◆ Human Brain Mapping 7:1–14(1999) ◆

# Mixture Model Mapping of Brain Activation in Functional Magnetic Resonance Images

**Brian S. Everitt* and Edward T. Bullmore**

*Department of Biostatistics and Computing, Institute of Psychiatry, London, UK*

**Abstract:** We report on a novel method of identifying brain regions activated by periodic experimental design in functional magnetic resonance imaging data. This involves fitting a mixture distribution with two components to a test statistic estimated at each voxel in an image. The two parameters of this distribution, the proportion of nonactivated voxels, and the effect size can be estimated using maximum likelihood methods. Standard errors of the parameters can also be estimated. The fitted distribution can be used to derive brain activation maps and two examples are described, one involving a visual stimulation task, the other an auditory stimulation task. The method appears to have some advantages over direct use of the *P*-values corresponding to each voxel's value of the test statistic. *Hum. Brain Mapping 7:1–14, 1999.* © 1999 Wiley-Liss, Inc.

**Key words:** finite mixture model; brain activation mapping; fMRI; alternative hypothesis

## INTRODUCTION

There is already an extensive literature on statistical methods for the analysis of functional magnetic resonance images of the brain [e.g., Rabe-Hesketh et al., 1997]. In general, these methods involve estimating some measure of the experimentally determined signal at each voxel of the image, and testing this statistic against its null distribution. The usual approach is then to consider a voxel ''activated'' by the experimental design if its test statistic has an associated *P*-value which is lower than some predetermined threshold. Within this general framework, opinions differ as to the most suitable model for the experimental effect; the most appropriate way of fitting a model to time series data that are likely to demonstrate residual, i.e., unmodelled, temporal autocorrelation; and how to choose the threshold *P*-value against which to judge departure from the null hypothesis of no activation.

In this paper, we introduce an alternative approach for identifying activation in the voxels of an image, which involves fitting a simple finite mixture model [Everitt and Hand, 1981] to the observed distribution of the estimated test statistics. The mixture has two components, one corresponding to the distribution of the test statistic under the null hypothesis of no activation, and the other to its distribution under the alternative hypothesis of activation. Each component is explicitly specified to have a particular form. The parameters in the proposed model can be estimated by maximum likelihood methods and their standard errors obtained simply from the inverse of a Hessian matrix. The results from fitting the model are estimates both of the proportion of activated voxels, and of the overall effect size, i.e., the parameter that characterizes the activation distribution. In addition, for each voxel, an estimated posterior probability that it is activated can be determined. These probabilities can be used to produce activation maps.

Contract grant sponsor: Wellcome Trust.

*Correspondence to: B.S. Everitt, Department of Biostatistics and Computing, Institute of Psychiatry, London SE5 8AF, UK.
E-mail: b.everitt@iop.bpmf.ac.uk

Received for publication 1 December 1997; accepted 1 May 1998

## FORMULATION OF A MIXTURE MODEL

Bullmore et al. [1996] showed that the magnetic resonance (MR) signal change induced by periodic experimental design can be modelled by the sum of a sine wave and a cosine wave at the experimentally determined frequency of alternation between contrasting cognitive or sensorimotor conditions. The amplitude of the sine wave $\gamma$ may be understood to measure the magnitude of MR signal change related to neural activation, and the amplitude of the cosine wave $\delta$ to measure hemodynamically mediated delay. From these two amplitudes, Bullmore et al. [1996] derived a test statistic, the fundamental power quotient (FPQ), which they use as the basis of separating voxels into activated and nonactivated. The statistic is given by

$$\mathrm{FPQ} = \frac{1}{2}\left[\left(\frac{\hat{\gamma}}{\hat{\sigma}}\right)^2 + \left(\frac{\hat{\delta}}{\hat{\sigma}}\right)^2\right], \tag{1}$$

where $\hat{\gamma}$ and $\hat{\delta}$ are estimates of $\gamma$ and $\delta$ and $\hat{\sigma}$ is an estimate of the assumed common standard error of each estimate. Under the null hypothesis of no experimentally determined signal change (nonactivation), TFPQ = 2 × FPQ has a chi-squared distribution with two degrees of freedom. In the presence of an experimental effect in a voxel (activation) there will be an increase in amplitude of the sine wave but no corresponding increase in the amplitude of the cosine wave. In this case then, $\hat{\sigma}$ is assumed to have a normal distribution with zero mean and unit variance, and $\hat{\sigma}$ a normal distribution with mean greater than zero (say $\mu$), but again with a variance of one. Consequently in the presence of an experimental effect the test statistic, TFPQ will have a *noncentral chi-squared distribution* with noncentrality parameter $\lambda = \mu^2$ and again two degrees of freedom.

It follows, therefore, that the distribution of the test statistic over all voxels in an image, both activated and not activated, can be modelled by a mixture of these two component distributions: chi-squared with two degrees of freedom and the appropriate noncentral chi-squared. If p denotes the proportion of nonactivated voxels in an image comprising N voxels in total, the mixture distribution can be written down more formally as follows:

$$f(x; \mu, p) = pf_1(x) + (1 - p)f_2(x; \mu) \tag{2}$$

where $f_1$ is the appropriate chi-squared (null) distribution and $f_2$ the appropriate noncentral chi-squared (alternative) distribution. Explicitly, $f_1$ and $f_2$ are as follows:

$$f_1(x) = \frac{1}{2} e^{-\frac{1}{2}x} \tag{3}$$

$$f_2(x) = \frac{1}{2} e^{-\frac{1}{2}(x+\lambda)} \sum_{r=0}^{\infty} \frac{\lambda^r x^r}{2^{2r}[r!]^2} \tag{4}$$

If the observed values of the test statistic are represented as $x_1, x_2, \ldots, x_N$, the log-likelihood of the data (L) is given by:

$$L = \sum_{i=1}^{N} \log f(x_i; \mu, p) \tag{5}$$

The parameters p and $\mu$ may be estimated in the usual way by maximizing L, using any one of a variety of optimization methods. The method used in the numerical examples given later is from Gay [1984] and is implemented in S-Plus as the function *nlminb*. This method maximizes L, constraining p to lie in the interval (0,1).

Having found estimates of p and $\mu$ ($\hat{p}$ and $\hat{\mu}$), an estimated posterior probability for activation can be assigned to each voxel using

$$\Pr(\text{activated}|x_i) = \frac{(1 - \hat{p})f_2(x_i; \hat{\mu}, \hat{p})}{f(x_i; \hat{\mu}, \hat{p})} \tag{6}$$

where $x_i$ represents the value of the test statistic TFPQ for voxel Pi. All voxels in the image which have an estimated posterior probability of activation greater than some arbitary threshold value–a value of 0.5 would appear to be an obvious choice, can then be classified as activated.

**TABLE I. Results from fitting the mixture model to three simulated data sets***

| | p | μ | $\hat{p}$ (SE) | $\hat{\mu}$ (SE) |
|---|---|---|---|---|
| Data set 1 | 0.5 | 4 | 0.487 (0.017) | 4.00 (0.052) |
| Data set 2 | 0.5 | 2 | 0.483 (0.045) | 2.01 (0.095) |
| Data set 3 | 0.2 | 2 | 0.258 (0.045) | 2.02 (0.074) |

* Initial parameter values used in all cases were the values used to generate the data. Each data set consisted of 1,000 observations.

