NeuroImage **11,** 313–325 (2000)
doi:10.1006/nimg.2000.0547, available online at http://www.idealibrary.com on IDEAL®

A Principal Components-Based Method for the Detection of Neuronal Activity Maps: Application to Optical Imaging

M. Gabbay,*,[1] C. Brennan,† E. Kaplan,‡ and L. Sirovich*,[2]

**Laboratory of Applied Mathematics, ‡Departments of Biophysics and Ophthalmology, Mount Sinai School of Medicine, Box 1012, New York, New York 10029; and †Department of Neurosurgery, Memorial Sloan Kettering, New York, New York 10021*

Received September 13, 1999

We present a novel analysis technique for the extraction of neuronal activity patterns from functional imaging data. We illustrate this technique on data from optical imaging. Optical imaging of the mammalian visual cortex probe the patterns in which the neuronal responses to various aspects of the visual world, such as orientation and color, are spatially organized within the cortex. Recovering these patterns from the image data is a challenging problem as the neuronal response signal is extremely weak in comparison to the background vegetative processes (e.g., circulation and respiration). The proposed technique obtains the neuronal activity pattern using a combination of principal component analysis and statistical significance testing. The performance of this method is compared with the results of existing analysis techniques. The comparison shows the new method to be more sensitive than previous methods.

Key Words: **optical imaging; functional imaging; principal component analysis.**

INTRODUCTION

Functional brain imaging methods such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and optical imaging have provided a wealth of information on how the brain responds to external stimuli. Optical imaging, in particular, has proven to be a fruitful probe for illuminating the functional architecture of the mammalian primary visual cortex having a superior spatial resolution to PET and fMRI. The spatial organization of the neuronal response in various species has been investigated with respect to ocular dominance and orientation (Blasdel and Salama, 1986; Blasdel, 1992), motion direction (Weliky *et al.*, 1996), and spatial frequency (Shoham *et al.*, 1997; Everson *et al.*, 1998). In typical optical imaging experiments, an animal is shown a visual stimulus on a screen while monochromatic (red) light shines on an exposed part of its cortex that is being imaged by a CCD camera. The intrinsic optical signal is an indirect measure of neuronal activity—active regions of the cortex get darker as the result of effects associated with increased metabolic demand, such as changes in blood flow, oxygen consumption, and cell swelling. This decrease in reflectance, however, is very slight, approximately 0.01–0.1% and is dwarfed by the variations in reflectance due to vegetative processes such as circulation and respiration.

As a consequence of the weakness of the signal, faithfully extracting the cortical response map is a difficult task and many (typically thousands) of images have to be taken to obtain good results. The simplest and most common analysis procedure is to calculate the difference between the two pictures that are the averages of the sets of images obtained under two different stimulus conditions. We refer to the activity map obtained in this manner as the "standard difference." The standard difference yields good results when the optical response signal is relatively strong and/or a sufficiently large amount of data has been taken. When these conditions are not met, the resultant activity map is a poor representation of the cortical response, often heavily contaminated by vascular artifact or other large spatial scale features that are usually not associated with the regions of neurons that preferentially respond to a given stimulus. Since the amount of data one can obtain from a given animal is limited, an improved analysis method can produce a combination of the following benefits: more accurate determination of the activity maps; access to weaker activity maps; and the use of a richer repertoire of stimulus conditions in a given experiment.

Efforts to improve signal recovery beyond that of the standard difference include Principal Component Analysis (PCA) (Sirovich and Everson, 1992; Sirovich *et al.*,

[1] Current Address: Information Systems Laboratories, 7047 Carroll Road, San Diego, CA 92121.

[2] To whom correspondence and reprint requests should be addressed. Fax: (212) 426-5037. E-mail: chico@camelot.mssm.edu.

1053-8119/00 $35.00

1996; Cannestra *et al.*, 1996), the indicator function approach (Everson *et al.*, 1997), and the application of independent component analysis (Bell and Sejnowski, 1995; Hyvärinen and Oja, 1997) and extended spatial decorrelation (Molgedy and Schuster, 1994) to optical imaging data (Schieffl *et al.*, 1999). Carmona *et al.* (1995) presented a method for the analysis of optical imaging data that uses wavelets to clean the images of vascular artifact in conjunction with PCA to obtain the response region. While the time course of the intrinsic signal seems to be captured by this technique, the method is very computationally intensive and its ability to produce highly resolved activity maps such as ocular dominance or orientation columns has not been shown in the literature.

In this paper, we present a novel analysis technique for activity map extraction, which we refer to as the "truncated difference." This technique is essentially a combination of the standard difference and PCA in that the activity map is determined by projecting the standard difference onto the basis spanned by the range of principal components of highest statistical significance. It will be seen that the truncated difference has the merits of being conceptually and computationally simple while still producing excellent results. The truncated difference method is described under Analysis Methods, along with the standard difference, PCA, and the indicator function. Its performance on an illustrative artificial data set and on real data is shown under Results, where it is compared with the other analysis techniques.

ANALYSIS METHODS

The Standard Difference

We will take our experimental data to consist of two sets of images, the "stimulated" data set and the "reference" data set. The stimulated data set contains images for which the animal was shown the stimulus of interest and, for simplicity, the reference data is taken to be the set of images where no stimulus was presented (in practice, the reference data set usually corresponds to a second stimulus condition). Our goal is to find the map that best characterizes the difference in neuronal activity between the stimulated and reference data.

Each image record consists of the intensity value of light reflected off the cortex received by each pixel $\mathbf{x}$ in a two-dimensional array of M pixels. We denote the mean-subtracted reflectance recorded in pixel $\mathbf{x}$ for a given image by $f(\mathbf{x}, t)$, where t is the index or time-stamp of that image. The mean that is subtracted is the average of all the image records obtained by grouping together the stimulated and reference data. When a distinction needs to be made, members of the stimulated data will be denoted by $\tilde{t}$ and reference set images by $\hat{t}$. The total number of images is N, and, for convenience, the number of images in each of the stimulated and reference sets is taken to be $N/2$. The order in which the stimulated and reference images are arranged is represented schematically by the stimulus presentation sequence, $w(t)$:

$$w(t) = \begin{cases} 1 & \text{for } t \in \{\tilde{t}\} \\ -1 & \text{for } t \in \{\hat{t}\}. \end{cases} \tag{1}$$

This is a boxcar function with a mean of zero.

We suppose that there exists a true activity map $p(\mathbf{x})$ that we are trying to extract from the background. This background is denoted by $\eta(\mathbf{x}, t)$ and is mean-subtracted so that it's average over time is zero for each pixel. We will make the assumption that the background and the signal of interest add linearly. This is reasonable given the weakness of the signal relative to the background. Accordingly, any image record can be written

$$f(\mathbf{x}, t) = \eta(\mathbf{x}, t) + \tfrac{1}{2}\, w(t)\, p(\mathbf{x}). \tag{2}$$

Note that the mean subtraction is responsible for the appearance of $p(\mathbf{x})$ in the reference data images. The use of a boxcar function for $w(t)$ assumes that the signal is always at full or zero strength, which is a fair approximation if one only uses data for which the equilibrium signal is expected to be nearly attained. Alternatively, if the typical time course of the signal is known then that response can be incorporated into $w(t)$.

