

# Thresholding in correlation analyses of magnetic resonance functional neuroimaging

J. Baudewig\*, P. Dechent, K.D. Merboldt, J. Frahm

*Biomedizinische NMR Forschungs GmbH am Max-Planck-Institut für biophysikalische Chemie, 37070 Göttingen, Germany*

Received 15 August 2003; accepted 22 August 2003

## Abstract

The definition of objective and effective thresholds in MRI of human brain function is a crucial step in the analysis of paradigm-related activations. This paper introduces a user-independent and robust procedure that calculates statistical parametric maps based on correlation coefficients. Thresholds are introduced as  $p$  values and defined with respect to the physiologic noise distribution of the individual maps. Experimental examples from the human visual and motor system rely on dynamic acquisitions of gradient-echo echo-planar images (2.0 T, TR = 2000 ms,  $96 \times 128$  matrix) with blood oxygenation level-dependent contrast. The results demonstrate the disadvantages of thresholding with fixed correlation coefficients. In contrast, taking the individual noise into account allows for a derivation of  $p$  values and a reliable identification of highly significant activation centers. An adequate delineation of the spatial extent of activation may be achieved by adding directly neighboring pixels provided their correlation coefficients comply with a second lower  $p$  value threshold. © 2003 Elsevier Inc. All rights reserved.

*Keywords:* Functional magnetic resonance imaging; Correlation analysis; Brain activation; Image processing; Human

## 1. Introduction

Since its inception about a decade ago, functional MRI (fMRI) of the human brain has developed into an important research tool in both basic and clinical neuroscience. The majority of studies relies on blood oxygenation level-dependent (BOLD) contrast [1], which exploits changes in the intravascular content of paramagnetic deoxyhemoglobin as an endogenous indicator of altered brain activity. While the exact nature of the coupling between (synaptic) brain activity and cerebral hemodynamics remains a subject under investigation [2], the enormous increase in the number of fMRI studies points to a rather pragmatic approach by most researchers. Unfortunately, this attitude bears the risk of ignoring some of the putative problems associated with data acquisition and data analysis. For example, apart from issues such as structural susceptibility artifacts [3], the appropriate statistical treatment of fMRI data sets is part of an ongoing discussion despite the availability of detailed theoretical descriptions [4,5]. In fact, most strategies face com-

mon problems: 1) they attempt to calculate a map of a defined statistical parameter; 2) they have to identify activated pixels via task-related fMRI signal increases (or decreases) in a noisy environment; and 3) they use some kind of “thresholding” at a particular step during processing. The purpose of this contribution is to present an effective data-driven approach to the analysis of fMRI data sets that is physiologically plausible, statistically justified, and of considerable practical value.

## 2. Methods and subjects

Altogether, the proposed evaluation strategy has evolved during the analysis of several thousands of fMRI data sets acquired in multiple studies of this laboratory. An early version of the initial concept has been described by Kleinschmidt et al. [6]. Experiments have been carried out at 2.0 T using the standard head coil (Siemens Vision, Erlangen, Germany). Subjects were mainly healthy volunteers but eventually included patients. In most cases, BOLD-sensitive MRI was based on echo-planar imaging (EPI) to provide adequate volume coverage at intermediate resolution, while selected applications employed (multi-echo) FLASH se-

\* Corresponding author. Tel.: +49-551-201-1728; fax: +49-551-201-1307.

*E-mail address:* jbaudew@gwdg.de (J. Baudewig).

