Inferior Frontal Regions Underlie the Perception of Phonetic Category Invariance

Emily B. Myers[1], **Sheila E. Blumstein**[1], **Edward Walsh**[2], and **James Eliassen**[3]

[1] Brown University, Department of Cognitive and Linguistic Sciences

[2] Brown University, Department of Neuroscience

[3] University of Cincinnati, Department of Psychiatry

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. At issue is identifying those neural mechanisms underlying this perceptual stability. A short-interval habituation fMRI paradigm was used to investigate neural sensitivity to within and between phonetic category acoustic changes. A region in the left inferior frontal sulcus showed an invariant pattern of activation: insensitivity to acoustic changes within a phonetic category in the context of sensitivity to changes between phonetic categories. Left superior temporal regions, in contrast, showed graded sensitivity to both within- and between-phonetic 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.

1. Introduction

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 such as 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 different speakers (Peterson & Barney, 1952), speech rate (Miller, 1981), and coarticulation effects from adjacent speech sounds (Liberman et al., 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, [IFG]) and posterior (left superior temporal gyrus, [STG], and superior temporal sulcus, [STS]) structures show sensitivity to acoustic variation within a phonetic category (Blumstein et al., 2005; Liebenthal et al., 2005), and temporo-parietal structures (left STS, left middle temporal gyrus [MTG], left angular gyrus [AG], and left supramarginal gyrus [SMG]) show sensitivity to acoustic variations that distinguish phonetic categories (Celsis et al., 1999; Joanisse et al., 2006; Zevin & McCandliss, 2005). Thus, the neural system shows sensitivity to differences within as well as between phonetic categories. At issue is identifying areas of the brain which reflect listeners' 'invariant' percept of a phonetic category despite variability in the signal.

Within the speech literature, four hypotheses have been proposed to address how the perceptual system solves the invariance problem, each implicating different candidate neural areas as sources of phonetic category invariance. While each of these theories proposes that phonetic categories are abstract, the crucial issue and the one that is the focus of this paper is determining how different acoustic inputs are treated as functionally equivalent. The first 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 and/or STS, areas involved in processing the acoustic properties of speech (Belin et al., 2002; Liebenthal et al., 2005; Scott & Johnsrude, 2003). A second hypothesis proposes that invariance is motor or gesturally-based 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 (BA44 of the IFG), the supplementary motor area (SMA), as well as ventral premotor areas (BA6), and possibly primary motor cortex (BA4) (Pulvermuller et al., 2006; Wilson et al., 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 upon 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; 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 et al., 1995) and phonological working memory (Paulesu et al., 1993).

In the current experiment, a short interval habituation (SIH) paradigm was utilized to investigate neural sensitivity to within and between phonetic category acoustic changes along a voice-onset time (VOT) acoustic-phonetic series ranging from [da] to [ta] (Celsis et al., 1999; Joanisse et al., 2006; Zevin & McCandliss, 2005). 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 BOLD signal) if that neural area is sensitive to the acoustic difference between the repeated token and the change token. The question is 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 acoustic changes.

Methods

Participants

Eighteen healthy subjects from the Brown University community (13 females) 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 were screened for MR compatibility. Subjects gave informed consent in accordance with the Human Subjects policies of Brown University.

Stimuli

Stimuli were tokens varying in 25-ms VOT steps, i.e. −15 and 10 (both 'da'), and 35 and 60 (both 'ta') ms VOT. These stimuli were from a larger series created using a parallel synthesizer at Haskins Laboratory. Each 230-ms stimulus contained five formants. Formant frequencies were 40-ms starting at 200 Hz (F1), 1350 Hz (F2), and 3100 Hz (F3) and ending at steady-state frequencies of 720, 1250, and 2500 Hz, respectively. F4 and F5 remained steady throughout at 3600 and 4500 Hz. Stimuli had an average fundamental frequency of 119 Hz. High-pitched 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 a 50-ms ISI in which: all stimuli were identical (Repeat trials), four repeated stimuli were followed by a different stimulus from the same phonetic category (Within trials) or from a different phonetic category (Between trials). The VOT interval between stimuli in both Within-category and Between-category trials was always 25-ms. Repeat, Within, and Between conditions were distributed equally across four runs (32 trials per condition). There were 36 Target trials in which one of the stimuli in the trial was replaced by its high-pitched version. Target stimuli occurred in every position within a trial and in trials of each experimental type.

