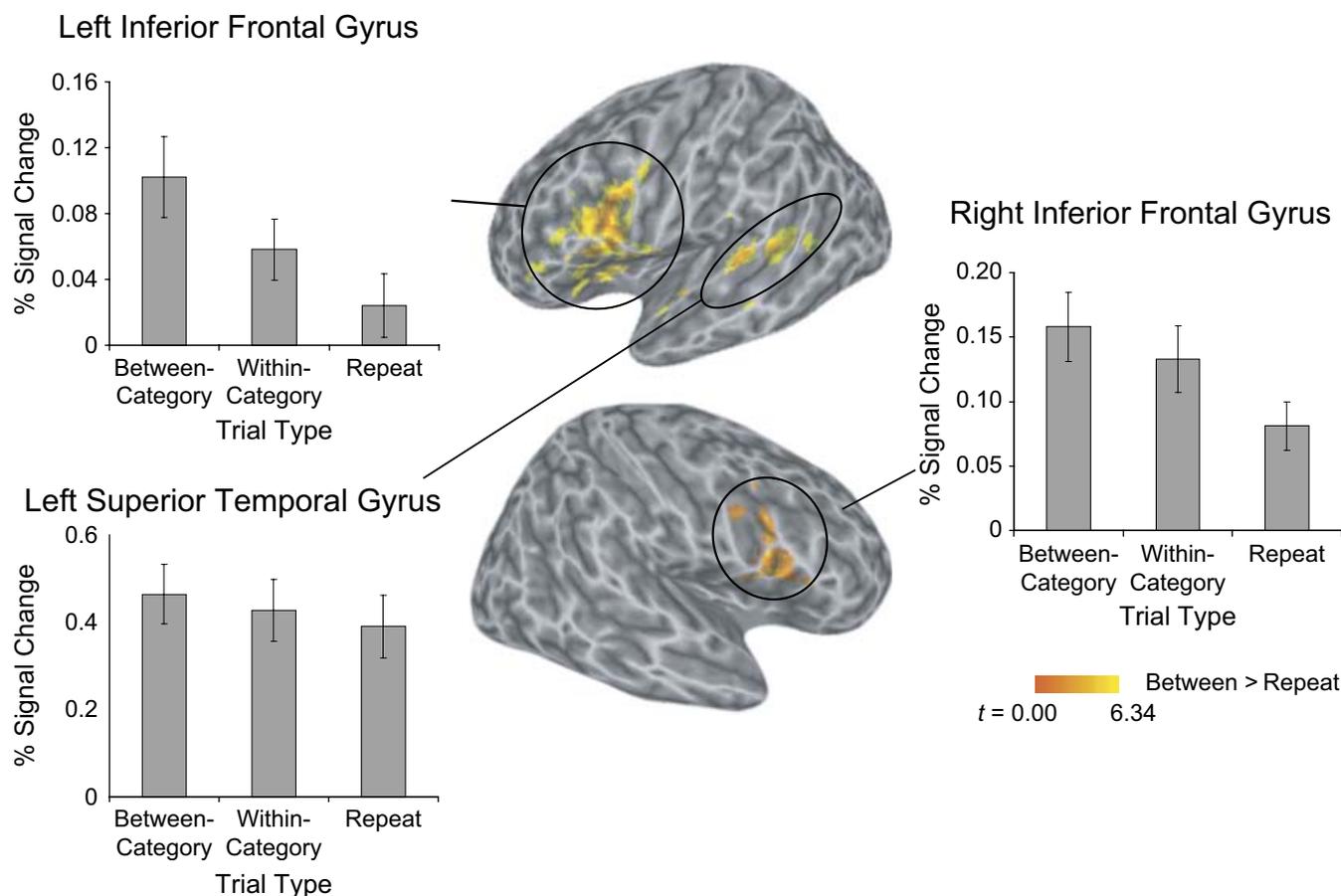


Fig. 2. Areas showing greater activation for between-category than repeat trials. All clusters were significant at a corrected threshold of $p < .05$ ($p < .025$ voxel-wise threshold, minimum cluster size of 56 voxels). The colors indicate the t value at each voxel for the comparison of between-category and repeat trials. Activation is displayed on a canonical inflated brain surface (Holmes et al., 1998), with the left (top) and right (bottom) hemispheres shown separately. The bar graphs present results of second-level analyses, showing the percentage of signal change within each cluster for the three trial types. Error bars indicate standard errors of the means.

of the functional maps onto a canonical inflated brain surface showed that the between-category/within-category cluster fell primarily within the inferior frontal sulcus, and the within-category/repeat cluster lay principally on the lateral surface of the pars opercularis (BA 44; Fig. 3).