We can take the temporal inner product of both sides of Eq. (2) with the stimulus presentation sequence, where the temporal inner product between two functions, $a(t)$ and $b(t)$ is $(a, b)_t = \Sigma_t a(t)b(t)$ and the sum is taken over all N times. Since the background is assumed to be independent of the signal, their correlation should vanish in the limit of an infinite amount of data, so that $(\eta, w)_t \to 0$ as $N \to \infty$. This yields an expression relating the activity map to the data,

$$p(\mathbf{x}) = \lim_{N\to\infty} \frac{2}{N} (f(\mathbf{x}, t), w(t))_t, \tag{3}$$

where we have used $(w, w)_t = N$. Given the particular functional form, (1), for $w(t)$, this expression can be rewritten as $p(\mathbf{x}) = \lim_{N\to\infty}(\langle f(\mathbf{x}, \tilde{t}\rangle - \langle f(\mathbf{x}, \hat{t})\rangle)$, where $\langle f(\mathbf{x}, \tilde{t})\rangle$ denotes the image obtained by averaging over all the records indexed by $\tilde{t}$ that comprise the stimulated data and $\langle f(\mathbf{x}, \hat{t})\rangle$ is the average over the index $\hat{t}$ that denotes reference set images. As infinitely long experiments are impractical, taking the difference between the means of the stimulated and reference data

yields only an approximate answer for the activity map. This is the standard difference which we write formally as

$$\phi_{SD}(\mathbf{x}) = \langle f(\mathbf{x}, \tilde{t})\rangle - \langle f(\mathbf{x}, \hat{t})\rangle. \tag{4}$$

A rough calculation can now be made to illustrate why the standard difference often performs poorly in a typical experiment. Using Eq. (2) the standard difference becomes

$$\phi_{SD}(\mathbf{x}) = p(\mathbf{x}) + \Delta\langle\eta(\mathbf{x})\rangle, \tag{5}$$

where $\Delta\langle\eta(\mathbf{x})\rangle = \langle\eta(\mathbf{x}, \tilde{t})\rangle - \langle\eta(\mathbf{x}, \hat{t})\rangle$ is the difference between the means of the background over the stimulated and reference data. The error in a measurement of $\Delta\langle\eta(\mathbf{x})\rangle$ is about $2\sigma_\eta/\sqrt{N}$ where σ_η is the standard deviation of $\eta(\mathbf{x})$. The ratio of the signal to the background standard deviation is $|p(\mathbf{x})|/\sigma_\eta \sim 0.01$. If we take $N = 2000$, then the ratio of the signal strength at $\mathbf{x}$ to the background fluctuation is

$$\left|\frac{p(\mathbf{x})}{\Delta\langle\eta(\mathbf{x})\rangle}\right| \sim \frac{|p(\mathbf{x})|\sqrt{N}}{2\sigma_\eta} \sim 0.2.$$

This shows that the contribution of the background to the standard difference can be several times greater than that of the signal.

Principal Component Analysis

A key part of the truncated difference analysis scheme is representing the standard difference in the basis obtained by principal component analysis. PCA is a scheme for compressing spatiotemporal data to provide a basis where each element is ranked by its importance in reconstructing the data. Compression is then achieved by omitting less important elements in the representation of the data (this compression is optimal in a least mean squares sense). If one is dealing with continuous functions rather than discrete data, then the procedure analogous to PCA goes under the name of the Karhunen–Loeve decomposition (see Sirovich and Everson 1992).

In PCA, each image record t is represented by the linear combination,

$$f(\mathbf{x}, t) = \sum_n a_n(t)\psi_n(\mathbf{x}). \tag{6}$$

(Recall that the notation $f(\mathbf{x}, t)$ is not a function of continuous variables in space and time but rather of discrete pixel locations $\mathbf{x}$ and image records t. If one is more comfortable with matrix notation then one can replace $f(\mathbf{x}, t)$ by the matrix f with elements f_{ij}, where i indexes the pixels and j the times.) The $\{\psi_n\}$ are images known as the principal components. We assume that the number of images N is less than the number of pixels M as is typical in optical imaging experiments. Since the mean has been subtracted from the data, the number of principal components will in general be $N - 1$. They form an orthonormal basis for the data,

$$(\psi_m(\mathbf{x}), \psi_n(\mathbf{x}))_{\mathbf{x}} \equiv \sum_{\mathbf{x}} \psi_m(\mathbf{x})\psi_n(\mathbf{x}) = \delta_{mn}, \tag{7}$$

where the sum is taken over all M pixels. Any image can be represented as a column vector in which the pixels are components. The principal components are then the eigenvectors obtained from the eigenvalue equation,

$$K\psi_m = \sigma_m^2\psi_m, \tag{8}$$

where K is the covariance matrix whose elements are given by

$$K(\mathbf{x}, \mathbf{y}) = \frac{1}{N}\sum_t f(\mathbf{x}, t) f(\mathbf{y}, t). \tag{9}$$

(In matrix notation this would be $K = ff^T/N$.) Since N is large we can ignore the distinction between N and $N - 1$. The $\{a_n\}$ are the coordinates corresponding to the basis set $\{\psi_n\}$. Each coordinate for a given image is the projection of that image onto the corresponding eigenvector,

$$a_m(t) = (f(\mathbf{x}, t), \psi_m(\mathbf{x}))_{\mathbf{x}} \tag{10}$$

From Eqs. (8), (9), and (10) we find

$$\frac{1}{N}(a_m(t), a_n(t))_t = \sigma_m^2\delta_{mn}. \tag{11}$$

This expression shows that the temporal activities of different eigenvectors are uncorrelated and that the eigenvalue σ_m^2 is the variance of the m^{th} mode. The principal components are ranked in order of descending eigenvalue, as modes that carry more variance are more important in representing the data. The term "power" is often used synonymously with variance.

In optical imaging data, the first few principal components usually contain over 95% of the variance. They generally correspond to so-called vegetative modes of activity like circulation and respiration and have a vascular appearance and/or large scale spatial structures that are not characteristic of neuronal response to stimuli. Midrange principal components appear patchy and the eigenvectors with the lowest power look like pure pixel noise. Typically, only the first 200 or so

principal components have useful information and the rest can be discarded.

We note that the PCA eigenvalues and eigenvectors can be efficiently computed using the *snapshot method* (Sirovich, 1987). This method exploits the fact that the eigenvalues of the covariance matrix K (size M^2) are the same as those for the smaller matrix (size N^2) with elements given by

$$C(t, s) = \frac{1}{N} \sum_{\mathbf{x}} f(\mathbf{x}, t) f(\mathbf{x}, s). \tag{12}$$

The $\{a_n\}$ are eigenvectors of C and the principal components can be calculated from

$$\psi_m(\mathbf{x}) = \frac{1}{\sigma_m^2} (f(\mathbf{x}, t), a_m(t))_t. \tag{13}$$

Since PCA decomposes the data into modes of uncorrelated activity, it is possible that the neuronal response map is well-captured by one of the principal components—the one that is most correlated with the stimulus presentation sequence $w(t)$. This is not usually the case, however, and even in the limit of infinite data a single principal component will not, in general, yield the true activity map $p(\mathbf{x})$ as we now show. Using Eq. (2) we can write the covariance matrix K as

$$K(\mathbf{x}, \mathbf{y}) = \frac{1}{N} \sum_t \Big\{ \eta(\mathbf{x}, t)\eta(\mathbf{y}, t) = \frac{1}{2} \eta(\mathbf{x}, t) w(t) p(\mathbf{y}) + \frac{1}{2} \eta(\mathbf{y}, t) w(t) p(\mathbf{x}) + \frac{1}{4} p(\mathbf{x}) p(\mathbf{y}). \tag{14}$$

Since the background is uncorrelated with the stimulus presentation sequence, $(\eta, w)_t \to 0$ as $N \to \infty$, the two cross terms in (14) vanish and, in the infinite data limit, K becomes

$$K(\mathbf{x}, \mathbf{y}) = K_0(\mathbf{x}, \mathbf{y}) + \tfrac{1}{4} p(\mathbf{x}) p(\mathbf{y}), \tag{15}$$

with K_0 defined as $K_0(\mathbf{x}, \mathbf{y}) = N^{-1} \Sigma_t \eta(\mathbf{x}, t)\eta(\mathbf{y}, t)$. If $p(\mathbf{x})$ is an element of the null space of K_0 then $\Sigma_{\mathbf{y}} K(\mathbf{x}, \mathbf{y}) p(\mathbf{y}) \propto p(\mathbf{x})$ making $p(\mathbf{x})$ an eigenvector of K. Since the principal components are also eigenvectors of the covariance matrix then one of them will be identified with the true activity map in the infinite data limit. This would also be the case if $p(\mathbf{x})$ were an eigenvector of K_0. In general, however, $p(\mathbf{x})$ could be some arbitrary map and neither of the above two possibilities need pertain; in which case, a lone principal component will not correspond to the actual neuronal response map.

