Robust analysis of event-related functional magnetic resonance imaging data using independent component analysis

Yasser M. Kadah*
Biomedical Engineering Department, Cairo University, Giza, Egypt

ABSTRACT

We propose a technique that enables robust use of blind source separation techniques in fMRI data analysis. The fMRI temporal signal is modeled as the summation of the true activation signal, a physiological baseline fluctuation component, and a random noise component. A preprocessing denoising is used to reduce the dimensionality of the random noise component in this mixture before applying the principal/independent component analysis (PCA/ICA) methods. The set of denoised time courses from a localized region are utilized to capture the region-specific activation patterns. We show a significant improvement in the convergence properties of the ICA iteration when the denoised time courses are used. We also demonstrate the advantage of using ICA over PCA to separate components due to physiological signals from those corresponding to actual activation. Moreover, we propose the use of ICA to analyze the magnitude of the Fourier domain of the time courses. This allows ICA to group signals with similar patterns and different delays together, which makes the iteration even more efficient. The proposed technique is verified using computer simulations as well as actual data from a healthy human volunteer. The results confirm the robustness of the new strategy and demonstrate its value for clinical use.

Keywords: Functional magnetic resonance imaging, denoising, independent component analysis.

1. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) provides a valuable noninvasive tool for investigating brain function[1]. It localizes brain activity during mental or physical activity by detecting the corresponding increase in average cerebral blood oxygenation or cerebral blood flow[2]. To observe these hemodynamic changes, rapid acquisition of a series of brain images is performed. The sequence of images is analyzed to detect such changes and the result is expressed in the form of a map of the activated regions in the brain[1].

Classically, most of fMRI studies were conducted using the so-called block design approach, whereby two sets of data are acquired. First, a number of frames are acquired while the subject is at rest or under some baseline condition, then another set is acquired during the stimulus[1]. This pattern is repeated for a number of cycles in order to improve the signal-to-noise ratio (SNR), which would otherwise be quite low. Recent advances in both data acquisition and analysis have improved the temporal resolution of fMRI and made it possible to observe transient hemodynamic changes with reasonable accuracy. A good example for that is a new experimental design, similar to that of evoked-response potential (ERP) protocol, called single trial or event-related fMRI (ER-fMRI). In this new design, the subject receives a short stimulus or performs a single instance task while the resultant transient response is measured[3]. Event-related fMRI offers many advantages over block design that include versatility, investigation of trial-to-trial variations, and extraction of epoch-dependent information and direct adaptation of the methods used for ERP to fMRI. The main drawback of ER-fMRI is the degradation in SNR due to the transient nature of the response. As a result, such studies now include epoch averaging[3]. Nevertheless, this comes at the expense of suppressing the information about intra-subject variations related to psychophysiological function with each execution of the task. Therefore, a processing method that can be used to suppress noise in the acquired data would be very useful to reduce the experiment duration and preserve the information within the acquired data[3,4].

* E-mail: ymk@ieee.org

 Medical Imaging 2002: Physiology and Function from Multidimensional Images, Anne V. Clough, Chin-Tu Chen, Editors, Proceedings of SPIE Vol. 4683 (2002)

Several methods of data analysis have been used to process the fMRI raw data. The ultimate goal of such analysis is to try to separate signal components due to true activation, physiological fluctuations and random noise. The latter two components are considered as nuisance and must be removed for correct results[7]. Several methods have been proposed to suppress physiological noise including the use of harmonic model[5] and noise subspace characterization[6]. Others attempted to use different strategies to suppress the effect of random noise in the analysis using finite impulse response (FIR) filter modeling[8], smart spatial averaging[9], inter-epoch averaging[3], and Wiener filtering[4]. These techniques suffer from at least one of the following limitations: the need for extended data acquisition for inter-epoch averaging that might not be practical, assumption of limited epoch-to-epoch variability, dependence on assumptions about the signal characteristics to build the denoising filter, and the inability to use conventional statistical detection approaches because of correlated noise among spatial and/or temporal points. Therefore, a denoising strategy that does not have the above limitations would be rather useful in the clinical practice.

