Permutation testing of orthogonal, factorial effects in a language processing experiment using FMRI

John Suckling, Matthew H. Davis, Cinly Ooi, Alle-Meije Wink, Jalal M. Fadili, Raymond Salvador, David Welchew, Levent Sendur, Voichita Maxim, Edward Bullmore

Permutation Testing of Orthogonal Factorial Effects in a Language-Processing Experiment Using fMRI

John Suckling,[1] Matthew H. Davis,[2] Cinly Ooi,[1] Alle Meije Wink,[1] Jalal Fadili,[3] Raymond Salvador,[4] David Welchew,[1] Levent Şendur,[1] Vochita Maxim,[5] and Edward T. Bullmore[1]

[1]*Brain Mapping Unit, Department of Psychiatry, University of Cambridge, UK*
[2]*Speech and Language Group, MRC Cognition and Brain Sciences Unit, Cambridge, UK*
[3]*Image Processing Group, GREYC UMR CNRS 6072 ENSICAEN, Caen, France*
[4]*Sant Joan de Déu – Serveis de Salut Mental, St. Boi de Llobregat, Barcelona, Spain*
[5]*Département de Mathématiques, Institut National des Sciences Appliquées, Toulouse, France*

Abstract: The block-paradigm of the Functional Image Analysis Contest (FIAC) dataset was analysed with the Brain Activation and Morphological Mapping software. Permutation methods in the wavelet domain were used for inference on cluster-based test statistics of orthogonal contrasts relevant to the factorial design of the study, namely: the average response across all active blocks, the main effect of speaker, the main effect of sentence, and the interaction between sentence and speaker. Extensive activation was seen with all these contrasts. In particular, different vs. same-speaker blocks produced elevated activation in bilateral regions of the superior temporal lobe and repetition suppression for linguistic materials (same vs. different-sentence blocks) in left inferior frontal regions. These are regions previously reported in the literature. Additional regions were detected in this study, perhaps due to the enhanced sensitivity of the methodology. Within-block sentence suppression was tested post-hoc by regression of an exponential decay model onto the extracted time series from the left inferior frontal gyrus, but no strong evidence of such an effect was found. The significance levels set for the activation maps are *P*-values at which we expect <1 false-positive cluster per image. Nominal type I error control was verified by empirical testing of a test statistic corresponding to a randomly ordered design matrix. The small size of the BOLD effect necessitates sensitive methods of detection of brain activation. Permutation methods permit the necessary flexibility to develop novel test statistics to meet this challenge.

Key words: FIAC; functional MRI; nonparametric; permutation; wavelet

INTRODUCTION

Permutation (or randomization) tests are methods of nonparametric statistical inference that use the observed dataset to simulate other, surrogate, datasets under conditions of the null-hypothesis. For example, rather than obtaining critical values from a parametric *t* distribution for the difference of a measurement made on two groups of independent subjects, a permutation test simulates the null-hypothesis by randomly repeating reassignments of the original measurements to either of the groups and recalculating *t*, thus "empirically" sampling the null-distribution against which the observed value can be compared.

Contract grant sponsor: Human Brain Project grant from the National Institute of Biomedical Imaging & Bioengineering and the National Institute of Mental Health.

*Correspondence to: J. Suckling, Brain Mapping Unit, Department of Psychiatry, University of Cambridge, Addenbrooke's Hospital, Box 255, Cambridge CB2 2QQ, United Kingdom. E-mail: js369@cam.ac.uk

Received for publication 20 October 2005; Accepted: 27 December 2005

DOI: 10.1002/hbm.20252

Published online 4 April 2006 in Wiley InterScience (www.interscience.wiley.com).

This simple case illustrates a number of issues with permutation tests. First, the null-hypothesis is that: *there are no differences between the groups*, and is independent of the test statistic and distinct from the exact mathematical requirement of a parametric test (e.g., means are equal). Second, for a permutation test to be valid the observations must be exchangeable. Formally, the n units of observation of the random variable X are exchangeable if the joint probability distribution is invariant under permutation [Lindley and Novick, 1981]. This criterion of exchangeability is more general than the alternative notion that observed data have been experimentally randomized to groups and can therefore be relabeled, as this is often unachievable for groups such as male/female or healthy/diseased states. Of course, exchangeability (independence) applies equally to parametric tests. In either case, inference on nonindependent observations requires modification of both test statistic and relabeling scheme. Third, the results from permutation testing only refer to the dataset under scrutiny and cannot strictly be generalized to the wider population. However, if recruitment was sufficiently scrupulous to have randomly sampled from the population(s) under study, then extrapolation of the results is equivalent to fulfilling similar requirements for a parametric test, without the need for additional assumptions on the form of the null-distribution.

The main motivating factor in our development of permutation tests for neuroimaging has been the considerable latitude available in the design of test statistics with enhanced sensitivity. We have previously described cluster-based statistics in the context of between-group designs [Bullmore et al., 1999a; Suckling and Bullmore, 2004] that have consistently outperformed voxel-based statistics with nominal type I error control. Permutation methods and cluster-statistics have been incorporated into the Brain Activation and Morphological Mapping software (BAMM: http://www-bmu.psychiatry.cam.ac.uk/software) [Bullmore et al., 1996, 1999a,b, 2000, 2003, 2004; Brammer et al., 1997; Suckling and Bullmore, 2004] and the Functional Image Analysis Contest (FIAC) offers the opportunity to demonstrate the capabilities of the software.