PCA has been used successfully to identify neuronal activity maps with individual principal components (Sirovich *et al.*, 1996). One can identify the appropriate principal component by finding the one whose time coordinate series is best correlated with the stimulus presentation sequence. Cannestra *et al.* (1996) also used a single principal component to effect improved analysis of the time course of the intrinsic optical signal and better noise reduction of reconstructed images. As argued above, there is no mathematical reason for restricting ourselves to a single principal component and, indeed, experimentally weak signals such as macaque orientation preference are usually not captured well by just one component. This is also seen to be the case for the test data set discussed under Results (see Fig. 3b). This leads us to seek a linear combination of principal components in recovering the neuronal response map.

The Truncated Difference

The first step in obtaining the truncated difference map is rewriting the standard difference in the basis of principal components. Using the definition of the standard difference (4) and the expansion (6), we have

$$\phi_{SD}(\mathbf{x}) = \sum_n \Delta_n \psi_n(\mathbf{x}), \tag{16}$$

where Δ_n is the difference in the means of the coordinate a_n over the stimulated and reference data sets, $\Delta_n = \langle a_n(\hat{t}) \rangle - \langle a_n(\hat{t}) \rangle$.

As noted above, it is the strong background that is responsible for the often poor quality of the standard difference map. The higher power principal components are typically associated with background processes and these modes dominate the standard difference. A remedy is to reduce or eliminate the contribution of the high power background modes. Of course, this truncation cannot be done capriciously and an objective method for doing so is detailed below.

The truncated difference can be written,

$$\phi_{TD}(\mathbf{x}) = \sum_n \mu(n) \Delta_n \psi_n(\mathbf{x}), \tag{17}$$

where $\mu(n)$ is the coefficient that reweights the contributions of the principal components. For example, we may choose the hard truncation,

$$\mu(n) = \begin{cases} 1 & \text{for } L \le n \le H \\ 0 & \text{for } n < L \text{ or } n > H. \end{cases} \tag{18}$$

Here, the truncated difference map is obtained by projecting the standard difference onto the range of eigenvectors with indices between the low cutoff L and the high cutoff H. The goal is to remove principal components that have little or nothing to do with the

neuronal response signal and primarily represent background processes or noise.

The method used for selecting the truncation essentially involves the application of statistical significance testing to the modal coordinates $\{a_n\}$. To calculate the significance of the n^{th} principal component, we first make the null hypothesis that its coordinate series $a_n(t)$ and the stimulus presentation sequence, $w(t)$, are uncorrelated. We denote by Δ the difference between the means of the stimulated and reference data sets that could be observed if the null hypothesis were true. This is distinct from the actual measured difference Δ_n. It is reasonable to take its probability distribution, $\rho_n(\Delta)$, to be a Gaussian centered about zero,

$$\rho_n(\Delta) = \frac{1}{\sigma_\Delta \sqrt{2\pi}} \exp\left(-\frac{\Delta^2}{2\sigma_\Delta^2}\right). \tag{19}$$

The standard deviation is given by $\sigma_\Delta = \sqrt{\sigma^2_{\langle \tilde{a}_n \rangle} + \sigma^2_{\langle \tilde{a}_n \rangle}}$, where $\sigma_{\langle \tilde{a}_n \rangle}$ and $\sigma_{\langle \tilde{a}_n \rangle}$ are, respectively, the standard deviations one would find for the means of the stimulated and reference data for the coordinate series $a_n(t)$. However, the division of $a_n(t)$ into stimulated and reference halves is irrelevant under the null hypothesis and we expect the standard deviation of $\langle a_n \rangle$ to scale as $\sigma_n/\sqrt{N}$ and so $\sigma_{\langle \tilde{a}_n \rangle} = \sigma_{\langle \hat{a}_n \rangle} = \sigma_n \sqrt{2/N}$. This yields $\sigma_\Delta = 2\sigma_n/\sqrt{N}$.

In assessing the significance of a principal component, it is convenient to calculate its correlation, r_n, with respect to the stimulus presentation sequence,

$$r_n = \frac{(a_n(t),\, w(t))_t}{\|a_n\|\, \|w(t)\|} = \frac{1}{2}\frac{\Delta_n}{\sigma_n}. \tag{20}$$

We see that apart from a numerical factor the correlation is essentially the difference Δ_n scaled by the square root of the power in the n^{th} mode. The probability distribution can now be expressed more simply in terms of the correlation,

$$\rho_n(r) = \sqrt{\frac{N}{2\pi}} \exp\left(-\frac{1}{2} N r^2\right), \tag{21}$$

and since N is typically on the order of a thousand, the function is vanishingly small for $r = 1$ and can therefore be assumed to range over $-\infty < r < \infty$. So for a given measured correlation r_n, the probability, $P_n = P(|\geq |r_n|)$, of observing a correlation whose absolute value is equal to or larger than $|r_n|$ can be calculated (assuming the null hypothesis):

$$P_n = 2 \int_{|r_n|}^{\infty} \rho_n(r)dr = 1 - \text{erf}\left(\sqrt{\frac{N}{2}}\,|r_n|\right). \tag{22}$$

This probability tells us the likelihood that the measured correlation could have arisen simply by chance due to the partition into two halves of a set of elements picked from the same distribution. Taking the complementary probability yields the confidence level, $\overline{P}_n = 1 - P_n$, which is a gauge of how assured we are that the principal component under scrutiny is genuinely related to the stimulus. It is also possible to more directly calculate the confidence levels by randomly shuffling the data, measuring the correlation with respect to the stimulus presentation sequence and then repeating this many times to obtain the probability distribution. For optical imaging data, the confidence levels obtained with this much more time-consuming method are very close to those that stem from Eq. (22), thereby validating the Gaussian assumption.

In optical imaging data, the principal components that are most significant in representing the neuronal response map often cluster near each other. When the correlation magnitude, $|r_n|$, is plotted as a function of eigenvector index, this is seen as a hump of relatively high correlations (see Fig. 5a). On a plot of the $\{\overline{P}_n\}$ the hump corresponds to a plateau of high confidence levels (Fig. 5b). Accordingly, for the truncated difference (17) with the hard truncation (18) we retain only those eigenvectors within the range of the high confidence level plateau whose edges are defined by the cutoff points, L and H. The purpose of the low cutoff L is to exclude high power principal components corresponding to vegetative processes. It can be chosen by finding the first principal component whose confidence level exceeds some threshold value, say 0.99. As it is possible that a given principal component could have a high confidence level by chance, the plots of correlation and confidence level should be visually inspected to make sure that the principal component indeed marks the start of the high confidence plateau. The upper cutoff H serves primarily to smooth the response map as it excludes low power principal components, which tend to be noisy in appearance. As such, H should be chosen liberally to ensure the inclusion of all the significant principal components.