## MIXTURE MODELLING OF SIMULATED DATA

Three data sets were simulated by sampling N=1,000 observations from the distribution f(x;μ,p), defined in

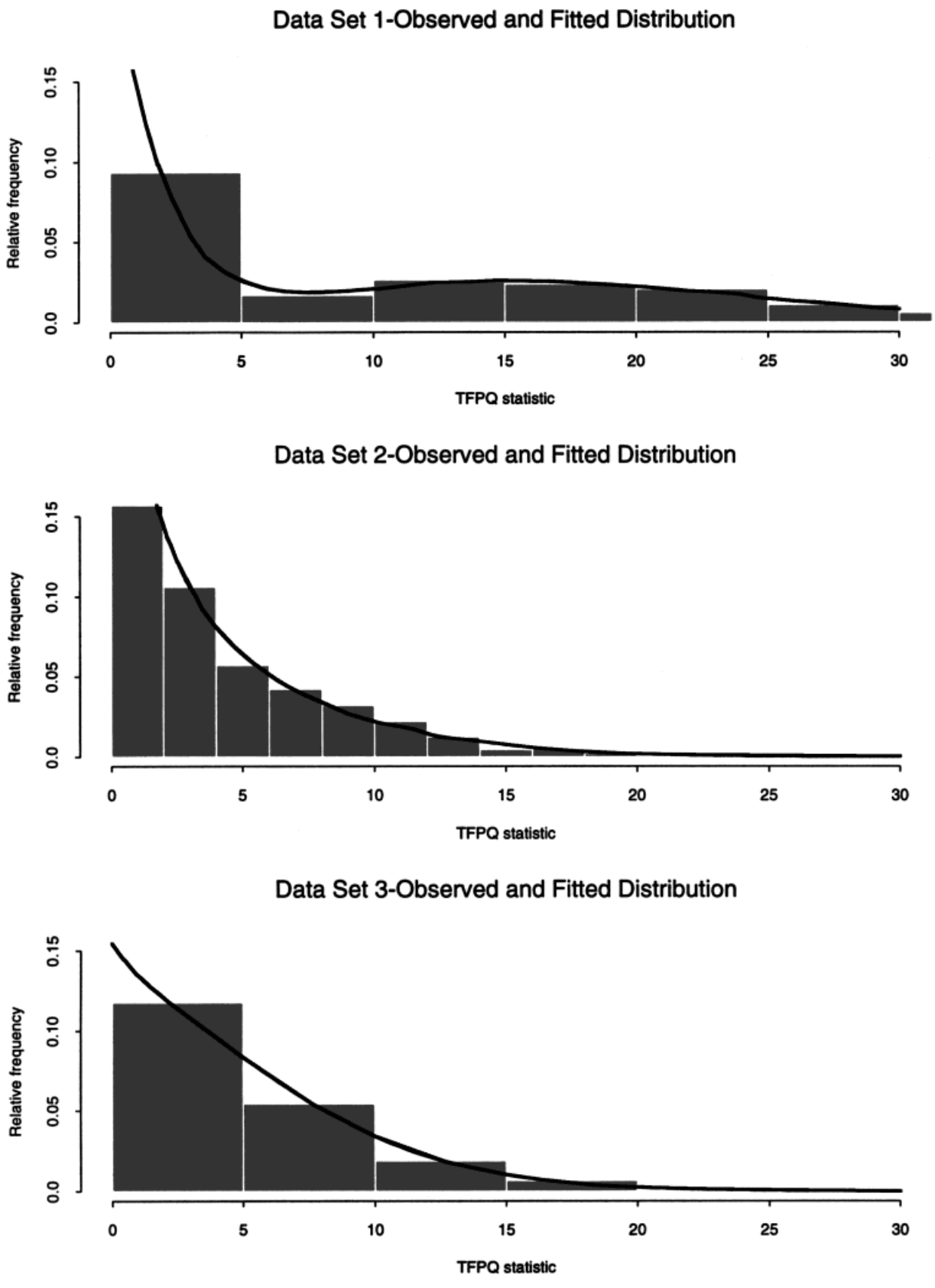


**Figure 1.**

Observed histograms and fitted (solid line) distributions for three sets of data, each with 1,000 observations simulated from the mixture distribution defined in Equation (2). Data set 1, $p = 0.5$, $\mu = 4$; data set 2, $p = 0.5$, $\mu = 2$; data set 3, $p = 0.2$, $\mu = 2$.

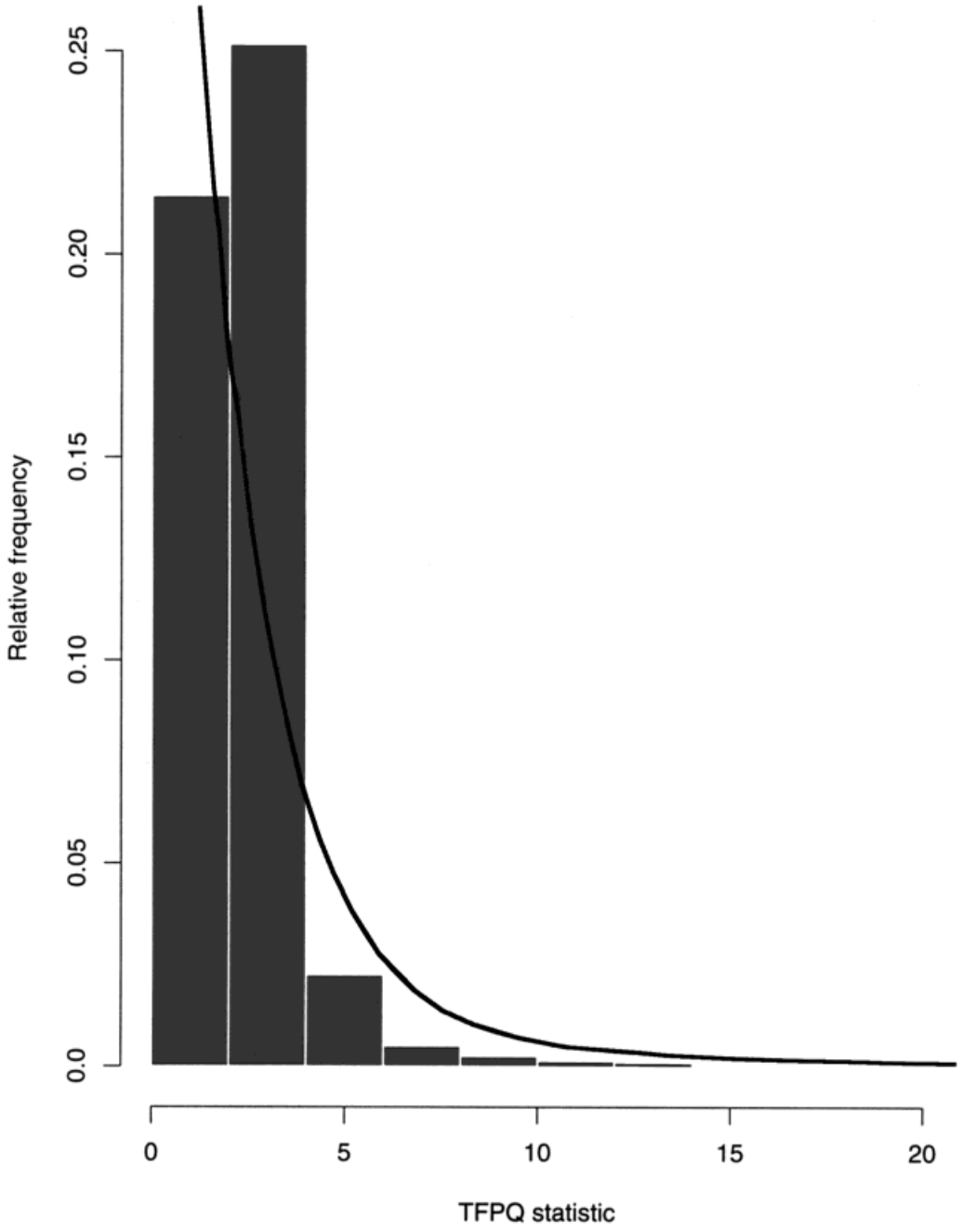


**Figure 2.**
Observed distribution (histogram) and fitted mixture distribution (solid line) for the statistic TFPQ = 2 × FPQ observed at 26,535 voxels in the visual stimulation data.