Further analysis examined the patterns of activation within these two functional clusters. The between-category/within-category cluster located in the inferior frontal sulcus contained voxels showing greater activation to between-category changes than to within-category changes. To determine if the pattern of activation in this cluster reflected phonetic category invariance, we examined potential differences in activation between within-category and repeat trials. Results revealed a pattern consistent with phonetic category invariance: Between-category trials showed significantly greater activation than either within-category or repeat trials—between-category versus within-category: $t(17) = 3.522$, $p_{\text{rep}} = .983$, $d = 0.830$; between-category versus repeat: $t(17) = 5.127$, $p_{\text{rep}} = .997$, $d = 1.209$; however, activation did not differ significantly between within-category trials and repeat trials, $t(17) = 1.877$. Thus, this area showed a

release from adaptation only for those trials in which there was a change in phonetic category membership.

The cluster that emerged in the within-category/repeat comparison showed sensitivity to acoustic changes within a phonetic category. To determine whether this area also showed greater activation for between-category changes than for within-category changes, we conducted further comparisons. Both between-category and within-category trials showed significantly more activation than repeat trials—between-category versus repeat: $t(17) = 4.913$, $p_{\text{rep}} = .996$, $d = 1.158$; within-category versus repeat: $t(17) = 3.991$, $p_{\text{rep}} = .990$, $d = 0.941$; however, between-category and within-category trials did not differ significantly in activation, $t(17) = 2.071$. Thus, this area in BA 44 showed sensitivity to acoustic changes irrespective of their phonetic relevance.

In addition to the left-hemisphere clusters, clusters emerged in the right STG and the right IFG. The right STG cluster emerged in the within-category/repeat comparison, but showed no significant difference between phonetically relevant versus nonrelevant changes (see Fig. 3)—between-category versus