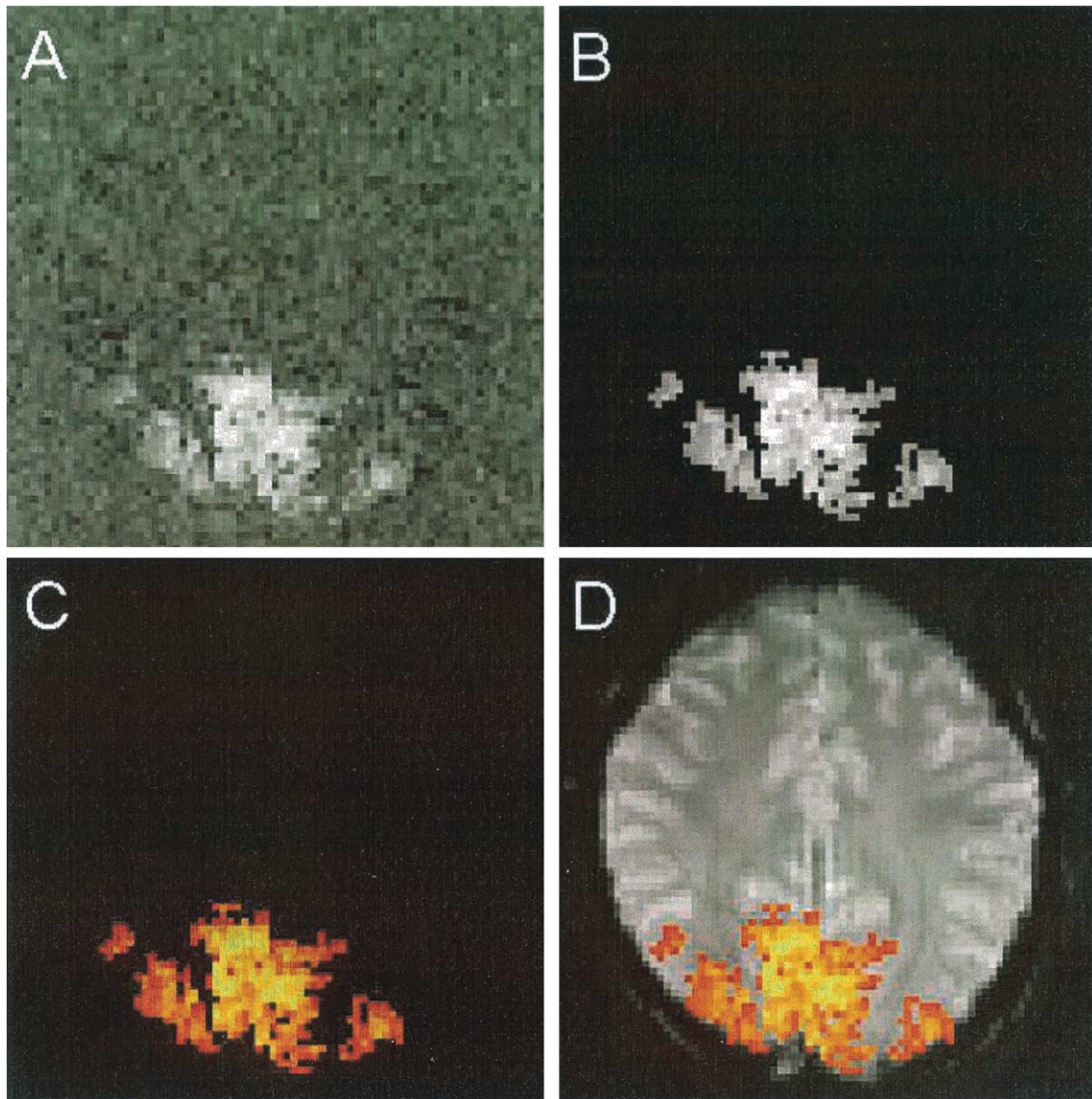


Fig. 1. Thresholding with a fixed correlation coefficient  $cc$ . Maps of correlation coefficients at fixed lag time (A) as gray-scale correlation image as well as after (B) thresholding with  $cc = 0.5$ , (C) color-coding, and (D) superposition with an EPI reference image.

quences in order to achieve high spatial resolution. The fMRI experiments performed here refer to an active motor task and a passive visual stimulation. Acquisitions used a standardized multi-slice EPI sequence (TR/TE = 2000/54 ms,  $96 \times 128$  acquisition matrix) covering a rectangular field-of-view at  $2 \times 2 \text{ mm}^2$  resolution and 4 mm section thickness.

The motor task contrasted self-paced bimanual finger movements with motor rest, while the visual study compared stimulation by a flickering checkerboard (5 Hz) with the presentation of a gray screen. In either case, the paradigm represented a simple block design alternating 12 s periods of enhanced (or attenuated) neural activity with 18 s periods of control. Each experiment started with a 20 s control period in order to establish steady-state conditions followed by 6 repetitions of the 30 s paradigm.

The dynamic fMRI data sets were processed using a temporal correlation analysis as originally proposed by Bandettini et al. [7]. A reference function with a rectangular box-car waveform served to represent the stimulation protocol. This function was shifted by 4 s (2 images) to account for hemodynamic latencies and rise times. Correlation coefficients were calculated on a pixel-by-pixel basis and visualized in a gray-scaled correlation image as shown in Fig. 1A.

Mathematically, the cross-correlation of two functions is, of course, defined as a *function* which represents the correlation coefficients of these functions for all possible lag times. Thus, also the cross-correlation analysis of an fMRI data set yields a complete function for each image pixel. However, this work follows the standard procedure of reducing this function to a single correlation coefficient  $cc$  for

a fixed lag time. Depending on the definition of the reference function mimicking the stimulus protocol, this lag time may be 4 s - reflecting the hemodynamic delay - or zero for a correspondingly shifted function.

### 2.1. Fixed correlation coefficients

A correlation image contains information not only about the functional anatomy and the strength of the paradigm-associated neural processing, but also about the quality of the fMRI experiment in terms of background noise and subject performance such as involuntary movements. Instead of taking advantage of this information, it is common practice to uniformly reduce the data by means of a fixed correlation coefficient and to generate apparently unambiguous spots of color-coded activation on top of a high-quality anatomic reference image. It is this part of the analysis that deserves further consideration.

The values of the correlation image shown in Fig. 1A range from +1 (perfect positive correlation) to 0 (no correlation) and -1 (inverted negative correlation). Noteworthy, visual inspection by an experienced experimenter immediately results in a subjective judgment of the quality of the gray-scaled activations. This effective cognitive process needs to be translated into an objective and user-independent mathematical strategy. In the most simple case, a separation of “activated” pixels or regions from other areas is achieved by introducing a fixed  $cc$  threshold, e.g.,  $cc = 0.5$  as in Fig. 1B. In this case, thresholding simply refers to the fact that all pixels with  $cc < 0.5$  will be dismissed, leaving only pixels with  $cc \geq 0.5$  for further processing (Fig. 1C) and visualization (Fig. 1D).

It should be emphasized that the common procedure of overlaying an EPI-derived activation map onto an anatomic image with much better spatial resolution is misleading and often unacceptable. This is because echo-planar images may exhibit serious spatial distortions which then translate into geometric inconsistencies between the functional map and the anatomy—and obviously hamper any reliable interpretation of the results. Conversely, the use of an EPI reference image as in Fig. 1D offers an inherent quality control to both the experimenter and the reader of a publication: the degree of susceptibility artifacts present in a particular section will become evident. Although trivial, a color-coded activation map can never be better than the quality of the original EPI acquisitions, and any attempt to hide this problem should be discouraged.