Among the most powerful techniques that can be used to separate signal components are those based on blind source separation such as principal component analysis (PCA) and independent component analysis (ICA). These techniques decompose the signal sources using either the second order statistics (as in PCA) or higher order statistics such as the kurtosis (as in ICA) to account for the non-Gaussian nature of the sources[12,13]. According to the assumptions of both techniques, the number of independent signal components must be less than or equal to the number of signals to be analyzed[13]. Otherwise, the separation of components yields incorrect results or even may not converge at all as in ICA. Unfortunately, this condition is not satisfied in fMRI data sets. Given the general assumption of uncorrelated noise, the number of components of random noise alone is equal to the number of signals. The total number of components has to add the number of components due to physiological fluctuations as well as the activation components. As a result, the use of PCA and ICA based techniques may not yield useful results in this case and this may account for the limited use of such techniques in related applications[13-16]. Therefore, a technique that suppresses random noise or removes some of its components would be rather useful for making the use of PCA and ICA more robust for clinical practice.

In this work, we study the problem of reducing the random noise while preserving the other deterministic components in fMRI signals. A new adaptive technique is proposed based on spectrum subtraction. This technique relies on the observation that accurate information about the random noise model parameters can be easily obtained adaptively from event-related fMRI data sets. Given this information, the new technique uses the uncorrelatedness of the random noise and the deterministic components of the signal to separate the two in the original power spectrum. The theoretical analysis of the new technique and the implementation details are presented. The new technique is tested using computer simulations as well as real data and the performance is analyzed. Finally, the value of the proposed method as a preprocessing stage for blind source separation techniques such as PCA and ICA techniques is demonstrated.

2. THEORY

Generally speaking, the fMRI temporal signal can be modeled as the summation of the true activation signal, a physiological baseline fluctuation component, and a random noise component. The physiological baseline fluctuation component can be considered as a deterministic yet unknown signal. Therefore, we will consider a model that is composed of the sum of one deterministic component *d(t)* incorporating both the true signal and the physiological noise and an uncorrelated stochastic component *n(t).* That is,

$$s(t) = d(t) + n(t) \quad . \tag{1}$$

Since these two component are assumed independent, the corresponding power spectrums are related by,

$$P_{ss}(\omega) = P_{dd}(\omega) + P_{nn}(\omega) \quad , \tag{2}$$

where cross-terms vanish because the two components are assumed uncorrelated. Hence, an estimate of the power spectrum of the deterministic component takes the form[10],

$$P_{dd}(\omega) = P_{ss}(\omega) - P_{nn}(\omega) \quad . \tag{3}$$

That is, the signal power spectrum is obtained by spectrum subtraction of the noisy signal and noise power spectra. In order to compute the deterministic signal component from its power spectrum, the magnitude of the Fourier transform can be obtained as the square root of the power spectrum. The problem now becomes that of reconstructing the signal using magnitude only information about its Fourier transform. Several techniques can be used to do that. The one used for this work relies on an estimate obtained from the phase of the Fourier transform of the original signal *S(ω).* Hence, the Fourier transform of the processed signal $S_d(\omega)$ can be expressed as,

$$S_d(\omega) = \sqrt{P_{dd}(\omega)} \cdot e^{j\,\mathrm{Phase}(S(\omega))} \quad . \tag{4}$$

The enhanced deterministic signal $s_d(t)$ is then computed as the real part of the inverse Fourier transformation of this expression.

3. FMRI NOISE POWER SPECTRUM MODEL

3.1. Rician Model Characteristics

The k-space data acquired in MRI consists of two quadrature components representing the real and imaginary parts of the signal. The samples of these components can be modeled as the summation of a deterministic part and a random noise part consisting of independent and identically distributed white Gaussian random noise. Given the orthogonality of the Fourier transform, the pixels within the reconstructed image will also have a similar description. Given that fMRI relies on magnitude information to represent the time course, the magnitude of points within the time course can be described as,

$$m(n) = \sqrt{R(n)^2 + I(n)^2} \ , \tag{5}$$

where *m(n)* is the magnitude of the time course, *R(n)* is the real part and *I(n)* is the imaginary part of the signal at time *n*. When both quadrature components belong to a zero-mean white Gaussian noise model, the resultant distribution for the magnitude is the Rayleigh distribution. On the other hand, in the case when the component distributions have nonzero means, the magnitude belongs to the Rician distribution. This distribution form depends in a complex manner on the value of the means of the component distributions[18]. Given that these mean values are unknown (in fact their values represent the solution for the denoising problem), the exact random distribution of each time point is also unknown. Nevertheless, in the following we will show that the derivation of the expression for the power spectrum in practical fMRI settings results in a rather simple expression that can be directly estimated from the data.

