NeuroImage **16,** 765–780 (2002)
doi:10.1006/nimg.2002.1131

Meta-Analysis of the Functional Neuroanatomy of Single-Word Reading: Method and Validation

Peter E. Turkeltaub, Guinevere F. Eden, Karen M. Jones, and Thomas A. Zeffiro

Center for the Study of Learning, Department of Neurology, and Department of Pediatrics, Georgetown University Medical Center, Washington, DC 20007

Received September 14, 2001

Intersubject variability and subtle differences in experimental design can lead to variable results in studies of cognitive processes such as reading. To accurately identify the neural processes associated with cognition and sensorimotor processing, meta-analytic methods capable of identifying areas of consistent activation among studies are useful. This paper describes a novel approach for combining published neuroimaging results from multiple studies, designed to maximize the quantification of interstudy concordance while minimizing the subjective aspects of meta-analysis. In this method, a localization probability distribution was modeled for each activation focus obtained from 11 PET studies of reading single words aloud, and the union of these distributions was taken to yield an activation likelihood estimate map for the brain. Significance was assessed via permutation analysis of randomly generated sets of foci. Regions of significant concordance were identified in bilateral motor and superior temporal cortices, pre-SMA, left fusiform gyrus, and the cerebellum. These meta-analytic results were validated by comparison with new fMRI data on aloud word reading in normal adult subjects. Excellent correspondence between the two statistical maps was observed, with fMRI maxima lying close to all meta-analysis peaks and statistical values at the peaks identified by the two techniques correlating strongly. This close correspondence between PET meta-analysis and fMRI results also demonstrates the validity of using fMRI for the study of language tasks involving overt speech responses. Advantages of this automated meta-analysis technique include quantification of the level of concordance at all brain locations and the provision for use of a threshold for statistical significance of concordance.

INTRODUCTION

The ability to compare and combine data from diverse studies is important in the interpretation of functional brain imaging findings and is essential for building consensus in the identification of neuroanatomical correlates of behavior. Especially in the study of cognitive functions such as reading, characterized by high degrees of between-subject variability, a quantitative technique to generate statistical maps of consistent findings from the literature could provide a valuable tool for the design of new studies and the interpretation of existing ones. Maps of consistent findings could principally aid in the identification of inconsistent findings, which are more likely to be related to the specific task parameters of an experiment. Alternatively, maps of consistent areas of activation could be used to generate regions of interest for the testing of new hypotheses. The methods described here represent a novel approach for combining neuroimaging results across studies. This technique was designed to maximize the quantification of interstudy concordance while minimizing the subjectivity of the analytic technique. Three major advantages of this technique over previous meta-analytic methods are (1) the automatization of the analysis, (2) the quantification of the level of concordance in addition to the location, and (3) the use of significance thresholds, providing statistically defensible conclusions.

The most common approach to combining data across studies has been to merge foci from several experiments into a single table or figure in an effort to describe patterns of task-related activity (Buckner and Petersen, 1996; Poeppel, 1996; Owen, 1997). Judgments of concordance or discordance between studies have been largely subjective, based on the proximity of activation peaks or the anatomical labels assigned them from the Talairach Atlas (Talairach and Tournoux, 1988). While these types of studies may be considered forms of meta-analysis, they are typically and more accurately referred to as reviews. This approach allows investigators to apply their knowledge and experience to the interpretation of previous findings and by this virtue is certainly a valuable tool for examination and interpretation of the literature. For quantitative description of previous findings, however, this subjectivity is undesirable.

1053-8119/02 $35.00

Subjectivity in the analysis of previous neuroimaging findings can be introduced at several points (for an excellent review of issues in imaging meta-analysis, see Fox *et al.* (1998)). The first, and perhaps most unavoidable, degree of subjectivity is introduced during the selection of studies to be examined. Thorough criteria for inclusion must be determined prior to literature selection and applied stringently to minimize bias (Fox *et al.*, 1998). Selection is not always a simple matter, as studies which walk the line between inclusion and exclusion or which present unpredicted variants of experimental design will inevitably arise. Further bias can be introduced during the analysis of the pooled data set. Again, clear procedures for analysis set forth prior to data set selection can reduce this bias. Techniques relying on anatomical descriptors or visual examination of data for analysis are necessarily associated with a high degree of subjectivity. It is expected that highly automated, mathematical approaches to analysis will incur the least subjectivity.

Few attempts have been made to analyze merged results of several imaging studies using quantitative and objective approaches (Markowitsch and Tulving, 1994; Picard and Strick, 1996; Grezes and Decety, 2001). While the focus of most of these has been the dissociation of related cognitive processes within individual brain structures (Lepage *et al.*, 1998), a handful have attempted to describe the neural substrates of a single task using meta-analytic techniques (Farah and Aguirre, 1999; Indefrey and Levelt, 2000). For several reasons, single word reading is a task ideally suited for a meta-analysis of this type. A relatively large body of literature examining the neural basis of reading through functional imaging exists, allowing for powerful meta-analyses. Additionally, reading recruits multiple brain regions during performance, but the activation of these regions reported in different studies has been somewhat variable. Thus, the literature has provided a somewhat cloudy picture of the neural systems involved in word reading. Quantitative meta-analyses of the results of these studies could reveal those regions that have been activated consistently and that are more likely to be related to the cognitive processes required for reading, rather than those that are specific to the particular task conditions of a single study. Once this basic reading network is identified, the wealth of task variability used to study reading allows for further meta-analyses, designed to elucidate the relationship between task parameters and activity to determine the specific processing roles of active regions in word reading.

In one such meta-analysis of aloud word reading studies, Fiez and Petersen (1998) eliminated inconsistent activation foci within their data set to reveal areas of convergent activation. The goal of this analysis was to evaluate the reliability of functional neuroimaging by exploring whether similar tasks activate similar brain structures. Nine papers using aloud single word reading tasks were gathered, from which 147 foci were extracted. While the task condition for all studies was aloud reading, the control condition varied widely, including both passive and active tasks. Foci were considered to represent common activation if half of the other papers reported at least 1 focus within 20 mm. The 104 foci meeting this criterion were assigned to clusters based on anatomical boundaries, and the mean location of foci within each cluster was calculated. Eighteen clusters were identified in this manner in the left inferior frontal gyrus, SMA, anterior cingulate cortex, bilateral motor and superior temporal cortex, left extrastriate cortex, and bilateral and midline cerebellum. While this method proved a simple technique for eliminating inconsistent findings, one methodological shortcoming was the reliance on anatomical labels for assignment of foci to clusters. More quantitative and objective meta-analytic techniques could potentially glean more information from published results and provide more statistically defensible findings.

One way to quantitatively evaluate convergence of neuroimaging findings is to model the probability distributions of the locations of activation foci. Although foci are reported as coordinate points, limitations and variations of the image acquisition and analysis techniques and intra- and intersubject variability result in uncertainty in the localization of foci. This is clearly demonstrated by the relative lack of agreement among similar studies and the scattering of foci reported in some meta-analyses and reviews (Poeppel, 1996; Farah and Aguirre, 1999). Given the inherent uncertainty in spatial coordinate determination, each focus is best viewed not as a single point, but as a probability distribution centered about a peak at the reported coordinates. By evaluating the union of these distributions for all brain locations, a map for the entire brain representing the differential likelihood of activation at all locations can be generated. This paper describes such a technique and its application to the evaluation of imaging studies involving aloud word reading. To demonstrate the predictive value of this technique, results were compared to new data from an fMRI study of aloud word reading in normal adult subjects.

EXPERIMENT 1: META-ANALYSIS OF SINGLE WORD READING

Methods

The goal of this analysis was to provide an objective, mathematical review of the corpus of published imaging data on single word reading. For preliminary searches of the literature, the only criterion for inclusion was the use of a brain imaging technique to elucidate the neural basis of single word reading. A preliminary list of articles for the meta-analysis was

obtained through several searches of the Medline database. In addition, the citations within reviews of the functional anatomy of word reading (Fiez and Petersen, 1998; Posner *et al.*, 1999; Price, 2000) were examined for relevant references. Verification of the merit of the meta-analysis methodology required that the findings be compared to those of a statistically powerful study of single word reading. For purposes of comparison with a new fMRI dataset, studies included in this analysis were restricted to those examining aloud single real-word reading. Only those publications reporting task-related activity obtained solely from normal subjects in a standardized coordinate space were included. Descriptions of the 11 papers meeting these criteria follow (See Fig. 1 for additional study parameters):

(1) In the first neuroimaging study on word reading, Petersen *et al.* (1988) investigated the anatomy of single word processing in both the auditory and the visual modalities. Seventeen (11 female, 6 male) right-handed subjects were studied with PET rCBF using seven tasks: fixation, passive viewing, passive listening, reading aloud, auditory repetition, and visual and auditory generate uses. Results were interpreted to support separate visual and auditory word processing networks.

(2) Howard *et al.* (1992) also sought to localize the cortical regions responsible for visual and auditory word recognition. Specifically, the authors hoped to resolve the discrepancy between the classical models of lexical anatomy and the findings of the Petersen *et al.* (1988) paper that placed the visual word form area in the left medial extrastriate cortex. PET-rCBF measurements were obtained from 12 subjects (7 male, 5 female) during aloud reading and auditory word repetition tasks. Control conditions for these tasks were a false-font viewing condition with an oral response of "crime" for each stimulus and a reversed-word listening task with an oral response of "crime" for each stimulus. The authors identified a spoken word lexicon in middle left superior and middle temporal gyri and a visual word lexicon in the posterior left middle temporal gyrus.