Although fixed correlation coefficients seem to be popular thresholds, even the original work on fMRI correlation analysis widely varied the actual values from 0.7 to 0.25 [7]. On the other hand,  $cc$  thresholds as low as 0.25 are often (mis-) interpreted as a hint that no activation could be found. However, correlation coefficients do not represent objective levels of confidence. Even worse, their values strongly depend on the length of the acquired fMRI time series, i.e., the number of images or data points contributing to the corre-

lation analysis. When the duration of the fMRI experiment is doubled while all other parameters remain the same—including the underlying noise—the resulting correlation coefficient of a selected, truly activated pixel decreases. Or conversely, for a shorter fMRI experiment with fewer images, such correlation coefficients increase. Accordingly, the use of identical correlation coefficient thresholds in fMRI experiments comprising different numbers of images is by no means statistically justified.

The apparently paradoxical result that longer acquisitions yield activation maps with lower correlation coefficients is demonstrated in Fig. 2, which compares activation maps for a motor task and for visual stimulation that were obtained for a single, fixed threshold of  $cc = 0.4$  and time series of 25, 50, and 100 images. The shorter time series were taken as subsets of the full acquisition comprising 100 images. While the resulting correlation images (not shown) are more or less similar to each other, the thresholded activation maps yield notable differences. The enhanced number of pixels with  $cc \geq 0.4$  for the series of 25 images causes considerable “activation” outside the brain, whereas the generally decreased correlation coefficients obtained for “activated” pixels in the experiment with 100 images results in a much cleaner map. The key to understanding this behavior lies in an analysis of the underlying distribution of correlation coefficients.

### 2.2. From correlation coefficients to $p$ values

Fig. 3 shows distributions of correlation coefficients for an fMRI study without any stimulation (null experiment) but otherwise identical acquisition and analysis parameters as in the experiments of Fig. 2. The values are obtained from respective correlation images calculated using similar reference functions as for the actual experiments. Because of the large number of pixels involved, i.e.,  $96 \times 128 = 12,288$ , the histograms may be represented as smoothed distributions of correlation coefficients. In the absence of a stimulus, i.e., without altered brain activity, the “noise” distributions in Fig. 3 emerge as symmetric functions with a Gaussian shape. Their widths increase when the number of images decreases. Thus, correlation coefficients with values above 0.4 by chance are more likely to occur when fewer images contribute to the analysis.

In order to overcome the aforementioned inadequacies, the procedure should involve a statistical test. Fortunately, this is equivalent to the application of a *probabilistic* threshold, which asks to which degree an observed high correlation coefficient occurs by chance or represents a significant effect. In line with earlier suggestions [7] and a wide acceptance in test statistics, it is therefore strongly recommended to use  $p$  values instead of correlation coefficients for thresholding in fMRI. As demonstrated in Fig. 4, a fixed threshold of  $p < 0.001$  leads to a more common behavior as far as activation maps are concerned. In comparison with Fig. 2 and a  $cc$  threshold of 0.4, the fixed  $p$  value corre-