The power spectrum can be computed as the Fourier transform of the autocorrelation function of the data *A(τ).* The autocorrelation function can be expressed as,

$$A(\tau) = E\{m(t) \cdot m(t+\tau)\} \tag{6}$$

In order to evaluate the above expression, consider the two cases when $\tau = 0$ and otherwise. In the first case, the autocorrelation function *A(0)* is the just the second moment of the data (which is a function of the variance of Gaussian noise components and the mean values of the components as defined by the Rician distribution). On the other hand, for all other autocorrelation values where $\tau \neq 0$, given the independence and identical distribution of samples, all values of τ other than 0 provide identical mathematical expressions and consequently are equal in value. Hence, the autocorrelation function will take the form of a flat distribution with a different value only at $\tau=0$. Consequently, the power spectrum takes the form of a delta function at the DC value corresponding to the level of the flat part, in addition to a flat function with a level corresponding to the difference between the second moment and the flat part in the autocorrelation function. Since the DC value of the spectrum does not generally contribute much to the denoising process, we will consider our noise power spectrum model as only the flat portion.

3.2. Practical Noise Model Estimation

Even though the derived noise power spectrum model looks similar to zero-mean white Gaussian noise power spectrum, the level of this flat function is not equal to the variance of the data in theory. It will depend in a complex manner on the mean values of the component distributions as well as the variance. When there is a true activation signal within a pixel time course, the mean value of the pixel changes with time and consequently it may not be possible in theory to obtain an expression for the noise power spectrum required to implement Eq. (3) in practice.

In order to solve this problem, we observe that the numerical evaluations of the Rician distribution in the literature exhibit special characteristics when the means of the components are much larger than their standard deviations[18]. Given that practical fMRI data show similar properties, we evaluated the power spectrum level under different values of the component in order to derive any special characteristics that may simplify our implementation. In particular, a simulation of the fMRI signal generation was performed whereby two quadrature components with variable means and similar standard deviations of unity were used and the resultant power spectrum was computed for all values of the mean (or equivalently in our case the mean over standard deviation). The time course lengths used were 64, 128, 256, 512, and 1024 points consecutively and the results were averaged over 128 time courses in each. The power spectrum is estimated using the periodogram method as the square of the absolute value of the discrete Fourier transform of the time course. The orthogonal discrete Fourier transformation matrix where $1/\sqrt{N}$ is included in both the forward and inverse transformations was used to avoid any unwanted factors[19]. The result of this simulation for the case of 512-point time courses suggested that the values of the mean above approximately five times the standard deviation result in nearly similar levels for the flat region. Given the contrast parameters used for clinical fMRI studies, this condition is always satisfied and hence the problem of estimating a general noise model is simplified. Moreover, if we look at the plot of the calculated variance of the data in (c) and the division of the mean power spectrum level over this variance in (d), we observe that the result is a flat curve over all values of the component means. The value of this flat curve varies in a very narrow range around the mean value. To confirm this result, the simulation from the results from other time course lengths of 64, 128, 256 and 1024 were considered as well. The linear regression of the data provided a formula for the noise power spectrum model as a linear curve with a slope of -5.7250e-6 and an intercept of 1.0053. The 95% confidence intervals for this regression are [-0.1518e-5 , 3.7279e-6] for the slope and [1.0003 , 1.0103] for the intercept. Given the very small value of the slope, we can consider only the flat portion for our model. Hence, all is needed to compute the model of the noise power spectrum is to compute the variance of an ensemble of the data that need not have the same mean as the time course of interest. Moreover, the variations of the time course signal mean due to either physiological baseline variations or true activations will not have any effect on this random noise power spectrum model. This conclusion substantially simplifies the implementation of the proposed denoising method.

4. INDEPENDENT COMPONENT ANALYSIS

Independent component analysis (ICA) is an extension of PCA whereby higher order statistics rather than second order moments are used to determine the basis vectors that are statistically as independent as possible[12]. Let v_i, i=1,…,n be measured signals and s_j, j=1,…,m independent components (ICs) with zero mean and unit variance. The basic problem in ICA is to estimate the mixing matrix A in $v = As$ and the realizations of ICs. One restriction for the problem is naturally $m \leq n$; usually m is assumed known and often m = n.