Behavioral Procedure and Apparatus

A discrimination pretest was conducted to confirm that subjects were sensitive to the 25-ms VOT steps. Twenty-two subjects discriminated pairs of stimuli that were either identical (Repeat), within the same phonetic category (Within) or from two phonetic categories (Between). Subjects were included if they exhibited sensitivity to Within compared to Repeat stimuli. "Sensitivity" was defined as either longer reaction time (RT) for Within pairs than Repeat pairs or more correct "different" judgments for Within pairs than Repeat pairs. On this basis, four subjects were excluded, leaving 18 subjects. Paired t-tests (N = 18) showed significant differences in error rate between Within-category and Repeat pairs: (mean percent 'different' responses: Within=8.78%, SD=5.9%; Repeat =3.67%, SD=4.15%; $t(17)=4.273$, p_{rep} =0.992, 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 percent 'different' responses =93.8%, SD=8.44%). RT differences between Within and Repeat approached significance (Within mean RT=848.8 ms SD=134.2 ms, Repeat mean RT= 817.4 ms SD=127.9 ms; $t(17)=1.532$, p_{rep} =0.849, d=0.361,)

Participants in the scanner listened to each trial, 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 3T Siemens Trio scanner. High-resolution 3D T1-weighted anatomical images were acquired for anatomical co-registration (TR=1900 ms, TE=4.15 ms, TI=1100 ms, 1mm^3 isotropic voxels, 256 × 256

matrix). Functional images consisted of 15 5mm-thick echo planar (EPI) axial slices with a 3mm isotropic in-plane resolution, acquired in an ascending, interleaved order. Functional slabs were positioned to image peri-sylvian cortex (TR=1 sec, TE=30 ms, flip angle = 90 degrees, FOV=192mm3, 64×64 matrix). Each functional volume was acquired with a 1-second TR. Auditory stimuli were presented during a 2-second silent interval after every 12th EPI volume. The first trial began 12 volumes into each run to avoid saturation effects. An auditory trial occurred in each successive silent gap (see Figure 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, two dummy volumes were inserted 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 due to T1 saturation effects. Functional datasets were corrected for slice acquisition time, and runs were concatenated and motion-corrected using a six-parameter rigid-body transform (Cox & Jesmanowicz, 1999). Functional datasets were resampled to 3mm^3, transformed to Talairach and Tournoux space, spatially smoothed with a 6mm Gaussian kernel, and converted to percent signal change units.

Statistical Analysis—Each subject's preprocessed functional data was submitted to a regression analysis by creating a vector for each stimulus type (Between, Within, Repeat, Target) in which the start time of each stimulus was convolved 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 condition.

A mixed-factor ANOVA was performed on the fit coefficients with subjects as a random factor and stimulus condition as a fixed factor. Planned comparisons were performed: Between vs. Repeat, Between vs. Within, and Within vs. Repeat. Statistical maps for each comparison were thresholded at $p<0.05$, corrected for multiple comparisons (56 contiguous voxels at a $p<0.025$). Within functionally-defined regions of interest the mean percent signal change for each subject was calculated for each condition, and these data were submitted to paired t-tests (Bonferroni-corrected alpha=0.0167, critical t=2.655). Group statistical maps were displayed on a canonical inflated brain surface in Talairach space (Holmes et al., 1998) (http://afni.nimh.nih.gov/afni/suma).

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 vs. Repeat condition indicating that this area showed sensitivity to phonetic category changes (see Figure 2, Table 1). Analysis of the patterns of activation across the three trial types in this cluster showed significant differences between all conditions (see Figure 2), with Between-category trials resulting in the greatest activation, Within-category trials with less activation, and Repeat trials with the least activation (Between-Repeat, $t(17)=6.480$, $p_{rep}=0.999$, $d=1.527$; Between-Within, $t(17)=2.993$, $p_{rep}=0.908$, $d=0.706$; Within-Repeat, $t(17)=3.053$, $p_{rep}=0.970$, $d=0.720$). This graded activation pattern suggests that the pSTG is not only sensitive to between category differences (Joanisse et al., 2006)

but also to within category differences. No other significant clusters emerged in the temporal areas.