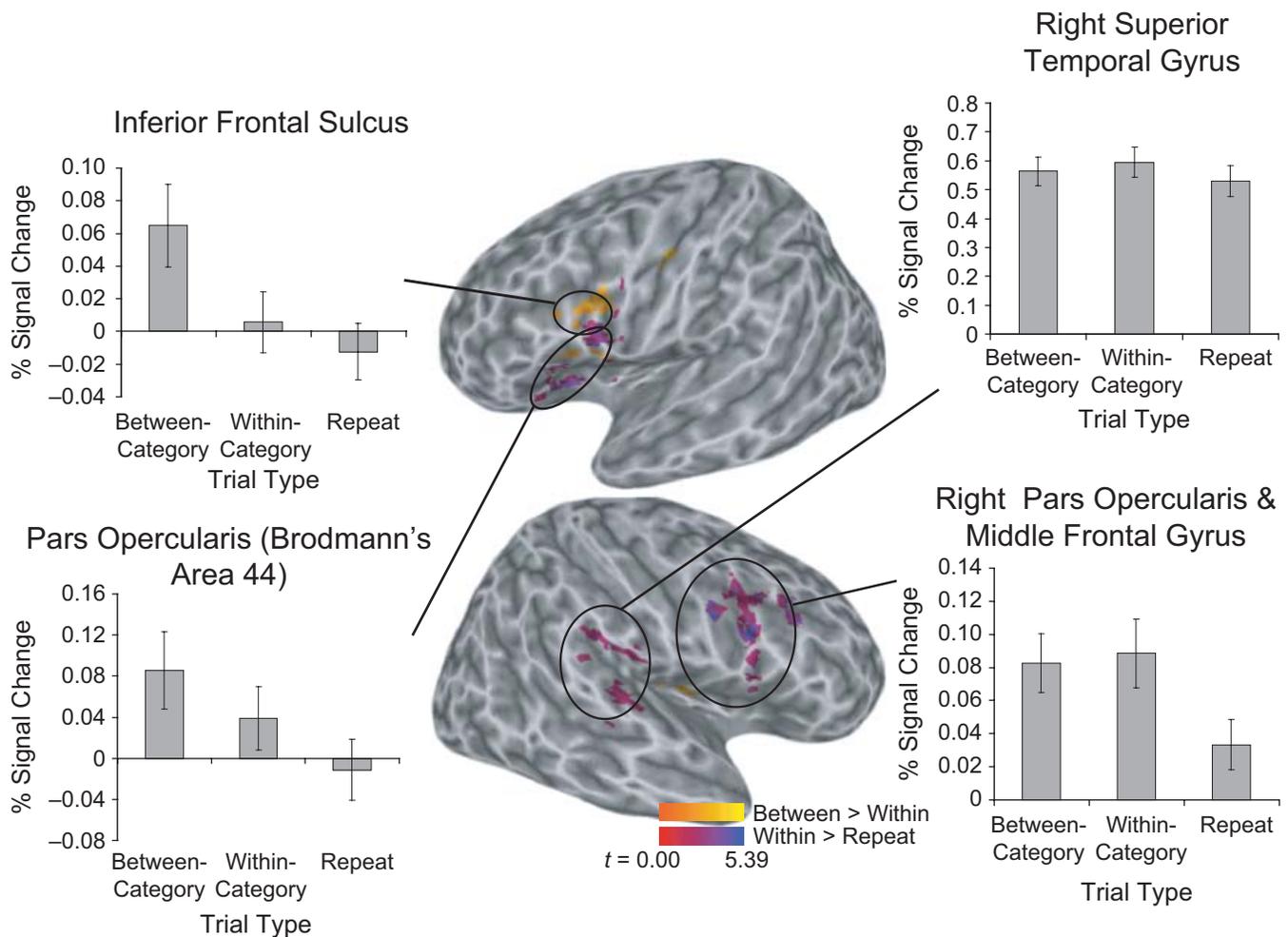


Fig. 3. Areas showing greater activation for between-category than within-category trials and greater activation for within-category than repeat trials. All clusters were significant at a corrected threshold of $p < .05$ ($p < .025$ voxel-wise threshold, minimum cluster size of 56 voxels). The colors indicate the t value at each voxel for the indicated comparison. Activation is displayed on a canonical inflated brain surface (Holmes et al., 1998), with the left (top) and right (bottom) hemispheres shown separately. The bar graphs present results of second-level analyses, showing the percentage of signal change within each cluster for the three trial types. Error bars indicate standard errors of the means.

within-category: $t(17) = 2.643$. Two clusters emerged in the right IFG (pars opercularis; see Table 1, Figs. 2 and 3), one in the between-category/repeat comparison and the other in the within-category/repeat comparison. Examination of the patterns of activation within these clusters revealed significant differences as a function of acoustic changes irrespective of their phonetic relevance. There was significantly greater activation in both between-category trials and within-category trials compared with repeat trials—between-category/repeat cluster: $t(17) \geq 2.881$, $p_{\text{rep}} \geq .964$, $d \geq 0.679$; within-category/repeat cluster: $t(17) \geq 4.534$, $p_{\text{rep}} \geq .994$, $d \geq 1.069$. However, no significant differences emerged in either cluster in the comparison of between-category and within-category trials, $t_s(17) \leq 1.175$.

DISCUSSION

The goal of the current study was to investigate the neural correlates of phonetic category invariance. A pattern of activation

consistent with phonetic invariance emerged in the left inferior frontal sulcus. This cluster showed release from adaptation for stimuli that crossed the phonetic category boundary and no release from adaptation for stimuli drawn from the same phonetic category. This pattern of response suggests that this neural area is involved in one's perceptual experience of functional equivalence for different sensory inputs.

Supporting evidence for these findings comes from a study by Hasson, Skipper, Nusbaum, and Small (2007). In this study, subjects were exposed to stimuli containing conflicting auditory and visual cues. Such stimuli are subject to the McGurk effect (McGurk & MacDonald, 1976), in which the presentation of a visual stimulus of a speaker saying “ka,” coupled with an auditory stimulus of the speaker saying “pa,” is typically perceived as “ta.” Hasson et al. showed that when the McGurk stimulus was preceded by an auditory exemplar of “ta,” repetition suppression was seen in the left pars opercularis, and this suppression was equivalent to that observed when this stimulus was

preceded by itself. Note that there was no sensory overlap between the adapting and target stimulus in either the auditory or the visual domain. These results suggest that invariance emerges in inferior frontal cortex as a function of perceptual rather than sensory overlap (i.e., the stimuli are perceived as belonging to a common category “t”).

A role for the inferior frontal lobes in computing categorical representations has been suggested by work in nonhuman primates (Freedman, Riesenhuber, Poggio, & Miller, 2001, 2003). This work has implicated inferior frontal areas in computing decisions necessary for action (Petrides, 2005). For example, using single-cell recordings, Freedman et al. (2001) showed invariant responses to exemplars from a learned visual category in lateral prefrontal cortex of monkeys. Freedman et al. also found individual cells in the same region that were sensitive to within-category variation. Similarly, in the current study, posterior prefrontal cortex also showed functional heterogeneity in its responsiveness to category representations; within this region, a cluster in the inferior frontal sulcus failed to show sensitivity to within-category variation and hence showed phonetic category invariance, and another cluster in the pars opercularis, BA 44, showed sensitivity to within-category variation. The fact that results for speech categories in humans are analogous to those for learned visual categories in nonhuman primates suggests that this region plays a domain-general role in computing category representations.