(3) Noting stimulus exposure duration differences between the Petersen *et al.* and the Howard *et al.* studies (150 ms vs 1000 ms per stimulus), Price *et al.* (1994) sought to examine whether these differences were sufficient to explain the disparate findings in the two studies. Two experiments were conducted, utilizing PET to measure rCBF patterns with six male subjects in each. The first experiment included six conditions: 150- and 1000-ms exposure duration versions of aloud reading, aloud lexical decision of words, and aloud lexical decision of words and pseudowords. The second experiment consisted of four conditions: 150- and 981-ms exposure duration versions of silent viewing of words and false-font strings. The authors concluded that left middle and superior temporal areas were involved in visual word processing, as these areas were active during both aloud and silent reading. Exposure duration was found to significantly affect task-related activity, with shorter duration conditions producing greater activity.

(4) Bookheimer *et al.* (1995) sought to identify the "direct" and "indirect" pathways of reading using PET rCBF measurements to identify cortical regions active during aloud and silent word reading and object naming. The authors surmised that object naming utilized only the direct route such that aloud object naming activity would resemble silent object naming with the addition of articulatory and auditory areas. Conversely, reading would utilize both direct and indirect pathways such that aloud reading might access a pathway different from that of silent reading, presumably an indirect phonological pathway. Sixteen subjects (8 male, 8 female) were scanned using the aforementioned four conditions plus a random line drawing baseline condition. From their findings, the authors concluded that separate pathways, representing the direct and indirect routes of reading, are accessed for silent and aloud reading, respectively. Silent reading utilized an inferior temporal pathway closely resembling that used for object naming, while aloud reading engaged a network for phonological decoding involving superior temporal and inferior parietal cortex in addition to sensorimotor and auditory association areas. The authors found no evidence for a specific visual word form area.

(5) Price *et al.* (1996a) designed a study to further explore the impact of presentation rate and duration of exposure on task-related activity. Six male subjects were scanned for PET rCBF patterns during aloud and silently mouthed word reading at several presentation rates (20, 40, 60, 120 words per minute) and two exposure durations (150 and 1000 ms). Including 2 rest conditions, a total of 10 conditions were acquired. Different regions of the brain were affected differently by the variations in task parameters. Activity within the posterior fusiform gyri, associated with early visual analysis, and within the cerebellum, SMA, and primary motor cortex, associated with response generation, increased with increasing stimulus presentation rate and duration. Activity within bilateral STG and IPL displayed the opposite effect, while lingual gyrus activity was independent of rate and duration. The authors concluded that the differing effects of rate and duration in different regions of the cortex indicated their functional segregation during word reading.

(6) Price *et al.* (1996b) used a series of five PET experiments to study the neuroanatomy of pronunciation and audition especially in Wemicke's and Broca's areas. Two of these experiments were reanalyses of previously published studies, while three represented

new experimental data. Tasks included different combinations of listening to words, listening to reversed words, repeating heard words, and aloud and silent reading. Task rate was varied to assess differential effects of stimulus presentation rate. The principal findings included a dissociation within the left inferior frontal cortex, with word perception in anterior regions (BA 45) and production in posterior regions (BA 44).

(7) Seeking to clarify the inconsistencies in previous findings of neuroimaging studies of reading, Herbster *et al.* (1997) aimed to explore the brain regions associated with various aspects of oral reading, including semantic processing and regularity effects. PET was utilized to measure rCBF changes in 10 healthy subjects (5 male, 5 female) as they performed six tasks: fixation, unpronounceable letter string viewing, unpronounceable letter string viewing with an oral response of "Hiya," pronounceable nonword reading, regular word reading, and irregular word reading. Regular and irregular word reading activated the left fusiform gyrus, while nonword and irregular word reading activated the left inferior frontal gyrus. Based on these results, which the authors viewed as largely consistent with previous findings, the potential roles of anterior left fusiform cortex in semantic processing and left inferior frontal gyrus in phonological processing were suggested.

(8) Rumsey *et al.* (1997) utilized pronunciation and lexical decision tasks to determine the neural substrates of orthographic and phonological processing. PET was used to measure rCBF while 14 healthy male subjects performed several tasks. Aloud reading of irregular or inconsistent real words was used to emphasize orthographic processing, while aloud reading of pseudowords was intended to emphasize phonological processing. A lexical decision task, in which a real word and a pseudohomophone (e.g., hoal hole) were simultaneously presented and the subjects were asked to determine which was the real word, was used to investigate orthographic processing. To emphasize phonological processing, a lexical decision task was used in which two pseudowords were presented simultaneously (e.g., bape baik) and subjects were asked to determine which sounded like a real word. Fixation was used as a common baseline for all tasks. Only the portion of the brain volume between $Z = -22$ and $Z = 44$ was scanned. While several areas were activated during performance of all four tasks, the left superior temporal gyrus was activated preferentially during pronunciation tasks, and the left inferior frontal cortex was activated during decision making tasks. For each of these areas, the activity observed was greater for phonological tasks than for their orthographic counterparts. Active regions within these areas were then correlated with other brain regions, yielding similar results for phonological and orthographic tasks. These results were considered supportive of connectionist models describing a single network capable of processing pseudowords and real words.

(9) Noting dissociations between explicit memory and priming effects in the literature and evidence that degrading visual word stimuli reduces this dissociation, Jernigan *et al.* (1998) sought to explore the neuroanatomical bases of these phenomenon using PET rCBF measurements. Eight healthy adult subjects (six male, two female) were scanned while performing 12 tasks, including 6 reading tasks, 4 recognition tasks, and 2 fixation tasks. Immediately prior to scanning, subjects were presented with a list of 30 words to read, some appearing normal and some visually degraded. The 6 reading conditions were as follows: undegraded new words, degraded new words, undegraded old words, degraded old words, undegraded old words (study stimuli undegraded), degraded old words (study words degraded). Only one peak of task-related activity was reported from each cluster. While significant differences between recognition and reading conditions were observed, with the former recruiting more prefrontal and cerebellar areas, no effect of repetition was observed in either task. The only significant difference between degraded and undegraded words was in bilateral occipital lobe, with maximal difference in the fusiform gyri. Because no foci of significant signal change for reading conditions were reported in fusiform gyri, and no difference was observed between new and old word sets, these divergences in experimental design from other reading studies did not warrant exclusion of this study from the meta-analysis.

(10) Fiez *et al.* (1999) used PET measures of rCBF during reading of several sets of words to evaluate the impact of word lexicality, frequency, and consistency on the neural substrates of reading. Eleven healthy subjects were scanned (6 male, 5 female) during fixation and during five different aloud reading conditions, each consisting of only one type of word: high-frequency consistent, low-frequency consistent, high-frequency inconsistent, low-frequency inconsistent, or pronounceable nonwords. A lexicality effect was demonstrated in an area of left inferior frontal cortex, while bilateral motor cortex displayed an effect of consistency. The authors discussed several interpretations of this consistency effect, each requiring revisions of current models of reading.

(11) Hagoort *et al.* (1999) used PET to measure rCBF changes during silent and aloud reading of German words and pseudowords. Eleven healthy subjects (8 male, 3 female) were scanned during performance of four reading tasks and rest and fixation control conditions. The middle temporal gyri were activated to a greater extent during real word reading than during pseudoword reading, indicating their possible role in phonological and semantic retrieval. A possible role for the left inferior frontal gyrus in sublexical phonological decoding was indicated by its preferential activation

during pseudoword reading. The authors further proposed that articulatory gestures are controlled by the SMA during reading of high-frequency syllables, while articulation of low-frequency syllables is supported by left premotor cortex.

A variety of different task and control conditions were used in these experiments, and many studies reported data for several comparisons between different task and control conditions. Only comparisons utilizing a task condition of aloud real word reading were included in the present study. Often, several different control conditions were subtracted from the same task condition, and each of these comparisons was reported separately. To avoid including the same task-related activity in the analysis multiple times, only the comparisons with the most basal control condition reported were included (e.g., if a paper separately reported task-related signal change obtained by subtracting a rest condition and a false-font viewing condition from the same word reading condition, only the comparison with the rest condition was included in the analysis). Based on the assumptions of the cognitive subtraction paradigm (Sergent *et al.*, 1992), this allows for the inclusion of data on all the cognitive processes examined in each study without unnecessarily overrepresenting particular processes. Following these guidelines, 172 coordinate points of significant reading-related activity were obtained from the 11 papers evaluated (Fig. 1). These points represent peaks of activity yielded during task performance without imparting any information on extent of the activation. While these maxima alone do not fully portray the patterns of activity observed in each experiment, they represent the best quantitative depiction of the data available for meta-analysis.

