The Role of Segmentation in Phonological Processing: An fMRI Investigation

Martha W. Burton and Steven L. Small
University of Maryland School of Medicine

Sheila E. Blumstein
Brown University

Abstract

■ Phonological processes map sound information onto higher levels of language processing and provide the mechanisms by which verbal information can be temporarily stored in working memory. Despite a strong convergence of data suggesting both left lateralization and distributed encoding in the anterior and posterior perisylvian language areas, the nature and brain encoding of phonological subprocesses remain ambiguous. The present study used functional magnetic resonance imaging (fMRI) to investigate the conditions under which anterior (lateral frontal) areas are activated during speech-discrimination tasks that differ in segmental processing demands. In two experiments, subjects performed "same/different" judgments on the first sound of pairs of words. In the first experiment, the speech stimuli did not require overt segmentation of the initial consonant from the rest of the word, since the "different" pairs only varied in the phonetic voicing of the initial consonant (e.g., *dip–tip*). In the second

experiment, the speech stimuli required segmentation since "different" pairs both varied in initial consonant voicing and contained different vowels and final consonants (e.g., dip–ten). These speech conditions were compared to a tone-discrimination control condition. Behavioral data showed that subjects were highly accurate in both experiments, but revealed different patterns of reaction-time latencies between the two experiments. The imaging data indicated that whereas both speech conditions showed superior temporal activation when compared to tone discrimination, only the second experiment showed consistent evidence of frontal activity. Taken together, the results of Experiments 1 and 2 suggest that phonological processing per se does not necessarily recruit frontal areas. We postulate that frontal activation is a product of segmentation processes in speech perception, or alternatively, working memory demands required for such processing. ■

INTRODUCTION

Phonology plays a crucial role in language processing. It is the medium by which sound information maps on to higher levels of language processing (e.g., words). It also provides the code in which verbal information can be temporarily stored in working memory. There has been a strong convergence of results from the neuropsychological literature and the neuroimaging literature to suggest that the neural basis of phonological processing is lateralized to the left hemisphere, and encompasses a distributed neural system that includes posterior brain structures (superior temporal gyrus) and anterior brain structures (inferior frontal gyrus) (Blumstein, 1998; Pugh et al., 1996; Fiez et al., 1995; Paulesu, Frith, & Frackowiak, 1993; Paulesu et al., 1996; Démonet et al., 1992; Démonet, Price, Wise, & Frackowiak, 1994; Démonet, Fiez, Paulesu, Petersen, & Zatorre, 1996; Sergent, Zuck, Lévesque, & MacDonald, 1992; Shaywitz et al., 1995; Zatorre, Evans, Meyer, & Gjedde, 1992; Zatorre, Meyer, Gjedde, & Evans, 1996, but cf Poeppel, 1996). In particular, studies with

aphasic patients have shown impairments in speech perception irrespective of their clinical type and underlying neuropathology. Such patients often have difficulties in discriminating phonological contrasts such as initial stop-consonant voicing (e.g., *pear–bear*), or in identification of speech sounds, or both (Blumstein, 1998). That patients with lesions in the temporal lobe show such deficits is expected given that the primary auditory areas surface in the temporal lobe and, as such, have direct connections to the auditory association areas. What may appear to be more surprising is that patients with anterior lesions, including Broca's aphasics, also manifest such impairments.

Consistent with the behavioral studies with brain-injured patients, neuroimaging studies with normal subjects using positron emission tomography (PET) and functional MRI have shown that the inferior frontal gyrus and the posterior superior temporal gyrus show increased activity during tasks that involve sounds and words (Binder et al., 1997; Fiez et al., 1995; Price et al.,

1994; Price et al., 1996; Paulesu et al., 1993; Paulesu et al., 1996; Démonet et al., 1992; Démonet et al., 1994; Sergent et al., 1992; Zatorre et al., 1992; Zatorre et al., 1996). Increased activation in the posterior regions especially on the left has emerged during passive listening tasks to either words or nonwords. Increased activity has emerged in Broca's area in the left hemisphere in phonetic-discrimination and phoneme-monitoring tasks (Démonet et al., 1992; Démonet et al., 1994; Zatorre et al., 1992; Zatorre et al., 1996).

It has been hypothesized that the different neural structures implicated in phonological processing have different functional roles. On the basis of their findings, Zatorre et al. (1992, 1996) have proposed that Broca's area is involved in speech processing in tasks that require an overt phonetic decision by subjects. These conclusions are based on the findings that Broca's area shows activation when subjects are required to discriminate final consonants in word and nonword CVC stimuli (e.g., *fat–tid*) compared to a pitch-discrimination task (Zatorre et al., 1992, 1996), but there is no such activation in Broca's area under passive listening conditions to the same stimuli. Thus, they propose that Broca's area activation occurs when more fine-grained phonetic analysis is required and subjects must make phonetic judgments. In their view, to accomplish such analysis, listeners must access articulatory representations. It is this recoding process from acoustics to articulatory gestures that is said to require Broca's area involvement. Such an articulatory recoding process is thought to take place in Broca's area because of its traditional association with articulatory deficits in neuropsychological studies of aphasic patients.

What is not clear from the findings of Zatorre et al. (1992, 1996) is whether Broca's area is necessarily involved in all phonetic judgment tasks or whether its role is more circumscribed. In particular, there is evidence from the speech perception literature that suggests that listeners can process the sound segments of language solely in terms of their acoustic properties and discriminate phonetic contrasts based on these properties (compare Stevens & Blumstein, 1981; Klatt, 1979). If this is the case, then Broca's area involvement may not be necessary for such processing. Instead, auditory association areas (i.e., superior temporal gyrus) would more likely be involved in extracting the acoustic properties associated with phonetic features.

It is the goal of this research to investigate this issue by exploring in more detail the neural basis of speech processing. In particular, we investigated whether Broca's area will always show activation in an overt speech-discrimination task in which subjects are required to make a same/different judgment about phonetic segments. It is expected that Broca's area will only be involved when the subject must perform a task that requires articulatory recoding, namely, segmentation. By segmentation, we refer to the process whereby a subject needs to separate the individual sounds from the whole stimulus in order to complete the task. If there are a number of differences in the segments in the stimuli, but the task focuses on, for example, the initial sound for discrimination, then it appears likely that the subject must overtly identify the initial segment of the first stimulus and compare it to the initial segment of the second stimulus in order to make a same/different judgment. In contrast, if there is only one phonetic difference in the discrimination pair, as is typically the case in traditional speech perception experiments, there is in theory no need for segmentation since the subject only has to perceive a single phonetic difference between the stimulus pairs to make a decision. With regard to posterior brain structures, it is expected that the left superior temporal gyrus will be activated in the discrimination tasks irrespective of the phonetic complexity of the stimuli, presumably because the acoustic input must ultimately be analyzed in terms of the acoustic properties corresponding to the sound structure of language. To test this hypothesis, we report the results of two functional magnetic resonance imaging (fMRI) studies examining the discrimination of voicing in stop consonants (e.g., *t–d, k–g*) in tasks that either do or do not require overt segmentation.