In contrast to the frontal clusters, the left STG cluster showed graded sensitivity to phonetic category membership, showing a significant release from adaptation for within-category differences and an even greater release from adaptation for between-category differences. At first glance, these findings appear to be at odds with those of Joanisse et al. (2007), who showed that the left STS was sensitive to phonetic differences between but not within categories. Several important differences between that study and ours might account for the divergent results. First, it is unclear from the behavioral data reported by Joanisse et al. whether their subjects showed perceptual sensitivity to within-category stimuli. Failure to find any cortical regions that showed sensitivity to within-category changes could reflect subjects' inability to perceptually resolve differences among these stimuli. In contrast, in the current study, subjects showed perceptual sensitivity to within-category stimuli, a fact reflected by a significant release from adaptation for these stimuli in the STG, and confirmed by a behavioral pretest.

Second, because the focus of their study was the preattentive processing of phonetic category information, Joanisse et al. (2007) did not require their subjects to attend to the speech stimuli or even to the auditory stream. Rather, subjects watched a subtitled movie while the speech stimuli were being presented. This lack of attention to the auditory input may have attenuated responses to phonetic category information. Consistent with this view, a recent study using a bimodal (auditory and visual) selective-attention task (Sabri et al.,

2008) showed reduced activation to speech and nonspeech stimuli in the STG when subjects performed a demanding visual task and were not required to attend to the auditory stream. In the current study, although participants did not have to explicitly process the speech stimuli, they were required to attend to the auditory stream in order to perform the low-level pitch-detection task. Differences between the findings of these two studies highlight the importance of attention to the auditory stream in processing the acoustic-phonetic details of speech.

The fact that patterns of activation consistent with phonetic invariance were seen in frontal areas, but not in areas involved in acoustic processing (i.e., temporal areas), suggests that perceptual invariance for speech categories does not arise through sets of shared acoustic patterns (Stevens & Blumstein, 1981). The failure to find patterns of activation consistent with phonetic invariance in motor areas involved in articulatory implementation (i.e., in BA 4) is evidence against gestural theories of phonetic invariance (Fowler, 1986). Taken together, the failure to find patterns of activation consistent with phonetic invariance in either the STG or motor areas suggests that phonetic invariance does not arise from invariant acoustic or motor properties, but instead arises from higher-order computations on that input.

A number of recent studies have implicated temporo-parietal areas, such as the SMG and AG, in phonetic category processing (Blumstein et al., 2005; Caplan et al., 1995; Hasson et al., 2007; Raizada & Poldrack, 2007; Zevin & McCandliss, 2005). In addition, there appears to be a tight link between the IFG and SMG in phonological processing (Gold & Buckner, 2002). In view of these results, it is perhaps surprising that no clusters emerged in the SMG in the current study. To explore the possibility that temporo-parietal areas play a role in phonetic category invariance, we further analyzed activation in the SMG and AG. At a much reduced threshold ($p < .025$ voxel-level threshold, 25 contiguous voxels), a cluster did emerge in the SMG for the between-category/within-category comparison (peak: $x = 62, y = 29, z = 27$). This cluster showed no release from adaptation for within-category stimuli and a release from adaptation for stimuli that were from two different categories. Although these results should be interpreted with caution, such a pattern suggests that there may be dual routes to phonetic invariance: one that emerges through probabilistic decisions on graded acoustic data and one that involves a mapping to abstract phonological codes. Alternatively, phonetic category invariance may arise through the interaction of decision-related mechanisms in the frontal lobes with a phonological code in the SMG (Gold & Buckner, 2002).

Much research suggests that one of the basic functions of the prefrontal cortex is to facilitate goal-directed action, and hence this area plays a critical role in mediating the transformation of perception into action (Freedman, Riesenhuber, Poggio, & Miller, 2002; E.K. Miller & Cohen, 2001). The ability to group a set of stimuli into categories facilitates this transformation by

providing a means of segmenting the world into meaningful units so that one can ultimately act on them. In the case of speech, phonetic categorization provides the basic building blocks used for communication. Nonetheless, the data from the current study indicate that category membership may be computed even when not required by the task. In particular, phonetic invariance was shown in a task that required subjects to attend to the auditory stream, but not to make an overt decision about the phonetic category to which a stimulus belonged. The implicit nature of phonetic categorization has been shown even in infants, who clearly were not responding to the stimuli with an explicit goal (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Thus, it appears that an explicit categorization or decision about speech stimuli is unnecessary for categorical-like neural responses to them. It would be of interest to determine whether similar categorical-like neural responses arise in the implicit processing of learned nonlanguage categories, or whether such responses have evolved as part of the biological substrates of language because of its functional importance for humans.