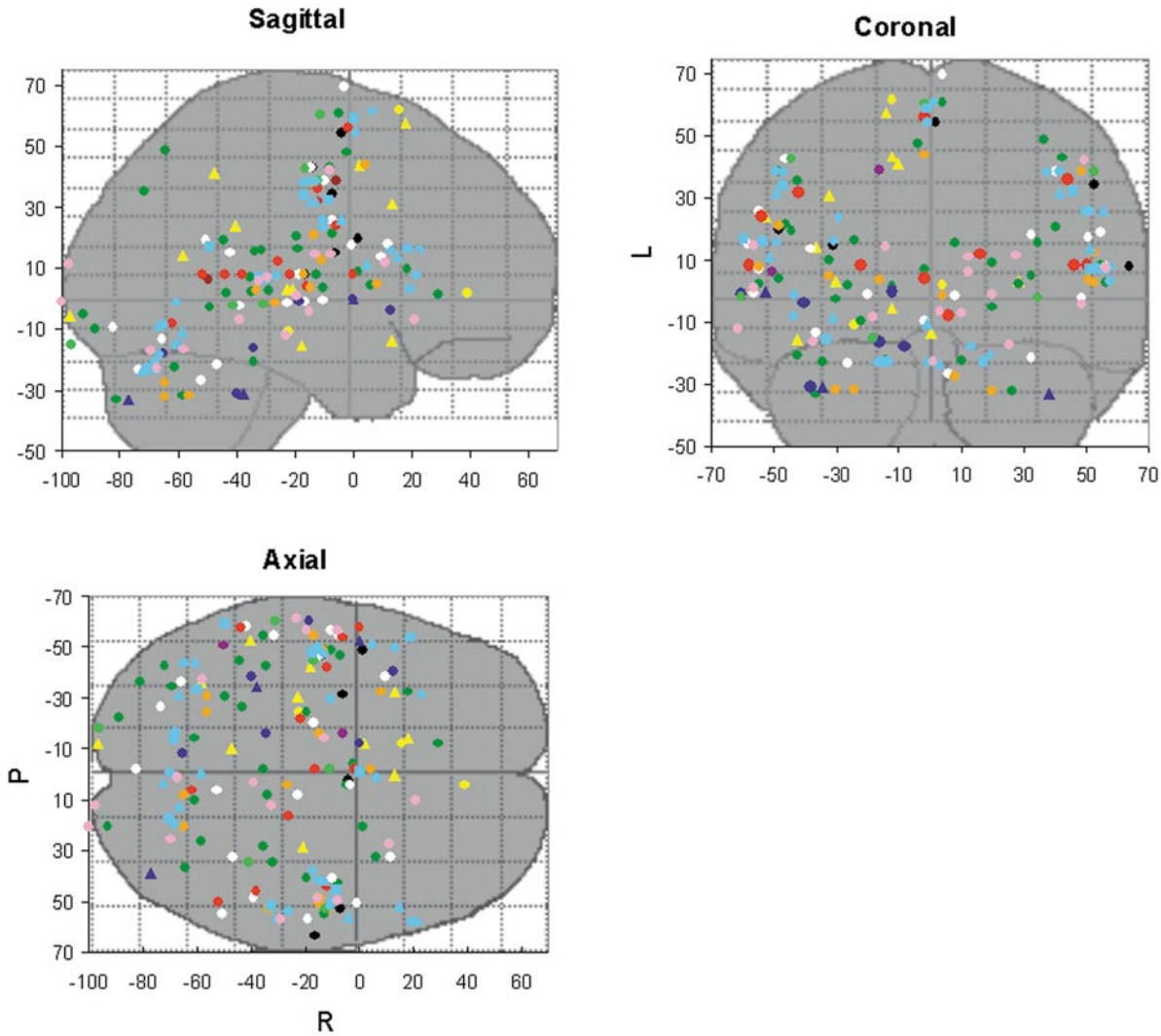
Procedure

To reveal those brain areas consistently activated during aloud single word reading tasks, the coordinates of each of the 172 activation foci were entered into a spreadsheet along with the identity of the study and the task comparison yielding the activity. Based on analysis methods described in the papers, we determined the anatomical template used for spatial normalization. Some controversy has recently emerged as to the correspondence between the Talairach atlas (Talairach and Toumoux, 1988) and the Montreal Neurological Institute (MNI) templates used for spatial normalization in versions of SPM dating from 1996. The MNI templates are somewhat larger than the Talairach brain in the y and z dimensions, leading to errors in anatomical labeling of data normalized to these templates (Brett *et al.*, 2001) using the Talairach atlas. In the case of individual coordinate points, this error can be partially corrected using a set of conversion equations to adjust MNI-normalized coordinates to better fit the dimensions of the Talairach atlas (Brett, 1999). Likewise, these equations can be reversed to convert Talairach coordinates to the "MNI-305 coordinate space." Most of the studies included in this analysis utilized versions of SPM preceding the MNI templates and were considered to be in alignment with the Talairach atlas. Data in Price *et al.* (1996b) had been spatially normalized to the MNI-305 template. Foci from Petersen *et al.* (1988), which utilized a coordinate system from an earlier Talairach atlas (Talairach *et al.*, 1967), with x coordinates reversed from the Talairach (1988) convention and y coordinates measured from the midpoint of the anterior and posterior commisure, were converted to Talairach (1988) coordinates by reversing the sign of the x coordinates and subtracting 11.5 mm from the y coordinates. All foci in Talairach coordinates were transformed to the MNI-305 coordinate space using the equations described above. Three bivariate scatter plots of the points, representing collapsed coronal, sagittal, and horizontal planes of view, revealed a diffuse pattern, with few obvious clusters of task-related activity (Fig. 1).

To reveal interstudy consistencies not obvious upon immediate inspection, a statistical map was generated by using the collection of observed maxima to estimate the likelihood of activation for each voxel in the brain. While specific coordinates for activation foci are provided in neuroimaging studies, technical issues lead to some uncertainty as to the actual location of these peaks. Accuracy and precision of localization is limited by the imaging technique used, smoothing performed on the data, intersubject anatomical variations, and imperfections in the algorithms used to convert the data into standardized anatomical space. For this reason, activation foci are more accurately viewed not as single points, but as localization probability distributions centered at the given coordinates. Following the transformation of the foci into these probability distributions, a whole-brain map can then be created by assigning each voxel within the brain a value equal to the probability that at least one of the points in the data set actually lay within the voxel. For each voxel, this value will be referred to as the "activation likelihood estimate" (ALE).

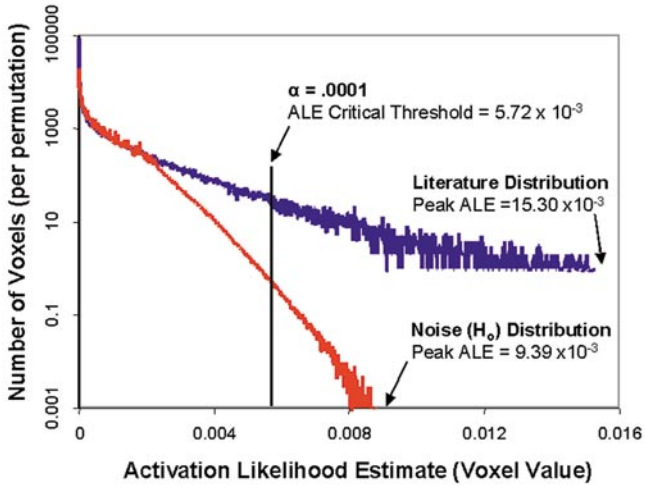
To create the ALE map, MNI-305 space was divided into 2-mm^3 voxels, and the localization probability distributions for the foci were modeled by three-dimensional (3-D) Gaussian functions. The probability that a given focus actually lay within a particular voxel was calculated using the formula for a three-dimensional Gaussian distribution,

$$p = \frac{e^{-d^2/2\sigma^2}}{(2\pi)^{1.5}\sigma^3},$$

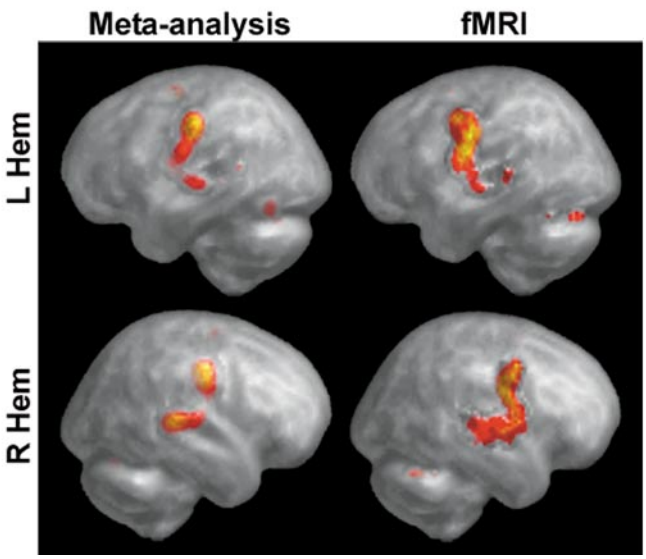


	Paper	Task	n	Within-Plane Res. (mm)	Between-Plane Res. (mm)	Filter (mm)	Critical Threshold	Foci
1	Petersen et al, 1988	read vs silent read	17	18	-	-	p< .03	6
2	Howard et al, 1992	read vs. falsefont aloud ("crime")	12	8	8.5	20	p< .001	2
3a	Price et al, 1994	read vs aloud false font feature det (1000ms)	6	8	8.5	20	p< .001	3
3b		read vs aloud false font feature det (150ms)						11
4	Bookheimer et al, 1995	read vs. random line drawing viewing	16	6.5	-	6^2x10	p< .001	33
5	Price et al, 1996a	read vs. rest (1000ms)	6	6	8.5	20	p< .001	20
6	Price et al, 1996b	read vs rest (40 wps)	6	8	8.5	16	p< .001	12
7a	Herbster et al, 1997	read irregular vs. aloud letter string ("hiya")	10	-	-	16	p< .001	6
7b		read regular vs. aloud letter string ("hiya")						3
8	Rumsey et al, 1997	read vs. fix (low freq, irregular)	14	6.5	5.5	20^2x12	p< .001 & >8 voxels	14
9	Jernigan et al, 1998	read (normal and degraded) vs fix	8	8.5	4.0	16	cor. p<.05 (Z or extent)	6
10a	Fiez et al, 1999	read vs fix (high freq consistent)	11	17	-	-	p< .0005	10
10b		read vs fix (high freq inconsistent)						9
10c		read vs fix (low freq consistent)						9
10d		read vs fix (low freq inconsistent)						11
11	Hagoort et al, 1999	read vs silent read (German)	11	9	9	18	p< .05 & >40 voxels	17
								172

FIG. 1. Plots of activation foci included in the analysis. Foci are shown on three orthogonal projections representing collapsed axial, sagittal, and coronal slices. Studies are coded by color, and different conditions within studies are coded by symbol. Although some level of concordance is apparent, the variability between studies obscures the consistency imbedded in these results.