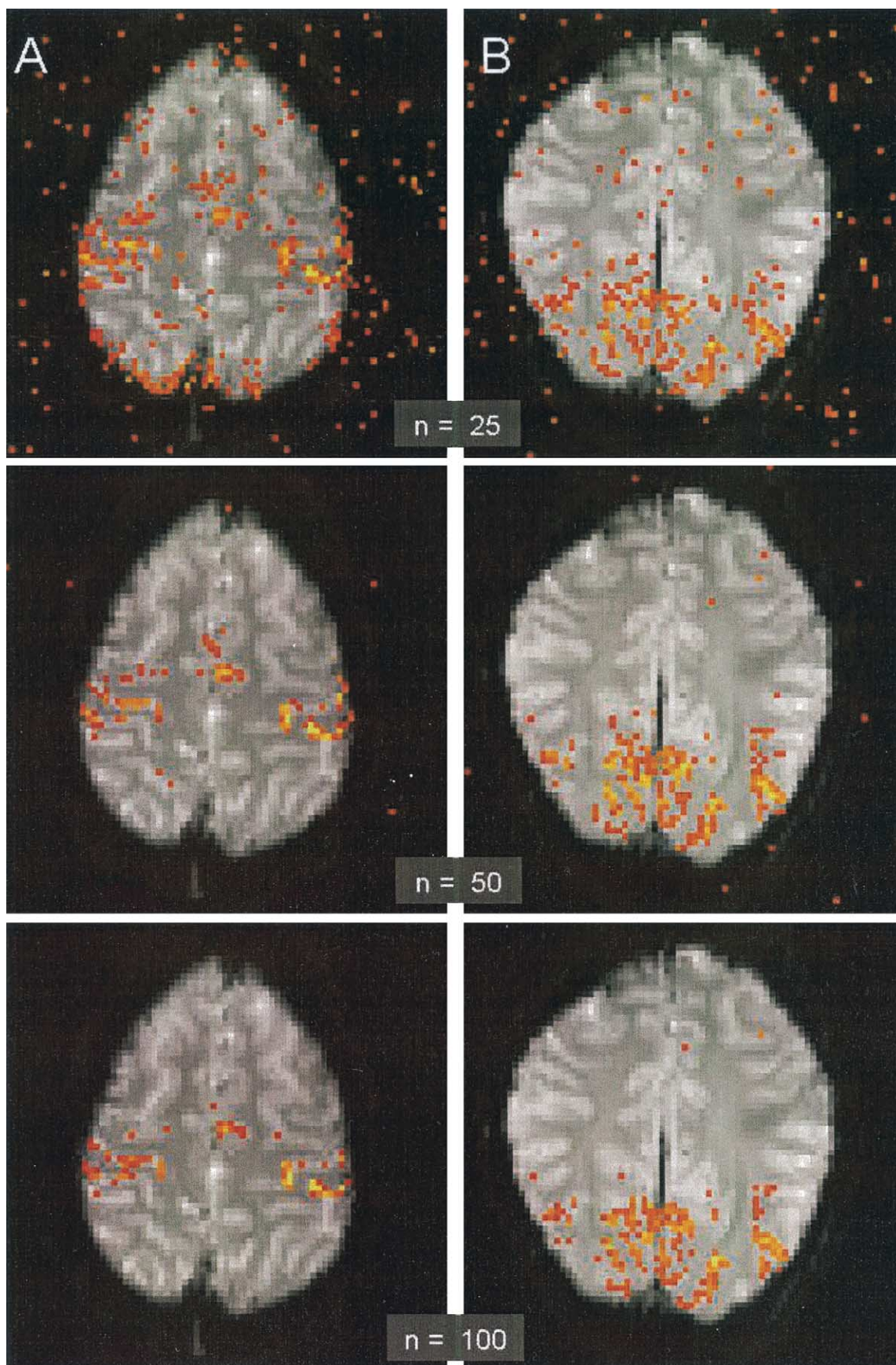


Fig. 2. Activation maps obtained for (A) a hand motor task and (B) visual stimulation using a fixed threshold of  $cc = 0.4$  but different numbers of images from the same time series: (top) 25 images, (middle) 50 images, and (bottom) 100 images (full acquisition).

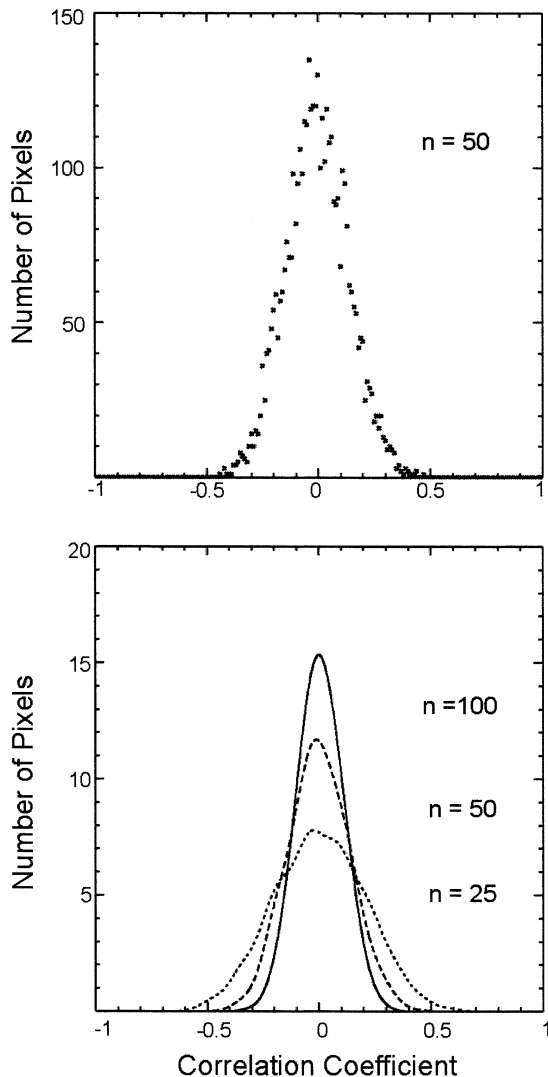


Fig. 3. Distributions of correlation coefficients obtained in the absence of a stimulus (null experiment). Original histogram (top,  $cc$  binning: 0.01 units) and smoothed representations (bottom,  $cc$  binning: 0.001 units) of serial acquisitions with different numbers of images.

sponds to  $cc$  values of 0.588 (25 images), 0.427 (50 images), and 0.305 (100 images), respectively. Thus, a reduction of the number of images decreases the number of pixels that are classified as activated by increasing the inherently used  $cc$  threshold. Or in statistical terms, reducing the sample size results in a loss of power.

### 2.3. Individualized noise distribution

Fig. 5 shows correlation images and corresponding distributions of correlation coefficients obtained for the same data as in Figs. 2 and 4. As theoretically predicted by statistics and demonstrated by the null experiment, the basic form of the distribution is adequately described by a Gaussian curve. Brain activations, or more precisely pixels representing paradigm-associated fMRI signal alterations, manifest themselves as aberration of the true distribution

from the estimated Gaussian noise distribution, especially for high positive-correlation coefficients. The actual histogram can therefore be understood as the sum of two independent distributions reflecting different subpopulations of pixels: a dominating noise distribution given by a large number of nonactivated pixels and a second much smaller distribution of activated pixels with high positive (or negative)  $cc$  values. Of course, deviations from the noise distribution may also occur simultaneously for both positive and negative correlation coefficients.

While part of the noise may arise from technical sources such as signal fluctuations due to the MRI sequence and instrumentation, an important if not dominant source is physiologic noise determined by the subject itself. Putative factors are stimulus-correlated movements, variations in respiration and cardiac pulsations as well as temporal instabilities in hemodynamic responsiveness and mental state. As a consequence, the width of the Gaussian noise distribution not only depends on the number of images (compare Fig. 3), but is also affected by unavoidable and largely uncontrollable biologic factors [6]. In order to account for these problems, one may take advantage of the additional information in the actual distribution of correlation coefficients and develop a suitable correction of the theoretically determined  $p$  values. This is achieved by the following procedure.