Equation (2). The data sets differed in terms of the proportion of nonactivated voxels, p, and the size of the noncentrality parameter, $\lambda = \mu^2$. The results are shown in Table I. Identical parameter estimates to those shown were obtained from a number of different starting values for the two parameters. For accurate estimates, large sample sizes are likely to be needed when p is far from 0.5 and/or μ is small. For imaging data, however, this is unlikely to be a problem since the number of voxels, N, will in general be large. The histograms of the observations and the fitted distributions are shown in Figure 1.

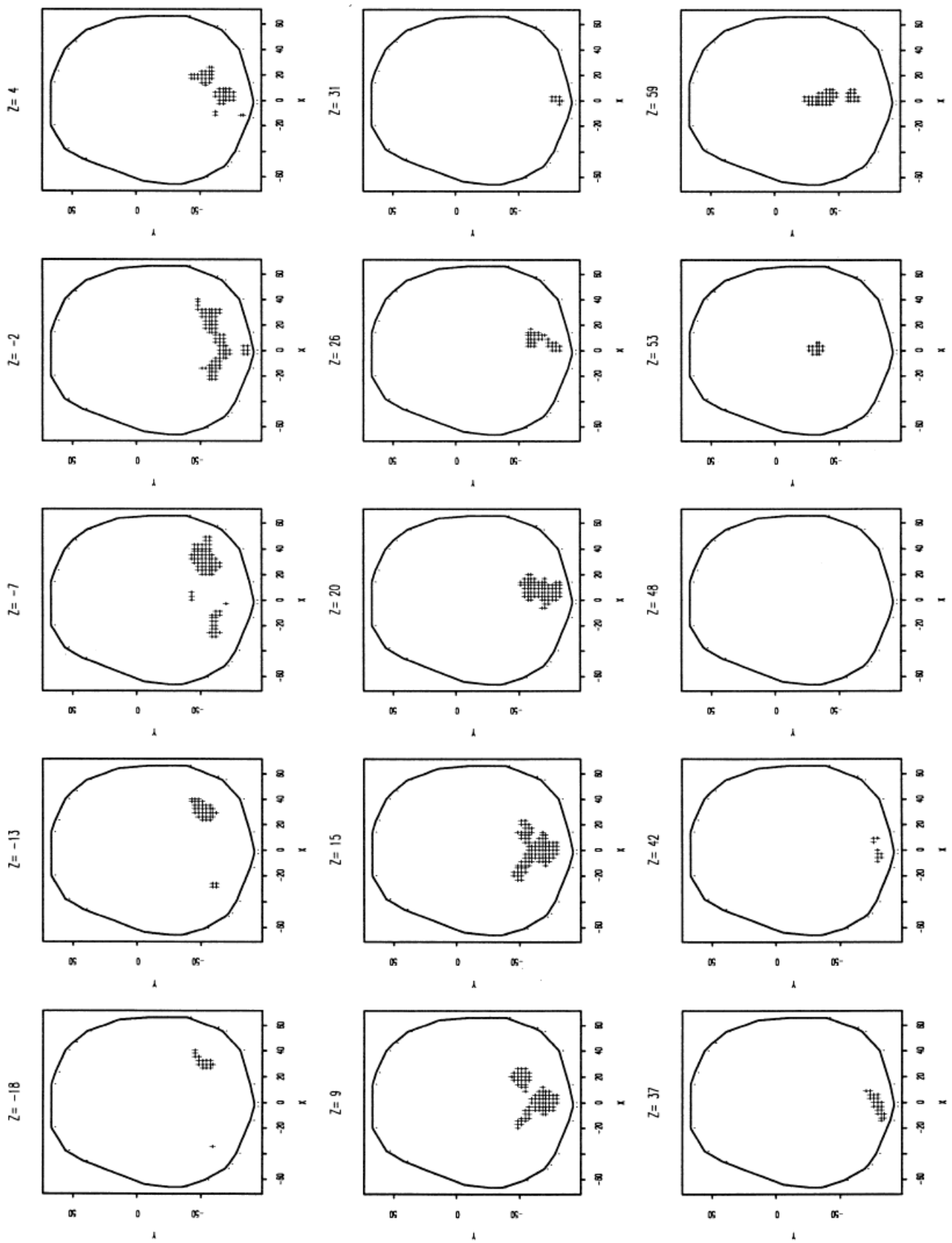


**Figure 3.**

Mixture model activation map of visual simulation data derived from estimated posterior probabilities of activation for the 26,535 voxels. Threshold for the posterior probabilities = 0.5. Each slice of data is displayed in the standard anatomical space of Talairach and Tournoux [1988].

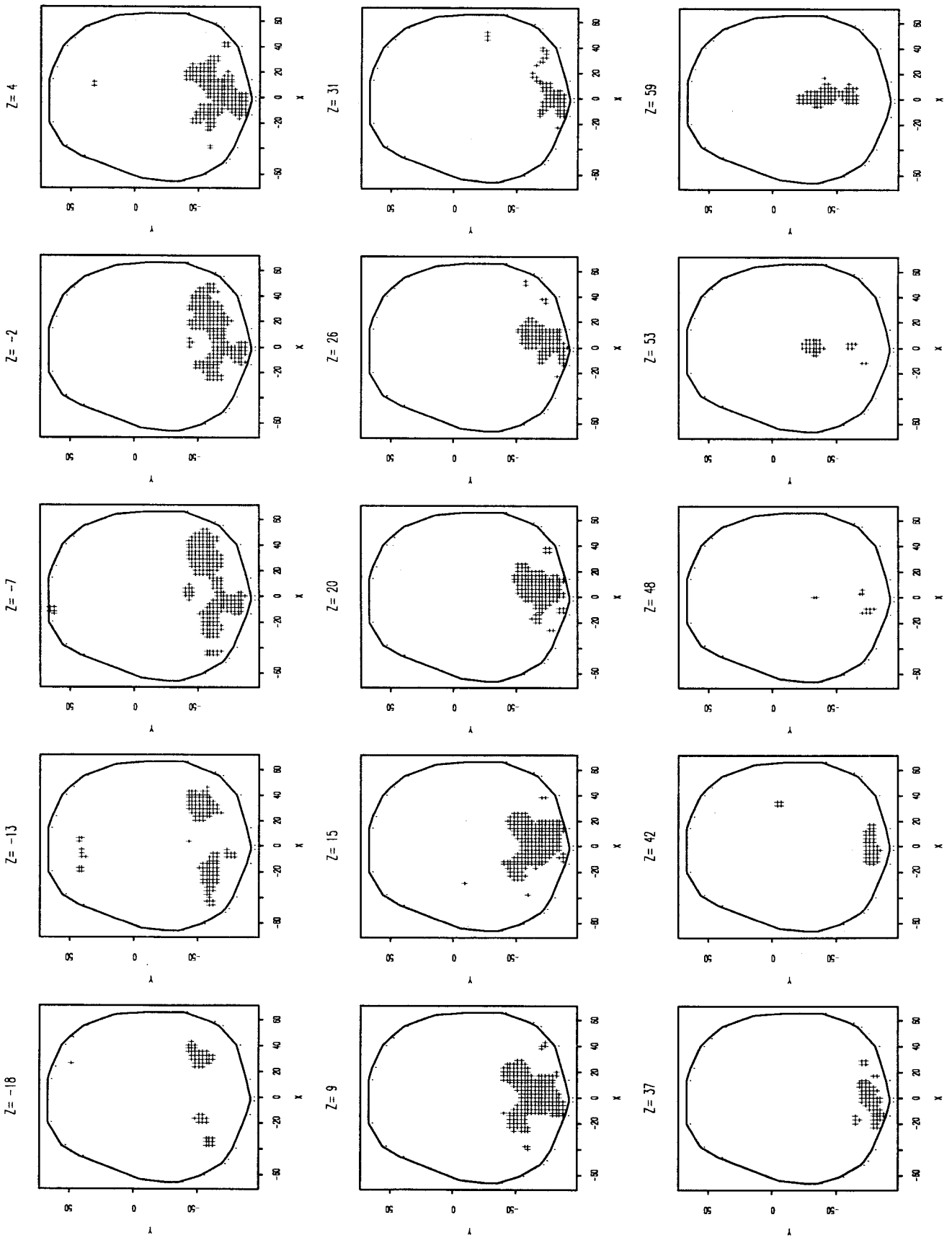


**Figure 4.**

Activation map for visual stimulation data derived from thresholding the *P*-values of the statistic TFPQ at 0.05.