Acknowledgments—This research was supported by Grant RO1 DC006220 from the National Institutes of Health (NIH), National Institute on Deafness and Other Communication Disorders (NIDCD), and by the Ittleson Foundation. The content is the responsibility of the authors and does not necessarily represent official views of NIH or NIDCD.

REFERENCES

- Badre, D., & Wagner, A.D. (2004). Selection, integration, and conflict monitoring: Assessing the nature and generality of prefrontal cognitive control mechanisms. *Neuron*, *41*, 473–487.
- Belin, P., Zatorre, R.J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*, 17–26.
- Blumstein, S.E., Myers, E.B., & Rissman, J. (2005). The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience*, *17*, 1353–1366.
- Blumstein, S.E., & Stevens, K.N. (1981). Phonetic features and acoustic invariance in speech. *Cognition*, *10*, 25–32.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, *5*, 341–345.
- Caplan, D., Gow, D., & Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology*, *45*, 293–298.
- Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J.P., Berry, I., Nespoulous, J.L., & Chollet, F. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *NeuroImage*, *9*, 135–144.
- Cox, R.W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–173.
- Cox, R.W., & Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magnetic Resonance in Medicine*, *42*, 1014–1018.
- Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306.
- Fowler, C.A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.
- Freedman, D.J., Riesenhuber, M., Poggio, T., & Miller, E.K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*, 312–316.
- Freedman, D.J., Riesenhuber, M., Poggio, T., & Miller, E.K. (2002). Visual categorization and the primate prefrontal cortex: Neurophysiology and behavior. *Journal of Neurophysiology*, *88*, 929–941.
- Freedman, D.J., Riesenhuber, M., Poggio, T., & Miller, E.K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience*, *23*, 5235–5246.
- Gold, B.T., & Buckner, R.L. (2002). Common prefrontal regions co-activate with dissociable posterior regions during controlled semantic and phonological tasks. *Neuron*, *35*, 803–812.
- Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, *107*, 293–321.
- Hasson, U., Skipper, J.I., Nusbaum, H.C., & Small, S.L. (2007). Abstract coding of audiovisual speech: Beyond sensory representation. *Neuron*, *56*, 1116–1126.
- Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W., & Evans, A.C. (1998). Enhancement of MR images using registration for signal averaging. *Journal of Computer Assisted Tomography*, *22*, 324–333.
- Joanisse, M.F., Zevin, J.D., & McCandliss, B.D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cerebral Cortex*, *17*, 2084–2093.
- Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461.
- Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., & Medler, D.A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, *15*, 1621–1631.
- Magnuson, J.S., & Nusbaum, H.C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 391–409.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.
- Miller, E.K., & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.
- Miller, J.L. (1981). Some effects of speaking rate on phonetic perception. *Phonetica*, *38*, 159–180.
- Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.
- Paulesu, E., Frith, C.D., & Frackowiak, R.S.J. (1993). The neural correlates of the verbal component of working memory. *Nature*, *362*, 342–345.
- Peterson, G.E., & Barney, H.L. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America*, *24*, 175.
- Petrides, M. (2005). Lateral prefrontal cortex: Architectonic and functional organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*, 781–795.
- Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, USA*, *103*, 7865–7870.

- Raizada, R.D., & Poldrack, R.A. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, *56*, 726–740.
- Sabri, M., Binder, J.R., Desai, R., Medler, D.A., Leitl, M.D., & Liebenthal, E. (2008). Attentional and linguistic interactions in speech perception. *NeuroImage*, *39*, 1444–1456.
- Scott, S.K., & Johnsrude, I.S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26*, 100–107.
- Stevens, K.N., & Blumstein, S.E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *64*, 1358–1368.
- Stevens, K.N., & Blumstein, S.E. (1981). The search for invariant acoustic correlates of phonetic features. In P.D. Eimas & J.L. Miller (Eds.), *Perspectives on the study of speech* (pp. 1–38). Hillsdale, NJ: Erlbaum.
- SUMA – AFNI surface mapper. (2006). Retrieved September 1, 2008, from AFNI Web site: <http://afni.nimh.nih.gov/afni/suma>
- Wilson, S.M., Saygin, A.P., Sereno, M.I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*, 701–702.
- Zevin, J.D., & McCandliss, B.D. (2005). Dishabituation of the BOLD response to speech sounds. *Behavioral and Brain Functions*, *1*, 4.

(RECEIVED 8/1/08; REVISION ACCEPTED 11/27/08)