In a first step, the histogram of correlation coefficients is used to estimate the underlying noise distribution of a particular correlation image. In order to obtain a reliable estimate, it is desirable to increase the sample size as much as possible. One may therefore consider the entire multislice fMRI data set as a possible optimum yielding rows  $\times$  lines  $\times$  slices (e.g.,  $96 \times 128 \times 24$ ) samples. However, this procedure has to assume a uniform noise distribution in all slices of the covered brain volume. Because multiple sections are usually acquired in an interleaved sequential manner, their noise structure may indeed be influenced by slice-dependent factors. It therefore seems advisable to derive the noise distribution from individual sections. A further restriction stems from the characteristics of the image noise outside the subject's head in the surrounding air. These regions provoke no relevant MRI signal but contribute noise with an often narrower distribution of correlation coefficients than found within brain tissue. Because such contributions may alter the statistical thresholds derived from the estimated noise distribution, which, for example, may lead to reduced specificity, this estimate includes only pixels above a certain MRI signal intensity threshold. Thus, the subsequent introduction of probabilistic thresholds relies on histograms that represent pixels from brain tissue in individual sections.

In a second step, a Gaussian curve is fitted to the central portion of the histogram. The rationale for using this part is the dominance of noise for low correlation coefficients representing nonactivated pixels. The reason for excluding the region around  $cc = 0$  is based on the eventual experi-

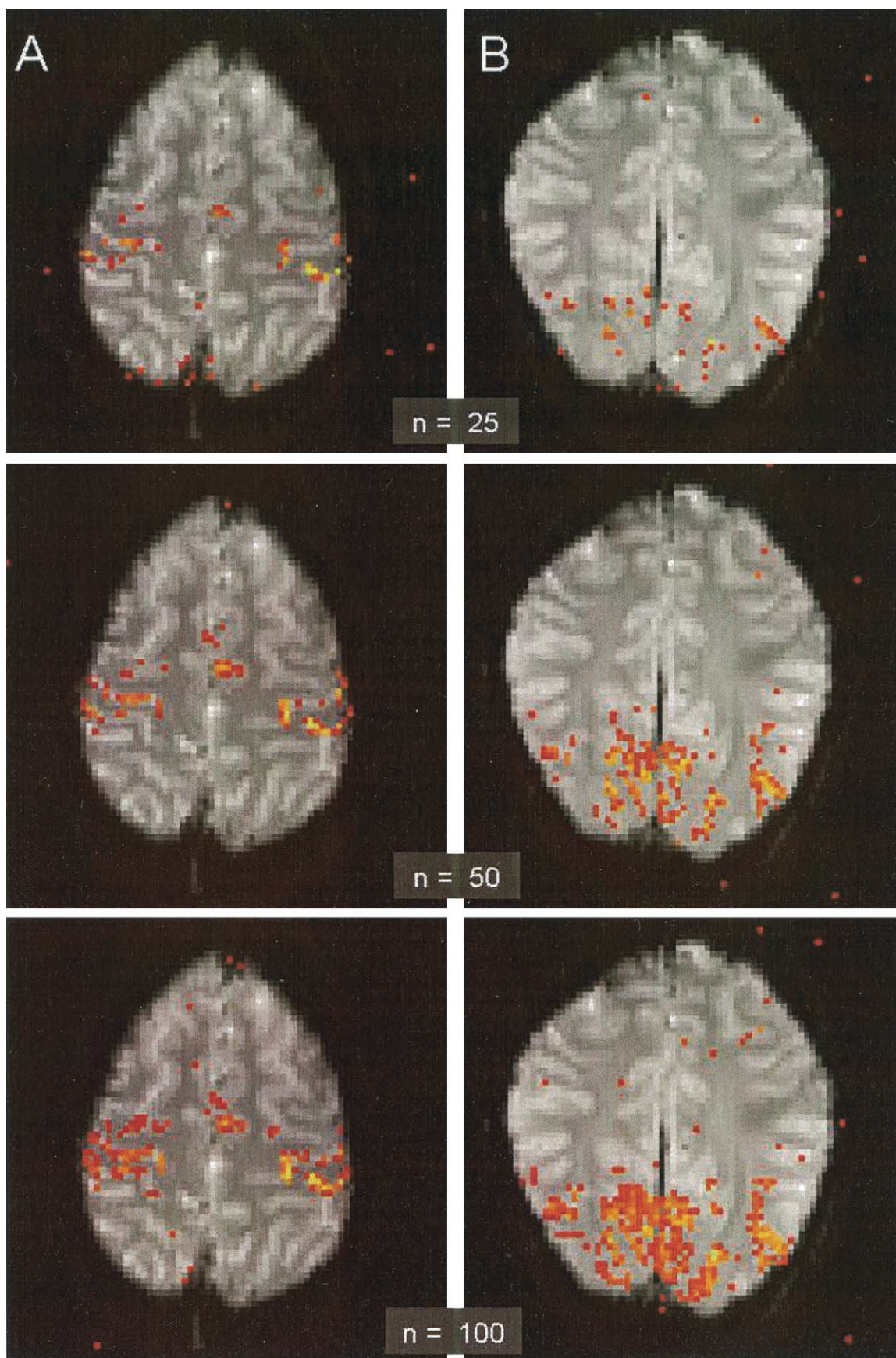


Fig. 4. Activation maps obtained for (A) a hand motor task and (B) visual stimulation (same data as in Fig. 2) using fixed p values rather than correlation coefficients for thresholding.