EXPERIMENT 1

The first study investigated the discrimination of segmental contrasts that do not necessarily require overt segmentation. The main hypothesis is that subjects can perform a discrimination task without activating Broca's area when that task does not require segmentation.

## Results

Behavioral Results

The behavioral results are shown in Table 1. The response times (RTs) for one subject were not recorded due to computer error, and thus, were not included in the response time analysis. In addition, incorrect responses were discarded. Both accuracy (percent correct) and reaction time latencies were then analyzed

Table 1. Mean Reaction Time (RT) in milliseconds and % Correct with Standard Deviations (SD) for Experiment 1 in the Speech and Tone Conditions

	Speech		Tone	
	Same	*Different*	*Same*	*Different*
% correct	94	98	97	99
SD	5	2	2	1
Mean RT	1,399	1,415	1,409	1,423
SD	97	35	77	49
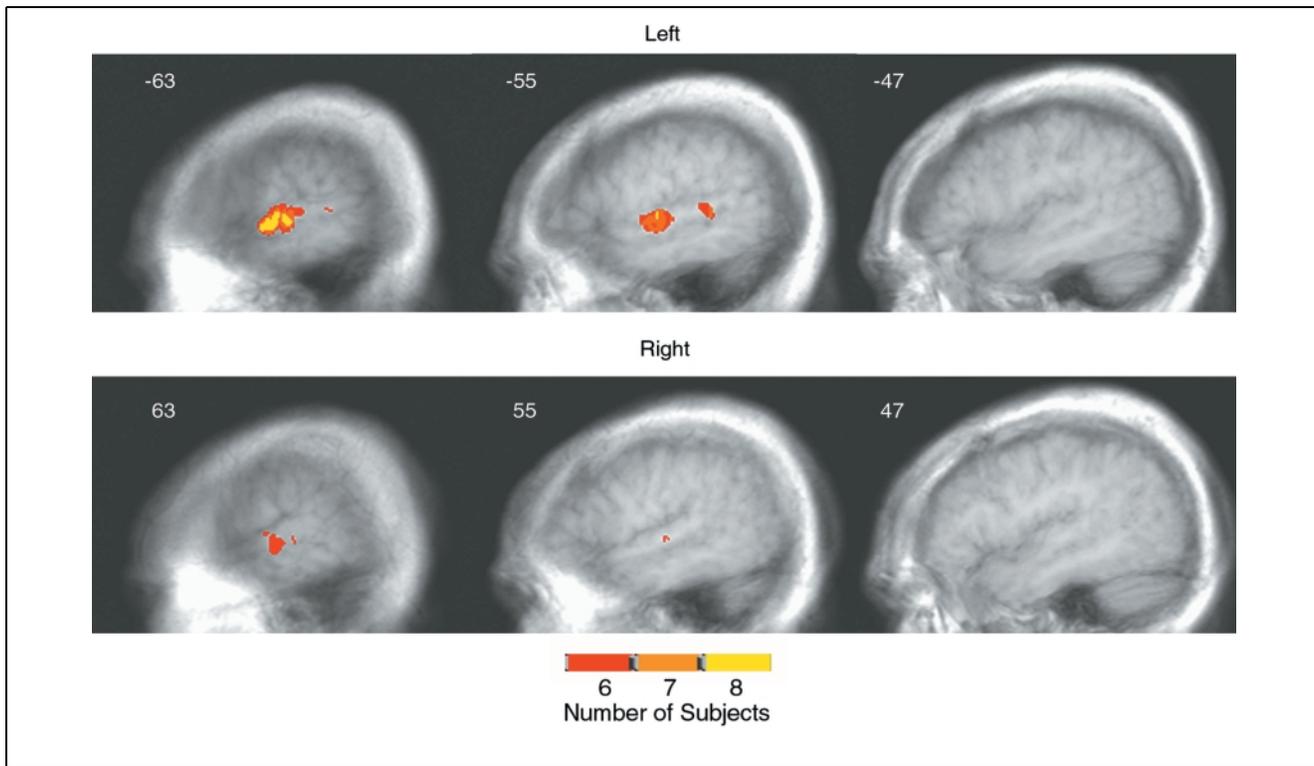
Figure 1. Speech discrimination minus tone discrimination in Experiment 1. Regions of overlap of six or more subjects are shown on averaged anatomical brain images in the sagittal plane. The number of subjects showing activation are indicated by the red-yellow color scale at the bottom of the figure. The top panels show patterns of activation in the left hemisphere and the bottom panels show right hemisphere activity. The numbers in each panel (in white) indicate the stereotaxic x coordinates. The left and middle panels show significant activity bilaterally in the superior temporal gyrus, although the activity is greater on the left than the right.
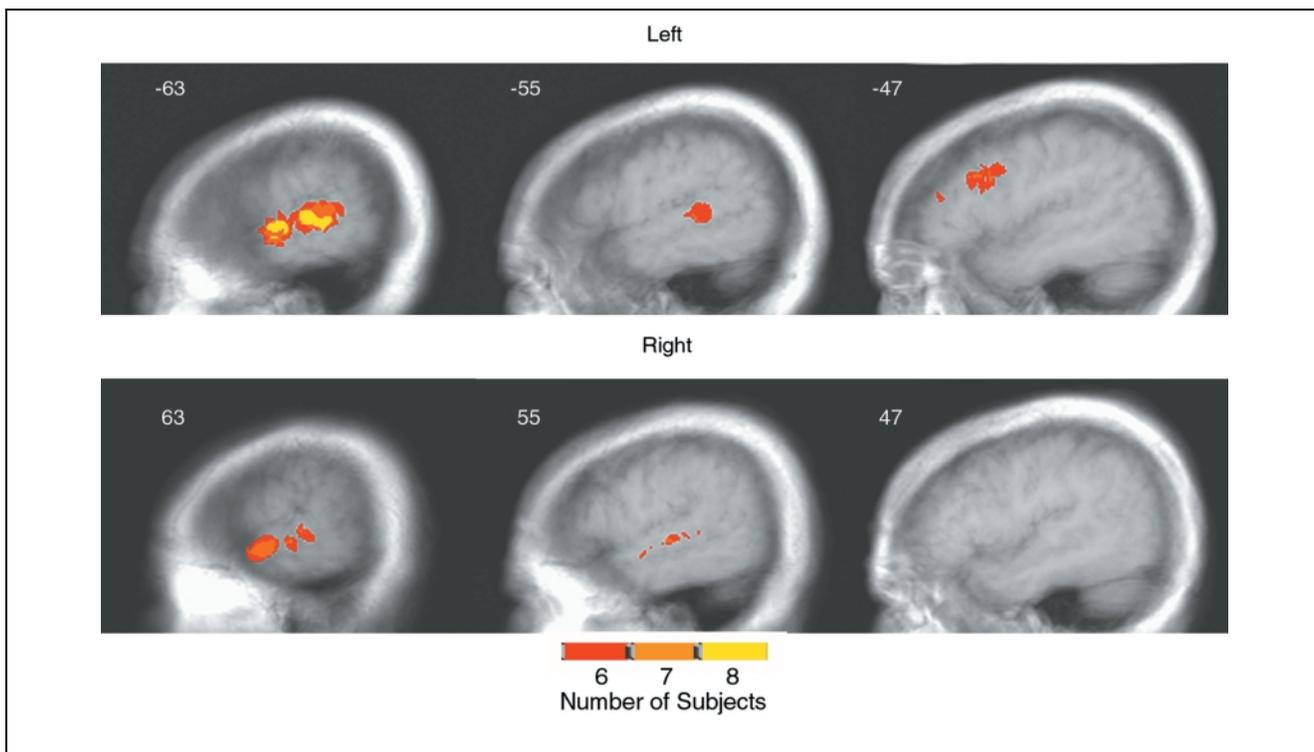
Figure 2. Speech discrimination minus tone discrimination in Experiment 2. The location of the structural images are the same as in Experiment 1. Areas showing significant activity are in the superior temporal gyrus bilaterally (see left and middle panels) and at the border of the left inferior and middle gyri (see top right panel).