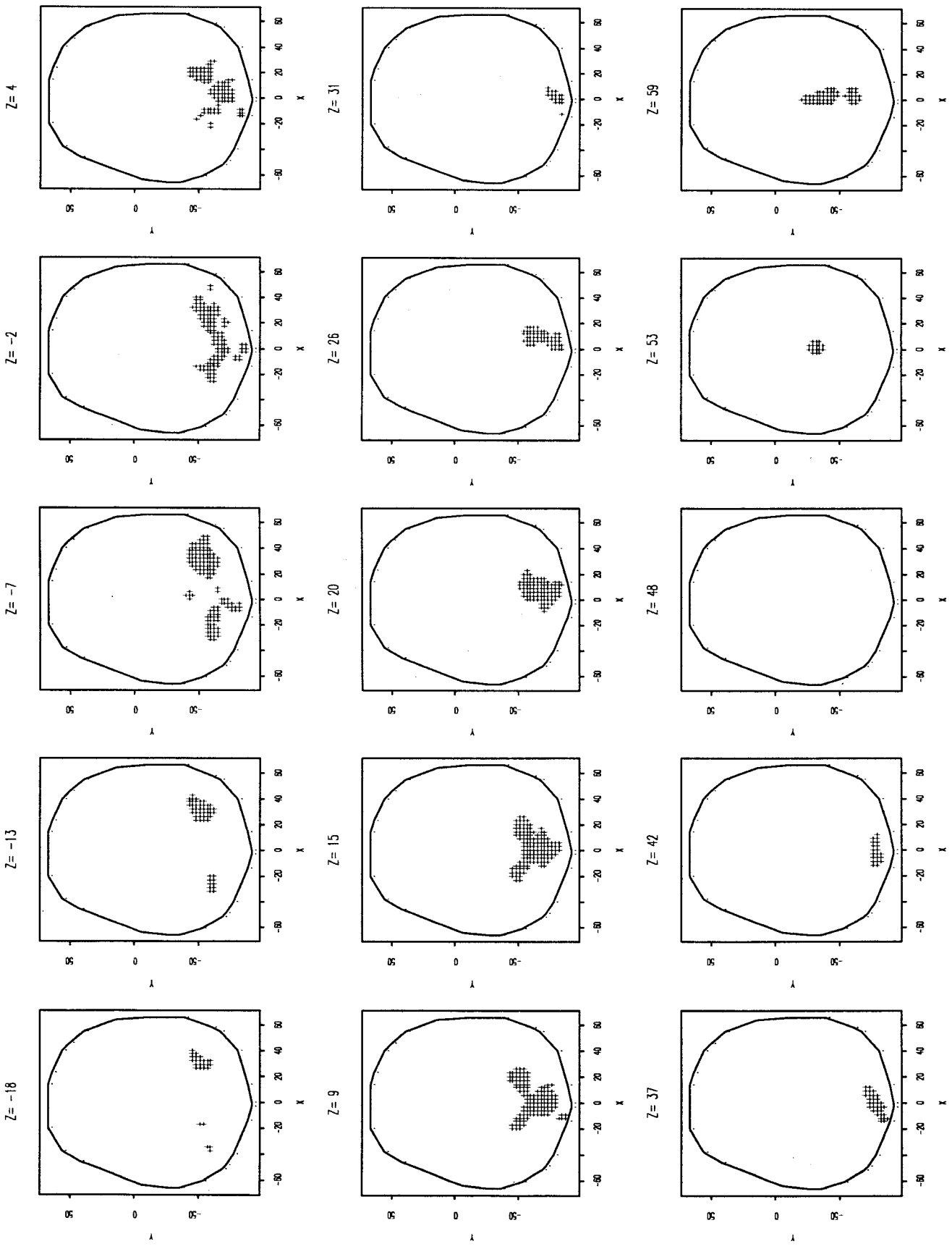


**Figure 5.**

Activation map for visual stimulation data derived from thresholding the *P*-values of the statistic TFPQ at 0.01.

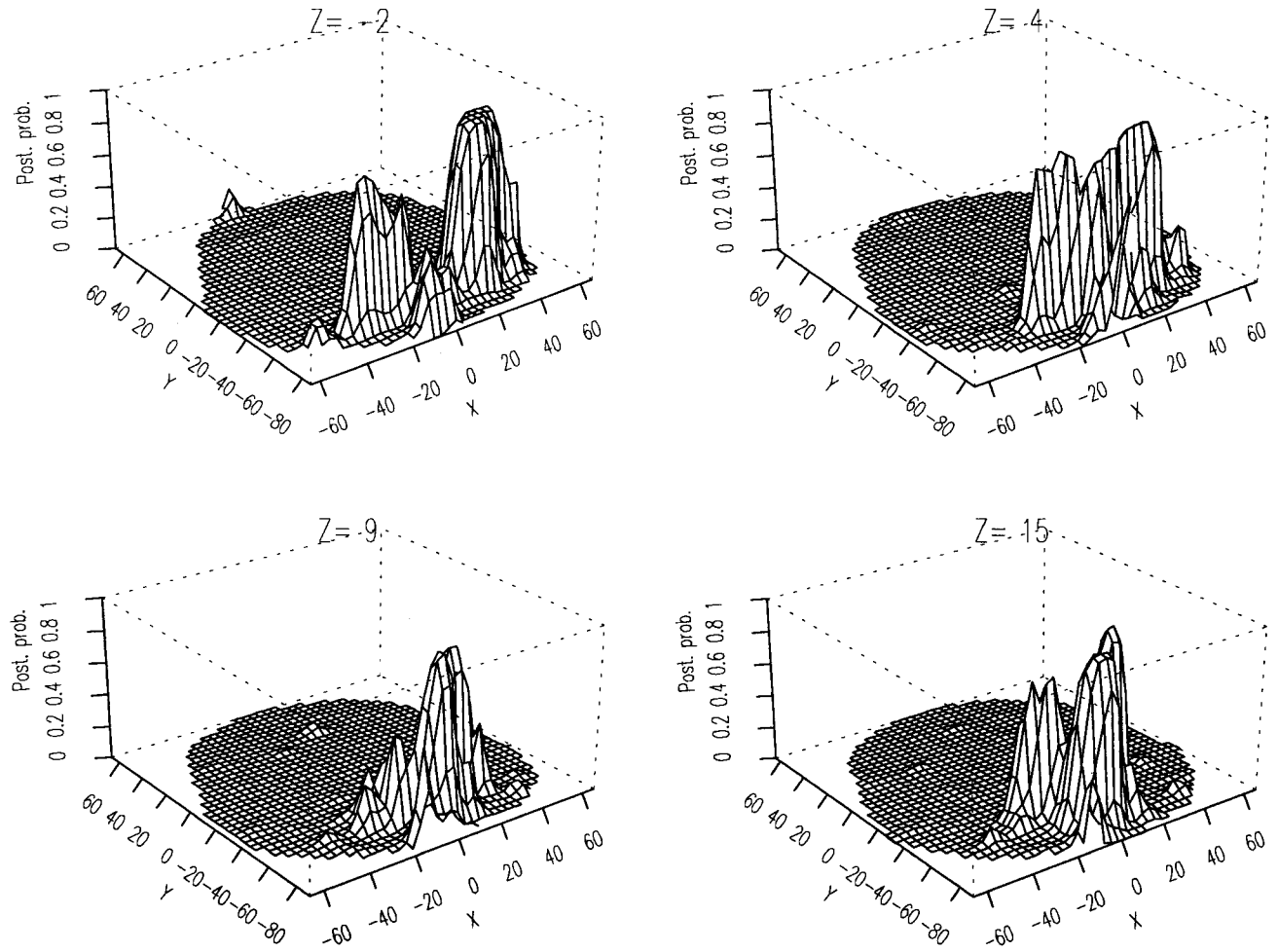


**Figure 6.**
Perspective plots of estimated posterior probabilities of activation for four slices of the visual stimulation data.