We note that the truncated difference method has the desirable property of agreeing with the standard difference in the limit of $N \to \infty$. As the number of image records increases, we gain greater confidence in those principal components whose correlation with the stimulus presentation sequence happens to be small. The range of retained eigenvectors increases until we recover the standard difference. It is also worth recalling that the principal components will themselves change as more images are added to the data set.

One might wonder as to why a contiguous range of principal components is chosen instead of simply screening for only those components whose confidence level exceeds a threshold value. As there is no *a priori*

method of setting the threshold confidence level, the screening method involves the following tradeoff: select a high threshold and exclude eigenvectors that are genuinely related to the activity map; or select a lower threshold and run the risk of including falsely correlated eigenvectors some of which may be of high power, thereby distorting significantly the recovered activity map. By using a natural feature of the data, that is, the clustering of significant principal components, the truncated difference method helps avoid the accidental inclusion of bogus, high power components as it is highly unlikely that multiple high confidence eigenvectors would be located near each other simply by chance. By using all the eigenvectors between the cutoff indices, we avoid excluding those principal components whose correlations with the stimulus presentation sequence happen to be low but genuine. We have tested both methods on the artificial data set described under Results and the truncated difference performs better than the simple screening technique.

Rather than completely discarding the principal components outside the cutoff limits as done in (18), it is also possible to use a softer truncation, in which these components are given a reduced but nonzero weight. For instance, the differences, Δ_n, from principal components before the low cutoff can be rescaled so that their power is pegged to the power of the eigenvector at the low cutoff, ψ_L. This corresponds to setting $\mu(n) = \sigma_L/\sigma_n$ for $n < L$. As we have found that this method produces at best minor improvements in activity map recovery, and as there are cases in which the truncation points can be set by criteria other than confidence level (see Discussion), we will use the hard truncation (18) in the results reported in this paper.

In summary, the truncated difference method for obtaining the neuronal response map can be outlined as follows:

1. Calculate the correlations and associated confidence levels of the mode coordinates $\{a_n\}$ with the stimulus presentation sequence $w(t)$.
2. Identify the eigenvector indices, L and H, that form the lower and upper boundaries of the high confidence level plateau.
3. Project the standard difference map onto the basis composed of those principal components lying between L and H.

The Indicator Function

When a set of data can be divided into two classes, an indicator function can be used to indicate to which class a given data element belongs. The application of indicator functions in conjunction with PCA as a method of obtaining activity maps in optical imaging data was proposed by Everson *et al.* (1997). Here, the two classes are the stimulated and reference data sets and we seek an indicator function, $\phi_I(\mathbf{x})$, which attempts to satisfy the conditions[1]:

$$\begin{aligned}(f(\mathbf{x}, t), \phi_I(\mathbf{x}))_{\mathbf{x}} &= 1, \; t \in \{\tilde{t}\} \\ (f(\mathbf{x}, t), \phi_I(\mathbf{x}))_{\mathbf{x}} &= -1, \; t \in \{\hat{t}\}.\end{aligned} \tag{23}$$

The image $\phi_I(\mathbf{x})$ is determined by minimizing the square difference between the left and righthand sides of the above conditions summed over all times t. In addition, $\phi_I(\mathbf{x})$ is calculated using the first T principal components, where T is chosen to limit the influence of noise and chance correlations between the stimulated and reference data sets. The indicator function can then be expressed in terms of the principal components,

$$\phi_I(\mathbf{x}) = \sum_{n=1}^{T} \frac{(a_n(t), w(t))_t}{\sigma_n^2} \psi_n(\mathbf{x}). \tag{24}$$

Given that $\Delta_n = 2(a_n, w)_t/N$, we see from Eq. (16) that $\phi_I(\mathbf{x})$ is not proportional to the standard difference as $N \to \infty$ and emphasizes low power principal components. However, the indicator function method produces significantly improved performance over the standard difference for typical signals of interest in optical imaging experiments (Everson *et al.*, 1997). A similar technique for determining activity maps from PET experiments, which uses PCA combined with the Fisher linear discriminant rather than the conditions (23), is described in Ardekani *et al.* (1998).

RESULTS

The true response map to a given stimulus in an actual experiment is unknown, so in this section, we assess the performance of the truncated difference as well as the other analysis techniques on an artificial data set. This test data set consists of a known test pattern, the checkerboard of Fig. 1a, which is digitally added to a data set of images taken from an actual experiment on a macaque visual cortex. The test pattern plays the role of the activity map which we attempt to recover and is added at a very weak level in order to mimic the low signal-to-noise ratios of real neuronal response maps. This test data set is split into two halves: the stimulated data consists of the half of the background macaque cortex images to which the checkerboard has been added; the reference data consists of only the remaining half of the background images, without the checkerboard. One can see from Figs. 1b and 1c that the stimulated and reference images look

[1] These are the conditions appropriate for mean-subtracted data. For data that has not been mean-subtracted, the right-hand side would be 1 for $t \in \{\tilde{t}\}$ and 0 for $t \in \{\hat{t}\}$. The intent is to find an image that is most parallel with the stimulated data and most orthogonal to the reference data.

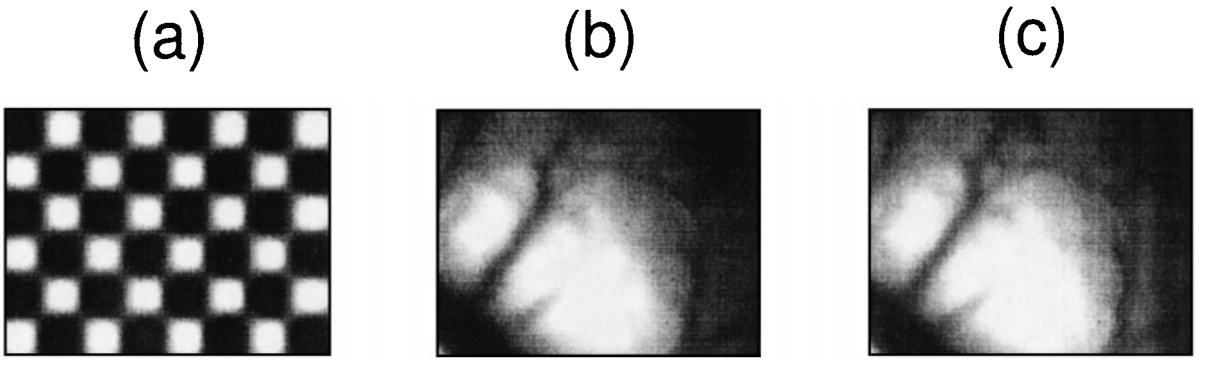

FIG. 1. The test data set: (a) the test pattern, (b) a member of the stimulated data set, and (c) a member of the reference data set. The grayscale in (a) is not the same as in (b) and (c). If it were, the checkerboard would visually appear to be a field of constant intensity.

virtually identical and, in particular, the test pattern cannot be visually discerned in the stimulated picture. This is because the difference in intensity between the center of a light square and that of a dark square is $1/2000^{th}$ of the mean reflectance level of the background. The stimulated and reference sets each have 1080 frames.