using repeated measures analysis of variance (ANOVA) with two factors, Stimulus type (speech versus tone) and Response (same versus different). The ANOVA on accuracy revealed a significant main effect of Response ($F(1, 6) = 11.118$, $p = .02$). Although subjects were highly accurate on both speech and tone discrimination (greater than 96% correct), "same" responses (96%) were less accurate than "different" responses (99%). The speech task did not differ from the tone task in accuracy ($F(1, 6) = 2.69$, $p = .16$), nor was there a significant interaction ($F < 1$). Although subjects showed slower reaction-time latencies for "different" responses, the analysis of variance on response times did not show any significant main effects or a significant interaction ($F < 1$).

fMRI Results

Intersubject analyses. Figure 1 shows the areas of overlap of voxels showing a correlation of greater than .30 ($p < .001$ Bonferroni corrected for multiple comparisons) for six or more of the eight subjects performing speech discrimination compared to tone discrimination. The comparison of these two tasks produced strong evidence of activation in all subjects in two areas of the left superior temporal gyrus (STG). Table 2 shows the location of the center of mass and volumes of this overlap. The centers of mass of these areas of activity are located in auditory association areas, one anterior and the other posterior to the primary auditory cortex (Heschl's gyrus). As in previous studies, the phonological task showed bilateral activation in homologous areas of the temporal lobe. However, there was greater activation on the left than on the right. There was no evidence of Broca's area activation in the overlap data.

In addition, the reverse comparison (tone discrimination compared to the speech discrimination) was analyzed. No areas of activation appeared using the same correlation thresholds and number of subject criteria as for the speech–tone comparison.

Intrasubject analyses. To quantify and compare the amount of relative activity for the speech and tone tasks, the data for individual subjects were analyzed. The results showed that all subjects showed large significant clusters of activation bilaterally in the superior temporal gyrus with the average volume of activation 1258 mm^3

on the left and 629 mm^3 on the right. Two of the subjects also activated the middle temporal gyrus on the left. No activation appeared in the right middle-temporal region for any of the subjects. The subjects showed a great deal of variability in the total amount of temporal activation (range = 506 to 4177 mm^3, $SD = 1202$ mm^3), mostly due to one subject who had approximately four times more activity than any other subject. Without that subject, the standard deviation was reduced to 396. Because of the large variation in total volume, the percent of volume of temporal activation in each hemisphere compared to the total temporal activation was calculated for each subject. A one-way ANOVA indicated a strong lateralization effect with greater activation on the left (mean = .78, $SD = .21$) than the right (mean = .22, $SD = 21$) ($F(1, 7) = 13.75$, $p = .008$).

As described above, there did not appear to be consistent frontal activation in the overlap data. Within-subject analysis indicated that three of the eight subjects showed some left inferior or middle frontal gyrus activity. For one of the subjects (the same one who showed the greatest amount of temporal activity), this activity was strong (volume = 3333 mm^3), but for the other two, the activity was small (84 mm^3 or 4 voxels). Thus, there did not appear to be consistent evidence of Broca's area activity for the discrimination task. Further within-subject analyses failed to reveal any other areas of consistent activation, that is activation in more than two subjects.

The tone–speech comparison was also performed for individual subjects. There was no consistent activation across subjects (i.e., any area in which there was activation for more than two subjects).

Discussion of Experiment 1

The results of Experiment 1 showed a pattern of greater activity in the left and right superior temporal gyri during a speech-discrimination task compared to a tone-discrimination task, with greater activation on the left than on the right. No other areas showed consistent activation. These results are in accord with more traditional views proposing that auditory association areas are involved in the auditory processing of language. They also suggest that phonological processing per se does not necessarily recruit Broca's area. The involve-

Table 2. Center of Mass Given in Talairach Coordinates and Volume of Areas of Overlap of 6 or More Subjects for Experiment 1

Region	x	y	z	Number of Subs	Volume
L STG	−61	−12	3	8	2847
L STG	−57	−36	8	7	327
R STG	62	−12	2	6	471

Abbreviations: L = left; R = right; STG = superior temporal gyrus.

ment of Broca's area in phonological processing may be a function of task demands requiring overt segmentation. Experiment 2 explores this question.

EXPERIMENT 2

As in Experiment 1, we focused on the discrimination of initial stop consonants distinguished by a single phonetic feature (voicing). Unlike Experiment 1, however, the vowel and consonant portion of the stimuli differed. As a consequence, for subjects to perform the discrimination task, they had to segment out the initial consonant of each member of the pair and compare them. Only then could they make the discrimination judgment.