A fast fixed-point algorithm (FastICA) is introduced for independent component analysis [12,17]. In FastICA, the data vectors v are first whitened. Whitening step is performed using PCA. Besides whitening the data, PCA is used to reduce dimensionality (e.g. from n to q). The whitened data x is defined by $x = Mv = MAs = Bs$. The problem of finding an arbitrary full-rank matrix A is reduced to the simpler problem of finding an orthogonal matrix B, which then gives $s = B^T x$. Thus, we are searching for an orthogonal matrix W^T so that $W^T x$ should be good estimates of the independent components. Several approaches can then be taken to utilize the higher-order information. A principled approach is given by finding linear combinations of maximum nongaussianity, as motivated by the central limit theorem. Sums of independent random variables tend to be closer to gaussian than the original ones. Therefore if we take a linear combination $y = \sum_i w_i^T x$ of the whitened variables, this will be maximally nongaussian if it equals one of the independent components. When reaching a solution, $w = b_i$, i.e. the orthogonal matrix B is constructed.

Nongaussianity is measured by many ways; most suggested solutions to the ICA problem use the fourth-order cumulant or kurtosis [12], which is defined for a random variable y as

$$\mathrm{kurt}(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad . \tag{7}$$

Kurtosis is zero for gaussian random vectors; for densities peaked at zero, it is positive, and negative for flatter densities. For $y = w^T x$, the kurtosis criterion is maximized by applying the iterative FastICA algorithm. The iteration is $w(k) = E\{x(w(k-1)^T x)^3\} - 3w(k-1)$ under the constraint of $\|w(k)\| = 1$ [17]. The iteration stops when $\left|w(k)^T w(k-1)\right|$ is close enough to one.

To estimate m independent components, the algorithm iterates m times. To ensure that a different independent component is estimated each time, an orthogonalizing projection is added to the iteration; $w(k) = w(k) - \overline{B}\overline{B}^T w(k)$, where $\overline{B}$ is a matrix whose columns are the previously found w vectors. At the end, the matrix B is constructed from all w vectors and s is estimated by $s = B^T x$ where x is the whitened data.

5. METHODS

4.1. Adaptive parameter estimation of the noise model

According to the above derivation, we need to compute the variance of the data in order to obtain the noise power spectrum model using the empirical formula. The simplest way to do that is to estimate the variance of background areas within the available data set. In order to avoid bias errors from baseline variations in calculating the noise variance, the variance is computed from the selected background pixels within each image in the sequence separately. Then, the average of the obtained variance values from all images is taken to be the noise model variance estimate. This simple procedure allows baseline variations among the images in the sequence to be eliminated thus providing a more accurate estimate for the noise model.

4.2. Signal power spectrum estimation

Since the proposed technique is applied to a single time course at a time, the periodogram estimate of signal power spectrum is expected to have a rather large variance[19]. As a result, the subtraction of power spectra in Eq. (3) may contain negative values in practical implementations. This causes a problem in trying to compute the square root to recover the processed signal. The approach used in our implementation to overcome this problem relies on replacing all negative values in the subtraction results by zero. This approach is justified because all values lower than the estimated power spectrum are more likely to be noise components within the variance limits of the periodogram estimate.

4.3. Statistical noise removal

Given the nature of the original signal, we observe that the variance in the power spectrum estimate may only result from the random component. Since the expected value of the noise variation is known from the derived model and given the statistical characteristics of the periodogram estimate, we can express the noise at each of the power spectrum frequency bins as a Gaussian random variable with mean and variance both equal to the noise model[10,19]. As a result, the subtraction in Eq. (3) would effectively remove only a part of the noise power spectrum. In other words, the upper half of the Gaussian distribution would still remain in the processed signal.