MATERIALS AND METHODS

Linguistic Processing Block Paradigm

Fifteen subjects of the FIAC dataset [Dehaene-Lambertz et al., 2006] were used, omitting Subject 5, who was only scanned once.

The analysis software is now described in detail, followed by a description of how processing of the FIAC dataset was approached.

BAMM Processing: Parenchymal Mask

A mask of the parenchymal region constrained subsequent processing. This binary image was generated by segmentation and morphological operations on the mean of the first 16 image volumes.

BAMM Processing: Subject Motion Correction

Correction was made for subject motion assuming the head to be a rigid body, with translations and rotations about its centre of mass. Each 3D dataset was registered to the mean, masked image with tri-cubic spline interpolation [Brammer et al., 1997; Bullmore et al., 1999b].

Residual spin excitation history effects [Friston et al., 1996; Jezzard and Clare, 1999] were corrected by regressing the current ($t = 1...T$) and lagged ($t-1$) first- and second-order displacements at each voxel onto the realigned time series, and the estimated signal change subtracted [Brammer et al., 1997].

BAMM Processing: Global Image Rescaling and Spatial Smoothing

Changes in global gray-level scaling during image acquisition were corrected by normalisation to the mean gray-level across all voxels, in all images. Linear trends were estimated and removed by least-squares regression onto the time series of corrected global means.

2-D spatial smoothing was applied to each corrected image volume with a Gaussian kernel of standard deviation = 0.5 (in-plane) voxels, convolved via the Fourier domain.

BAMM Processing: Time Series Modeling

Prior to response estimation, the design matrix was convolved with a canonical hemodynamic response function (HRF) modeling the delay and dispersion of the BOLD effect [Friston et al., 1998]. Estimates of the BOLD response to the stimuli were made by regression of the GLM onto each mean-zeroed time series:

$$\mathbf{Y} = \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{Y}$ is the preprocessed time series, $\boldsymbol{\beta}$ the coefficients to be found, and $\boldsymbol{\varepsilon}$ the residuals. Estimates of the coefficients and their standard errors (SE) were made from least-squares minimization using singular valued decomposition [Press et al., 1992]. The voxelwise standardized test statistic for each contrast ($k = 1 \cdots K$) of the design matrix, $\mathbf{X}$, was then calculated:

$$F_k^* = \left(\frac{\hat{\beta}_k}{\mathrm{SE}(\hat{\beta}_k)}\right)^2 \tag{2}$$

The test statistic was signed by the direction of the correlation of the time series with the fitted model, $\boldsymbol{\beta}\mathbf{X}$. For a single component HRF this operation reduced simply to the assignment of the sign of $\hat{\beta}k$. The image of the test statistic estimated at each intracerebral voxel is known as the *observed response map*.

BAMM Processing: Time Series Permutation

Statistical testing of the observed response map was made against a null-distribution sampled using permutation

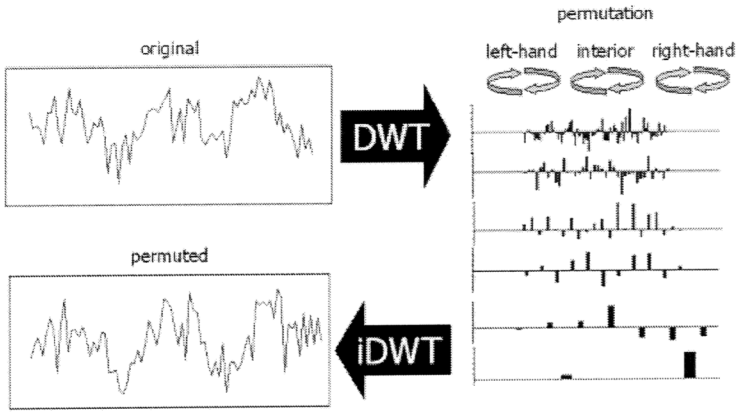

Figure 1.

Schematic overview of the permutation method. Coefficients of the wavelet transform of an observed time series were permutated scale-by-scale, accounting for edge effects, before the inverse transform generated a time series with a disrupted relationship to the applied stimuli, but similar autocorrelation properties as the original.

methods. Under the null-hypothesis, the power spectrum of the signal is adequately represented as $1/f$-like noise [Bullmore et al., 2001, 2004; Zahran et al., 1997]. In the time domain this is manifest by strong positive autocorrelation or long memory. Under such conditions, the residuals of GLM (Eq. 1) are correlated and thus render biased estimates of the standard errors of the coefficients. Our permutation approach generates time series under a simulated null-hypothesis with similar spectral properties as the observed data. Thus, whilst estimates are biased, they are equally so under observed and permuted conditions, leading to nominal type I error control.

Signals with long-memory properties have coefficients of the discrete wavelet transform (DWT), which are approximately decorrelated [Dijkerman and Mazumdar, 1994]. These coefficients are accordingly postulated to be the exchangeable elements of the algorithm.