Several frontal clusters emerged in the comparisons across conditions (see Table 1). The Between vs. Repeat comparison yielded a large cluster in left inferior frontal regions (Figure 2). Two separate clusters emerged within this larger inferior frontal region in the comparisons of the Between vs. Within and Within vs. Repeat conditions (see Figure 3, Table 1). Overlap between these two clusters was minimal (18 voxels). Projection of the functional maps on to a canonical inflated brain surface showed that the Between vs. Within cluster fell primarily within the inferior frontal sulcus and the Within vs. Repeat cluster lay principally on the lateral surface of the pars opercularis (BA44, Figure 3).

Further analysis examined the patterns of activation within these two functional clusters. The Between vs. Within cluster located in the inferior frontal sulcus contained voxels showing greater activation to between phonetic category changes than to within phonetic category changes. To determine if the pattern of activation in this cluster reflected phonetic category invariance, potential differences in activation between Within-category and Repeat trials were examined. Results revealed a pattern consistent with phonetic category invariance; Between-category trials showed significantly greater activation than either the Within-category or Repeat trials (Between-Within, $t(17)=3.522$, $p_{rep} =0.983$, $d=0.830$; Between-Repeat, $t(17)=5.127$, $p_{rep} = 0.997$, $d=1.209$), but Within-category trials were not significantly different from 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 vs. 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 within category changes, comparisons were made between the stimulus conditions. Both Between-category and Within-category trials showed significantly more activation than Repeat trials (Between-Repeat $t(17)=4.913$, $p_{rep} =0.996$, $d=1.158$; Within-Repeat $t(17)=3.991$, $p_{rep} =0.990$, $d=0.941$); however, Between-category and Within-category trials were not significantly different from each other ($t(17)=2.071$). 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 vs. Repeat comparison, but showed no significant difference between phonetically-relevant vs. non-relevant changes (Between vs. Within, $t(17)=2.643$). Two clusters emerged in the right IFG (pars opercularis, see Table 1, Figures 1 and 2), one in the Between vs. Repeat comparison and the other in the Within vs. Repeat comparison. Examination of the patterns of activation within these clusters revealed significant differences only 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 to Repeat trials (Between-Repeat cluster $t(17) \geq 2.881$, prep≥ 0.964, $d \geq 0.679$; Within-Repeat cluster $t(17) \geq 4.534$, prep≥ 0.994 $d \geq 1.069$). However, no significant differences emerged in either cluster in the Between-category compared to the Within-category trials ($t(17) \leq 1.175$).

Discussion

Role of inferior frontal areas in computing phonetic category invariance

The goal of the current study was to investigate the neural correlates of phonetic category invariance. An invariant pattern of activation emerged in the left inferior frontal sulcus. This

cluster showed release from adaptation for stimuli which 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 subjects' perceptual experience of functional equivalence for different sensory inputs consistent with category invariance.