After subtraction of the mean image from all the records, the test data set is decomposed via principal component analysis. The resulting spectrum of the first 250 eigenvalues (of 2159 total) is shown in Fig. 2. By retaining only the first 250 eigenvectors in the PCA representation of the image set, we capture 99.99% of the variance in the data. The spectrum has been labeled by three regions and its basic structure is the same as that observed in unaugmented experimental data. The early, high power principal components correspond to vegetative effects such as those of circulation and respiration. These eigenvectors are typically vascular in nature and contain relatively broad swaths of high and low reflectance as can be seen in the first principal component (Fig. 3a). The eigenvectors that are of most consequence in representing the neuronal activity map (the checkerboard in this case) lie within the "response" region of the spectrum (although there are eigenvectors here that are unrelated to the neuronal response). They are usually patchy in appearance and are not appreciably marked by blood vessels (Fig. 3b). As an indication of how weak the signal of interest is, note that the beginning of the response range is

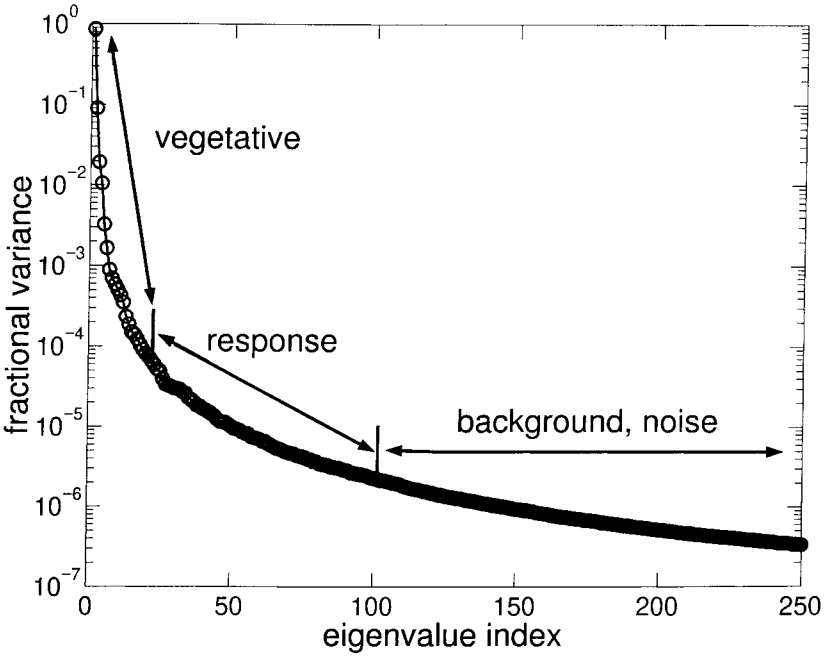

FIG. 2. Eigenvalue spectrum of the test data set. On the vertical axis is plotted the ratio of the given eigenvalue to the sum of all the eigenvalues. Only the first 250 eigenvalues are shown.

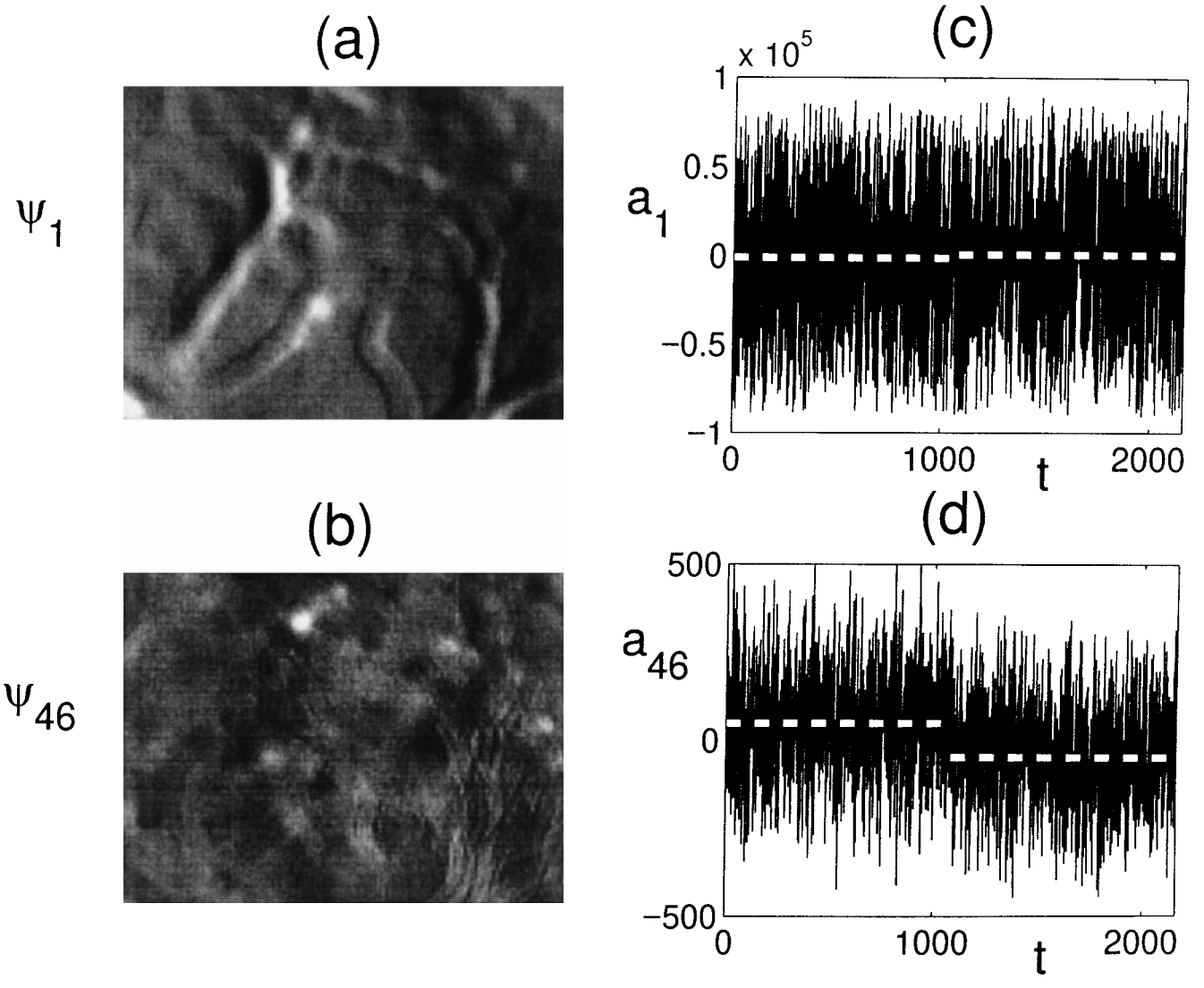

FIG. 3. Sample principal components and their associated time coordinates. (a) The highest power principal component, $\psi_1(\mathbf{x})$, (b) $\psi_{46}(\mathbf{x})$, the principal component best correlated with the stimulus presentation sequence, (c) $a_1(t)$, and (d) $a_{46}(t)$.

about four orders of magnitude lower in power than the first principal component. The final zone of the spectrum consists of low power principal components that correspond to background processes, chance correlations of activity, and are increasingly marked by pixel noise the deeper in the spectrum one goes.