2



3

FIG. 2. Histograms of activation likelihood estimate (ALE) voxel values within the brain for the word reading literature ALE map (blue) and 1000 noise maps generated from randomly generated foci (red). The omnibus critical ALE threshold was determined by setting the fraction of voxels in the random (null hypothesis) histogram greater than the critical threshold equal to α (0.0001).

FIG. 3. Rendered hemispheres displaying significant voxels for meta-analysis ($P < 0.0001$) and fMRI ($Z > 4.89$, matched to meta-analysis for sensitivity) results. Activation is overlaid on an EPI image transformed to the MNI-305 template and projected toward the cortex for visualization.

where d is the Euclidean distance between the center of the voxel and the focus, and σ is the standard deviation of the distribution. A value of 6 mm was chosen for σ, giving a FWHM of 15 mm. This distribution width was selected somewhat arbitrarily to fall in the range of the smoothing filters used in the PET studies included in this analysis. Results of the analysis appeared robust to variation of this width between 10 and 20 mm. As the above equation gives the point probability of localization at the center of a given voxel, this value was multiplied by a factor of 8 mm^3 to yield an estimate of the probability that a given focus lay within the voxel. For each voxel, the probability associated with each of the 172 data points was calculated using this formula and the union of these probabilities was calculated to give the activation likelihood estimate. This ALE score represents the probability that at least one of the activation foci lies within a given voxel. By designating a value equal to this probability for each voxel in space, an activation likelihood estimation map was created for all of MNI-305 space.

The ALE map was imported into MEDx (Sensor Systems, Sterling, VA) for further analysis. The map was masked using a thresholded EPI template normalized to MNI-305 space. To differentiate the voxels within

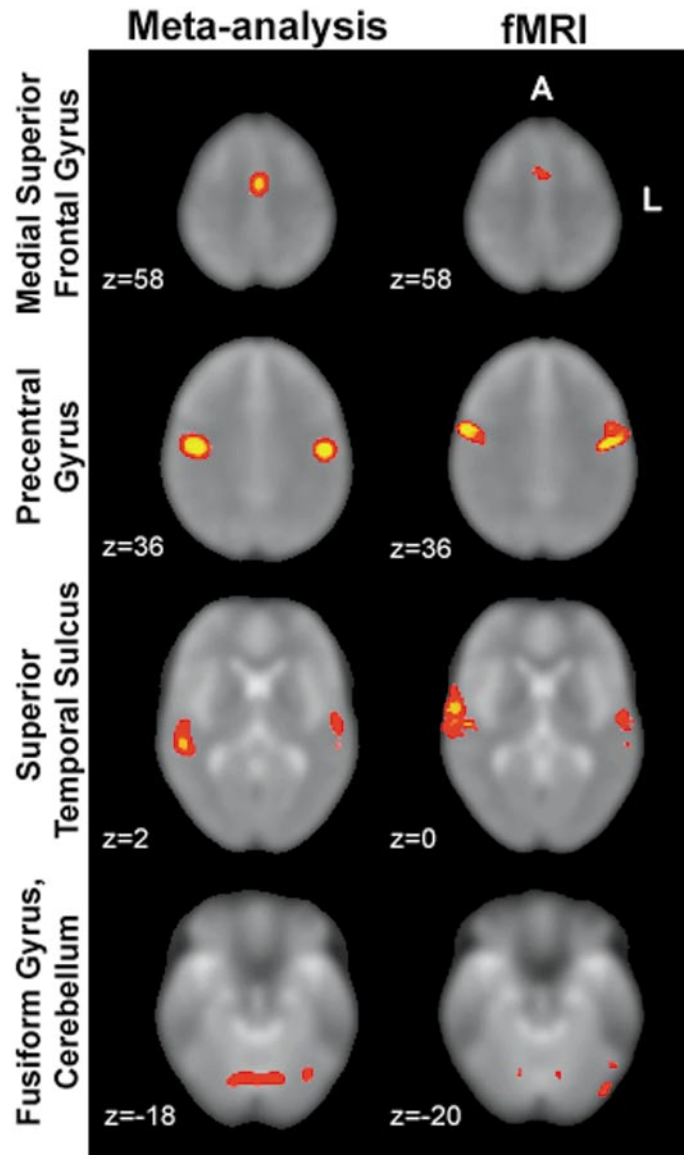


FIG. 4. Selected slices illustrating patterns of significant activation for meta-analysis ($P < 0.0001$) and fMRI ($Z > 4.89$, matched to meta-analysis for sensitivity). Cortical activation seen in meta-analysis tends to lie farther from the brain surface than fMRI activation.

the map that represented signal (i.e., nonrandom clustering of foci) from those that represented noise (i.e., random clustering), ALE values were compared to those obtained from random sets of foci. For this permutation analysis, 1000 sets of 172 random coordinates within the mask volume were generated and processed identically to the literature data. Histograms of all voxel values obtained throughout the brain volume were calculated for each random data set and the 1000 histograms were averaged to obtain a single histogram representing the noise distribution of ALE values. This histogram was utilized as a null hypothesis distribution against which the significance of ALE values within the single word reading localization probability map were evaluated (Fig. 2). Using the omnibus single threshold test described by Holmes *et al.* (1996) the null hypothesis (random distribution of foci) was rejected ($P < 0.001$) for the ALE map obtained from the single word reading literature. This statistical test obtains *P* values for a statistical map against permutations of null hypothesis maps by ranking the maximal voxel value in the statistical map against the maximal voxel values obtained from the null hypothesis simulations. In this case, the maximal ALE value in the statistical map derived from the literature (0.015) was greater than the maximal values of all 1000 null hypothesis simulations (the maximum of all 1000 was 0.009). To identify the particular voxels within the statistical map for which the null hypothesis could be rejected, the probability of obtaining ALE values greater than a given threshold under the null hypothesis was calculated as the fraction of the total voxels in the noise histogram greater than that threshold. Thus, the ALE value corresponding to a chosen α for the statistical test was the value with 100*α% of the area under the noise histogram to its right. The null hypothesis (random distribution of foci) was rejected for any voxels in the statistical map with ALE values greater than the α threshold. This omnibus method of significance thresholding is similar to the nonparametric single threshold test described by Holmes *et al.* (1996). Whereas the single threshold test compares voxel values only to the maximal values obtained from each randomization, this method compares them to all values obtained from all randomizations. The omnibus significance thresholding also bears similarities to change-distribution analysis (Fox *et al.*, 1988; Fox and Mintun, 1989) in that it determines significance against a noise distribution by evaluating the likelihood of obtaining statistic values independent of their locations within the brain. Both techniques obtain histograms of the values of voxels of interest and determine a significance threshold throughout the brain based on the difference between the data histogram and a noise histogram. Whereas the population of values considered in change-distribution analysis consists of only local maxima within the brain, this method considers all voxels within the brain.

A conservative α of 0.0001 was selected to reduce type I error, and the corresponding ALE threshold value of 5.72×10^{-3} was identified using the thresholding technique described above. Local maxima meeting this significance threshold were identified also, and coordinates were transformed to Talairach coordinate space (Brett, 1999).

Results

Because the results were derived from a reanalysis of published foci, typically representing peaks of task-related activity, the discussion and interpretation of those results will be limited to the peaks in the activation likelihood estimation map. Omnibus significance thresholding could allow discussion of the extent of significantly nonrandom ALE values within the statistical map. However, interpretation of spatial extent would be premature until the influence of parameters such as the FWHM of the Gaussian model on extent is more fully evaluated. Thus, discussion of extent is here limited to areas not exhibiting maxima for which an a priori hypothesis of activation during reading exists.

The highest activation likelihood estimate values in the brain volume were observed in primary motor cortex. The ALE values at the local maxima within these regions were approximately 0.015. This represents a 1.5% chance that 1 of the 172 foci included in the analysis actually lay within the voxel in question. This small absolute likelihood is due to the extremely small volume of brain tissue contained within each voxel. By comparison, the highest ALE value identified in the 1000 noise distributions used to generate the null hypothesis histogram was 0.009. Thus, while the absolute activation likelihood estimates associated with the individual voxels were very small, it could be said with confidence that they represented nonrandom clustering of activation foci within the data set.