To solve this problem, a slight modification to the technique is added to allow direct control over the extent of noise removed. The modified equation takes the form,

$$P_{dd}(\omega) = P_{ss}(\omega) - \alpha \cdot P_{nn}(\omega) \quad . \tag{8}$$

Here, the factor α is added to control the confidence of noise removal. This problem can be expressed in the form of a statistical z-test where α controls the p-value of the test. That is, the larger the value of α, the less the probability that the output power spectrum contains a noise component. On the other hand, increasing this value would increase the likelihood that some parts of the signal may also be removed. Therefore, the selection of the value of α is useful to fine-tune the results of the new technique. Several optimization criteria can be used to select the value of this parameter. An example of these is the use of entropy based objective function optimized over the autocorrelation function of the difference between the original and processed signals for different α values. This favors the values of α that give an autocorrelation function with narrow extent around zero and of minimal side peaks. This tends to preserve the components of the true signal, which give rise to periodic peaks in the autocorrelation function. Another approach is to optimize the kurtosis of the difference as close as possible to zero to make sure that the removed signal is only the

Gaussian random noise component. In this work, we used a fixed value of this parameter that is equal to 1 to make it easier to compare the results and assess the improvement after using this technique as a preprocessing stage and the optimization of this parameter will be left for further investigation.

4.4. Analysis of Results Using ICA

To show the improvement in using ICA on the processed signal, the new technique is applied to process all pixel time courses in the acquired data set independently and then the processed data set is used for subsequent ICA. The PCA and ICA techniques were performed on the time courses of local regions using a Matlab (Math Works, Inc.) program based on FASTICA[17]. The goal of this analysis is to assess the performance of the new technique in enhancing the results of PCA and ICA and stabilizing the convergence characteristics of the ICA. Moreover, the difference signals between the original and filtered data sets were also analyzed using these techniques. This helps verify the absence of signal components within this discarded part of the original signal.

6. RESULTS

The proposed technique was verified using actual data from a human volunteer. The data were obtained from an event-related fMRI study performed on a normal human volunteer using a Siemens 1.5T Magnetom Vision clinical scanner[4]. In this study, an oblique slice through the motor and the visual cortices was imaged using a $T2^*$-weighted EPI sequence (TE/TR=60/300ms, flip angle=55^o, Matrix=64$\times$64, FOV=22cm$\times$22cm, slice thickness=5mm). The subject performed rapid finger movement cued by flashing LED goggles. The study consists of 32 epochs with 64 images per epoch. Temporal data from only 4 or 8 epochs of pixels in both the motor and visual cortices were processed using the new method and compared to the case when the remainder of the acquired epochs are used for averaging. The PCA and ICA techniques were applied to decompose the signal into its basic components. Both techniques were used to process time course signals before and after the new technique is applied on pixels within a window selected by the user. Moreover, the difference signals were also analyzed from the same window.

In Figures1 and 2, the results of applying the proposed denoising technique to process real fMRI data from a 4$\times$4 region with time course length of 512 are presented. The results in Figure 2 look significantly improved compared to the original in Figure 1. We also notice that the baseline variations also remained unaltered. In Figures 3-6, the results of applying the PCA and ICA techniques before and after processing with the new technique are presented. The PCA and ICA were applied to the time courses of pixels within a local region of size 4$\times$4 in the acquired data set. The figures show the first five components of the PCA and five selected components from the ICA results (notice that ICA involves a whitening step that removes any preference of one independent component over the other unlike PCA). The results appear very noisy before applying the new technique in both methods. The results after it was used appear significantly clearer. We notice that the baseline variation component now appears in the ICA results, which was not present in the analysis before denoising. To verify the efficiency of the new denoising technique, the difference signals between the original and the denoised time courses are shown in Figure 7. The results of applying PCA on the difference signal are shown in Figure 8. As can be observed, no periodic components appear on the first four principal components indicating the random nature of the removed signal. Moreover, severe instability (i.e., no convergence of the ICA iteration) was observed when using the ICA iteration on this difference data. In fact, the analysis of the same difference signal using ICA never reached convergence in our experiments. This supports our hypothesis of the need to remove noise components to make the number of components less than the number of signals. Also, this shows that the proposed method was indeed successful to remove such components without affecting the true signal form.

7. DISCUSSION

The results of using the new technique suggest that it suppresses random noise while preserving deterministic signal components. Because it relies on subtracting the noise component (if we ignore the small nonlinear effect of clipping the negative values in the power spectrum), it does not generally affect independence of data points within the time course. This allows the technique to be virtually transparent to conventional statistical analysis methods, which assume statistical independence of samples. This means that no constraints are imposed on the data analysis when the new method is used as a preprocessing step.