Results

Behavioral Results

The results of Experiment 2 are shown in Table 3. Reaction times and accuracy were again analyzed as in Experiment 1, with incorrect responses eliminated from the RT data. The ANOVA on accuracy showed no significant main effects or interactions. As in the previous experiment, subjects were highly accurate (> 97% in all conditions). However, in contrast to Experiment 1, the ANOVA on response times showed significant main effects of Stimulus type ($F(1, 7) = 5.6$, $p = .05$) and Response ($F(1, 7) = 9.397$, $p = .02$). Overall, mean response times for tone judgments were 36 msec faster than for speech, and "same" judgments were 46 msec faster than "different" judgments. There was also a significant interaction between these factors due to a difference in reaction times for "same" and "different" stimuli in the speech condition (89 msec), but not in the tone task ($F(1, 7) = 40.96$, $p = .0004$). Newman–Keuls post hoc tests showed the speech-different condition was significantly slower than all other conditions ($p < .05$). Thus, the change in speech stimuli from Experiment 1 to Experiment 2 produced a different pattern of results. This difference emerged only for "different" responses and was due to slowed reaction-time

Table 3. Mean Reaction Time (RT) and % Correct with Standard Deviations (SD) for Experiment 2 in the Speech and Tone Conditions

	Speech		Tone	
	Same	Different	Same	Different
% correct	97	99	99	99
SD	4	1	1	1
Mean RT	1,350	1,441	1,360	1,360
SD	73	80	70	92

latencies for these responses compared to "same" responses for the stimuli in Experiment 2. Importantly, as in Experiment 1, the tone stimuli, which were exactly the same in both experiments, did not show a change in pattern of responses.

fMRI Results

Intersubject analyses. As in Experiment 1, speech discrimination compared to tone discrimination produced strong activation in the left superior temporal gyrus for all subjects (see Figure 2, Table 4). The location of the center of the temporal areas, where all subjects revealed overlapping activation, was similar in the two experiments. Although activity on the left appeared to be greater than on the right, all eight subjects showed activation in the superior temporal gyrus of the right hemisphere. In contrast to the previous experiment, activity also appeared on the border of the left inferior and middle frontal gyrus for seven of the eight subjects. No other areas showed activation across subjects for the speech–tone or tone–speech comparisons.

To determine whether the pattern of activation between the subjects in the two experiments significantly differed, a voxel-wise t test was conducted comparing the mean difference in signal intensities between the speech and tone tasks from individual analyses for the two groups of subjects at each voxel. The results of this analysis are shown in Table 5. Importantly, there was a significant difference between the two groups of subjects in an area of the left inferior frontal gyrus, located very close to the middle frontal gyrus location found in Experiment 2. Two other regions showed significant differences between the two groups of subjects, the anterior cingulate, and a region in the inferior temporal gyrus.

Areas where there was significantly more activation for the subjects in the first experiment included a left anterior fronto-temporal region, left middle frontal gyrus and left middle temporal gyrus, and the right cerebellum. The middle frontal gyrus was at the anterior edge of the frontal lobe near the eye, and may be due to artifact rather than to differences between the two tasks. The left temporal-frontal activation was particularly strong in two of the subjects and was not strong across the entire group of subjects in Experiment 1.

Intrasubject analyses. The results of the individual subject analyses on the speech versus tone comparison again showed strong bilateral activation in the superior temporal gyrus with the average volume of activation 1282 mm^3 on the left and 698 mm^3 on the right. Four of the subjects also activated the middle temporal gyrus on the left and one on the right. Although the percent volume of left-temporal activation (mean = .62, SD = .32) was greater than on the right (mean = .38, SD = .32), the difference was not significant ($F(1,7) = 1.178$, $p = .31$).

Table 4. Centers of Mass Given in Talairach Coordinates and Volume of Areas of Overlap of Six or More Subjects for Experiment 2

			Speech–Tone Comparison		
Region	*x*	*y*	*z*	*Number of Subs*	*Volume*
L STG	−64	−23	6	8	6,060
L MTG	−51	−43	3	6	47
L MFG	−47	15	29	7	1,113
L MFG	−47	37	19	6	47
R STG	61	−10	1	8	2,208
R STG	51	−30	8	6	81

Abbreviations: L = left; R = right; STG = superior temporal gyrus; MFG = middle frontal gyrus; MTG = middle temporal gyrus.

Seven of the eight subjects showed strong evidence of significant activity in the left inferior and middle frontal gyrus both in terms of volume of activity and in multiple clusters of activity. This activity was primarily on the left (average volume of left = 702 mm^3, right = 401 mm^3), though four subjects did show some right frontal activity. An analysis of variance on the percent of volume in each hemisphere (mean left = .86, mean right = .14, *SD* = .19) confirmed that this hemispheric difference was significant ($F(1,7) = 26.434$, $p = .002$).

The pattern of activity in this experiment was strongly localized to the temporal and frontal regions described above with one exception. An additional posterior parietal region in the left hemisphere appeared in six subjects (average volume = 914 mm^3, standard deviation = 1506 mm^3), though there was a great deal of variation in the location and size of activation within the parietal lobe. Because of the variation in location, it did

not appear as a single cluster for the six subjects and is not reported in Table 4. For four of the subjects, the activity was more inferior than the other two subjects. No other areas appeared consistently (i.e., in more than two subjects).

For the tone–speech comparison, five subjects showed significant activity in the right hemisphere. The activity in the individual subjects was distributed along the superior and middle temporal gyri, although an area posterior to the auditory cortex appeared to be present in all five of the subjects showing superior temporal gyrus activation (centered at $x = 62$, $y = −42$, $z = 15$). Temporal activity was strongly lateralized to the right (average volume of left = 130 mm^3, right = 563 mm^3). The hemispheric difference in percent volume (mean left = .83, mean right = .17, *SD* = .24) was significant ($F(1,4) = 8.907$, $p = .04$), according to an analysis of variance.

Table 5. Centers of Mass Given in Talairach Coordinates and Volume of Areas Showing Significant Differences in *t* Test Comparing Groups of Subjects

Region	*x*	*y*	*z*	*t value*	*p value*	*Volume*
Experiment 2–Experiment 1 Comparison						
L IFG	−47	17	24	3.209	.006	274
IT	−58	−56	−8	3.45	.004	167
Ant. Cing	0	21	−6	2.275	.04	2480
Experiment 1–Experiment 2 Comparison						
L frontal temporal	−60	13	5	−4.8	.00028	1329
L MFG	−40	50	−9	−3.6	.0029	345
L MTG	−59	−32	−3	−3.36	.0047	112
R CRB	33	−80	−29	−3.39	.0044	311

The top half of the table shows areas where there was significantly greater activity for subjects in Experiment 2 than in Experiment 1. The bottom panel shows areas where subjects in Experiment 1 show greater activation. The *t* value and corresponding *p* value at the center of mass are given in the fifth and sixth columns. Abbreviations: L = left; R = right; IFG = inferior frontal gyrus; IT = inferior temporal lobe; Ant. Cing = anterior cingulate; MFG = middle frontal gyrus; MTG = middle temporal gyrus; R CRB = right cerebellum.