(Experience on simulated data sets suggested that the number of terms necessary to accurately evaluate $f_2$ given by Equation (4), for the values of μ likely to be encountered in practice, is at least 20.)

Standard errors of the parameters can be found relatively simply from the inverse of the estimated Hessian matrix (for details see Appendix). The estimated standard errors for the three simulated data sets are shown in Table I.

**TABLE II. Parameter estimates for visual stimulation fMRI data***

| | Starting values | | Final values | |
|---|---|---|---|---|
| | $\hat{p}$ | $\hat{\mu}$ | $\hat{p}$ (SE) | $\hat{\mu}$ (SE) |
| 1 | 0.8 | 3 | 0.9659 (0.0018) | 3.467 (0.0559) |
| 2 | 0.8 | 5 | 0.9659 (0.0018) | 3.467 (0.0559) |

* Upper limit for r in Equation (4) was set at 25.

## MIXTURE MODELLING OF FMRI DATA

Functional MRI data were acquired on a GE Signa 1.5 T system (General Electric, Milwaukee, WI) retrofitted with an ANMR operating console (ANMR, Woburn, MA) at the Maudsley Hospital, London. One hundred single-shot $T_2^*$-weighted gradient echo echo-planar images depicting blood oxygen level-dependent (BOLD) contrast [Ogawa et al., 1990] were acquired at each of 14 near-axial noncontiguous planes parallel to the AC-PC line: TE = 40 msec, TR = 3 sec, in-plane resolution = 3 mm, slice thickness = 7 mm, interslice skip = 0.7 mm, number of signal averages = 1.

Images were acquired from a healthy male volunteer during the following periodically designed experi-

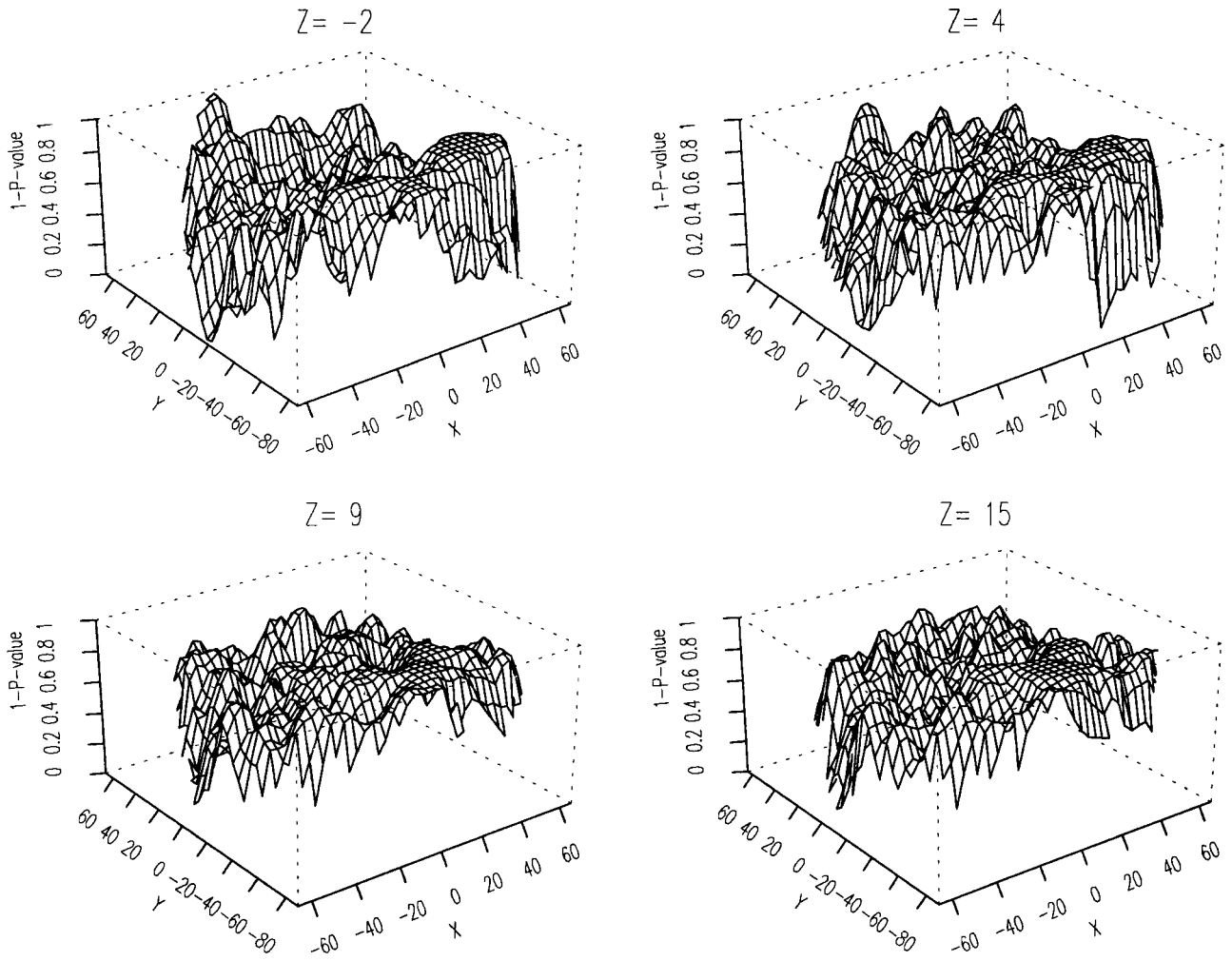


**Figure 7.**
Perspective plots of 1 - *P*-value for the statistic TFPQ for four slices of the visual stimulation data.

ments, each of which involved a regular and repeated contrast between two (A and B) contrasting sensory conditions:

- Visual stimulation: A, the subject was exposed to 30 sec of 8 Hz photic stimulation; B, the subject was exposed to 30 sec of darkness. This cycle was repeated 5 times in the course of a 5-min experiment. The subject was instructed to lie quietly in the scanner with his eyes open.
- Auditory-verbal stimulation: A, the subject heard a written narrative read aloud to him for 39 sec; B, the subject heard nothing but the continuous background noise of the scanner for 39 sec. The cycle was repeated <5 times in the course of a 5-min experiment. The subject was instructed to lie quietly in the scanner.

After motion correction [Brammer et al., 1997], the fMRI time series, $\{Y_t\}$, t = 1, 2, 3, . . . , 99, at each voxel was fitted to the following sinusoidal regression model:

$$Y_t = \gamma \sin(\omega t) + \delta \cos(\omega t) + \gamma' \sin(2\omega t) + \delta' \cos(2\omega t) + \gamma'' \sin(3\omega t) + \delta''(3\omega t) + \alpha + \beta t + \rho_t \quad (7)$$

where ω is the angular frequency of alternation between experimental conditions ($2\alpha/60$ radians in these data), and $\alpha + \beta t$ represents a (nuisance) linear trend. Since the residual terms $\{\rho_t\}$ are temporally autocorrelated, the model is fitted by pseudogeneralized least squares, modelling the residuals as a first-

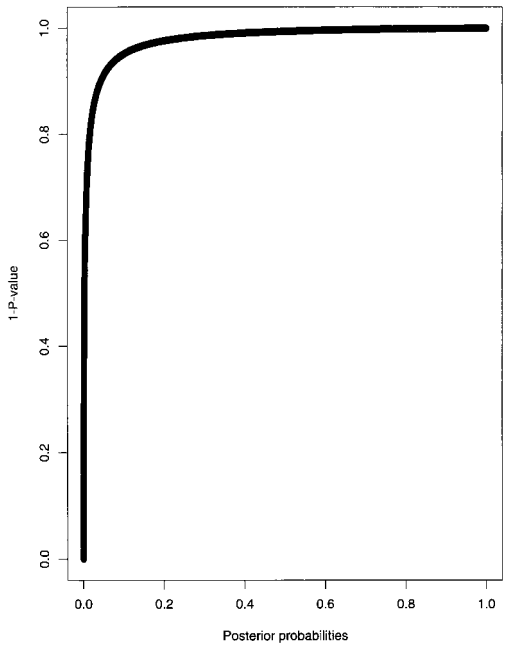


**Figure 8.**
Plot of 1 - *P*-values against posterior probabilities for all voxels of the visual stimulation data.

order autoregressive process [Bullmore et al., 1996]. An estimate of the standardized power at fundamental frequency ω was derived at each voxel from the estimated coefficients $\hat{\gamma}$, $\hat{\delta}$ as in Equation (1), and represented in a parametric FPQ map. FPQ maps were interpolated to 21 slices and registered in the standard space of Talairach and Tournoux [1988] as described by Brammer et al. [1997].