The technique also has a very small computational complexity. Assuming a data set of L N×N images, the process of noise model estimation requires a computational complexity of nearly $O(L \cdot N)$. The actual denoising of a single time course requires $O(L \cdot \log(L))$ computations. This means that the whole data set can be denoised using $O(N^2 \cdot L \cdot \log(L))$ computations. For efficient implementation, the noise model is estimated once and the denoising is implemented only for the regions of interest. Given that the denoising of a particular time course is performed completely independently from the others, the proposed method can readily take advantage of parallel processing when available.

In developing the proposed technique, very few assumptions about the nature of the noise model and no assumptions about the deterministic signal components are made. This is an obvious advantage compared to previous techniques that rely on signal models to construct the denoising filter[4]. Given that the noise model parameters can be adaptively estimated from background areas within the fMRI data set, the implementation of the new technique is expected to maintain robustness.

The new technique provides a statistical control over noise removal using a single parameter α. This allows the user to customize its use to the specific data analysis tool of his/her choice. At the same time, it allows the comparison of implementations between different groups using this unambiguous reference.

The results indicate that new technique enables robust use of PCA and ICA. The main advantages of using the new technique come as a direct result of the conditioning of the input signals to match the requirements of these analysis methods. That is, to make sure that the number of independent sources is less than the number of signals. This is reflected on the results as better convergence characteristics as well as less mixing between remaining noise components and the deterministic components in the output of such analyses. The results emphasize the non-Gaussian nature of the deterministic fMRI signal components. This is evident from comparing the outputs from ICA as compared to those from PCA. This suggests that a study of this characteristic of the signal components should be addressed in a future study using the new technique as a preprocessing step.

The use of the proposed analysis to detect phase differences between activations can be done by analyzing the magnitude of Fourier-transformed time courses instead of the actual time course data. Given the characteristics of ICA, instead of having several ICs representing the different shifts, this method enables the detection of the basic activation shape as one IC and the phase differences can be computed from the phase components of the individual time courses. This part is trivial to demonstrate on simulated data. On the other hand, we could not acquire a data set with enough temporal resolution to demonstrate this on real data (basically all activations appear as one IC in our data sets). With the current pace of improvement on imaging sequences, it will not probably be too long before this become possible. At this time, this part of the work will remain open for future work.

The limitations of the new technique include two parts, namely, the noise in phase component, and the effect of the clipping in the subtracted power spectrum. The first is partially removed when only the real part of the signal is used as the denoised signal, which effectively means that the phase values in the negative and positive frequency are averaged. On the other hand, the effect of spectral clipping can be modeled as an addition of another random signal that accounts for the clipped parts. This results in a degradation of the results that was deemed insignificant by our experiments. Further investigation of these limitations is required to improve on the results of the proposed technique.

8. CONCLUSIONS

A new signal denoising technique was proposed for fMRI signals. The new strategy based on spectrum subtraction method is adaptive and simple to implement while offering a substantial improvement of the signal-to-noise ratio. The implementation was described and its performance was demonstrated using computer simulations and real data. The use of the proposed technique as a preprocessing step was also shown to substantially improve the performance of PCA and ICA in analyzing fMRI data sets. Further work is needed to investigate the potential of the new technique in different clinical applications.

ACKNOWLEDGMENTS

This work is supported in part by International Electronics – Biomedical Division, Egypt. The author would like to thank Drs. X. Hu, S. Ngan, and S. Laconte for sharing their data and for helpful discussions.