Supporting evidence for these findings comes from a study by Hasson and colleagues (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 examplar of 'ta,' repetition suppression was seen in the left pars opercularis which was equivalent to that observed when this stimulus was preceded by itself. Of importance, there was no sensory overlap in either the auditory or visual domain between the adapting and target stimulus. These results suggest that invariance emerges in inferior frontal cortex as a function of perceptual rather than sensory overlap, i.e. between stimuli that 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 non-human primates (Freedman et al., 2001, 2003). This work has implicated inferior frontal areas in computing decisions necessary for action (Petrides, 2005). For example, using single-cell recordings, Freedman and colleagues (2001) showed invariant responses in lateral pre-frontal cortex of monkeys to exemplars from a learned visual category. Of interest, Freedman et al. also showed 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 phonetic category variation. The fact that analogous results are shown for speech categories in humans as well as learned visual categories in non-human 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. This finding would appear to be at odds with those of Joanisse et al (2006) who showed that the left STS was sensitive to between phonetic category differences but not to within phonetic category differences. Several important differences between the Joanisse et al. study and the current one might account for these differences. First, it is unclear from the behavioral data in Joanisse et al. whether their subjects showed perceptual sensitivity to within phonetic category stimuli. A failure to show 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 phonetic category stimuli, a fact which is reflected by a significant release from adaptation for these stimuli in the STG, and confirmed by a behavioral pre-test. Second, because the focus of their study was the pre-attentive processing of phonetic category information, Joanisse and colleagues 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, 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 invariant patterns of activation 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 a set of shared acoustic patterns (Stevens & Blumstein, 1981), but instead emerges through higher-order computations on graded input, presumably from temporal areas. The failure to find invariant patterns of activation in primary motor areas involved in the articulatory implementation (i.e. BA4, M1) argues against gestural theories of phonetic invariance (Fowler, 1986). Taken together, the failure to show invariant patterns of activation in either the STG or motor areas suggests that phonetic invariance does not arise from invariant acoustic/motor properties but instead arises from higher order computations on that input.

Nonetheless, while the results of the current study suggests that perceptual invariance does not rely on activation of a motor code, two clusters were activated in left inferior frontal areas (inferior frontal sulcus/precentral gyrus and pars opercularis) which have been suggested to contain mirror neurons (Craighero et al., 2007; Rizzolatti & Craighero, 2004). Some have speculated that mirror neurons, by providing a neural substrate for modeling the motor gestures necessary for speech, also provide a route to perceptual stability of phonetic category percepts across multiple types of variance. This claim cannot be either confirmed or ruled out by the results of the present study. However, pertinent to this question is recent research suggesting that BA6 and not BA44 (pars opercularis) is most similar to F5 of the macaque brain where mirror neurons have been found. In particular, BA6 responds consistently to goal-directed actions whereas BA44 does not (Morin & Grezes, 2008). As such, the cluster falling in the inferior frontal sulcus/precentral gyrus which showed an "invariant" pattern of activation is more likely to reflect the existence of putative mirror neuron activity.

A number of recent studies have implicated temporoparietal 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. In order to explore the possibility that temporoparietal areas play a potential role in phonetic category invariance, further analysis was undertaken to investigate activation in the SMG and AG. At a much reduced threshold ($p < 0.025$ voxel-level threshold, 25 contiguous voxels), a cluster did emerge in the SMG for the Between-Category vs. Within comparison (peak, $x = -62$, $y = -29$, $z = 27$). This cluster showed an invariant pattern with no release from adaptation for within category stimuli and a release from adaptation for those stimuli which were from two different categories. While these results should be interpreted with caution, such a pattern suggests that there may be dual routes to phonetic invariance: one that emerges due to probabilistic decisions on graded acoustic data, and one which 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 in the service of goals and hence this area plays a critical role in mediating the transformation of perception into action (Freedman et al., 2002; 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 and ultimately acting on them. In the case of speech, phonetic categorization provides the basic building blocks used for language communication. Nonetheless, the data from the current study indicates that category membership may be computed even when not required by the task. In particular, phonetic invariance was shown in the current experiment in a task which 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 are not responding to the stimuli in the service of an explicit goal (Eimas et al., 1971). Thus, it appears that an explicit categorization or decision about speech stimuli is unnecessary for categorical-like neural responses in the context of speech sounds. What is of interest is whether similar categorical-like neural responses would be found in the implicit processing of learned non-language categories or whether such have evolved as part of the biological substrates of language because of its functional importance for humans.

Acknowledgments

References

Badre D, Wagner AD. Selection, integration, and conflict monitoring: assessing the nature and generality of prefrontal cognitive control mechanisms. Neuron. 2004; 41:473–487. [PubMed: 14766185]

Belin P, Zatorre RJ, Ahad P. Human temporal-lobe response to vocal sounds. Cognitive Brain Research. 2002; 13:17–26. [PubMed: 11867247]

Blumstein SE, Myers EB, Rissman J. The perception of voice onset time: an fMRI investigation of phonetic category structure. Journal of Cognitive Neuroscience. 2005; 17:1353–1366. [PubMed: 16197689]

Blumstein SE, Stevens KN. Phonetic features and acoustic invariance in speech. Cognition. 1981; 10:25. [PubMed: 7198546]

Boersma P. Praat, a system for doing phonetics by computer. Glot International. 2001; 5:341–345.