Eleven significant local maxima were identified in the ALE map (Table 1). These were located in bilateral primary motor cortex, midline superior frontal gyrus, bilateral superior temporal sulci, left fusiform gyrus, and bilateral cerebellum. The peaks in primary motor cortex lay approximately 20 mm superior to the Sylvian fissure. While no peaks were identified in the left inferior frontal gyrus, significant activity contiguous with the left motor cortex extended anteriorly into the posterior portion of Brodmann's area (BA) 44. This pattern was not observed in the right hemisphere, where motor cortical activity remained confined to BA 4 and 6. The midline superior frontal gyrus peak was just anterior to the level of the anterior commisure, on the border between the pre-SMA and the SMA (Picard and Strick, 1996). This peak was located exactly on the midline with a virtually symmetric pattern of activity

TABLE 1

Locations of Significant ALE Maxima ($P < 0.0001$)

	x	y	z	ALE ($*10^{-3}$)	Petersen *et al.*, 1988	Howard *et al.*, 1992	Price *et al.*, 1994	Bookheimer *et al.*, 1995	Price *et al.*, 1996a	Price *et al.*, 1996b	Herbster *et al.*, 1997	Rumsey *et al.*, 1997	Jernigan *et al.*, 1998	Fiez *et al.*, 1999	Hagoort *et al.*, 1999
L Precent G (4/6)	−48	−12	36	15.29	X			X	X					X	
L Sup Temp S (22/21)	−55	−15	3	7.14				X	X		X	X			X
L Sup Temp S (22/21)	−55	−33	5	5.79				X	X				X		
L Post STG	−57	−42	13	5.91					X	X				X	
L Fus G (19/37)	−36	−61	−9	7.68					X					X	X
L Thalamus (VPL)	*−20*	*−17*	*5*	*5.60*					X	X		X			
L CB	−14	−65	−15	9.45				X			X			X	
R Med Sup Fr G (6)	0	1	52	11.58	X			X		X				X	
R Precent G (6)	44	−10	34	15.02					X	X		X		X	
R Sup Temp S (22/21)	53	−29	7	12.75				X				X		X	X
R CB	12	−65	−15	8.46				X				X		X	
Med CB	4	−67	−17	8.49								X		X	X

Note. For each maximum, studies contributing greater than 1/11th of the total ALE value of that maximum are marked with an X.

and thus could not be said with any certainty to lie in one hemisphere or the other. Within the temporal lobe, maxima were identified bilaterally within the posterior superior temporal sulci. Additional peaks were identified anteriorly and farther posteriorly in the left superior temporal cortex. The left fusiform gyrus peak lay at the border of BA 19 and 37. Three peaks lay within cerebellar lobule VI, one medial and two somewhat more lateral. A local maximum in the left thalamus just failed to reach significance ($P = 1.4 \times 10^{-3}$). While discrimination between thalamic subnuclei is likely beyond the spatial resolving power of this technique, this peak was localized to the left ventral posterolateral nucleus. No significant voxels were located in the region of the angular gyrus.

As the purpose of the analysis was to reveal locations of consistent reading-related activity within the literature, the level of contribution of each of the 11 studies included in the analysis to each of the significant maxima was of interest. Optimally, each peak would have derived from equal contributions from all papers, each providing 1/11th of the total probability at the peak. Were this the case, the regions exhibiting consistent activity would be obvious, making this sort of analysis unwarranted. Indeed, nothing approaching this extreme of interstudy concordance was observed. Neither were any of the significant maxima overwhelmingly derived from a single study. In fact, contributions of greater than 1/11th of the total ALE value were made by at least 3 papers for all 11 significant peaks and the subthreshold thalamic peak (Table 1). Applying the criterion for interstudy concordance used by Fiez *et al.* (1998), 11 of the 12 peaks exhibited foci within 20 mm from at least half the studies. The left fusiform gyrus peak, with foci from 5 papers within 20 mm, was the only one not to meet this criterion. Two papers, Howard *et al.* (1992) and Price *et al.* (1994) did not contribute major portions of any of the significant maxima. Only 2 foci from the Howard *et al.* (1992) study were included in the analysis, making its small contribution to the significant maxima not entirely surprising. The Price *et al.* (1994) study, however, contributed 14 foci to the analysis. The somewhat unconventional control condition used in this study, a false-font feature detection task, may have resulted in activation patterns dissimilar to those observed in other studies.

EXPERIMENT 2: fMRI STUDY OF ALOUD WORD READING

Methods

To evaluate the predictive value of the meta-analysis, results were compared to those from a large fMRI study using an aloud single word reading task. Thirty-two adult subjects (15 female, 17 male) were included in this study. Age range was 19 to 55 years and mean age was 33.0 years. Subjects had good reading and phonological awareness skills defined as scores greater than 18 on the Word Attack subtest and greater than 42 on the Letter-Word identification subtests of the Woodcock Johnson Reading Mastery III-Revised (Woodcock and Johnson, 1987) and greater than 82 on the Lindamood Auditory Conceptualization Task (Lindamood and Lindamood, 1979). All subjects were right-handed based on scores greater than 70 on the Edinburgh Handedness Scale (Oldfield, 1971). All participants had negative neurological and psychiatric histories and no individual had a familial history of learning disability. Participants were medication-free at the time of the study and were monolingual native English speakers. Normal vision was assessed prior to participation using tests of visual acuity, depth perception, and color vision.

Sixteen subjects participated in four experimental runs, while the remaining 16 subjects participated in only one. Each run consisted of alternating blocks of aloud single word reading, a fixation baseline condi-

tion, and a phonological manipulation task not considered here. All words presented were high-frequency, single-syllable words. The words were presented at the center of the screen for 600 ms and then replaced by a fixation cross. Words were presented in large, lowercase, high-contrast print. Under all task conditions, subjects gave a spoken response, which was recorded via a microphone and scored for accuracy. To avoid susceptibility artifacts associated with speech responses, we used the "behavior-interleaved gradient" (BIG) technique (Eden *et al.*, 1999), allowing data acquisition following completion of the task. To minimize acoustic contamination, the subjects read words during a period of gradient silence and heard the gradient noise only after completing the task. Subjects performed the task for a 9-s period during which no data were acquired, followed by a 4-s interval of no task performance during which one whole-head volume was acquired. This condition resulted in data acquisition after the initiation and maximization of the hemodynamic response (Eden *et al.*, 1999). Three stimuli (written words) were presented for 600 ms each within the 9-s stimulus presentation period. In each task block, this 13-s trial was repeated five times, yielding five brain volumes (data points) in 65 s. A 26-s rest period occurred between each task (phonological manipulation or read) during which subjects viewed a fixation cross. With one brain volume acquired every 13 s, two fixation (rest) volumes were collected during this period. Fixation epochs alternated with phonological manipulation and reading epochs. Thus, the sequence of epochs during each run was as follows: fixation, read, fixation, phonological manipulation, fixation, read, etc. A functional run consisted of four epochs of the phonological manipulation task, four epochs of word reading, and eight epochs of fixation, making the total run time 13 min. In this manner, 20 whole-head volumes per condition were collected for each run. Over the four runs, 80 head volumes were collected from each subject for the reading condition and 80 for the fixation condition.

MRI scans were performed on a 1.5 Tesla Siemens Vision MRI System with a circularly polarized head coil, employing multislice echo-planar image (EPI) acquisition. EPI acquisition parameters were 40 ms echo time (TE), 13 s repetition time (TR), 64 $\times$ 64 matrix, 192 mm field of view (FOV), with 50 ascending slices of 3 mm thickness. These parameters allowed coverage of the entire brain with 3-mm^3 voxels.

All runs from each subject underwent the following processing using MEDx (Sensor Systems): head motion detection and correction (Woods *et al.*, 1998a,b), volume intensity normalization, spatial filtering (9 mm FWHM), high-pass temporal filtering (390-s cutoff), and statistical map construction (for details see Eden *et al.* (1999)). Task-related changes in signal were detected by contrasting reading and fixation conditions. We spatially registered the statistical maps to a high-resolution structural volume from the same subject, to identify the neuroanatomical localization of the observed activity. For each subject, the time courses of significant voxels were examined for confirmation of the task-related nature of the activity identified by the statistical techniques. After spatial normalization to an EPI template in the MNI-305 coordinate system, a mixed-effects analysis was performed using a single group *t* test of reading-minus-fixation mean-difference images. Statistical map probability values were converted to corresponding *Z* values with respect to the standard normal distribution. Local maxima were identified within the resulting statistical map.

Results

Discussion of the results of the fMRI experiment will be limited to local maxima for the purpose of comparison with the meta-analysis results. Because the purpose of the fMRI study was for comparison with the meta-analysis results, a significance threshold was selected for the fMRI data to yield the same number of significant voxels as the meta-analysis, thus matching the sensitivity of the two techniques. This value corresponded to an uncorrected *Z* score of 4.89 (corrected $P = 1.7 \times 10^{-3}$).

Areas displaying significant task-related activity included left inferior frontal gyrus, right pre-SMA, bilateral motor cortex, bilateral superior and middle temporal cortex, left inferior temporal and fusiform gyri, the superior colliculus, and the cerebellum (Table 2). Multiple local maxima were identified in many of these regions, 28 in all. One maximum was identified within the left inferior frontal gyrus at the posterior edge of BA 44. The activity in the medial frontal cortex lay 10 mm anterior to the anterior commisure, clearly in the right pre-SMA. Both the right and the left motor cortex were activated, each exhibiting several local maxima within this extent. In both hemispheres, maxima were identified in primary motor cortex (BA 4) and premotor cortex (BA 6). Maxima were scattered through long strips of superior and middle temporal cortex bilaterally also. Anteriorly, a left hemisphere maximum was located in BA 38, in the superior temporal gyrus. One peak lay in the right middle temporal gyrus inferior to primary motor and somatosensory cortex. The superior temporal sulcus was also active more posteriorly in both hemispheres, extending to Wernicke's area in the left hemisphere. On the ventral surface of the temporal lobe, the left ventral stream cortex was active at the border of the fusiform and inferior temporal cortex between BA 37 and 19. One maximum was also observed at the right inferior temporal gyrus (BA 37). The cerebellum exhibited bilateral paravermal activity, and a local maximum in the left thalamus just failed to reach significance ($Z = 4.71$).