In Fig. 3, along with the principal components themselves are shown their respective time coordinates (the $a_n(t)$ of Eq. (6)). The images are ordered so that the stimulated data constitutes the first half of the records and the reference data is the second half. The corresponding stimulus presentation sequence, $w(t)$, will be a downward step function. Any principal component that is associated with the neuronal response should have a coordinate time series with a step function appearance as is seen in Fig. 3d for $a_{46}(t)$. In fact, $\psi_{46}(\mathbf{x})$ is the principal component that is best correlated with $w(t)$ and hence is the activity map we would obtain from the application of PCA alone, although the checkerboard is not discernible. Hence, we see that for this test data set, a sole principal component is inadequate to recover the activity map as is often the case experimentally. It is clear from visual inspection that, given the variance in the data, the size of the step in Fig. 3d is much larger than one would expect merely by a chance grouping of data points unrelated to the presence or absence of the test pattern. On the other hand, the difference between the stimulated and reference means for $a_1(t)$ indicated by the slight misalignment of the dashed lines could easily have occurred by chance (using Eq. (22), $P_1 = 0.34$). However, a glance at the vertical scales of Figs. 3c and 3d alerts one to the following problem: the ratio $|\Delta_1/\Delta_{46}| = 18.6$, and so the first principal component is a much greater contributor to

the standard difference than is the principal component most correlated with the stimulus presentation sequence, even though the former's contribution is essentially determined at random.

The first principal component is not the only one that is disproportionately represented in the standard difference map. A plot of the projections of the test pattern and the standard difference onto the PCA basis shown in Fig. 4 reveals that many of the early, high power eigenvectors are overweighted in the standard difference relative to the values for the actual checkerboard components. However, the agreement for eigenvectors with indices larger than about $n = 20$ is good. We therefore seek to eliminate the offending high power principal components using the truncation procedure described under The Truncated Difference.

The correlations of the eigenvector coordinates $\{a_n\}$ with $w(t)$ are calculated and the magnitudes plotted in Fig. 5a. The plot displays a peak where principal components with relatively high correlations cluster together. On the associated confidence level plot, Fig. 5b, this peak manifests itself as a plateau of confidence levels very near to one. The region between the dashed lines is the basis of eigenvectors onto which the standard difference map is projected to obtain the truncated difference map. The left truncation point is the first principal component with $\overline{P}_n > 0.99$, which corresponds nicely with the very beginning of the plateau, although this need not always be the case. The plateau tapers off towards the right and the righthand truncation point is chosen a little beyond the last clustering of high confidence levels.

Having chosen our truncation points, the truncated difference map can now be calculated and compared with the actual test pattern. The result is shown in Fig. 6 along with the maps obtained using the standard difference and indicator function methods. The truncated difference reveals the cleanest checkerboard where the checkers are very uniform and distinct. In the standard difference map, the checkerboard is barely perceptible and the blood vessels are the single most dominant feature. In fact, the standard difference bears a strong resemblance to the negative of the first principal component, (Fig. 3a), which is not surprising given the discussion above. The test pattern is clearly visible in the indicator function map but the checkers look irregular and often not uniform. If these maps are considered to be vectors by treating the pixels as individual components, then the angle that each one makes with respect to the test pattern can be calculated, providing a rough measure of the performance of each method. An angle of 0° indicates identical maps. The angles made by each map with the checkerboard are: truncated difference, 17.3°; standard difference, 79.7°; indicator function, 31.7°. This confirms the visual impression that the truncated difference most faithfully extracts the test pattern. The strength at which the checkerboard is added to the background was varied over a range spanning two orders of magnitude

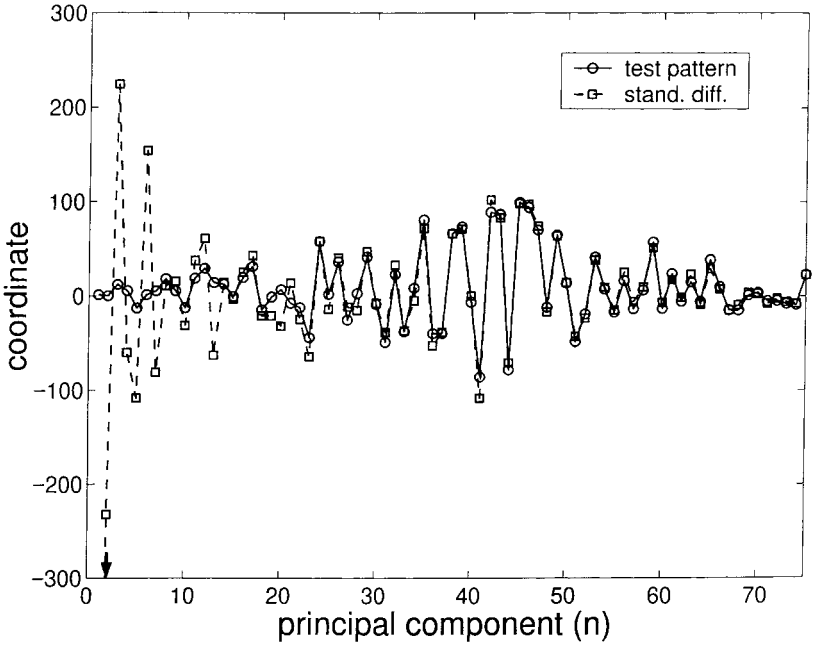

FIG. 4. Projection of the test pattern and the standard difference onto the basis of principal components. The arrow in the lower lefthand corner indicates that the value for the coordinate of the first principal component in the standard difference (−1802) extends below the low edge of the vertical scale.