Caplan D, Gow D, Makris N. Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. Neurology. 1995; 45:293–298. [PubMed: 7854528]

Celsis P, Boulanouar K, Doyon B, Ranjeva JP, Berry I, Nespoulous JL, Chollet F. Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. Neuroimage. 1999; 9:135–144. [PubMed: 9918735]

Cox RW. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research. 1996; 29

Cox RW, Jesmanowicz A. Real-time 3D image registration for functional MRI. Magnetic Resonance in Medicine. 1999; 42:1014–1018. [PubMed: 10571921]

Craighero L, Metta G, Sandini G, Fadiga L. The mirror-neurons system: data and models. Progress in Brain Research. 2007; 164:39–59. [PubMed: 17920425]

Eimas PD, Siqueland ER, Jusczyk P, Vigorito J. Speech perception in infants. Science. 1971; 171:303–306. [PubMed: 5538846]

Fowler CA. An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics. 1986; 14:3.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Categorical representation of visual stimuli in the primate prefrontal cortex. Science. 2001; 291:312–316. [PubMed: 11209083]

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. Journal of Neurophysiology. 2002; 88:929–941. [PubMed: 12163542]

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. Journal of Neuroscience. 2003; 23:5235–5246. [PubMed: 12832548]

Gold BT, Buckner RL. Common prefrontal regions coactivate with dissociable posterior regions during controlled semantic and phonological tasks. Neuron. 2002; 35:803–812. [PubMed: 12194878]

Grill-Spector K, Malach R. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychologica (Amsterdam). 2001; 107:293–321.

Hasson U, Skipper JI, Nusbaum HC, Small SL. Abstract coding of audiovisual speech: beyond sensory representation. Neuron. 2007; 56:1116–1126. [PubMed: 18093531]

Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC. Enhancement of MR images using registration for signal averaging. Journal of Computer Assisted Tomography. 1998; 22:324–333. [PubMed: 9530404]

Joanisse MF, Zevin JD, McCandliss BD. Brain Mechanisms Implicated in the Preattentive Categorization of Speech Sounds Revealed Using fMRI and a Short-Interval Habituation Trial Paradigm. Cerebral Cortex. 2006

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychological Review. 1967; 74:431. [PubMed: 4170865]

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. Cerebral Cortex. 2005; 15:1621–1631. [PubMed: 15703256]

Magnuson JS, Nusbaum HC. Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. Journal of Experimental Psychology: Human Perception and Performance. 2007; 33:391–409. [PubMed: 17469975]

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264:746–748. [PubMed: 1012311]

Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annual Review of Neuroscience. 2001; 24:167–202.

Miller JL. Some effects of speaking rate on phonetic perception. Phonetica. 1981; 38:159–180. [PubMed: 7267718]

Morin O, Grezes J. What is "mirror" in the premotor cortex? A review. Clincal Neuropsychology. 2008; 38:189–195.

Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 1971; 9:97–113. [PubMed: 5146491]

Paulesu E, Frith CD, Frackowiak RSJ. The neural correlates of the verbal component of working memory. Nature. 1993; 362:342–345. [PubMed: 8455719]

Peterson GE, Barney HL. Control methods used in a study of vowels. Journal of the Acoustical Society of America. 1952; 24:175.

Petrides M. Lateral prefrontal cortex: architectonic and functional organization. Philosophical Transactions of the Royal Society of London: B Biological Sciences. 2005; 360:781–795.

Pulvermuller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y. Motor cortex maps articulatory features of speech sounds. Procedings of the National Academy of Sciences U S A. 2006; 103:7865–7870.

Raizada RD, Poldrack RA. Selective amplification of stimulus differences during categorical processing of speech. Neuron. 2007; 56:726–740. [PubMed: 18031688]

Rizzolatti G, Craighero L. The mirror-neuron system. Annual Review of Neuroscience. 2004; 27:169–192.