Discussion of Experiment 2

For the comparison of speech versus tone tasks, the results of Experiment 2 showed activation patterns in temporal (superior temporal gyrus), frontal (inferior and middle frontal gyrus) and parietal regions (inferior and superior parietal lobule). For the temporal and frontal activation, there was generally greater activation on the left than in homologous areas on the right, but this difference was statistically reliable only in frontal areas. This frontal activation was significantly greater in Experiment 2 than in Experiment 1, as demonstrated by the voxel-wise t test. Tones compared to speech revealed increased activity in the right temporal lobe in this experiment. However, the group t test comparison did not indicate that this increased activity differed from Experiment 1.

GENERAL DISCUSSION

The results of Experiments 1 and 2 suggest that phonological processing during speech-discrimination tasks strongly activates superior-temporal regions in the left hemisphere and to a lesser degree in the right hemisphere. Superior temporal activation in language tasks is consistent with many other studies of speech stimuli (Zatorre et al., 1992; Wise et al., 1991; Petersen, Fox, Posner, Mintun, & Raichle, 1988). Such bilateral activation near the auditory cortex has been attributed to perceptual analysis of the speech signal. More posterior superior-temporal activation, which tends to be more strongly left-lateralized (and was significantly so in Experiment 1 but not in Experiment 2), is more generally associated with analysis of speech sounds for mapping onto higher levels of language processing (e.g., syllable, word) (Zatorre et al., 1996; Price et al., 1992). These findings are also consistent with behavioral studies based on lesion analysis methods (Blumstein, 1998).

Nonetheless, both the fMRI results and the behavioral results showed differential patterns between the two experiments. With respect to the behavioral results, subjects showed significantly slowed responses to "different" responses compared to "same" responses in Experiment 2, whereas there was no difference in reaction-time latencies between the two conditions in Experiment 1. As to the imaging data, the major difference between the activation patterns of Experiment 1 and Experiment 2 was the emergence of left hemisphere activity in the inferior and middle frontal gyri in Experiment 2 and the absence of consistent activation in these areas in Experiment 1. This result was further supported by the significant difference in that region in a between group voxel-wise t test. Taken together, the behavioral and imaging data indicate that phonological processing, and in this case, phonetic judgments, may invoke different neural mechanisms depending on the task demands. Importantly, they show that frontal areas

are not necessarily involved in all aspects of phonological processing.

But what are the conditions under which frontal activation occurs in phonological processing? In the present study, the speech stimuli in the two experiments differed in the final vowel-consonant (VC) portion of the pairs of syllables. In Experiment 1, subjects were required to distinguish a phonetic contrast (voicing) in initial position in a CVC word stimulus in which there were no other phonological differences (i.e., they were required to perceive the difference between the unvoiced [t] in *tip* and the voiced [d] in *dip*). In contrast, in Experiment 2, subjects were required to distinguish the same phonetic contrast (voicing) in initial position as in Experiment 1, but in a CVC word stimulus in which the medial vowel and final consonant also differed. Thus, subjects could not make a phonetic judgment (same/different) based on merely perceiving phonetic/phonological differences between the two stimuli (e.g., *tip* versus *doom*). Instead, they had to separate out the initial consonants from the whole stimulus and compare them in order to make a same/different judgment. It is this process of segmentation of the initial consonant from the following vowel, probably requiring articulatory recoding, that appears to involve left anterior brain structures, and in particular, the inferior and middle frontal gyrus. Furthermore, it is under these stimulus conditions that subjects also showed significantly slowed reaction-time latencies in the behavioral measures.

Such segmentation processes may also increase attentional demands by requiring subjects to attend selectively to the first phoneme. Such an increase in attentional demands in the second experiment receives support from the significant differences in activation between the two groups of subjects in the anterior cingulate, which has been associated with attention processes in a variety of tasks, including Stroop tasks (Smith & Jonides, 1999; Carter et al., 1998). More specifically, anterior cingulate activation emerges in experiments that induce a conflict between cognitive processes or response tendencies. The segmentation task in the present study produces such a conflict. Subjects are required to respond "same" to pairs such as *dip–doom* despite the differences in the rest of stimuli. In contrast, discrimination of the initial consonant in the first experiment does not involve such a conflict. Instead, subjects may base their decision on whether the stimuli are identical as in the "same" trials or "different" in any manner (including possible differences in intonation, amplitude, and allophone realization as well as initial voicing).

Nonetheless, whether the frontal activation found in Experiment 2 is a product of the segmentation processes themselves or the working memory demands required to segment a stimulus, retain its phonetic form for

comparison, and make a response decision is unclear and cannot be determined from the results of the current experiment. Recent neuroimaging studies are consistent with both possibilities. In particular, recent studies of phonological processing (e.g., visual rhyme judgment, auditory phoneme monitoring, and auditory detection) (Pugh et al., 1996; Fiez et al., 1995; Shaywitz et al., 1995; Démonet et al., 1992; Démonet et al., 1994; Zatorre et al., 1992; Zatorre et al., 1996) as well as recent studies of verbal working memory (e.g., n-back) (Jonides et al., 1998; Braver et al., 1997; Schumacher et al., 1996; Awh, Smith, & Jonides, 1995; Cohen et al., 1994; Cohen et al., 1997; Paulesu et al., 1993; Paulesu et al., 1996) have all shown activation of the inferior frontal gyrus (i.e., in portions of Broca's area), in some cases extending superiorly to the border with the middle frontal gyrus.

The location of frontal activation in the present study (compare Table 4) is consistent with those reported by Zatorre et al. studies (Zatorre et al., 1992; Zatorre et al., 1996), with the peak activity located on the border of the inferior and middle frontal gyri. Furthermore, the frontal activation was strongly left lateralized (even for female subjects in the second experiment (compare Shaywitz et al., 1995). This location differs from other studies, which have found left frontal-opercular activity for auditory-detection tasks (Fiez et al., 1995). It is of interest that Fiez et al. proposed that the left frontal opercular region is important for certain types of auditory/temporal analysis, and, in particular, for processing stimuli containing rapid temporal changes. The stimuli in Experiments 1 and 2 in the current study contained phonetic contrasts that also are characterized by rapid temporal changes. And yet, Experiment 1 showed only temporal lobe activation, and Experiment 2 showed frontal activation in the inferior and middle frontal gyrus and not in the frontal operculum. It is possible that the differences that have emerged across experiments reflect the different task demands made on the subjects, and in particular, the extent to which the tasks require verbal rehearsal processes to perform monitoring or detection tasks such as those used by Fiez et al. (1995). However, because the tasks are difficult to compare in terms of the degree to which processing of stimuli require verbal rehearsal, further research is necessary to settle this issue.