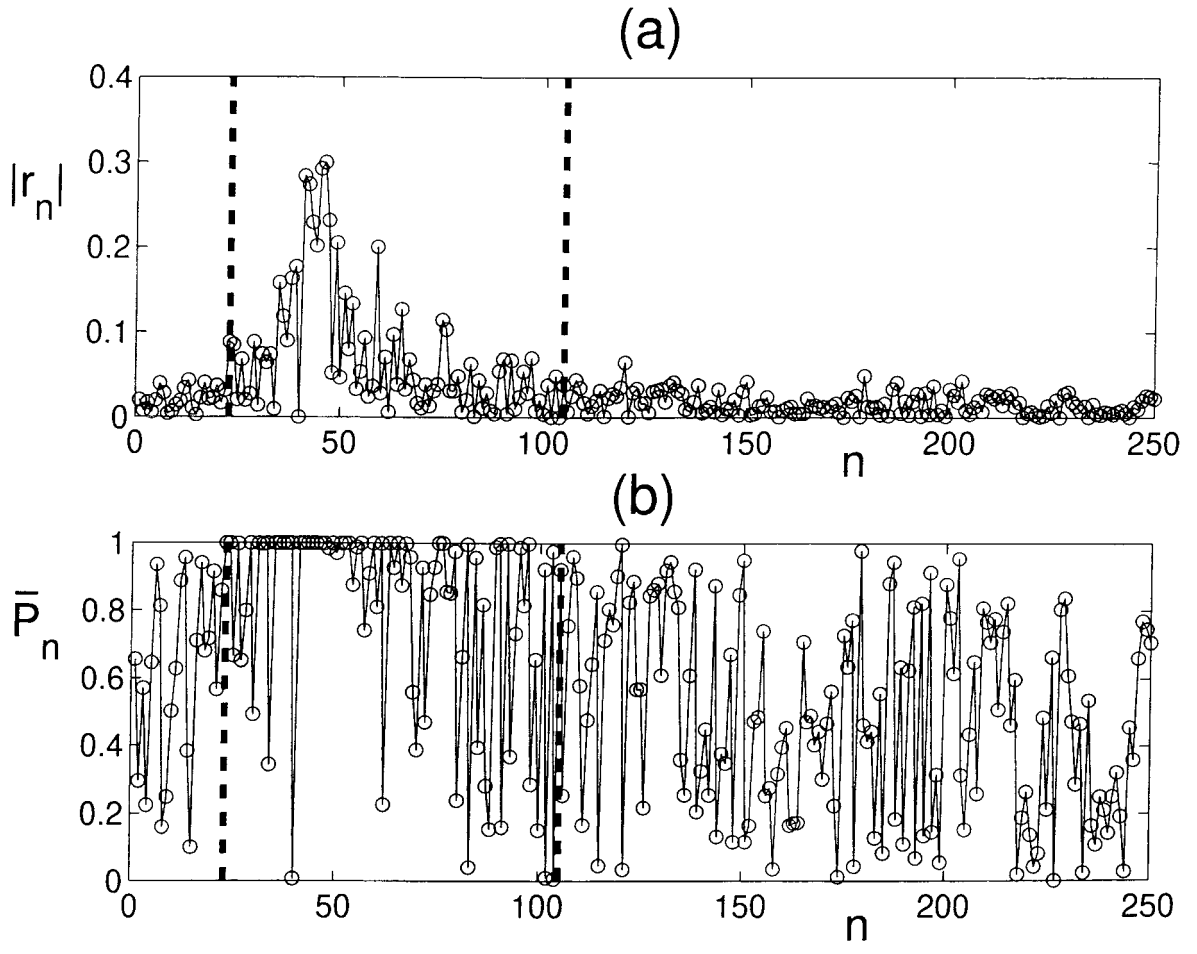

FIG. 5. (a) Magnitude of the correlations of the principal component time coordinates with the stimulus presentation sequence and (b) the associated confidence levels. The dashed lines indicate the truncation points.

and the truncated difference always yielded the best map.

Figure 7 gives an example of the application of the analysis techniques to obtain activity maps for real visual stimuli, in this case, oriented gratings. A cat was shown horizontally and vertically oriented luminance gratings and we seek the activity map that represents the difference between the horizontal and vertical responses. Similar features of activity appear in all three maps in Fig. 7. The standard difference, however,

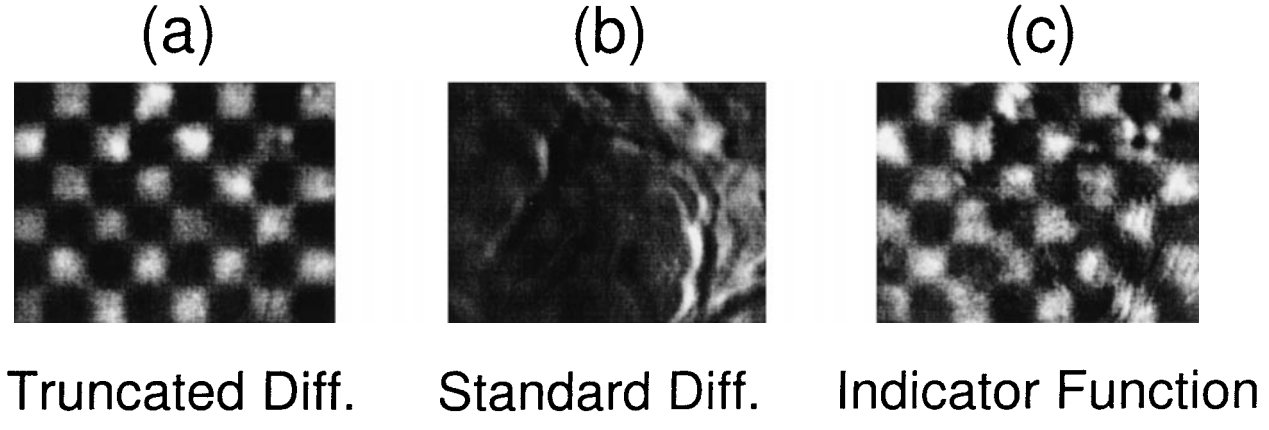

FIG. 6. Activity maps extracted from the test data by: (a) the truncated difference, (b) the standard difference, and (c) the indicator function.

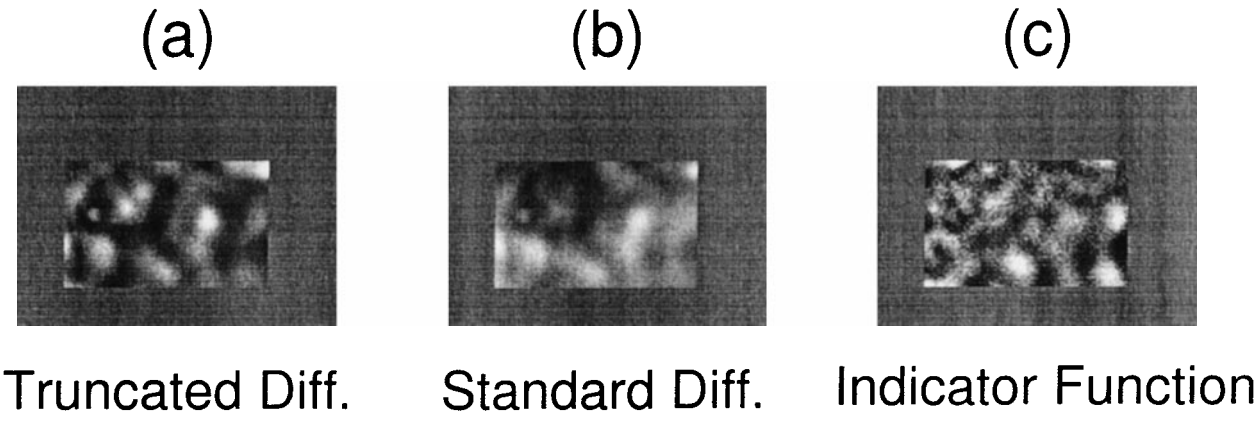

FIG. 7. Activity maps corresponding to the difference between neuronal responses to horizontal and vertical gratings for a cat primary visual cortex. (a) truncated difference, (b) standard difference, and (c) indicator function.

has a washed out appearance as the patches of activity are indistinct. The indicator function has more well-defined patches but the grainy appearance indicates an excessive amount of noise. The truncated difference captures the best of both worlds—it contains smooth and distinct activity regions as one would expect of orientation response maps.

DISCUSSION

We have applied the truncated difference method towards obtaining various types of response maps from optical imaging data such as ocular dominance, orientation, spatial frequency, and color and it has produced excellent results. The correlation and confidence level plots give a direct way of assessing the quality of the data. If the signal is relatively strong, there will be a clear peak in the correlation plot and a distinct plateau of confidence levels close to unity. Poor signals generally will not have a distinct correlation peak and the high confidence level principal components will be scattered rather than clustered. We remark that signals of similar strength but different spatial frequencies are not recovered with the same efficiency. Response patterns characterized by low spatial frequencies can be extracted more faithfully than those marked by higher spatial frequency variations. This is due to the fact that at high spatial frequencies there are a greater number of modes at high spatial frequencies across which the response can be distributed. This effect has been verified in test data sets by varying the size of the checkerboard squares.

As PCA forms the basis of our method, a cautionary word about the sensitivity of PCA to outlying data points is in order. It is observed that freak events that perhaps are found in only one image record are sometimes allotted their own principal component. For instance, due to a technical glitch, one or more camera pixels may fail to record any reflected light for a given image. After the dataset has been subjected to PCA, a principal component with these blanked out pixels may be generated. Supposing that this failure occurred during either the stimulated or reference conditions but not both, the time series of this principal component will be highly correlated with the stimulus and therefore, may be included in the truncated difference. Those principal components whose genesis lies in one-time events can be readily identified by examination of the corresponding time series and thus can be screened out if such freak events are of concern.