The DWT [Mallat, 1989] decomposes the time series into a hierarchy of frequency scales with different time resolutions, from fine-to-coarse; the number of coefficients reduced by a factor of two in consecutive scales. A surrogate time series under simulated conditions of the null-hypothesis was obtained by permuting (reordering) the wavelet coefficients within each scale of the DWT (Fig. 1). On reconstituting the signal by the inverse DWT, a new time series was generated with a disrupted relationship to the applied stimuli, but with similar autocorrelation properties to the observed signal.

In practical terms, the DWT used the fourth-order Daubechies wavelet supported over eight timepoints. This basis function is a compromise between minimising the interco-efficient correlation and generating artefactual correlations due to periodic boundary corrections. For time series with T a power of 2 in length, the edge-filtering scheme of Cohen et al. [1993] was adopted to ameliorate this effect. The coefficients can be considered in three "zones": left-hand edge-filter, interior, and right-hand edge-filter (Fig. 1). The permutation procedure was performed separately within each zone at every scale.

For time series for which T is not a power of 2, the vector length was zero-padded to the next largest power of 2 and the permutation of coefficients restricted to only those locations that were originally nonzero. Breakspear et al. [2004] noted that there was a systematic loss in overall signal power before and after permutation in a similar spatial resampling scheme. As with our temporal permutation method, losses were regained by renormalization.

Responses under the null-hypothesis were estimated from the surrogate time series with the GLM (Eqs. 1, 2). The order of permutation of the coefficients was maintained at each intracerebral voxel to retain the spatial autocorrelation in the *permuted response maps*, essential in the subsequent calculation of cluster statistics. The permutation was repeated $\Pi = 10$ times to adequately sample the null-distributions.

BAMM Processing: Standard Space Mapping

Observed and permuted response maps were transformed into the Montreal Neurological Institute (MNI)/Talairach [Talairach and Tournoux, 1988] standard coordinate system. Mappings were calculated from the affine registration of the mean image onto the EPI template (http://www.fil.ion.ucl.ac.uk/spm), maximizing the grey-scale correlation using the Fletcher-Davison-Powell search algorithm [Press et al., 1992]. The optimized mapping parameters were subsequently applied to all response maps for all subjects in the study.

BAMM Processing: Statistical Inference

Within-subject mean responses were first calculated by voxelwise averaging of the two sets of response maps from each repeat of the paradigm. At each voxel in template space, the median (chosen for its robustness to outlying

values) of the observed and permuted within-subject mean responses was taken across the group. This generated a median observed response map of the activation of the group and Π median permuted response maps representing activation under simulated conditions of the null-hypothesis.

Probabilistic thresholding of the median observed response map was performed in two stages: voxel- and cluster-level. For each contrast, all responses at all intracerebral voxels from the Π median permuted response maps were pooled to sample the two-tailed null-distribution. For positive and negative values separately, all voxels less than the critical value at the 2.5% level, CV_{vox}, were set to zero. This procedure resulted in sets of C_l voxel clusters in the median observed, $l = 0$, and permuted, $l = 1 \ldots \Pi$, response maps. For any given map, the clusters $c = 1 \ldots C_l$ were each comprised of V_c voxels, spatially contiguous (nearest and next-nearest neighbors) in three dimensions. Cluster-level statistics, Λ_c, were then computed as the sum of suprathreshold voxel statistics for all clusters in all maps as:

$$\Lambda_c = \sum_{m=1}^{V_c} (F_m^* - CV_{vox}) \tag{3}$$

Those values obtained from the median permuted response maps were pooled to sample the null-distribution of clusters with $\Lambda_c > 0$. Critical values for the cluster-level statistics, CV_{clu}, were calculated for the number of type I errors, η, expected under the null-hypothesis. This is related to the corresponding two-tailed P value by:

$$p = \frac{\eta}{\Xi}, \text{ where } \Xi = \sum_{l=1}^{\pi} C_l \tag{4}$$

where Ξ is the total number of clusters in all permuted response maps, and was typically ~11,000 ($P \asymp 9.0 \times 10^{-5}$ for $\eta = 1$). CV_{clu} was used as a threshold for the observed clusters with surviving regions displayed as a color overlay on template slices in group activation maps.

Clusters from this method can be considered spatially independent. With this assumption, the 95% confidence intervals were calculated from the sampling distribution of a proportion [Agresti and Coull, 1998].

FIAC Analysis: Cluster Statistics Nominal Type I Errors

A design matrix was composed with each data volume randomly assigned to one of two values, with sum zero. Following processing, cluster statistic η values were systematically varied and the resulting number of suprathreshold clusters at each threshold was plotted with its 95% confidence interval against the number expected by chance, $p.C_{l=0}$, to assess type I error control.

FIAC Analysis: Mean Response, Main Effects, and Interaction

The factorial effects addressing the primary goals of the contest were assessed with a design matrix with four orthogonal contrasts: (1) Mean activation across all "active" blocks compared to periods of silence; (2) Main effect of speaker; (3) Main effect of sentence; and (4) Interaction between speaker and sentence.