The activity in the parietal lobe seen in many of the subjects of Experiment 2 is consistent with involvement of working-memory components in tasks requiring segmentation. Several studies of working memory have reported parietal activity during tasks that require temporary storage of verbal materials (Jonides et al., 1998; Awh et al., 1995; Paulesu et al., 1993; Paulesu et al., 1996). Furthermore, Zatorre et al. (1996) also reported evidence of activity in this region in some of their phonetic task comparisons (discrimination in the 1992 study and monitoring in the 1996 study). This activation appears more consistently with tasks that involve letters

and nonwords that require more extensive phonological coding than with word tasks where subjects may use a combination of semantic and phonological coding to store verbal material (Jonides et al., 1998). The fact that only word stimuli were used in the current study and in some of Zatorre et al. (1992, 1996) comparisons may explain why the parietal activation appears less consistently than other sites of activation.

Other findings of the study are consistent with previous results. The group t test revealed significantly more activation in Experiment 2 in the inferior temporal lobe in Brodmann's Area 37 than in Experiment 1. This area, thought to be specialized for visual object processing including processing of letters and words, has shown also activity in tactile (Braille) tasks (Buchel, Price, Frackowiak, & Friston, 1998; Buchel, Price, & Friston, 1998; Sadato et al., 1996). Further research is needed to determine the role of this area in phonological processing. Recent studies have indicated the presence of multimodal representations in primary sensory areas during language processing (Buchel et al., 1998; Calvert et al., 1997). Thus, it would not be surprising for literate subjects to activate visual language processing areas during the segmentation of speech. In particular, subjects may be identifying sound segments and comparing them to letter representations. Such activation of letter representations may not be required when discrimination of phonetic differences in stimuli that do not require segmentation.

In addition, there was greater activation in the middle temporal gyrus and the right cerebellum in Experiment 1 than in Experiment 2. This greater activation could reflect the nature of the speech discrimination requirements in Experiment 1 compared to those in Experiment 2. In particular, evidence has suggested that the cerebellum is involved in temporal computations in not only the motor domain, but also in the perceptual domain (Keele & Ivry, 1990). Consistent with this view are some recent findings showing that the cerebellum is involved in the computation of temporal parameters in speech perception (compare Ackermann, Graber, Hertrich, & Daum, 1997). Experiment 1 of the current study required the discrimination of voicing of initial stop consonants. The most salient cue to voicing in this phonetic position is voice-onset time, the temporal relation between the release of the stop closure and the onset of vocal cord vibration (Kent & Read, 1992). In contrast, in Experiment 2, all of the segments in the test stimuli differed. Thus, subjects could not rely solely on the temporal parameters of the initial consonant to successfully perform the discrimination task, which we postulate to have led to the significant increase in right cerebellar activity in Experiment 1 compared to Experiment 2.

The comparison of tone–speech produced relatively little activation. This is not surprising since the complexity of the acoustic properties of the pairs of speech stimuli was greater than that of the tone stimuli. None-

theless, some right lateralized activity was observed in the second experiment. However, this difference did not emerge when the two experiments were statistically compared.

In summary, the present results support the traditional view that posterior structures of the left hemisphere (namely, auditory association areas) are involved in the auditory and phonological processing of speech. The involvement of frontal areas in phonological processes appears to be more circumscribed and arises when the subject must perform a task that requires articulatory recoding, i.e., overt segmentation of the phonetic units of the stimulus, or verbal rehearsal, i.e., holding a model or monitoring for a particular stimulus or characteristic of a stimulus. Thus, it appears that a distinction can be drawn between the functional role of anterior and posterior brain structures in phonological processing. Namely, perceptual analysis of the speech input itself appears to be largely a function of the temporal lobe, and in particular, of the superior temporal gyrus. In contrast, the involvement of frontal areas which have been reported in a number of imaging studies of phonological processing appears to reflect task demands for output rather than the perceptual analysis of the speech input itself. The relevant-task demands are those that require the subject to extract a phonetic segment and retain it and/or compare it to a phonetic label for response.

It is of interest to consider whether the processing mechanisms used for auditory language comprehension and for the mapping from sound structure to meaning are the same as those used in the types of phonological processing tasks considered in this paper. It may be that the processing of speech for purposes of auditory language comprehension does not require segmentation of the individual sounds to accomplish word recognition or lexical access, but rather requires a mapping of more global auditory/phonetic patterns on to meaning. If this is the case, then posterior brain structures may be involved in processing mechanisms for auditory language comprehension processes, whereas frontal areas may not typically be recruited in such processing.

METHODS

Experiment 1

Subjects

Eight monolingual native speakers of American English (four female, four male ranging in age from 23 to 42, mean = 33) participated in the study. Six of the subjects were strongly right handed with scores greater than +60 on the Edinburgh Handedness Inventory (Oldfield, 1971). The other subjects' scores were +6.7 and +41.2. The study was approved by the University of Maryland School of Medicine Institutional Review Board and all subjects provided written informed consent prior to participation.

Stimulus Materials

Subjects performed two tasks, speech discrimination and tone discrimination. In the speech-discrimination task, subjects were presented with eight different pairs of CVC real word stimuli (e.g., *dip–tip*) in which the initial consonant differed in voicing. For each pair, three matched pairs were created, one with the order of words reversed, (e.g., *tip–dip*), and two with either the voiced or voiceless member of the pair repeated (e.g., *dip–dip*, *tip–tip*). Thus, for half of the pairs, the first sound was different; for the other half, the initial consonants were the same. For comparison with the test condition, a control condition was needed that required discrimination of auditory nonspeech stimuli. To that end, a tone condition was created in which pairs of tones had either the same pitch or different pitch. The tones were either high (750 Hz) or low (500 Hz) and were matched in duration to the average duration of the speech stimuli.

Behavioral Testing Procedure

During the scan, subjects performed a forced choice auditory discrimination task, pressing a button to indicate whether the initial consonant of the stimulus pairs was the same or different. Subjects were asked to keep their eyes closed during the functional portion of the scan to minimize eye movement artifacts. To remind subjects of the button position, instructions (''same-left, different-right'') were presented 8 sec prior to the beginning of each stimulus block. Subjects' performance was scored for both accuracy and reaction time.