REFERENCES

1. W.W. Orrison, J.D. Lewine, J.A. Sanders, and M.F. Hartshorne, *Functional brain imaging*, Mosby-Year Book, St. Louis, 1995.
2. S. Ogawa, D.W. Tank, R. Menon, J.M. Ellermann, S.-G. Kim, H. Merkle, and K. Ugurbil, "Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging," *Proc Natl Acad Sci USA* **89**, pp. 5951–5955, 1992.
3. R.L. Buckner, P.A. Bandettini, K.M. O'Craven, R.L. Savoy, S.E. Petersen, M.E. Raichle, and B.R. Rosen, "Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging," *Proc. Natl. Acad. Sci. USA* **93**, pp. 14878–14883, 1996.
4. S.M. LaConte, S. Ngan, and X. Hu, "Wavelet transform-based Wiener filtering of event-related fMRI data," *Magn. Res. Med.* **44**, pp. 746-757, 2000.
5. T.H. Le and X. Hu, "Retrospective estimation and correction of physiological artifacts in fMRI by direct extraction of physiological activity from MR data," *Magn. Reson. Med*. **35**, pp. 290-298, 1996.
6. B.A. Ardekani, J. Kershaw, K. Kashikura, and I. Kanno, "Activation detection in functional MRI using subspace modeling and maximum likelihood estimation," *IEEE Trans. Med. Imag.* **18**, no. 2, pp. 101-114, 1999.
7. M. Svensen, F. Kruggel, and Y. von Cramon, "Probabilistic modeling of single-trial fMRI data," *IEEE Trans. Med. Imag.* **19**, no. 1, pp. 25-35, 2000.
8. C. Goutte, F.A. Nielsen, and L.K. Hansen, "Modeling the haemodynamic response in fMRI using smooth FIR filters," *IEEE Trans. Med. Imag.* **19**, no. 2, pp. 1188-1201, 2000.
9. A.F. Sole, S. Ngan, G. Sapiro, X. Hu, and A. Lopez, "Anisotropic 2-D and 3-D averaging of fMRI signals," *IEEE Trans. Med. Imag.* **20**, no. 2, pp. 86-93, 2001.
10. S.V. Vaseghi, *Advanced Signal Processing and Digital Noise Reduction*, John Wiley & Sons, New York, 1996.
11. A.M. Law and W.D. Kelton, *Simulation, Modeling & Analysis*, 2nd ed., McGraw-Hill, New York, 1991.
12. A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, 2001.
13. S. Makeig, T. Jung, D. Ghahremani, and T.J. Sejnowski, *Independent component analysis of simulated ERP data*, Institute for Neural Computation Technical Report No. INC-9606.
14. S. Makeig, M. Westerfield, T. Jung, J. Covington, J. Townsend, T.J. Sejnowski, and E. Courchesne, "Functionally independent components of the late positive event-related potential during spatial attention," *J. Neuroscience* **19**, no. 7, pp. 2665-2680, 1999.
15. M.J. Mckeown, T. Jung, S. Makeig, G. Brown, S.S. Kindermann, T. Lee, and T.J. Sejnowski, "Spatially independent activity patterns in functional MRI data during the Stroop color-naming task," *Proc. Natl. Acad. Sci., USA* **95**, pp. 803-810, 1998.
16. M.J. McKeown, S. Makeig, G.G. Brown, T. Jung, S.S. Kindermann, A.J. Bell, and T.J. Sejnowski, "Analysis of fMRI Data by Blind Separation into Independent Spatial Components," *Human Brain Mapping* **6**, pp. 160-188, 1998.
17. http://www.cis.hut.fi/projects/ica/fastica.
18. B.P. Lathi, *Modern Digital and Analog Communication Systems, 2nd ed.*, Oxford University Press, Oxford, 1989.
19. C.S. Burrus, J.H. McClellan, A.V. Oppenheim, T.W. Parks, R.W. Schafer, H.W. Schuessler, *Computer-Based Excercises for Signal Processing Using Matlab*, Prentice Hall, New Jersey, 1994.