Group activation maps for each contrast were made at thresholds for the cluster-level statistics of $\eta = 1$, with equivalent P values: mean response, $P = 9.1 \times 10^{-5}$; main effect of speaker, $P = 8.9 \times 10^{-5}$; main effect of sentence, $P = 9.2 \times 10^{-5}$; interaction of speaker and sentence, $P = 8.6 \times 10^{-5}$.

FIAC Analysis: Within-Block Sentence Suppression in the Inferior Frontal Gyrus

The mean time series was extracted from a region of significant main effect of sentence in the inferior frontal gyrus (see Results). For the four active blocks of same-speaker condition, the parameters of the model: $A_q \exp(-B_q (t - t_{q0}))$, where t_{q0} is the onset time for block q of a given condition, convolved with the HRF, were estimated by minimizing the least-squares difference. This was repeated for the different-speaker condition. The eight values of B_q from each subject were then used in a repeated-measures, two-way analysis of variance with same or different speaker as the levels of one factor and q as the levels of the second. Permuting the extracted time series 1,000 times and reestimating the model determined the significance of the observed data.

FIAC Analysis: Processing Times

The FIAC dataset was processed in its entirety in 14.5 hours (2.8-GHz Xenon, 1 GB RAM, Linux OS).

RESULTS

Cluster Statistics Nominal Type I Errors

Nominal type I error control is demonstrated in the plot of observed versus expected number of clusters for a range of values of η (Fig. 2).

Mean Response, Main Effects, and Interaction

The map of mean response over all active conditions vs. baseline (rest) condition (Fig. 3A) shows extensive activation in bilateral temporal and prefrontal cortex as well as significant deactivation in anterior and posterior cingulate cortex and large areas of parietal lobe.

Figure 3B shows the effect of changing the speaker on successive sentences within a single block. This contrast produced elevated activation in bilateral regions of the superior temporal lobe that are known to be particularly responsive to the human voice [see Belin et al., 2000, 2004, for review]. In particular, we see activation in a region of the

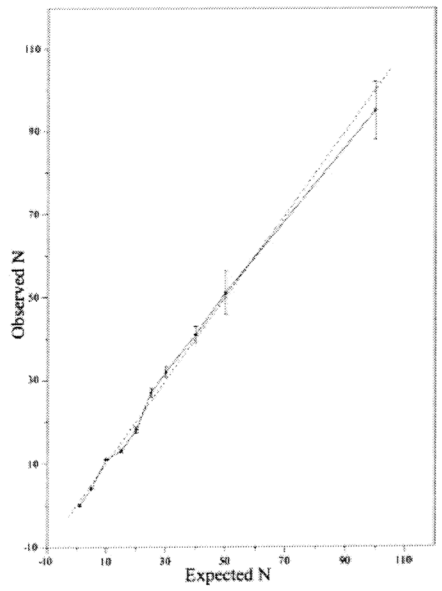

Figure 2.

Type I error calibration curve for group cluster statistics with the observed number of positive tests vs. the expected number of positive tests for the randomly designed model of activation. The points represent the observed number of surviving tests in the group activation map with error bars indicating calculated 95% confidence limits. The solid line $y = x$ indicates ideal agreement between observation and expectation.

right temporal lobe (Talairach coordinates: $x = 54$ mm, $y = -15$ mm, $z = -12$ mm; Fig. 3B).

Figure 3C shows the effect of changing the sentence that was produced during each block. This contrast highlights a number of brain regions that have been shown to be involved in linguistic processing of speech: for instance, bilateral regions of the anterior temporal lobe, including the middle temporal gyrus. A left-lateralized response was also observed in frontal brain regions, including the left inferior frontal gyrus (Talairach coordinates: $x = -54$ mm, $y = 18$ mm, $z = 12$ mm; Fig. 3C) and anterior insula (Talairach coordinates: $x = -40$ mm, $y = 24$ mm, $z = -4$ mm; Fig. 3C).

We further considered the nature of the interactions between sentence and speaker changes (Fig. 3D) and the overlap between some of these interactions and the main effects of speaker and sentence repetition: for instance, the region of the left middle temporal gyrus (Talairach coordinates: $x = -57$ mm, $y = -22$ mm, $z = -4$ mm; Fig. 3D). Inspection of the BOLD effect from this cluster reveals an activation profile dominated by reduced activity for the same-speaker, same-sentence condition (Fig. 4A). It is this reduction in activation that drives both main effects and interactions, since the other conditions (speaker changes, sentence changes, or both speaker and sentence changes) did not differ markedly. In contrast, other brain areas showed an interaction that appeared in the absence of either main effect (e.g., the right inferior frontal gyrus, Talairach coordinates: $x = 50$ mm, $y = 17$ mm, $z = 12$ mm; Fig. 3D). This crossover interaction (Fig. 4B) is more difficult to interpret since it appears to reflect a change in the direction of sentence (or speaker) repetition suppression depending on whether the same or different speaker (or sentence) was presented.