Test trials consisted of pairs of stimuli presented over sound-attenuating air conduction headphones (Resonance Technologies™, Northridge, CA) with a stimulus onset asynchrony (SOA) of 3 sec and a 50-msec interval between the first and second members of the pair. Stimuli were assigned to 24-sec blocks consisting of either eight pairs of words or tones. Within each block, the stimuli were presented in random order. Alternating blocks of speech and tone stimuli were repeated eight times during each scan, and 32 times during the course of a functional scanning session that contained four scans. Each stimulus pair (e.g., *dip–tip*) appeared eight times during the course of a scanning session, twice during each scan. In addition, four 32-sec rest blocks were placed after two cycles of the alternating speech/tone blocks. Thus, each scan lasted a total of 10 min and 40 sec (16 instruction blocks lasting 8 sec each, eight 24-sec speech blocks, eight 24-sec tone blocks, and four 32 sec rest blocks).

MRI Procedures

Image Collection

The functional imaging studies were carried out in the Department of Radiology at the University of Maryland Medical Center on a 1.5-T Signa scanner (GE Medical Systems, Milwaukee, WI) with a standard GE head coil. Twenty-four 6-mm structural T1-weighted anatomical images (500 ms repetition time (TR), 9-ms gradient echo time (TE), spin echo pulse sequence) were acquired in the sagittal plane to determine the anatomy of the functional slices, starting from the left temporal lobe and ending with the right temporal lobe. Functional data were acquired using the spiral k-space method (Noll & Schneider, 1994; Noll, Cohen, Meyer, & Schneider, 1995). A single shot gradient echo spiral scan pulse sequence provided 3.2×3.2 mm resolution over a 24-cm field-of-view (FOV). T2*-weighted imaging was accomplished with a TE of 35 ms, and a TR of 4000 ms with a flip angle of 60°. A complete set of the 24 slice locations was generated every repetition time cycle (4 sec). Each of the 24 slices was acquired six times during each 24-sec task interval. Since each task was repeated eight times during each experimental trial, the total number of images obtained for each condition for each experimental trial was 48. Following the end of each trial, a brief (< 1 min) break occurred during which the next trial was prepared. Each trial was repeated four times to improve the statistical power and reliability, yielding four time series of data, consisting of 48 time points at each voxel location. Thus, for each subject, there were 192 images for the speech task and 192 images for the tone task.

Intersubject Analysis

Because of possible scanner signal drift and movement by the subject, the first analysis step was to normalize each time series to a zero mean and to apply a linear detrending procedure to remove changes in the baseline signal. The second step was to co-register all the images for each subject together using the 3-D version of the automated image registration algorithm (Woods, Grafton, Holmes, Cherry, & Mazziotta, 1998). The four resulting time series were concatenated together. Due to computer error during data acquisition, only two time series were collected for one subject.

Cross-correlation analysis was then performed on the concatenated time series (Bandettini, Jesmanowicz, Wong, & Hyde, 1993) using AFNI software (Cox, 1996) to determine correlation coefficients for each subject. In this analysis, the concatenated waveform for each voxel was correlated with a reference waveform corresponding to the temporal pattern of stimulus presentation of the speech and tone conditions, ignoring all instruction and rest periods. The cross-correlation analysis was

performed with sine waves offset by 0°, 7.5°, 15°, 22.5°, and 30° corresponding to time delays of 0, 2.5, 5, 7.5, and 10 sec. The best match was retained to account for hemodynamic lag.

To assess the activation pattern of the group, cross-correlation data were analyzed as a count of the subjects who had activation at a particular voxel with a correlation coefficient of .3 or greater. To this end, the images were transformed into standard stereotactic space (Talairach & Tournoux, 1988). The images were then blurred by using a full-width-half-maximum gaussian filter 4 mm wide. Peak areas of activation were located by using a 3-D clustering program with a connectivity radius of 1.9 mm, a volume threshold of 2 voxels (40 mm^3) . The functional data were plotted on an average of the structural images from the eight subjects after transformation into Talairach coordinates.

The group data were further analyzed as individual cases to assess the variability in pattern and location of activation. The correlation coefficient data for each subject were thresholded at .30 ($p = .001$ with Bonferroni correction for multiple comparisons for single voxels) and cluster sizes were measured after applying a minimum threshold (using a 3-D neighborhood of 26 voxels) of greater than 4 voxels (volume = 84 mm^3).

Experiment 2

Subjects

A different group of eight, right-handed native speakers of English (five female, three male ranging in age from 22 to 48, mean = 28) participated in the experiment. Six of the subjects were strongly right-handed with scores greater than +60 on the Edinburgh Handedness Inventory (Oldfield, 1971). The other subjects' score were +50 and +30. The study was approved by the University of Maryland School of Medicine Institutional Review Board and all subjects provided written informed consent prior to participation.

Stimulus Material

The stimulus materials were the same as in the previous experiment with the following exception. The words of the original speech stimulus pairs used in Experiment 1 were rematched such that the pairs of CVC stimuli no longer contained the same medial vowel and final consonant, but as in Experiment 1, the initial consonants differed only in voicing. For example, stimulus pairs *tip–dip* and *tomb–doom* of Experiment 1 were rematched in Experiment 2 as *tip–doom* and *tomb–dip*. As in Experiment 1, a set of 32 stimulus pairs was created in which for half of the pairs, the initial consonants differed in voicing (*dip–tomb*, *tomb–dip*) and for the other half, the initial consonants were the same

(*dip–doom*, *tip–tomb*). The tone stimuli were identical to Experiment 1.

Procedures

All behavioral and fMRI testing procedures were identical to Experiment 1.

Image Analysis

All group and individual subject data analyses were performed using the same methods in both experiments. To determine whether there were significant differences between the patterns of activation between the two experiments, an additional voxel-wise t test was performed comparing the two groups of subjects. The mean differences in intensities between the speech and tone conditions, for each voxel in each of the individual subject datasets, were calculated. The functional data were then transformed into Talairach coordinates (Talairach & Tournoux, 1988). Using an 8-mm full-width half maximum filter to smooth the data, between group t tests were conducted on the mean difference in voxel intensities from the individual analyses, comparing subjects in Experiment 1 to subjects in Experiment 2. Peak areas of activation were located by using a 3-D clustering program with a connectivity radius of 1.9 mm and a volume threshold of 100 mm^3. The functional data were plotted on an average of the structural images after transformation into Talairach coordinates.