For the visual stimulation data, this analysis yielded 26,535 estimates of the test statistic TFPQ at 21 near-axial slices through the brain. The overall distribution is illustrated by a histogram in Figure 2.

Fitting the mixture model given by Equation (2) to these data with two different sets of starting values for the two parameters, p and μ, gives the results shown in Table II. The final values obtained are identical in each case, although the number of iterations of the optimization technique needed to reach these values differed. The estimated standard error of p is very small and leads to an approximate 95% confidence interval for the parameter of (0.9623, 0.9695). The corresponding confidence interval for the parameter μ calculated from its estimated standard error is (3.3452, 3.5788). (The estimated covariance between the parameter estimates was 0.000059; it must be remembered that because of spatial correlations of the TFPQ statistic in fMRI data, that the standard errors of the two parameters in the mixture are likely to be underestimated. Correcting the resulting confidence intervals to allow for spatial correlation will be taken up in a later paper.) The fitted distribution is superimposed on the observed histogram in Figure 2.

Using the estimated parameter values in Equation (6), estimated posterior probabilities of activation were calculated. Voxels having values above 0.5 were colored black, against a white background of nonactivated intracerebral voxels, to form a simple mixture model map of brain activation; see Figure 3 for selected slices. This may be compared with the corresponding maps obtained from thresholding the *P*-values of the test statistic directly at 0.05 (Fig. 4) and 0.01 (Fig. 5). The latter is very similar to the mixture model map.

A more dramatic image of the activation can be obtained from a perspective plot of the posterior probabilities of activation for particular slices, as shown in Figure 6. (Bivariate interpolation, using the method described in Akima [1978], was used to produce these plots.) It is of interest to compare these with perspective plots of one minus the *P*-value for each pixel, for the same slices (Fig. 7); the areas of activation are now not so immediately apparent. The explanation of the difference is seen if 1 - *P*-values are plotted against posterior probabilities for all the voxels (see Fig. 8). The initial steepness of the graph shows that posterior probabilities close to zero map to 1 - *P*-values from 0–0.9 and above.

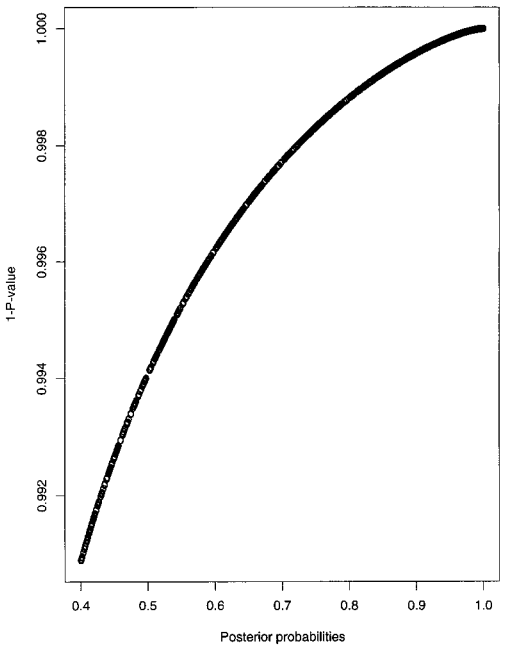


**Figure 9.**
Plot of 1 - *P*-values against posterior probabilities for voxels for which the latter is greater than 0.4 (visual stimulation data).

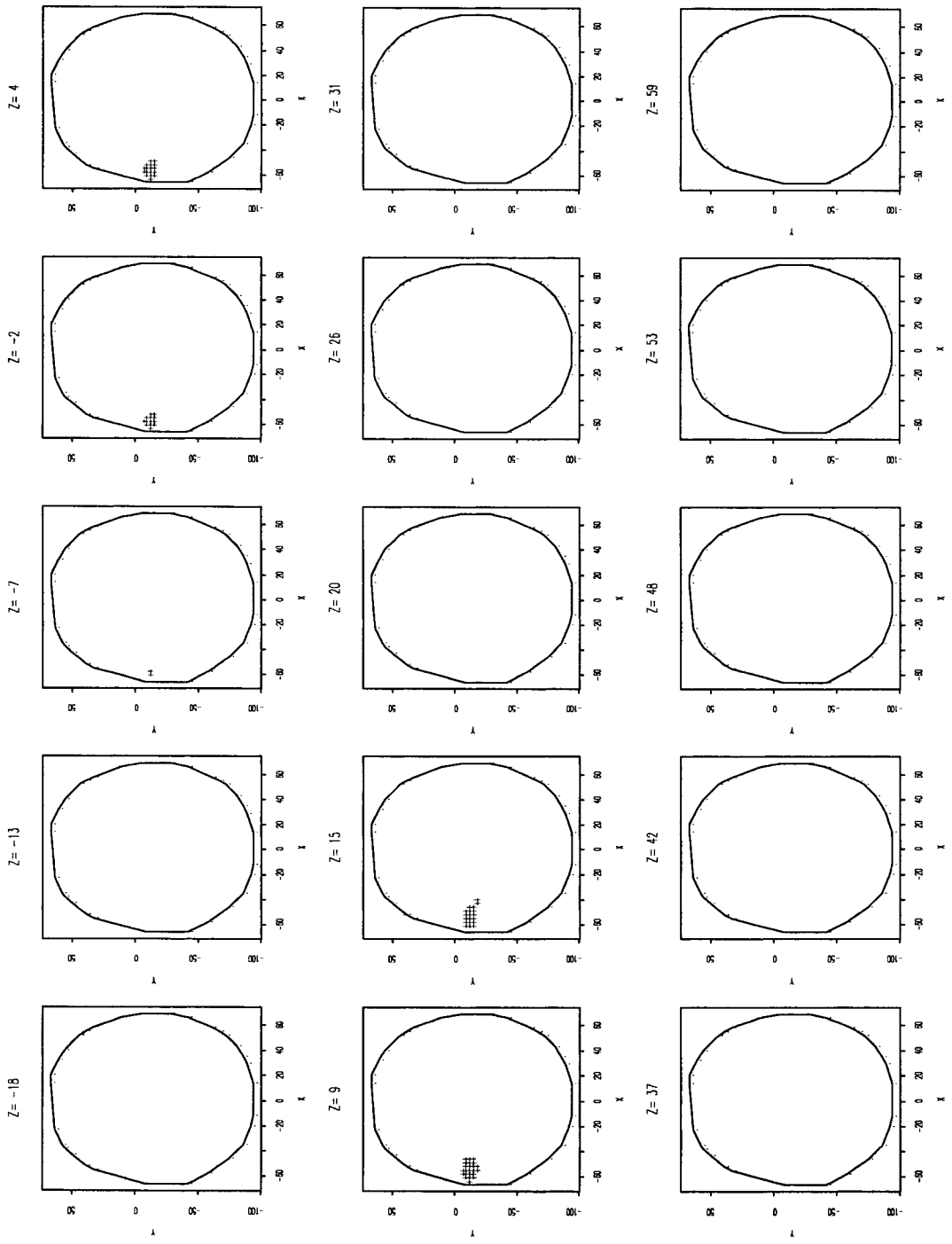


**Figure 10.**
Mixture model activation map of auditory stimulation data derived from the estimated posterior probabilities of activation for the 28,317 voxels, thresholded at 0.5.