Within-Block Sentence Suppression in the Inferior Frontal Gyrus

The post-hoc factorial analysis on the exponential model parameters, B_q, estimated from the active conditions of same sentence gave: main effect of speaker (same or different), $F = 0.93$, $P = 0.37$; main effect of block number, $F = 1.98$, $P = 0.18$; speaker by block number interaction, $F = 1.92$, $P = 0.14$. Thus, there is only weak evidence of within-block suppression of activation during repetition of sentences.

DISCUSSION

The primary goals of the FIAC have been addressed by methods of inference using permutation of a wavelet representation of functional MRI (fMRI) time series. Main effects and interactions were observed as extensive activation of language-related cortex (Fig. 3). The cluster-based statistics of the within-group median responses were shown to have nominal type I error control across a wide range of probability thresholds (Fig. 2). This replicates previous work in the context of between-group studies [Bullmore et al., 1999a, 2001; Suckling and Bullmore, 2004]. The use of cluster tests reduces the problem of multiple comparisons, acute with more than 250,000 voxel locations in the template, which arise in voxelwise significance testing. The spatial correlation of the response images renders P value correction schemes overconservative. Even small P values result in a large number of estimated type I errors, making interpretation of the observed activated regions difficult. More generally, brain activation in any but the most specific of sensory paradigms is often distributed rather than focal and tests based on magnitude alone potentially lose much of this information. Cluster-based statistics address both these concerns and other studies [Hayasaka and Nichols, 2003] have demonstrated that permutation methods of a cluster size statistic perform well over a range of image smoothness and are robust when assumptions required for a parametric test are violated.

The mean response map (Fig. 3A) gave a pattern of deactivations consistent with the default state network [Greicius et al., 2003; Gusnard and Raichle, 2001]. While it is predicted that those blocks in which speaker or sentence change give rise to additional activation in brain regions critical for sentence or voice processing [cf. repetition suppression, Henson, 2003; Schacter and Buckner, 1998], it is of interest to examine the reverse contrasts. In a blocked-design experiment in which stimulus repetition is rather predictable, the same-speaker and same-sentence blocks are likely to be rather uninteresting to subjects and might therefore produce

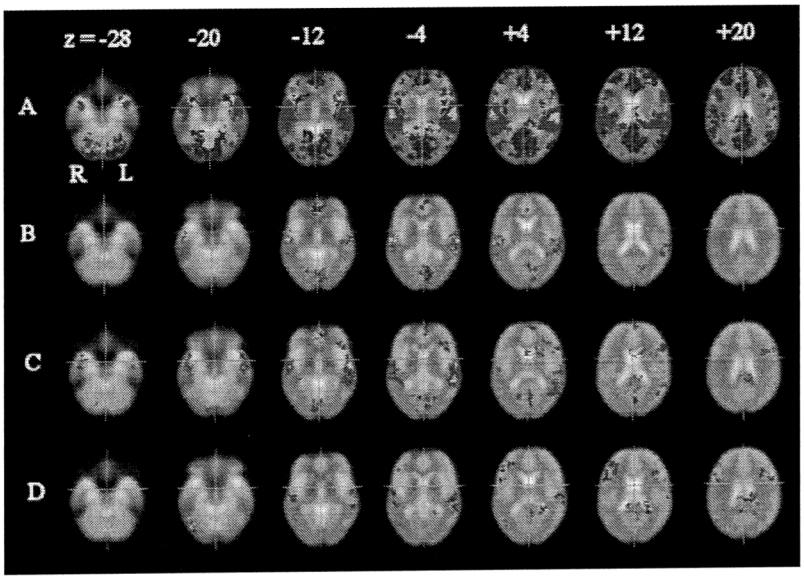

Figure 3.

Selected axial slices (z values of the Talairach atlas of −28 mm, −20 mm, −12 mm, −4 mm, +4 mm, +12 mm, and +20 mm) showing significant (**A**) mean response, (**B**) main effects of speaker, (**C**) main effect of sentence, and (**D**) sentence × speaker interaction at a cluster-statistic threshold of $\eta = 1$. The left-hand side of the image is the right-hand side of the brain and the crosshairs show the Talairach origin. Red regions of A are where active conditions > baseline and blue regions vice versa. Red regions of B refer to different-speaker > same-speaker and blue regions the reverse. Red regions of C refer to different-sentence > same-sentence and blue regions the reverse. Red regions of D refer to an interaction in which activation for same-sentence increases from same-speaker to different-speaker, and different-sentence decreases from same-speaker to different-speaker and vice versa for blue regions.

elevated activation in a set of brain areas designated the default state network.

In considering the main effect of speaker changes (Fig. 3B), we observed bilateral temporal lobe activation that in part replicates similar studies on voice changes [Belin and Zatorre, 2003; von Kriegstein et al., 2003]. However, in contrast to these earlier studies the current dataset also reveals activation in homologous regions of the left hemisphere. Potentially, the data analysis used has provided the necessary sensitivity to reveal these areas. Alternatively, it may be that some more general change in the degree of attentional allocation is responsible for this bilateral response in the temporal lobe [Hall et al., 2000; Petkov et al., 2004]. One piece of evidence consistent with an attentional explanation is that activation is also observed in more posterior temporal lobe regions as well as those areas reported by Belin and Zatorre [2003] and von Kriegstein et al. [2003].