TABLE 2

Locations of Significant Maxima for Meta-Analysis and fMRI Study

	Left hemisphere								Right hemisphere							
	Meta-analysis				fMRI				Meta-analysis				fMRI			
	x	y	z	ALE ($*10^{-3}$)	x	y	z	*Z* score	x	y	z	ALE ($*10^{-3}$)	x	y	z	*Z* score
Frontal																
IFG (44)					−42	13	21	4.90								
Med Sup Front G (6)									0	1	52	11.58	4	11	55	5.84
Med Sup Front G (6)													−2	12	45	5.21
Precentral G (6)					−50	4	42	6.39					55	2	33	6.95
Precentral G (4/6)					−51	−3	22	7.32					53	−1	18	6.78
Precentral G (4/6)	−48	−12	36	15.29	−46	−8	32	7.18	44	−10	34	15.02	48	−3	24	7.31
Temporal																
Sup Temp S (21/38)					−48	9	−11	5.00								
Sup Temp S (21/22)													61	−4	0	6.25
Mid Temp G (21)													67	−6	−8	4.90
Sup Temp S (21/22)	−55	−15	3	7.14	−55	−12	−1	5.72					63	−14	1	5.93
Sup Temp G (22)													53	−16	1	6.18
Sup Temp S (21/22)									53	−29	7	12.75	55	−23	5	5.65
Trans Temp G (41)					−40	−30	13	4.96								
Sup Temp S (21/22)	−55	−33	5	5.79	−59	−27	5	5.35								
Sup Temp G (22)	−57	−42	13	5.91	−63	−31	9	5.42								
Inf Temp Sulc (37)					−46	−51	−13	5.12					57	−50	−19	5.10
Inf Temp Sulc (37)					−50	−57	−12	5.18								
Fusiform G (19/37)	−36	−61	−9	7.68	−36	−61	−19	5.11								
Fusiform G (19)					−44	−74	−13	5.82								
Other																
Thalamus	*−20*	*−17*	*5*	*5.60*	*−16*	*−22*	*−6*	*4.71*								
Cerebellum					−36	−52	−21	5.03								
Cerebellum	−14	−65	−15	9.45	−12	−63	−14	5.20	12	−65	−15	8.46	28	−57	−21	6.38
Cerebellum									4	−67	−17	8.49				

Note. $P < 0.0001$ for meta-analysis; $Z > 4.89$ for fMRI. The significance threshold for fMRI data was selected to match the meta-analysis results for sensitivity (i.e., number of significant voxels). The thalamic maxima just failed to reach significance in both techniques.

META-ANALYSIS AND fMRI CONCORDANCE

Significant maxima within the statistical maps for the meta-analysis and fMRI are presented together in Table 2. The activation patterns for the two techniques appear very similar upon visual inspection (Figs. 3 and 4). Indeed, the major difference between the two sets of results was a greater number of peaks within active regions for the fMRI data. For every region exhibiting significant peaks in the ALE map, peaks were also identified in the fMRI results. The only regions displaying fMRI but not meta-analysis peaks were the left inferior frontal gyrus, the left anterior superior temporal gyrus (BA 38), and the right inferior temporal gyrus (BA 37). Of these regions, significant nonpeak voxels in the ALE map were observed only in the left posterior inferior frontal gyrus. Peaks in this region were likely smoothed over by the high ALE values in the adjacent motor strip. The mean distance from all meta-analysis maxima to their closest fMRI maxima was 10.15 mm (SD = 4.7 mm) (Table 3). The largest distance between a meta-analysis peak and its nearest fMRI peak was 18.3 mm, observed in the right paravermal region of cerebellum. The meta-analysis peak lay at an x coordinate of 12, while the nearest fMRI peak lay more laterally, at x = 28. The smallest distance between meta-analysis and fMRI peaks was 3.2 mm, observed in the left cerebellum. To determine whether the most proximate meta-analysis and fMRI maxima did indeed represent analogous peaks, the ALE values and *Z* scores of these peaks were submitted to linear regression (Fig. 5). A strong correlation was observed ($r = 0.797$, $P = 1.9 \times 10^{-3}$), indicating that meta-analysis peak ALE values are reliable predictors of *Z* scores of the nearest fMRI peaks.

Compared to the nearest fMRI maxima, the cortical ALE peaks lay approximately 2 mm closer to the center of the brain volume, a trend which just failed to achieve significance (ALE M = 55.6 mm; fMRI M = 57.3 mm; $P = 0.06$). The extent of significant meta-analysis activity appeared shifted a small distance toward the center of the brain also (see Fig. 4). This inward drift of meta-analysis peaks may be due to differences in the

TABLE 3

Significant Meta-Analysis Maxima and Nearest Significant fMRI Maxima

	Meta-analysis				fMRI				
	x	y	z	Prob (*10^{-3})	x	y	z	*Z* score	Dist (mm)
L Precent G (4/6)	−48	−12	36	15.29	−46	−8	32	7.18	5.70
R Precent G (6)	44	−10	34	15.02	48	−3	24	7.31	12.67
R Sup Temp S (22/21)	53	−29	7	12.75	55	−23	5	5.65	6.42
R Med Sup Fr G (6)	0	1	52	11.58	4	11	55	5.84	11.10
L CB	−14	−65	−15	9.45	−12	−63	−14	5.20	3.24
Med CB	4	−67	−17	8.49	−12	−63	−14	5.20	16.65
R CB	12	−65	−15	8.46	28	−57	−21	6.38	18.34
L Fus G (19/37)	−36	−61	−9	7.68	−36	−61	−19	5.11	10.08
L Sup Temp S (22/21)	−55	−15	3	7.14	−55	−12	−1	5.72	5.24
L Post STG	−57	−42	13	5.91	−63	−31	9	5.42	13.58
L Sup Temp S (22/21)	−55	−33	5	5.79	−59	−27	5	5.35	7.04
L Thalamus (VPL)	*−20*	*−17*	*5*	*5.60*	*−16*	*−22*	*−6*	*4.71*	*11.79*
								Mean	10.15
								St Dev	4.73

Note. Dist (mm) refers to the Euclidean distance between the meta-analysis maxima and the nearest fMRI maxima. The left thalamus maxima just failed to achieve significance in both techniques.

PET-rCBF and fMRI-BOLD signals, or it may be an artifact of the meta-analysis procedure (see Discussion).

DISCUSSION

This analysis was successful in identifying regions of consistent activation in PET studies of single word reading and predicting the location and magnitude of activity in a new fMRI study. The correspondence between the magnitude of the significant peaks in the two statistical maps is interesting as the magnitudes of activation foci from the literature were not used in generation of the ALE map. A logical interpretation of this finding is that certain brain regions reliably activate more strongly than others during aloud word reading, that the relative magnitude of the activity is similar for PET and fMRI, and that the more strongly an area activates, the more likely it is to be reported by multiple studies.

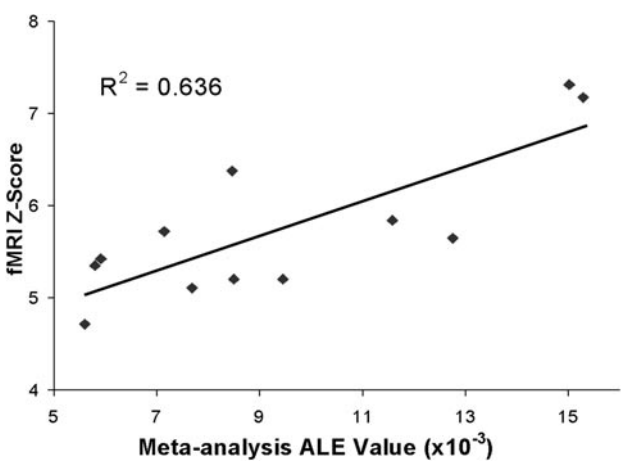


FIG. 5. Linear regression between activation likelihood estimate (ALE) values of significant meta-analysis maxima and *Z* scores of the nearest fMRI maxima (see Table 3). The Pearson correlation coefficient of the relationship between these measures is 0.797. The slope of the regression line is significantly nonzero at $P = 0.002$.

While the location and level of the most significant maxima in nearly all brain regions activated in the fMRI study were predicted by the meta-analysis, many less significant peaks within those regions were not specifically predicted. This may have been due to the lower resolution of rCBF PET compared to fMRI, the difference in the smoothness of the statistical maps (15 mm vs 9 mm), or the variation in task parameters and statistical methods among the included PET studies. These same sources of variance, which produced the wide scatter of foci illustrated in Fig. 1, are likely responsible for much of the difference between the fMRI and the meta-analysis results. It is, in fact, somewhat remarkable that reported findings from PET studies using a wide variety of control conditions yielded accurate predictions of activation maxima from an fMRI study using only a reading vs fixation comparison. Five of the studies contrasted the reading task with a resting or fixation baseline, and resultant activity would be expected to represent all the subprocesses required for reading single words. One study controlled only for visual input, with a random line drawing control, while two others controlled for visual information and vocal responses, isolating orthographic analysis and phonological decoding processes. Two studies used silent reading controls, isolating phonological assembly and articulatory processes. The one remaining study utilized a false-font feature detection control task, which controlled for visual input, and visual anal-

ysis, but was not closely matched to the reading task. Overall, this collection of contrasts likely reduced the weighting of areas responsible for the initial visual analysis processes required for reading.

Still, the close correspondence between the ALE and the fMRI *Z* maps serves to validate not only the meta-analysis methods presented here, but also the use of fMRI to study brain function during performance of overt speech tasks. While there has been some concern that fMRI is poorly suited for use in language studies, these results indicate that, through use of paradigms such as the BIG technique to alleviate motion susceptibility artifact and gradient noise interference during scanning (Eden *et al.*, 1999), fMRI tasks requiring spoken responses can yield blood oxygen level-dependent (BOLD) activation patterns which closely resemble activation from PET-rCBF studies.