Optical imaging of high spatial resolution activity maps generally requires the differential imaging of two "orthogonal" stimulus conditions, i.e., conditions that yield regions of cortical response that are roughly complementary. For ocular dominance columns, these conditions correspond to stimulus presentation to the left or right eye individually. In the case of orientation columns, the stimuli are a pair of gratings with orthogonal orientations such as those used to obtain Fig. 7. The use of orthogonal conditions is dictated by the physiological processes underlying the intrinsic signal and by the sensitivity of the imaging apparatus—one needs to eliminate the diffuse, global response of the intrinsic signal in order to reveal the "local" intrinsic signal associated with the specific stimulus (Vanzetta and Grinvald, 1999). The truncated difference method, however, is not bound by these orthogonal stimulus constraints, and so can be equally well used if greater understanding of the intrinsic signal dynamics and improved imaging techniques obviate the need for orthogonal conditions. In this case, the reference dataset would consist of the images taken just prior to stimulus onset. As mentioned previously, the stimulus presentation sequence, $w(t)$, need not be a simple step or boxcar function but can be a smooth function intended to mimic the time course of the intrinsic signal.

We note that in some experimental conditions it is possible to observe multiple clusters of significant

principal components. In this case the truncation cutoffs should encompass the entire range including all the clusters, although there are situations in which it is desirable to exclude early, high power clusters of significant principal components. This can occur when contrasting stimuli to which the cortex responds unequally such as gratings of different spatial frequencies or colors and is most likely related to the existence of local and global components of the intrinsic signal as noted above. In such cases, if one includes all the clusters of significant principal component, the resulting activity map is characterized by large regions of roughly constant reflectance. This stands in contrast to the local, patchy that one expects of a map which represents the specific regions of neurons that respond to a given stimulus. Fortunately, in such situations, the plot of correlation magnitudes often displays two peaks—a smaller, narrow one followed by a much stronger and broader peak. The truncation point can be chosen at the beginning of the second peak thereby eliminating the contribution of the principal components located in the first and weaker peak. If this is done, a map whose appearance is in greater accord with what one expects of a neuronal activity map is revealed. The origin of the first peak is not well-understood and so one must have a strong rationale for its elimination in calculating response maps.

We have presented a technique for extracting neuronal activity maps from optical imaging data which is simple, intuitive, and effective. It teams PCA with statistical significance testing in a straightforward way and removes the deleterious effect of high power vegetative modes on the activity map. The use of significance testing allows these modes to be eliminated in an objective manner rather than by the arbitrary application of high-pass spatial filtering to remove vegetative artifacts from the standard difference map. PCA provides a natural basis for this technique in that it decomposes the data into modes of uncorrelated activity that are ranked by power. It is the application of PCA to the data that is the most time-consuming step in this method, and this can be done in a computationally efficient way.

We note that there is nothing in the truncated difference technique that is specific to optical imaging data and so the method is applicable to other functional imaging techniques and other types of signals. In most current fMRI experiments, the signal is relatively strong and results are primarily reported simply as regions of activation or no activation so that the truncated difference may be of only marginal benefit. Recent work by Vanzetta and Grinvald (1998) shows that the initial dip seen in some high field fMRI studies (Yacoub *et al.*, 1999) represents changes in the oxidative metabolism that are colocalized with neuronal electrical activity, thereby holding the promise of greatly improved fMRI spatial resolution, nearing that of optical imaging. Consequently, the signal processing issues confronting high resolution fMRI experiments would be similar to those found in optical imaging. This is so because (i) the initial dip is relatively weak; and (ii) if one desires to map out functional modules such as hypercolumns, it is best to resolve the activity into finer grades than binary on/off regions of activation. Accordingly, as spatial resolution improves, the technique set forth in this paper may be of great utility in the analysis of fMRI data and perhaps PET data as well.

ACKNOWLEDGMENTS

This work was supported by ONR Grant N00014-96-1-0492, NIH/NEI EY11276, NIH/NIMH MH50166. Ehud Kaplan is the Jules and Doris Stein Research to Prevent Blindness Professor at the Ophthalmology Department at Mt. Sinai School of Medicine.

REFERENCES

Ardekani, B. A., Strother, S. C., Anderson, J. R., Law, I., Paulson, O. B., Kanno, I., and Rottenberg, D. A. 1998. On the Detection of Activation Patterns Using Principal Components Analysis. In *Quantitative Functional Brain Imaging with Positron Emission Tomography* (E. Carson, M. E. Daube-Witherspoon, and P. Herscovitch, Eds.) Academic Press.

Bell, A. J., and Sejnowski, T. J. 1995. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Comput.* **7:**1129–1159.

Blasdel, G. G. 1992. Differential imaging of ocular dominance columns and orientation selectivity in monkey striate cortex. *J. Neurosci.* **12:**3115–3138.

Blasdel, G. G., and Salama, G. 1986. Voltage-sensitive dyes reveal a modular organization in monkey-striate cortex. *Nature* **321:**579–585.

Cannestra, A. F., Blood, A. J., Black, K. L., and Toga, A. W. 1996. The evolution of optical signals in human and rodent cortex. *NeuroImage* **3:**202–208.

Carmona, R. A., Hwang, W. L., and Frostig, R. D. 1995. Wavelet analysis for brain-function imaging. *IEEE Trans. Med. Imaging* **14:**556–564.

Everson, R. M., Knight, B. W., and Sirovich, L. 1997. Separating spatially distributed response to stimulation from background. I. Optical Imaging. *Biol. Cybern.* **77:**407–417.

Everson, R. M., Prashanth, A. K., Gabbay, M., Knight, B. W., Sirovich, L., and Kaplan, E. 1998. Representation of spatial frequency and orientation in the visual cortex. *Proc. Natl. Acad. Sci. USA* **95:**8334–8338.

Hyvärinen, A., and Oja, E. 1997. A fast fixed-point algorithm for independent component analysis. *Neural Comput.* **9:**1483–1492.

Molgedey, L., and Schuster, H. G. 1994. Separation of a mixture of independent signals using time-delayed correlations. *Phys. Rev. Lett.* **72:**3634–3637.

Schieffl, I., Stetter, M., Mayhew, J. E. W., Askew, A., McLoughlin, N., Levitt, J. B., Lund, J. S., and Obermayer, K. 1999. Blind separation of spatial signal patterns from optical imaging records. In *ICA99—International Workshop on Independent Component Analysis and Blind Source Separation, Aussois, France, Jan. 11–15, 1999*, pp. 179–184.

Shoham, D., Hubener, M., Schulze, S., Grinvald, A., and Bonhoeffer, T. 1997. Spatio-temporal frequency domains and their relation to cytochrome oxidase staining in cat visual cortex. *Nature* **385:**529–532.

Sirovich, L. 1987. Turbulence and the dynamics of coherent structures. I: Coherent Structures. II: Symmetries and transformations. III: Dynamics and scaling. *Q. Appl. Math*. **45:**561–590.

Sirovich, L., and Everson, R. M. 1992. Management and Analysis of Large Scientific Datasets. *Int. J. Supercomp. Appl.* **6:**50–68.

Sirovich, L., Everson, R. M., Kaplan, E., Knight, B. W., O'Brien, E. V., and Orbach, D. 1996. Modeling the functional organization of the visual cortex. *Physica D* **96:**355–366.

Vanzetta, I., and Grinvald, A. 1999. Increased Cortical Oxidative Metabolism Due to Sensory Stimulation: Implications for Functional Brain Imaging. *Science* **286:**1555–1558.

Weliky, M., Boskin, W. H., and Fitzpatrick, D. 1996. A systematic map of direction preference in primary visual cortex. *Nature* **379:**725–728.

Yacoub, E., Le, T. H., Ugurbil, K., and Hu, X. 1999. Further evaluation of the initial negative response in functional magnetic resonance imaging. *Magn. Res. Med*. **41:**436–441.