The present work also replicates studies in the observation of repetition suppression for linguistic materials (i.e., same- and different-sentence stimuli) in left inferior frontal regions [Buckner et al., 2000; Macotta and Buckner, 2004] that are historically linked to linguistic functions [Broca, 1865; Dronkers, 1996]. This activation also overlaps with that observed in a number of studies that contrast intelligible and unintelligible sentences [Crinion et al., 2003; Davis and Johnsrude, 2003; Rodd et al., 2005; Scott et al., 2000], although the bilateral temporal lobe response seen here has not always been observed.

The observation of repetition suppression in frontal regions is often interpreted as evidence for task-specific response facilitation [cf. Dobbins, et al., 2004; Schacter et al., 2004]. It is at present unclear whether this same explanation can hold in the current study, where no task was required of subjects. One possibility to reconcile this tension is that activation in frontal regions reflects the "listening effort" that is required for sentence comprehension. Recent studies have shown that comprehending distorted or noisy speech can increase activation in left inferior frontal regions when compared to clearly perceived speech [Binder et al., 2004; Davis and Johnsrude, 2003]. It may therefore be that similar systems are activated in nonsparse fMRI studies (such as the FIAC dataset) in which ongoing scanner noise presents a substantial obstacle to comprehension (we note that some of the participants reported difficulties understanding some words). Sentence repetition will reduce the listening effort required for comprehension; it might therefore explain the observation of frontal repetition suppression in the absence of a task [see Giraud et al., 2004, for related results]. Investigation of the within-block, time course of differences between same- and different-sentence blocks did not uniquely determine the cause of the repetition suppression response.

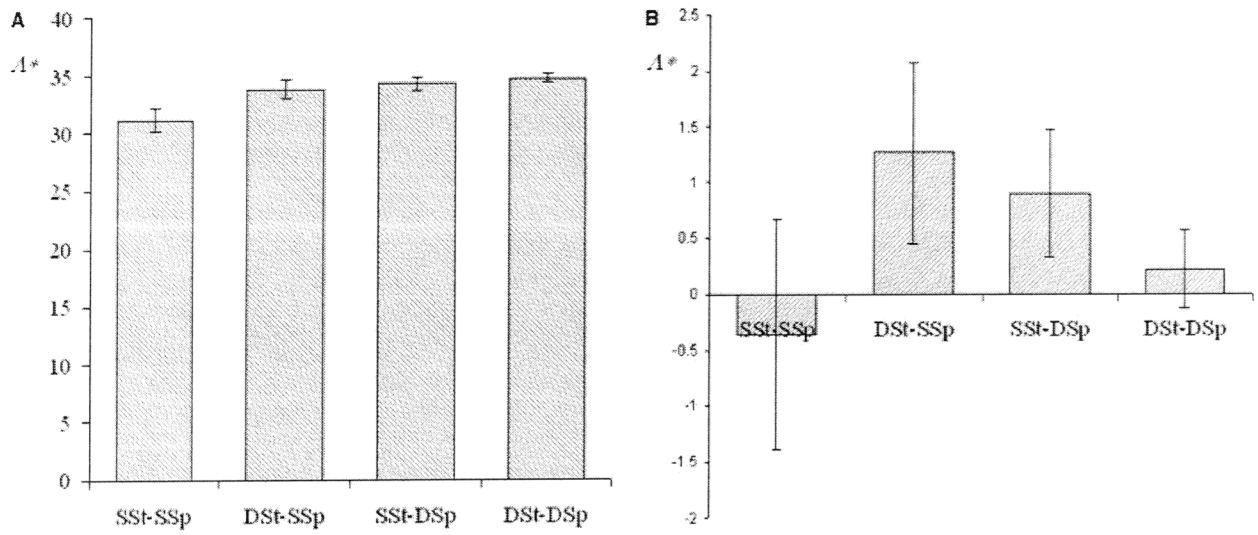

Figure 4.

Sum of simple effects extracted from significant interaction in (**A**) left middle temporal gyrus (Talairach coordinates: $x = -57$ mm, $y = -22$ mm, $z = -4$ mm) and (**B**) right inferior frontal gyrus (Talairach coordinates: $x = 50$ mm, $y = 17$ mm, $z = 12$ mm). Error bars were calculated for repeated measures [Loftus and Masson, 1994].

Several of the elements in the BAMM data processing pipeline are common to the many software packages for fMRI analysis. The novel contribution in the BAMM software stems from permutation methods that yield the necessary flexibility for the development of new test statistics for neuroimaging [Hayasaka and Nichols, 2004]. The approach is not new [Bullmore et al., 1996; Holmes et al., 1996], but over an extended period we have demonstrated its utility and power in analyses of individual, within-group, and between-group neuroimaging experiments. The deficiency of these methods has historically been identified as the large requirement of processing time for their implementation. It is clear now that computational speed, continuously rising, has rendered this issue inconsequential. In fact, the wavelet-based reassignment procedure only takes 1% to 2% of the total processing time, with subject motion correction and intersubject registration by far the largest contribution to the 14.5 hours taken for data analysis.