REFERENCES

Ackermann, H., Graber, S., Hertrich, I., & Daum, I. (1997). Categorical speech perception in cerebellar disorders. *Brain and Language, 60,* 323–331.

Awh, E., Smith, E. E., & Jonides, J. (1995). Human rehearsal processes and the frontal lobes: PET evidence. *Annals of the New York Academy of Sciences, 769,* 97–117.

Bandettini, P. A., Jesmanowicz, A., Wong, E. C., & Hyde, J. S. (1993). Processing strategies for time-course data sets in functional MRI of the human brain. *Magnetic Resonance in Medicine, 30,* 161–173.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience, 17,* 353–362.

Blumstein, S. E. (1998). Phonological aspects of aphasia. In M. Sarno (Ed.), *Acquired aphasia* (pp. 157–185). New York: Academic Press.

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage, 5,* 49–62.

Buchel, C., Price, C., Frackowiak, R. S., & Friston, K. (1998). Different activation patterns in the visual cortex of late and congenitally blind subjects. *Brain, 121,* 409–419.

Buchel, C., Price, C., & Friston, K. (1998). A multimodal language region in the ventral visual pathway. *Nature, 394,* 274–277.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D., & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science, 276,* 593–596.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science, 280,* 747–749.

Cohen, J., Forman, D., Braver, T., Casey, B., Servan-Schreiber, D., & Noll, D. (1994). Activation of prefrontal cortex in a non-spatial working memory task with functional MRI. *Human Brain Mapping, 1,* 293–304.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature, 386,* 604–608.

Cox, R. W. (1996). Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers in Biomedical Research, 29,* 162–173.

Démonet, J.-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.-L., Wise, R., Rascol, A., & Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain, 115,* 1753–1768.

Démonet, J.-F., Price, C., Wise, R., & Frackowiak, R. J. (1994). A PET study of cognitive strategies in normal subjects during language tasks: Influence of phonetic ambiguity and sequence processing on phoneme monitoring. *Brain, 117,* 671–682.

Démonet, J. F., Fiez, J. A., Paulesu, E., Petersen, S. E., & Zatorre, R. J. (1996). PET studies of phonological processing: A critical reply to Poeppel. *Brain and Language, 55,* 352–379.

Fiez, J., Raichle, M., Miezen, F., Petersen, S., Tallal, P., & Katz, W. (1995). PET studies of auditory and phonological processing: Effects of stimulus characteristics and task demands. *Journal of Cognitive Neuroscience, 7,* 357–375.

Jonides, J., Schumacher, E. H., Smith, E. E., Koeppe, R. A., Awh, E., Reuter-Lorenz, P. A., Marshuetz, C., & Willis, C. R. (1998). The role of parietal cortex in verbal working memory. *Journal of Neuroscience, 18,* 5026–5034.

Keele, S. W. & Ivry, R. (1990). Does the cerebellum provide a common computation for diverse tasks? A timing hypothesis. *Annals of the New York Academy of Sciences, 608,* 179–207.

Kent, R.D. & Read, C. (1992). *The acoustic analysis of speech*. San Diego, CA: Singular Publishing.

Klatt, D. H. (1979). Speech perception: A model of acoustic–phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, NJ: Erlbaum.

Noll, D. C., Cohen, J. D., Meyer, C. H., & Schneider, W. (1995).

Spiral K-Space MRI of cortical activation. *Journal of Magnetic Resonance Imaging, 5,* 49–56.

Noll, D. C. & Schneider, W. (1994). *Theory, simulation, and compensation of physiological motion artifacts in functional MRI*. Paper presented at the IEEE International Conference on Image Processing, Austin, TX.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113.

Paulesu, E., Frith, C. D., & Frackowiak, R. J. (1993). The neural correlates of the verbal component of working memory. *Nature, 362,* 342–345.

Paulesu, E., Frith, U., Snowling, M., Gallagher, A., Morton, J., Frackowiak, R. S. J., & Frith, C. D. (1996). Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain, 119,* 143–157.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M. A., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331,* 585–589.

Poeppel, D. (1996). A critical review of PET studies of phonological processing. *Brain and Language, 55,* 317–351.

Price, C., Wise, R., Ramsey, S., Friston, K., Howard, D., & Patterson, K. (1992). Regional response differences within the human auditory cortex when listening to words. *Neuroscience Letters, 146,* 179–182.

Price, C., Wise, R., Warburton, E., Moore, C., Howard, D., Patterson, K., Frackowiak, R., & Friston, K. (1996). Hearing and saying: the functional neuro-anatomy of auditory word processing. *Brain, 119,* 919–931.

Price, C., Wise, R., Watson, J., Patterson, K., Howard, D., & Frackowiak, R. (1994). Brain activity during reading: The effects of exposure duration and task. *Brain, 117,* 1255–1269.

Pugh, K. R., Shaywitz, B. A., Shaywitz, S. E., Constable, R. T., Skudlarski, P., Fulbright, R. K., Bronen, R. A., Shankweiler, D. P., Katz, L., Fletcher, J. M., & Gore, J. C. (1996). Cerebral organization of component processes in reading. *Brain, 119,* 1221–1238.

Sadato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Deiber, M. P., Dold, G., & Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature, 380,* 526–528.

Schumacher, E. H., Lauber, E., Awh, E., Jonides, J., Smith, E. E., & Koeppe, R. A. (1996). PET evidence for an amodal verbal working memory system. *Neuroimage, 3,* 79–88.

Sergent, J., Zuck, E., Lévesque, M., & MacDonald, B. (1992). Positron emission tomography study of letter and object processing: Empirical findings and methodological considerations. *Cerebral Cortex, 2,* 68–80.

Shaywitz, B., Shaywitz, S., Pugh, K., Constable, R., Skudlarski, P., & Fulbright, R. (1995). Sex differences in the functional organization of the brain for language. *Nature, 373,* 607–609.

Smith, E. E. & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283,* 1657–1661.

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.

Talairach, J. & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain: 3D proportional system: An approach to cerebral imaging*. New York: Georg Thieme Verlag.

Wise, R., Chollet, F., Hadar, U., Friston, K., Hoffner, E., & Frackowiak, R. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain, 114,* 1803–1817.

Woods, R., Grafton, S., Holmes, C., Cherry, S., & Mazziotta, J. (1998). Automated image registration: I. General methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography, 22,* 141–154.

Zatorre, R., Evans, A., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256,* 846–849.

Zatorre, R., Meyer, E., Gjedde, A., & Evans, A. (1996). PET studies of phonetic processing of speech: Review, replication and reanalysis. *Cerebral Cortex, 6,* 21–30.