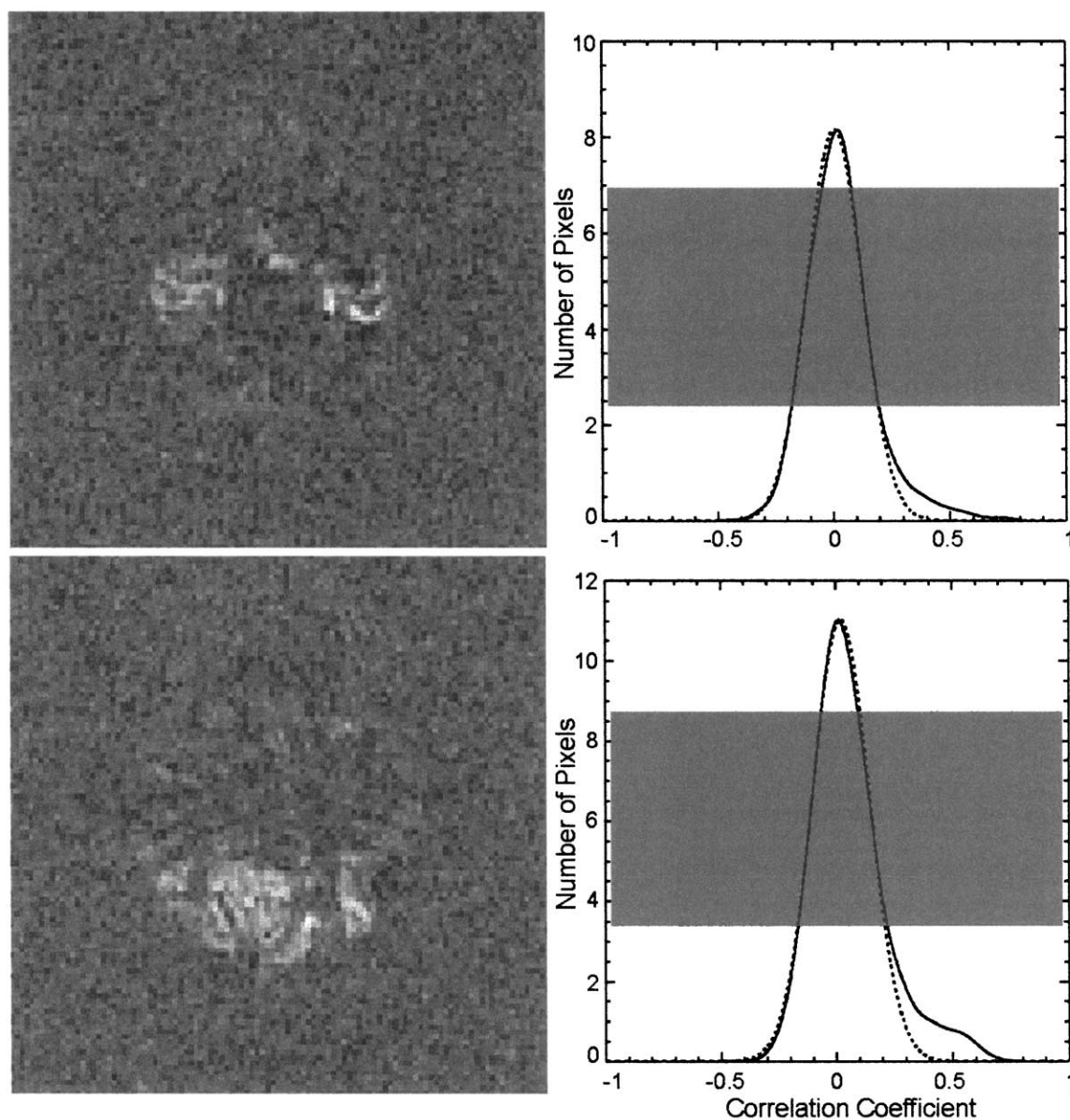


Fig. 5. Correlation images (left) and corresponding histograms of correlation coefficients (right) obtained for a hand motor task (top) and visual stimulation (bottom). While solid lines refer to the actual histogram, dotted lines represent Gaussian curves which are fitted to the central portion of the histogram (shaded) to estimate the underlying noise distribution.

mental observation of a perturbed peak structure of the histogram in cases where the total number of samples is not very high, for example, in transverse brain sections close to the apex. It turns out that a robust definition of the Gaussian-shaped noise distribution may be achieved by fitting those parts of the histogram that cover a 50% range of the observed peak height extending from the 30% level to the 80% level (shaded zones in Fig. 5).

In a final step, demonstrated in Fig. 6, the distribution of correlation coefficients is rescaled into percentile ranks of the individual noise distribution. The subsequent choice of a probabilistic threshold should be optimized to ensure both specificity for the identification of highly significant activations and sensitivity to properly define the spatial extent of

individual activation spots. The rationale behind this strategy is based on the observation that—at least for adequate spatial resolution—correlation images (compare Fig. 5) reveal activated *areas* rather than individual pixels. Activations may therefore be described as ‘mountains’ with a central peak for easy identification and an area around defining the actual size.

#### 2.4. Spatial response delineation

Even when defined with respect to individual noise levels, the use of only a *single* statistical threshold is often unable to yield a good compromise between specificity and sensitivity. For example, as shown in the top part of Fig. 7,