The BAMM software makes use of the DWT and its rich description of the spectral properties of the time series. They are used as the exchangeable elements in this permutation method, but elsewhere we have reported an extension of the software to estimate the Hurst exponent (related to Hausdorff fractal dimension) in the wavelet domain of fMRI time series of the resting state. Both disease and pharmaceutical effects have been demonstrated with this parameter [Maxim et al., 2004; Wink et al., 2006]. Furthermore, we are implementing the DWT in the spatial domain, using the variable localization of the coefficients in the transform hierarchy as a method of detecting extended regions of activation in fMRI response maps [Sendur et al., 2005]. Further development will introduce permutation methods.

REFERENCES

Agresti A, Coull BA (1998): Approximate is better than "exact" for interval estimation of binomial proportions. Am Stat 52:119–126.

Belin P, Zatorre RJ (2003): Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14:2105–2109.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000): Voice-selective areas in human auditory cortex. Nature 403:309–312.

Belin P, Fecteau S, Bedard C (2004): Thinking the voice: neural correlates of voice perception. Trends Cogn Sci 8:129–135.

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004): Neural correlates of sensory and decision processes in auditory object identification. Nat Neurosci 7:295–301.

Brammer MJ, Bullmore ET, Simmons A, Williams SCR, Grasby PM, Howard RJ, Woodruff PWR, Rabe-Hesketh S (1997): Generic brain activation mapping in functional magnetic resonance imaging: a non-parametric approach. Magn Reson Imaging 15:763–770.

Braver TS, Cohen JD, Nystrom LE, Jonides J, Smith EE, Noll DC (1997): A parametric study of prefrontal cortex involvement in human working memory. Neuroimage 5:49–62.

Breakspear M, Brammer MJ, Bullmore ET, Das P, Williams LM (2004): Spatiotemporal wavelet resampling for functional neuroimaging data. Hum Brain Mapp 23:1–25.

Broca P (1865): Sur la faculté du langage articule. Bull Soc d'Anthropol Paris 6:337–393.

Buchel C, Holmes AP, Rees G, Friston KJ (1998): Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. Neuroimage 8:140–148.

Buckner RL, Koutstaal W, Schacter DL, Rosen BR (2000): Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. Brain 123:620–640.

Bullmore E, Brammer M, Williams S, Rabe-Hesketh S, Janot N, David A, Mellers J, Howard R, Sham P (1996): Statistical meth-

ods of estimation and inference for functional MR image analysis. Magn Reson Med 35:261–277.

Bullmore ET, Suckling J, Overmeyer S, Rabe-Hesketh S, Taylor E, Brammer MJ (1999a): Global, voxel and cluster tests, by theory and permutation for a difference between two groups of structural MR images of the brain. IEEE Trans Med Imaging 18:32–42.

Bullmore ET, Brammer MJ, Rabe-Hesketh S, Curtis VA, Morris RG, Williams SC, Sharma T, McGuire PK (1999b): Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fMRI. Hum Brain Mapp 1999;7: 38–48.

Bullmore E, Long C, Suckling J, Fadili J, Calvert G, Zelaya F, Carpenter TA, Brammer M (2001): Colored noise and computational inference in neurophysiological (fMRI) time series analysis: resampling methods in time and wavelet domains. Hum Brain Mapp 12:61–78.

Bullmore E, Suckling J, Zelaya F, Long C, Honey G, Reed L, Routledge C, Ng V, Fletcher P, Brown J, Williams SC (2002): Practice and difficulty evoke anatomically and pharmacologically dissociable brain activation dynamics. Cereb Cortex 13:144–154.

Bullmore E, Fadili J, Breakspear M, Salvador R, Suckling J, Brammer M (2003): Wavelets and statistical analysis of functional magnetic resonance images of the human brain. Stat Methods Med Res 12:375–399.

Bullmore E, Fadili J, Maxim V, Sendur L, Whitcher B, Suckling J, Brammer M, Breakspear M (2004): Wavelets and functional magnetic resonance imaging of the human brain. Neuroimage 23(Suppl 1):S234–249.

Callicott JH, Mattay VS, Bertolino A, Finn K, Coppola R, Frank JA, Goldberg TE, Weinberger DR (1999): Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. Cereb Cortex 9:20–26.

Clark VP (2002): Orthogonal polynomial regression for the detection of response variability in event-related fMRI. Neuroimage 17: 344–363.

Clark VP, Fannon S, Lai S, Benson R, Bauer L (2000): Responses to rare visual target and distractor stimuli using event-related fMRI. J Neurophysiol 83:3133–3139.

Cohen A, Daubechies I, Vial P (1993): Wavelets on the interval and fast wavelet transforms. Appl Comput Harmon Anal 1:54–81.

Cox RW (1994): Analysis of functional neuroimages, v. 1.01. Milwaukee: Medical College of Wisconsin.

Crinion JT, Lambon-Ralph MA, Warburton EA, Howard D, Wise RJ (2003): Temporal lobe regions engaged during normal speech comprehension. Brain 126:1193–1201.

Davis MH, Johnsrude IS (2003): Hierarchical processing in spoken language comprehension. J Neurosci 23:3423–3431.