By replotting Figure 8 for only those voxels with posterior probabilities of activation above 0.4 (see Fig. 9), the equivalence of *P*-values and posterior probabilities in the area of most interest is displayed. For these data, a posterior probability of 0.5 is seen to be equivalent to a *P*-value of approximately 0.008; this is the reason, of course, why Figures 3 and 5 are very similar.

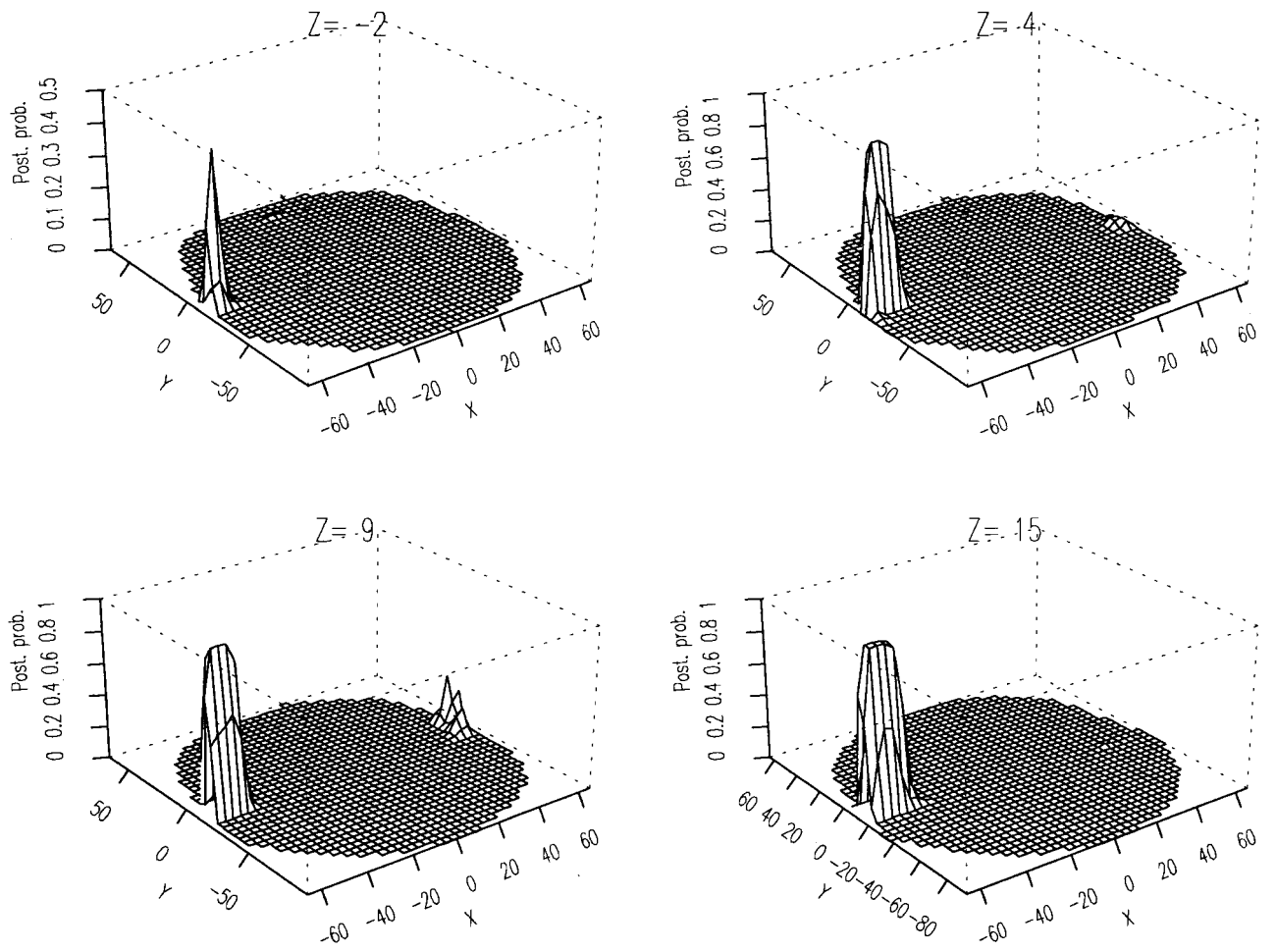


**Figure 11.**
Perspective plots of estimated posterior probabilities of activation for four slices of the auditory stimulation data.

Moving on to the auditory data which consisted of 28,317 estimates of TFPQ at 22 near-axial slices through the brain, the parameter estimates and their estimated standard errors were $\hat{p} = 0.997$ (0.000387), $\hat{\mu} = 4.852$ (0.144563). Here the estimated proportion of activated pixels is very small. The resulting mixture model map for 15 slices is shown in Figure 10, and the perspective plots of posterior probabilities of activation for the most interesting slices in Figure 11.

For this example, the plot corresponding to Figure 9 is shown in Figure 12.

In this case, the *P*-value equivalent to a posterior probability of 0.5 is approximately 0.00023.

## DISCUSSION

The proposed method is novel in the context of existing statistical methods for functional MR image analysis because it entails explicit consideration of the alternative distribution of a test statistic, as well as its null distribution. The particular mixture model we have specified in this paper is applicable to the distribution of standardized power at the (fundamental) frequency of alternation between baseline and activation conditions in a periodic experimental design. Under the null hypothesis (of no experimentally determined effect), we have modelled the distribution of the fundamental power quotient (FPQ), multiplied by a factor of two, as chi-squared with two degrees of freedom. Under the alternative hypothesis (of experimentally determined signal change or ''activation''), we have modelled the distribution of $2 \times$ FPQ as noncentral chi-squared with two degrees of freedom. The technique could easily be extended to other relevant test statistics having null and alternative distributions different from those used here. By using a

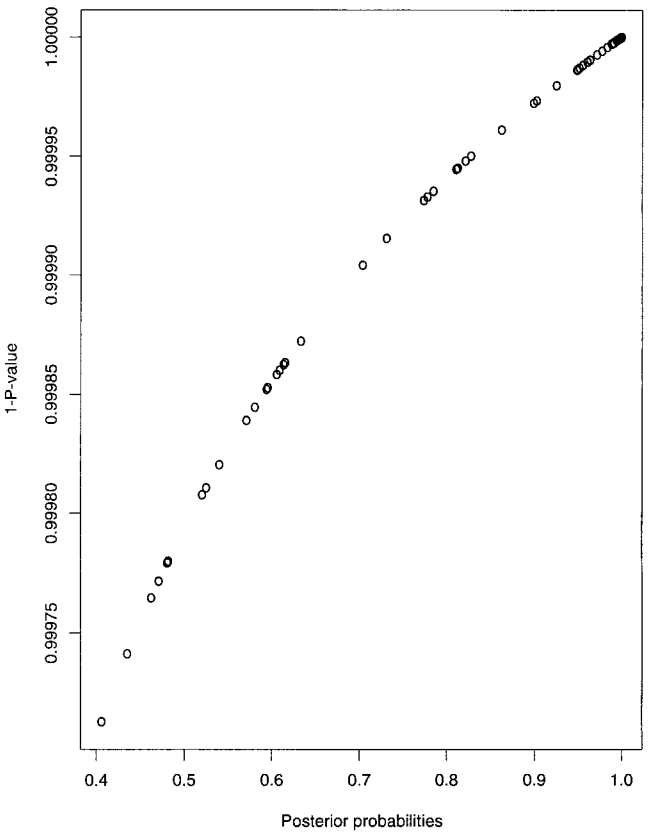


**Figure 12.**
Plot of 1 - *P*-values against posterior probabilties for voxels for which the latter is greater than 0.4 (auditory stimulation data).

mixture with more than two components, the method could also be extended to situations where it was believed that voxels activated in several, widely separated regions of the brain were not distributed under a single alternative distribution, but under two or possibly more alternative distributions.