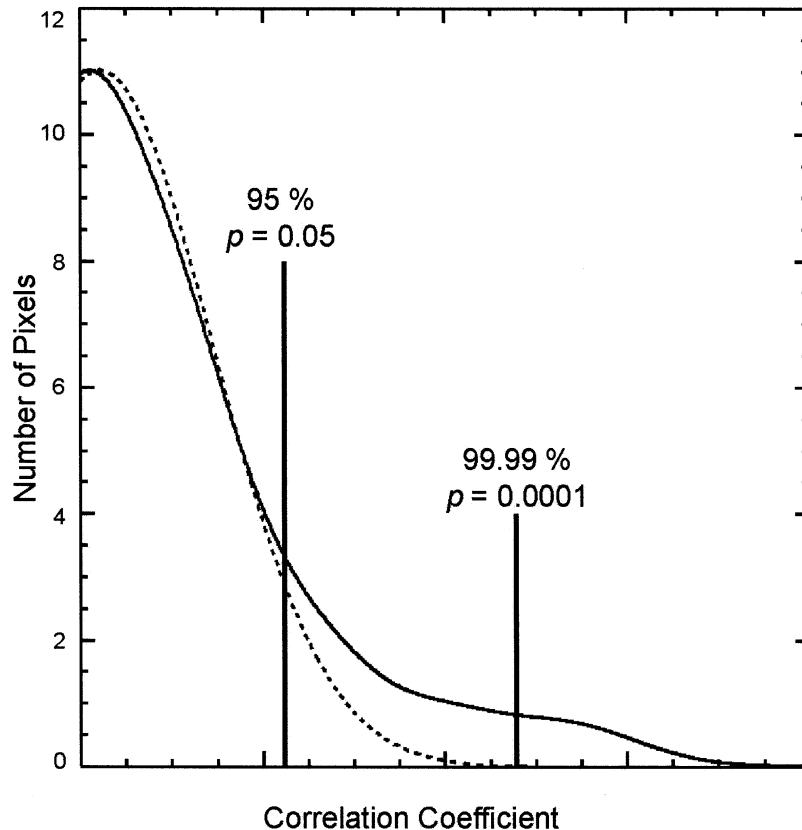


Fig. 6. Rescaling of correlation coefficients to percentile ranks of the estimated noise distribution (broken line). The solid line is a smoothed representation of the histogram ( $cc$  binning: 0.001 units). The percentile ranks of 99.99% ( $p = 0.0001$ ;  $cc = 0.478$ ) and 95% ( $p = 0.05$ ;  $cc = 0.223$ ) served as thresholds for the identification of activation foci and area delineation, respectively (for details, see text).

the common assumption of  $p < 0.05$ , i.e., a 95% threshold of the noise distribution, results in noisy activation maps for both the motor task and visual stimulation. The obviously wrong classification of activated pixels outside the brain indicates the presence of a large number of false positive “activations.” Their fraction is controlled by the chosen significance level, or in other words, by the  $p$  value which served for thresholding. In this example,  $p = 0.05$  means that 5% or  $0.05 \times 96 \times 128 = 614$  pixels represent false positive activations.

A possible way of avoiding this problem is a noticeable increase of the significance level. In the middle part of Fig. 7, the choice of  $p = 0.0001$  reduces the number of false positive pixels to 1.2. The corresponding activation maps exhibit no activated pixels outside the brain and therefore support the absence of false-positive activations in the entire map. On the other hand, these maps clearly demonstrate that the increased specificity is at the expense of a severe loss of sensitivity. The far too conservative nature of this high threshold is further evidenced by a comparison with the correlation maps shown in Fig. 5 which depict much more spatially extended and contiguous areas of pixels with high correlation coefficients. In this case, the specificity of the procedure is not adequate, or in other words, the number of false-negative pixels is too high.

The spatial information contained in the clustering of activated pixels as revealed in correlation images may be exploited to improve the sensitivity without losing specificity. Under the assumption that pixels within a truly activated area bear a tight topographic relationship to a smaller number of pixels which exceed a threshold related to a very low type-one error probability (false positives), specificity can be preserved by using a high threshold criterion, while an adequate area delineation may be achieved by a subsequent iterative nearest-neighbor analysis. This latter strategy includes pixels around predefined activation foci, provided their correlation coefficients comply with a second lower probabilistic threshold. The successful application of this approach is demonstrated in the bottom part of Fig. 7, which presents activation maps where activation foci were obtained with  $p < 0.0001$  and directly neighboring pixels were iteratively added as long as their correlation coefficients fulfilled  $p < 0.05$ .

### 3. Discussion

This work describes a practical approach to the analysis of fMRI data sets that is based on the physiologic characteristics of the underlying individual noise distribution. It