Dehaene-Lambertz G, Dehaene S, Anton JL, Campagne A, Ciuciu P, Dehaene GP, Denghien I, Jobert A, LeBihan D, Sigman M, Pallier C, Poline JB (2006): Functional segregation of cortical language areas by sentence repetition. Hum Brain Mapp 27:360–371.

Dijkerman RW, Mazumdar RR (1994): On the correlation structure of the wavelet coefficients of fractional Brownian motion. IEEE Trans Info Theory 40:1609–1612.

Dobbins IG, Schnyer DM, Verfaellie M, Schacter DL (2004): Cortical activity reductions during repetition priming can result from rapid response learning. Nature 428:316–319.

Dronkers NF (1996): A new brain region for coordinating speech articulation. Nature 384:159–161.

Friston KJ, Firth CD, Tuner R Frackowiak RSJ (1995): Characterizing evoked hemodynamics with fMRI. Neuroimage 2:157–165.

Friston KJ, Williams S, Howard R, Frackowiak RSJ, Turner R (1996): Movement-related effects in fMRI time-series. Magn Reson Med 35:346–355.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997): Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218–229.

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R (1998): Event-related fMRI: characterizing differential responses. Neuroimage 7:30–40.

Giraud AL, Kell C, Thierfelder C, Sterzer P, Russ MO, Preibisch C, Kleinschmidt A (2004): Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. Cereb Cortex 14:247–255.

Greicius, MD, Krasnow B, Reiss AL, Menon V (2003): Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc Natl Acad Sci U S A 100:253–258.

Gusnard DA, Raichle ME (2001): Searching for a baseline: functional imaging and the resting human brain. Nat Rev Neurosci 2:685–694.

Hall DA, Haggard MP, Akeroyd MA, Summerfield AQ, Palmer AR, Elliott MR, Bowtell RW (2000): Modulation and task effects in auditory processing measured using fMRI. Hum Brain Mapp 10:107–119.

Hayasaka S, Nichols TE (2003): Validating cluster size inference: random field and permutation methods. Neuroimage 20:2343–2356.

Hayasaka S, Nichols TE (2004): Combining voxel intensity and cluster extent with permutation test framework. Neuroimage 23:54–63.

Henson RNA (2003): Neuroimaging studies of priming. Prog Neurobiol 70:53–81.

Holmes AP, Blair RC, Watson JDG, Ford I (1996): Nonparametric analysis of statistic images from functional mapping experiments. J Cereb Blood Flow Metab 16:7–22.

Jezzard P, Clare S (1999): Sources of distortion in functional MRI data. Hum Brain Mapp 8:80–85.

Joshi DD (1987): Linear estimation and design of experiments. New Delhi: Wiley Eastern.

Lindley DV, Novick MR (1981): The role of exchangeability in inference. Ann Stat 9:45–58.

Loftus GR, Masson MEJ (1994): Using confidence-intervals in within-subject designs. Psychonom Bull Rev 1:476–490.

Maccotta L, Buckner RL (2004): Evidence for neural effects of repetition that directly correlate with behavioral priming. J Cogn Neurosci 16:1625–1632.

Mallat S (1989): A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Machine Intel 11:674–693.

Maxim V, Sendur L, Fadili J, Suckling J, Gould R, Howard R, Bullmore E (2005): Fractional Gaussian noise, functional MRI and Alzheimer's disease. Neuroimage 25:141–158.

Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL (2004): Attentional modulation of human auditory cortex. Nat Neurosci 7:658–663.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992): Numerical recipes in C: the art of scientific computing. Cambridge, UK: Cambridge University Press.

Rodd JM, Davis MH, Johnsrude IS (2005): The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cereb Cortex 15:1261–1269.

Schacter DL, Buckner RL (1998): Priming and the brain. Neuron 20:185–195.

Schacter DL, Dobbins IG, Schnyer DM (2004): Specificity of priming: a cognitive neuroscience perspective. Nat Rev Neurosci 5:853–862.

Scott SK, Blank CC, Rosen S, Wise RJ (2000): Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400–2406.

Sendur L, Maxim V, Whitcher B, Bullmore E (2005): Multiple hypothesis mapping of functional MRI data in orthogonal and complex wavelet domains. IEEE Trans Signal Proc 53:3413–3342.

Suckling J, Bullmore E (2004): Permutation tests for factorially designed neuroimaging experiments. Hum Brain Mapp 22:193–205.

Talairach J, Tournoux P (1988): A coplanar stereotaxic atlas of the human brain. Stuttgart: Thieme.

Venables WN, Ripley BD (1999): Modern applied statistics with S-PLUS. New York: Springer.

von Kriegstein, K, Eger E, Kleinschmidt A, Giraud, AL (2003): Modulation of neural responses to speech by directing attention to voices or verbal content. Cogn Brain Res 17:48–55.

Wink AM, Bernard F, Salvador R, Bullmore E, Suckling J (2006): Age and cholinergic effects on hemodynamics and functional coherence of human hippocampus. Neurobiol Aging (in press).

Zarahn E, Aguirre GK, D'Esposito M (1997): Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null hypothesis conditions. Neuroimage 5:179–197.