The good agreement between the results of this analysis and those of the Fiez and Petersen (1998) analysis of aloud word reading studies serves as further validation of the ALE method. This agreement is not surprising considering the substantial overlap between the studies included in the two analyses. In fact, all nine studies included in the Fiez and Petersen analysis were used for this study also. Additional foci included in our analysis were obtained from two other studies and from published data from Fiez *et al.* (1999) rather than the prepublication foci used for the Fiez and Petersen analysis. The Fiez and Petersen analysis revealed 18 clusters, while the analysis presented here yielded 11 significant maxima. Ten of the 11 significant maxima identified by the current analysis were also identified as centers of clusters in the Fiez and Petersen analysis. The subthreshold left thalamus peak was identified by the Fiez and Petersen analysis also. The distances between the maxima and the cluster centers were extremely small (mean = 4.9 mm), supporting the reliability of both techniques. The 11 analogous Fiez and Petersen cluster centers were slightly farther from the fMRI maxima than were the ALE maxima (mean = 12.0 mm vs 10.2 mm). Of the 7 additional cluster centers identified by Fiez and Petersen, 2 corresponded to significant maxima in the fMRI data. One of these lay in the inferior portion of the left motor strip, and the other lay in the left inferior frontal gyrus. Voxels of significant probabilities were observed in the ALE map in both of these locations. Another cluster center lay near a subthreshold fMRI maximum in the superior colliculus. A subthreshold peak in the ALE map was also located in the same region. Reducing the significance threshold of the fMRI *Z* map further did not reveal additional local maxima near the 4 additional cluster centers identified by Fiez and Petersen. The ALE peak not identified as a cluster center lay close to an fMRI peak in the left superior temporal sulcus. As the Fiez and Petersen analysis identified a greater number of active regions than the analysis presented here, a main difference between these two analyses could be stated as a greater specificity of the activation likelihood estimate analysis and a greater sensitivity of the analysis described in Fiez and Petersen (1998).

The regions identified by this meta-analysis are similar to those recognized by other reviews of the functional anatomy of reading (Friedman *et al.*, 1993; Fiez and Petersen, 1998; Price, 2000). We refer readers to these papers for discussions of the neural circuitry responsible for reading and the roles of these areas within that circuitry. The lack of consistent activation of the left angular gyrus, a replication of the Fiez and Petersen (1998) results, is worthy of note, however, as this area was long associated with memory for word forms due to results from studies involving the effects of focal brain injury on reading (Friedman *et al.*, 1993; Price, 2000). While studies of aloud real word reading have not activated this area, a study of functional connectivity during various reading tasks (Horwitz *et al.*, 1998) found greater connectivity between the left angular gyrus and other classical left hemisphere language and visual areas for normal readers than for developmental dyslexics. These findings are consistent with a role for the angular gyrus in deeper levels of word processing such as access to semantic content (Price, 2000), which may not be engaged to a large degree during simple recognition and pronunciation of known words.

Despite the correspondence between the findings of the ALE meta-analysis and those of the Fiez and Petersen (1998) study and the new fMRI study reported here, the overlap was not perfect. Several sources of error undoubtedly contaminated the findings of the meta-analysis, some inherent in the current technique and others resulting from study selection criteria. Two sources of bias in selection of the activation foci were differences in the number of reported foci in the studies and inclusion of multiple tasks from single studies. For cases in which a study reported findings from multiple tasks performed by the same subjects, all foci derived from different task conditions were included. Because the tasks were only subtly different in most cases, the regions activated by these studies were given disproportionate sway over the outcome of the analysis. The studies included in this analysis also varied widely in the number of foci reported, from 2 to 39. Because all foci were given equal influence over the ALE map, studies reporting more foci had more leverage over the resulting map. Design parameters, such as number of subjects, statistical models, and critical thresholds, vary greatly between imaging studies, and differences in these parameters contribute to the number of foci reported. Larger sample sizes will generate more statistically powerful results and hence a greater number of foci. Alternatively, studies employing fixed-effects models, small-cluster-size thresholds, or liberal signif-

icance thresholds will report more foci and will influence the results of the ALE analysis more than those employing mixed-effects models or more conservative parameters.

Concerns over bias introduced by differences in study parameters could be addressed in several ways. More stringent criteria for inclusion could be applied. Alternatively, to resolve the issue of variance in statistical thresholds between studies, the meta-analytic technique could be adjusted to give each study an equal contribution to the total probability within the ALE map. Currently, each focus included in the analysis is assigned a probability of 1 that it lies somewhere within the brain, and that probability is distributed throughout the brain volume with Gaussian weighting toward the reported coordinates. To control for differences in critical thresholds, each study could be assigned a probability of 1, and this probability could be distributed among its reported foci. The noise distributions could easily be adjusted to reflect this weighting also, yielding a result which would not give undeserving weight to studies employing liberal methods, but would also not give appropriate weight to more powerful studies. Another adjustment to this technique could weight the contribution of each study to the ALE map based on design parameters such as number of subjects. However, as Fig. 1 reveals, the relationship of study parameters such as resolution, sample size, or critical threshold to number of reported foci is not always predictable. The two 1996 Price *et al.* studies employed the same sample sizes and critical thresholds and used similar tasks, but the study utilizing a wider spatial filter reported eight more foci than the other. Other differences in number of reported foci seem to be based mostly on the control condition used, with more active controls resulting in fewer foci, as in Howard *et al.* (1992). Providing additional weight to foci from these studies would unduely emphasize certain aspects of aloud reading, such as articulation. As the influence of these parameters on number of reported maxima is difficult to quantify, this analysis was conducted with the assumption that all studies were roughly equal in quality and rigor. Thus all reported foci were treated equally. Table 1 reveals that nine of the studies included made relatively large contributions to the significant activation likelihood maxima, and some studies with relatively few reported foci contributed to several of the maxima. Given the correspondence of these findings with the new fMRI data reported here, bias based on this assumption did not likely invalidate the results to any large degree.

One consistent difference between the ALE and the fMRI statistical maps is the tendency for maxima to be more centrally located in the meta-analysis. This difference is small (2 mm), but very nearly significant ($P = 0.06$). One explanation for this discrepancy is the suggestion that PET tends to preferentially detect task-related activity in the fundi of cortical sulci, rather than gyral surfaces (Markowitsch and Tulving, 1994). The same tendency has not been demonstrated for the BOLD fMRI signal, and the influence of draining veins on the location of BOLD activity is still under investigation (Boxerman *et al.*, 1995). A study comparing single-subject PET and BOLD fMRI activation during a unimanual motor task (Kinahan and Noll, 1999) demonstrated a consistent dorsal shift in motor cortex fMRI activity compared to PET with no shift in transverse directions, a difference attributed to the weighting of fMRI activation toward the large veins draining dorsally into the sagittal sinus. While the motor cortex activity in the fMRI data presented here does appear to extend more dorsally than the meta-analysis findings, it is more obviously shifted laterally and anteriorly, as demonstrated by the locations of the local maxima.

Another possible explanation for the discrepancy is that a systematic source of error in the analysis method results in bias toward centrality of high values in the ALE map. Published foci of activation will necessarily be located within the brain volume and could theoretically arise from any voxel within the brain. Voxels deep within the brain are thus surrounded in all directions by large numbers of candidate locations for published foci. In contrast, for voxels nearer the cortex, relatively few candidate locations exist in the direction of the nearest brain surface. Because of this, the likelihood of achieving a high ALE value in this analysis increases with movement from the cortex toward the center of the brain. This might result in a tendency for peaks within the ALE map to drift inward from the cortex. Several possible correction techniques of varying theoretical rigor and computational demand could reasonably compensate for this bias. One correction technique would utilize different probability models for foci in different locations of the brain. Rather than the current 3-D Gaussian applied uniformly throughout the brain, a more accurate if considerably more complex model would be asymmetric, with the degree of asymmetry increasing nearer the cortex and ventricles. Foci surrounded by more brain tissue would be modeled by highly Gaussian functions, while those nearer the cortex or ventricles would conform to the contours of these surfaces.

Another correction method would employ voxelwise null distributions rather than omnibus significance thresholding. Multiple permutations of the noise condition could be used to generate individual null hypothesis histograms for each voxel in the brain volume. The omnibus null hypothesis histogram used in the current analysis is uniform throughout the brain, and while this serves as an adequate approximation, more accurate significance estimates would be measured against null hypothesis models specific to each brain location. This gain in accuracy, however, requires a significant loss of precision. Using the omnibus null histogram

allows measurement of all 243,000 voxels to be included in the estimation of a single distribution's shape. Each individual null distribution in the voxelwise analysis would benefit from only a single measurement from each permutation of the noise condition. To achieve the level of precision of the significance estimates in the current analysis, one would need to evaluate 243,000,000 permutations. A high level of precision in the null distribution estimate is essential for significance measurements at low *P* value thresholds. Highly precise estimates would be essential for a voxelwise analysis, requiring a threshold correction for multiple comparisons. A more feasible solution to inward drift would apply a correction coefficient to ALE value calculations for each voxel. A simple way to generate these coefficients would be to calculate the average ALE value for each voxel from the noise permutations. The coefficient in this method would be related to the reciprocal of the average ALE value for the voxel.

A related issue involves the method of creating noise distributions to estimate a null hypothesis histogram for omnibus thresholding. For this analysis noise estimates were generated by distributing activation foci randomly within the brain volume, but a more theoretically founded noise distribution would be nonrandom, with higher probabilities in brain regions more likely to demonstrate activity regardless of the task. No imaging paradigm will activate ventricular voxels as often as gray matter voxels. Thus, completely random noise distributions underestimate the probability of foci arising in gray matter by chance and overestimate the probability of foci arising in white matter or CSF by chance. An omnibus threshold based on the conglomeration of these probability estimates yields increased type I error in gray matter and increased type II error in white matter and CSF. The most rigorous solution to this problem would apply voxelwise critical thresholds, but this solution is infeasible for reasons outlined above. Alternatively, noise distributions could be generated in some nonrandom manner, with higher probabilities of focus localization associated with voxels in and near gray matter. The precise probabilistic weighting of brain regions could reasonably be based on the distribution of foci within in a large body of imaging studies probing diverse functions. This would have the effect of raising the omnibus threshold, which would reduce the type I error in gray matter. While this solution would also exacerbate the high type II error in other tissues, the trade-off would be worthwhile. Although this issue certainly warrants attention in future analyses, the inaccuracy of the noise distributions employed in this analysis did not likely contaminate the findings in any meaningful way. Assuming that a conservative noise distribution would restrict foci to voxels containing only gray matter, and estimating the average proportion of brain volume consisting of gray matter at around 50% (Grabowski *et al.*, 2000), the critical threshold could not reasonably be expected to increase from the current model by more than an order of magnitude. Because the significance threshold was set at very stringent level for this analysis ($\alpha = 0.0001$), the likelihood of type 1 error in the results is still very small.