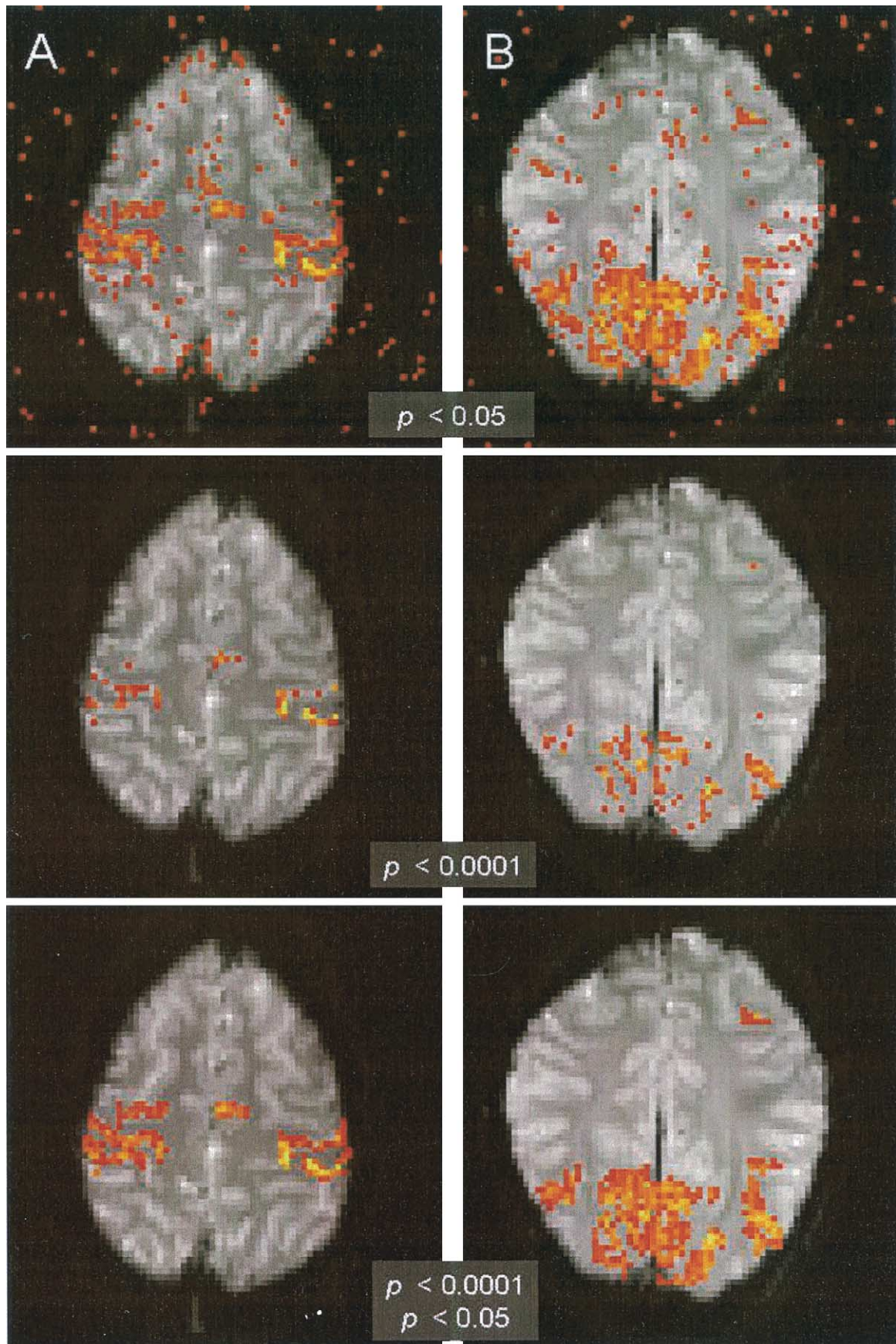


Fig. 7. Activation maps obtained for (A) a hand motor task and (B) visual stimulation (same data as in Figs. 2 and 4). Thresholding with  $p < 0.05$  (top) identifies most activations but includes a large number of false positive “activations.” The use of  $p < 0.0001$  (middle) increases the specificity at the expense of a serious loss of sensitivity. A combination of both  $p$  values (bottom) results in activation maps with both high specificity and sensitivity (for details see text).

results in a user-independent definition of objective thresholds and yields robust statistical parametric maps of activated pixels with both high specificity and sensitivity. Although the examples were based on a correlation analysis, the principles apply to other statistical procedures proposed for fMRI as well. This particularly holds true for the use of probabilistic thresholds, i.e.,  $p$  values, rather than the use of thresholding with fixed correlation coefficients. For example, when using a  $t$  test, the thresholds should not be given in  $t$  but in  $p$  values.

Beside these general considerations the present approach compensates for a variety of putative influences on the histogram of correlation coefficients. Several factors such as systemic alterations of the hemodynamic responsiveness (“arousal”), respiration, perfusion, flow-induced tissue pulsations or stimulus-correlated motion are likely to alter the noise structure of the acquired data sets and therefore affect the individual shape (“broadness”) of the distribution of correlation coefficients. The proposed strategy for estimating individual  $p$  values takes these effects into account and thereby results in a robust and reliable procedure independent of any user interference.

A further important issue addressed in this work is the improvement in sensitivity without essential loss of specificity. This is accomplished by using a combination of two distinct  $p$  values for identifying activation foci and defining the full spatial extent through neighborhood relations, respectively. The method should not be mistaken for a “cluster analysis” that relies on assumptions about the cluster size in relation to the image resolution. Specifically, the proposed spatial response delineation does not exclude small—

even single-pixel—activations necessarily omitted in a conventional cluster analysis.

In summary, proper statistical definitions for thresholding are essential for a meaningful comparison of brain activation maps across paradigms and subjects. Beyond this basic requirement, the present approach compensates for putative intra- and intersubject differences in performance characteristics and therefore emerges as a simple, automatic, and robust tool for high-quality analyses in magnetic resonance functional neuroimaging.

## References

- [1] Ogawa S, Lee TM, Kay AR, Tank DW. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *PNAS* 1990; 87:9868–72.
- [2] Arthurs OJ, Boniface S. How well do we understand the neural origins of the fMRI BOLD signal? *Trends Neurosci* 2002;25:27–31.
- [3] Merboldt KD, Fransson P, Bruhn H, Frahm J. Functional MRI of the human amygdala? *Neuroimage* 2001;14:253–7.
- [4] Biswal BB, Taylor PA, Ulmer JL. Use of jackknife resampling techniques to estimate the confidence intervals of fMRI parameters. *J Comput Assist Tomogr* 2001;25:113–20.
- [5] Genovese CR, Lazar NA, Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 2002;15:870–8.
- [6] Kleinschmidt A, Requardt M, Merboldt KD, Frahm J. On the use of temporal correlation coefficients for magnetic resonance mapping of functional brain activation. Individualized thresholds and spatial response delineation. *Int J Imag Sys Technol* 1995;6:238–44.
- [7] Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS. Processing strategies for time-course data sets in functional MRI of the human brain. *Magn Reson Med* 1993;30:161–73.