The results of fitting a mixture distribution to fMRI data are of interest for two main reasons.

First, mixture model parameters summarize the experimental effect over the whole image in a holistic and unconditional way. The estimated parameters, 1 - $\hat{p}$ and $\hat{\mu}$, have a ready interpretation in terms of the proportion of activated voxels in the image and the "average" size of the experimental effect over all activated voxels. They could therefore potentially be used as omnibus statistics to summarize the experimental response over an entire image, or to compare the extent and / or strength of response between two groups of images. As omnibus statistics, estimated mixture model parameters have a number of attractive properties. They have, for example, an associated measure of their variability, namely the standard errors estimated from the Hessian matrix (although note previous caveat about exaggerated precision). These can be used to produce confidence intervals for each parameter. A confidence interval for p that included the value one would be indicative of no activation. Additionally, the values at the upper ends of the two confidence intervals could be used to produce "conservative" activation maps if required. To compare two groups of images, estimated p and μ values for each image could be used to calculate group means, and then a confidence interval for their difference could be calculated from the estimated standard errors associated with each image.

Second, the estimated posterior probabilities of activation assigned to each voxel by the mixture model approach appear to identify activated regions far more distinctly than the *P*-values associated with the test statistic of each voxel. In Figure 6, for example, the peaks of activation in the occipital cortex rise steeply from a flat plateau of unactivated brain regions; this picture contrasts dramatically with the much less informative perspective plot of 1 - *P*-values for the same data shown in Figure 7.

Both the mixture approach and the direct use of *P*-values are, in essence, "thresholding" techniques, and from plots such as those shown in Figures 8 and 9 it is possible to assess the equivalence between particular posterior probability values and *P*-values. So, for example, if a posterior probability threshold of 0.5 was used for most data sets (and in the absence of information about the costs of identifying activated voxels as nonactivated and vice versa, this value seems sensible), it could, by applying the mixture method, be converted into the corresponding *P*-value for each particular data set. Some investigators (including the current authors) might find this a more appealing approach than searching for an acceptable correction to apply to the *P*-values.

## CONCLUSIONS

The mixture model approach described in this paper appears to be a simple but relatively effective approach to the problem of identifying areas of activation in fMRI experiments, given the values of a particular statistic estimated at each voxel. Activation maps derived from the estimated posterior probabilities are equivalent to those derived directly from the *P*-values of the test statistic for some particular threshold significance level. The advantages of the mixture approach are that it avoids the need to consider how to correct *P*-values and it provides standard errors for the esti-

mated parameters that might be used to provide a range of activation maps from "generous" to "conservative."

In addition to deriving activation maps for individual subjects, the estimated parameters that result from the mixture model might also be useful for comparing activation between different groups of subjects, and for characterizing individual images. By introducing different and / or more component distributions into the mixture, the approach should be capable of modelling other aspects of fMRI data.

## ACKNOWLEDGMENTS

E.T.B. is supported by the Wellcome Trust. We thank our colleagues in the Brain Image Analysis Unit, Institute of Psychiatry (London, UK), particularly Drs. Mick Brammer and Ian Wright, for their material support in developing these ideas.

## REFERENCES

Akima H (1978): A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. ACM Trans Mathematical Software 4:148–164.

Brammer M, Bullmore ET, Simmons A, Williams SCR, Grasby PM, Howard RJ, Woodruff PWR, Rabe-Hesketh S (1997): Generic brain activation mapping in fMRI: A nonparametric approach. Magn Reson Imaging 15:763–770.

Bullmore E, Brammer M, Williams SCR, Rabe-Hesketh S, Janot N, David A, Mellers J, Howard R, Sham P (1996): Statistical methods of estimation and inference for functional MR image analysis. Magn Reson Med 35:261–277.

Everitt BS, Hand DJ (1981): Finite Mixture Distributions, London: Chapman and Hall.

Gay DM (1984): A trust region approach to linearly constrained optimization. In: Lootsma FA (ed): Numerical Analysis. Berlin: Springer, pp 171–189.

Ogawa S, Lee TM, Kay AR, Tank DW (1990): Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci USA 87:9868–9872.

Rabe-Hesketh S, Bullmore ET, Brammer MJ (1997): Analysis of functional magnetic resonance images. Stat Methods Med Res 6:215–237.

Talairach J, Tournoux P (1988): A Coplanar Stereotactic Atlas of the Human Brain, Stuttgart: Thieme Verlag.

## APPENDIX

### Calculating standard errors for the mixture model parameters

The Hessian matrix **H** that leads to standard errors for the parameter estimates in the proposed mixture model is defined as follows:

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 L}{\partial p^2} & \frac{\partial^2 L}{\partial p \partial \mu} \\ \frac{\partial^2 L}{\partial \mu \partial p} & \frac{\partial^2 L}{\partial \mu^2} \end{pmatrix}$$

where L is the log-likelihood. Substituting estimated parameter values in the elements of $H^{-1}$ gives the estimated covariance matrix of the two parameters. Standard errors are found from the square roots of the diagonal elements.

The elements of **H** are obtained from the following:

$$L = \sum_{i=1}^{N} \log f(x_i; p, \mu)$$

$$\frac{\partial L}{\partial p} = \sum_{i=1}^{N} \frac{f_1(x_i) - f_2(x_i)}{f(x_i)}$$

$$\frac{\partial^2 L}{\partial p^2} = -\sum_{i=1}^{N} \frac{[f_1(x_i) - f_2(x_i)]^2}{[f(x_i)]^2}$$

$$\frac{\partial^2 L}{\partial \mu \partial p} = \sum_{i=1}^{N} \left\{ -\frac{1}{f(x_i)} \frac{\partial f_2(x_i)}{\partial \mu} - \frac{(1-p)}{[f(x_i)]^2} \frac{\partial f_2(x_i)}{\partial \mu} [f_1(x_i) - f_2(x_i)] \right\}$$

$$\frac{\partial L}{\partial \mu} = \sum_{i=1}^{N} \frac{(1-p)}{f(x_i)} \frac{\partial f_2(x_i)}{\partial \mu}$$

$$\frac{\partial^2 L}{\partial \mu^2} = (1-p) \sum_{i=1}^{N} \left\{ \frac{1}{f(x_i)} \frac{\partial^2 f_2(x_i)}{\partial \mu^2} - \frac{(1-p)}{[f(x_i)]^2} \left[ \frac{\partial f_2(x_i)}{\partial \mu} \right]^2 \right\}$$

The terms $[\partial f_2(x_i)] / \partial \mu$ and $[\partial^2 f_2(x_i)] / \partial \mu^2$ needed to evaluate each of the above are given by

$$\frac{\partial f_2(x_i)}{\partial \mu} = -\mu f_2(x_i) + \frac{1}{2} e^{-\frac{1}{2}(x+\lambda)} I_1$$

$$\frac{\partial^2 f_2(x_i)}{\partial \mu} = -\mu \frac{\partial f_2(x_i)}{\partial \mu} - f_2(x_i) - \mu \left[ \frac{\partial f_2(x_i)}{\partial \mu} + \mu f_2(x_i) \right] + \frac{1}{2} e^{-\frac{1}{2}(x+\lambda)} I_2$$

where

$$I_1 = \sum_{r=0}^{\infty} \frac{2r\mu^{2r-1} x^r}{2^{2r}[r!]^2}$$

$$I_2 = \sum_{r=0}^{\infty} \frac{2r(2r-1)\mu^{2r-2} x^r}{2^{2r}[r!]^2}.$$