This study has demonstrated the utility of a novel, automated meta-analysis technique for revealing areas of consistent activation within the single word reading literature. The similarity between the findings of this meta-analysis and those of new fMRI data validate the use of this technique in consolidating results of multiple published studies to identify the location and magnitude of task-related activity. The correspondence between the two experiments presented also serves to validate the use of the behavior-interleaved gradients fMRI acquisition paradigm for the study of language tasks requiring spoken responses.

The activation likelihood estimation method has several advantages over previous meta-analysis techniques. The use of a theoretically founded significance threshold provides a means to defend the results statistically and to fine tune the specificity or sensitivity of analyses. Even though the ultimate test for significance is through an omnibus threshold, the ALE map contains voxelwise statistics, allowing for development of relatively straightforward methods to compare multiple ALE maps from different groups of studies, a valuable tool for meta-analyses designed to compare related cognitive systems or to examine the effect of task parameters on outcome. This will be essential to more focused meta-analyses designed to identify specific subprocesses of complex cognitive skills. For example, the group of studies utilized for this analysis could be parsed to examine the difference between the neuronal systems subserving recognition of written words and those subserving pronunciation of words. After further examination and correction of any systematic bias in the statistical map, this method could also be employed in the definition of regions of interest for new imaging studies. Additionally, all methods presented here can be automated following extraction of relevant foci from the literature, maximizing efficiency and objectivity.

ACKNOWLEDGMENTS

We thank John Van Meter, Ph.D., who provided several TCL scripts that were essential to this analysis. We also thank two anonymous reviewers for their insightful comments, which certainly helped us to improve this work. This research was supported by the NICHD (Grant Nos. HD40095 and HD37890). For updates on availability of ALE meta-analysis tools, please go to http://csl.georgetown.edu/software.

REFERENCES

Bookheimer, S. Y., Zeffiro, T. A., *et al.* 1995. Regional cerebral blood flow during object naming and word reading. *Hum. Brain Mapp.* **3:** 93–106.

Boxerman, J. L., Bandettini, P. A., *et al.* 1995. The intravascular contribution to fMRI signal change: Monte Carlo modeling and diffusion-weighted studies in vivo. *Magn. Reson. Med.* **34:** 4–10.

Brett, M. (1999). The MNI Brain and The Talairach Atlas, Cambridge Imagers. http://www.mrc-cbu.cam.ac.uk/Imaging/mnispace.html.

Brett, M., Christoff, K., *et al.* 2001. Using the Talairach atlas with the MNI template. Human Brain Mapping, Brighton, U.K. *NeuroImage* **13:** S85.

Buckner, R. L., and Petersen, S. E. 1996. What does neuroimaging tell us about the role of prefrontal cortex in memory retrieval? *Sem. Neurosci.* **8:** 47–55.

Eden, G. F., Joseph, J. E., *et al.* 1999. Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: The behavior interleaved gradients technique. *Magn. Reson. Med.* **41:** 13–20.

Farah, M. J., and Aguirre, G. K. 1999. Imaging visual recognition: PET and fMRI studies of the functional anatomy of human visual recognition. *Trends Cogn. Sci.* **3:** 179–186.

Fiez, J. A., Balota, D. A., *et al.* 1999. Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. *Neuron* **24:** 205–218.

Fiez, J. A., and Petersen, S. E. 1998. Neuroimaging studies of word reading. *Proc. Natl. Acad. Sci. USA* **95:** 914–921.

Fox, P. T., and Mintun, M. A. 1989. Noninvasive functional brain mapping by change-distribution analysis of averaged PET images of $H_2{}^{15}O$ tissue activity. *J. Nuclear Med.* **30:** 141–149.

Fox, P. T., Mintun, M. A., *et al.* 1988. Enhanced detection of focal brain responses using intersubject averaging and change-distribution analysis of subtracted PET images. *J. Cereb. Blood Flow Metab.* **8:** 642–653.

Fox, P. T., Parsons, L. M., *et al.* 1998. Beyond the single study: Function/location metanalysis in cognitive neuroimaging. *Curr. Opin. Neurobiol.* **8:** 178–187.

Friedman, R. F., Ween, J. E., *et al.* 1993. Alexia. In *Clinical Neuropsychology* (K. M. Heilman and E. Valenstein, Eds.), pp. 37–62. Oxford Univ. Press, New York.

Grabowski, T. J., Frank, R. J., *et al.* 2000. Validation of partial tissue segmentation of single-channel magnetic resonance images of the brain. *NeuroImage* **12:** 640–656.

Grezes, J., and Decety, J. 2001. Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Hum. Brain Mapp.* **12:** 1–19.

Hagoort, P., Indefrey, P., *et al.* 1999. The neural circuitry involved in the reading of German words and pseudowords: A PET study. *J. Cogn. Neurosci.* **11:** 383–398.

Herbster, A. N., Mintun, M. A., *et al.* 1997. Regional cerebral blood flow during word and nonword reading. *Hum. Brain Mapp.* **5:** 84–92.

Holmes, A. P., Blair, R. C., *et al.* 1996. Nonparametric analysis of statistic images from functional mapping experiments. *J. Cereb. Blood Flow Metab.* **16:** 7–22.

Horwitz, B., Rumsey, J. M., *et al.* 1998. Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc. Natl. Acad. Sci. USA* **95:** 8939–8944.

Howard, D., Patterson, K., *et al.* 1992. The cortical localization of the lexicons. *Brain* **115:** 1769–1782.

Indefrey, P., and Levelt, W. J. 2000. The neural correlates of language production. In *The New Cognitive Neurosciences* (M. Gazzaniga, Ed.). MIT Press, Cambridge, MA.

Jernigan, T. L., Ostergaard, A. L., *et al.* 1998. Brain activation during word identification and word recognition. *NeuroImage* **8:** 93–105.

Kinahan, P. E., and Noll, D. C. 1999. A direct comparison between whole-brain PET and BOLD fMRI measurements of single-subject activation response. *NeuroImage* **9:** 430–438.

Lepage, M., Habib, R., *et al.* 1998. Hippocampal PET activations of memory encoding and retrieval: The HIPER model. *Hippocampus* **8:** 313–322.

Lindamood, C., and Lindamood, P. 1979. *Lindamood Auditory Conceptualization (LAC) Test*. Pro-Ed., Austin, TX.

Markowitsch, H. J., and Tulving, E. 1994. Cognitive processes and cerebral cortical fundi: Findings from positron-emission tomography studies. *Proc. Natl. Acad. Sci. USA* **91:** 10507–10511.

Oldfield, R. C. 1971. The assesment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia* **9:** 97–113.

Owen, A. M. 1997. The functional organization of working memory processes within human lateral frontal cortex: The contribution of functional neuroimaging. *Eur. J. Neurosci.* **9:** 1329–1339.

Petersen, S. E., Fox, P. T., *et al.* 1988. Positron emission tomographic studies of the cortical anatomy of single word processing. *Nature* **331:** 385–389.

Picard, N., and Strick, P. L. 1996. Motor areas of the medial wall: A review of their location and functional activation. *Cereb. Cortex* **6:** 342–353.

Poeppel, D. 1996. A critical review of PET studies of phonological processing. *Brain Lang.* **55:** 317–351.

Posner, M. I., Abdullaev, Y. G., *et al.* 1999. Neuroanatomy, circuitry and plasticity of word reading, *Neuroreport* **10:** R12–R23. [published erratum appears in *Neuroreport* 1999 June 23; **10**(9): inside back cover]

Price, C. J. 2000. The anatomy of language: Contributions from functional neuroimaging. *J. Anat.* **197:** 335–359.

Price, C. J., Moore, C. J., *et al.* 1996a. The effect of varying stimulus rate and duration on brain activity during reading. *NeuroImage* **3:** 40–52.

Price, C. J., Wise, R. J., *et al.* 1994. Brain activity during reading: The effects of exposure duration and task. *Brain* **117:** 1255–1269.

Price, C. J., Wise, R. J. S., *et al.* 1996b. Hearing and saying: The functional neuro-anatomy of auditory word processing. *Brain* **119:** 919–931.

Rumsey, J. M., Horwitz, B., *et al.* 1997. Phonological and orthographic components of word recognition: A PET-rCBF study. *Brain* **120:** 739–759.

Sergent, J., Zuck, E., *et al.* 1992. Positron emission tomography study of letter and object processing: Empirical findings and methodological considerations. *Cereb. Cortex* **2:** 68–80.

Talairach, J., Szikla, G., *et al.* 1967. *Atlas d'Anatomie Stereotaxique du Telencephale.* Masson, Paris.

Talairach, J., and Tournoux, P. 1988. *Co-planar Stereotaxic Atlas of the Human Brain*. Thieme, New York.

Woodcock, R. W., and Johnson, M. B. 1987. *Woodcock–Johnson Psychoeducational Battery*. Teaching Resources, Hingham, MA.

Woods, R. P., Grafton, S. T., *et al.* 1998a. Automated image registration: I. *J. Comput. Assist. Tomogr.* **22:** 139–152.

Woods, R. P., Grafton, S. T., *et al.* 1998b. Automated image registration: II. *J. Comput. Assist. Tomogr.* **22:** 153–165.