Sabri M, Binder JR, Desai R, Medler DA, Leitl MD, Liebenthal E. Attentional and linguistic interactions in speech perception. Neuroimage. 2008; 39:1444–1456. [PubMed: 17996463]

Scott SK, Johnsrude IS. The neuroanatomical and functional organization of speech perception. Trends in Neuroscience. 2003; 26:100–107.

Stevens KN, Blumstein SE. Invariant cues for place of articulation in stop consonants. Journal of the Acoustical Society of America. 1978; 64:1358–1368. [PubMed: 744836]

Stevens, KN.; Blumstein, SE. The search for invariant acoustic correlates of phonetic features. In: Eimas, PD.; Miller, JL., editors. Perspectives on the Study of Speech. Hillsdale, NJ: Lawrence Erlbaum; 1981.

Wilson SM, Saygin AP, Sereno MI, Iacoboni M. Listening to speech activates motor areas involved in speech production. Nature Neuroscience. 2004; 7:701–702.

Zevin JD, McCandliss BD. Dishabituation of the BOLD response to speech sounds. Behavioral and Brain Functions. 2005; 1:4. [PubMed: 15953396]
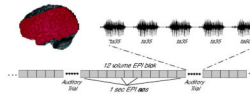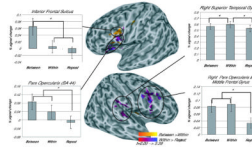
Figure 1.
Diagram of stimulus presentation during functional data acquisition. Red area on the left indicates the extent of EPI coverage for a typical subject.
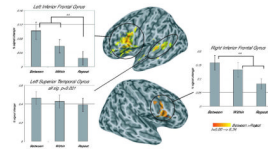
Figure 2.
Areas showing greater activation for Between-category than Repeat trials. All clusters are significant at a corrected threshold of p<0.05 (p<0.025 voxel-wise threshold, minimum cluster size 56 voxels), activation displayed on a canonical inflated brain surface (Holmes et al., 1998), with left (top) and right hemispheres (bottom) shown separately.

Figure 3.
Areas showing greater activation for Between-category than Within-category trials (purples) and greater activation for Within-category than Repeat trials (oranges). All clusters are significant at a corrected threshold of p<0.05 (p<0.025 voxel-wise threshold, minimum cluster size 56 voxels), activation displayed on a canonical inflated brain surface (Holmes et al., 1998), with left (top) and right hemispheres (bottom) shown separately.

Table 1

Clusters revealed in planned comparisons of syllable types. All clusters are significant at p<0.05, corrected (voxel-wise p<0.025, minimum 56 contiguous voxels per cluster).

Area	# Voxels	[x,y,z]	max t-stat	P value
		Coordinates of Maximum t value		
Between > Repeat				
LIFG (pars opercularis, pars triangularis)	465	[47, −5, 15]	6.3410	p<0.00001
RIFG (pars opercularis)	138	[−47, −5, 6]	4.7640	p<0.0005
L post STG, MTG	92	[62, 29, 12]	5.3400	p<0.001
Between < Repeat				
R Fusiform, R Lingual Gyrus	62	[−29, 53, −1]	−5.4150	p<0.025
Between > Within				
LIFG (pars opercularis)	105	[41, 8, 27]	4.0090	p<0.0012
Between < Within				
R Mid Orbital Gyrus, ACC	138	[−5, −29, −1]	−4.2000	p<0.0005
Within > Repeat				
R IFG (pars Opercularis), R Precentral	149	[−47, −5, 30]	5.1940	p<0.0003
R STG, R SMG	72	[−47, 32, 30]	3.7960	p<0.007
L Insula	60	[26, −23, 3]	3.7750	p<0.033
LIFG (pars Opercularis), L Precentral	59	[47, −5, 21]	3.8210	p<0.036
Within < Repeat				
L Fusiform, L Inferior Occipital	57	[38, 56, −16]	−5.0680	p<0.046
R Thalamus, Putamen	120	[−8, 20, 9]	−5.3760	p<0.0042