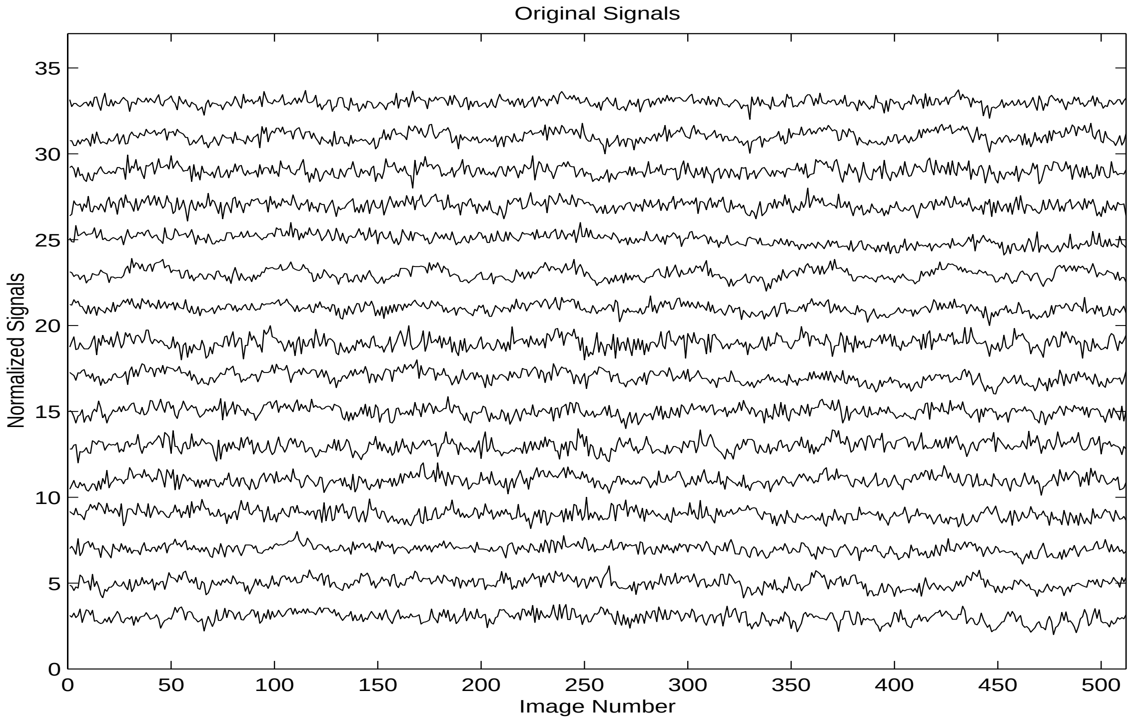

Figure 1. Original time course signals within a 4×4 region within the brain.

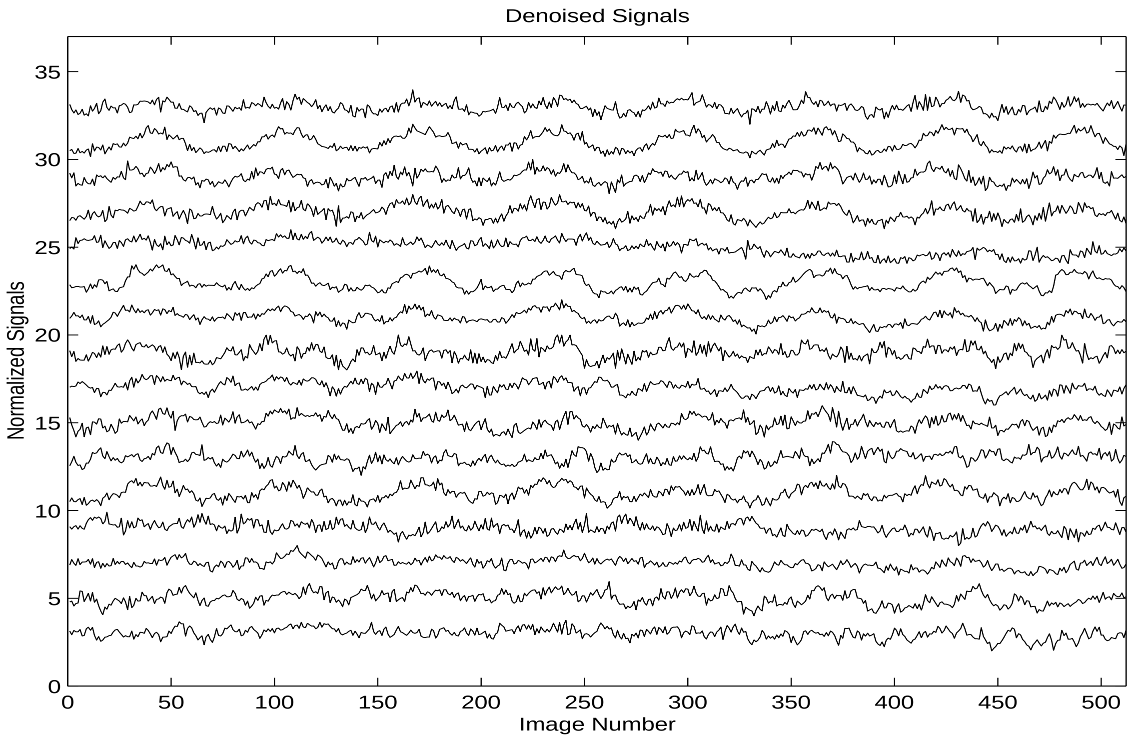

Figure 2. Denoised time course signals within a 4×4 region within the brain.

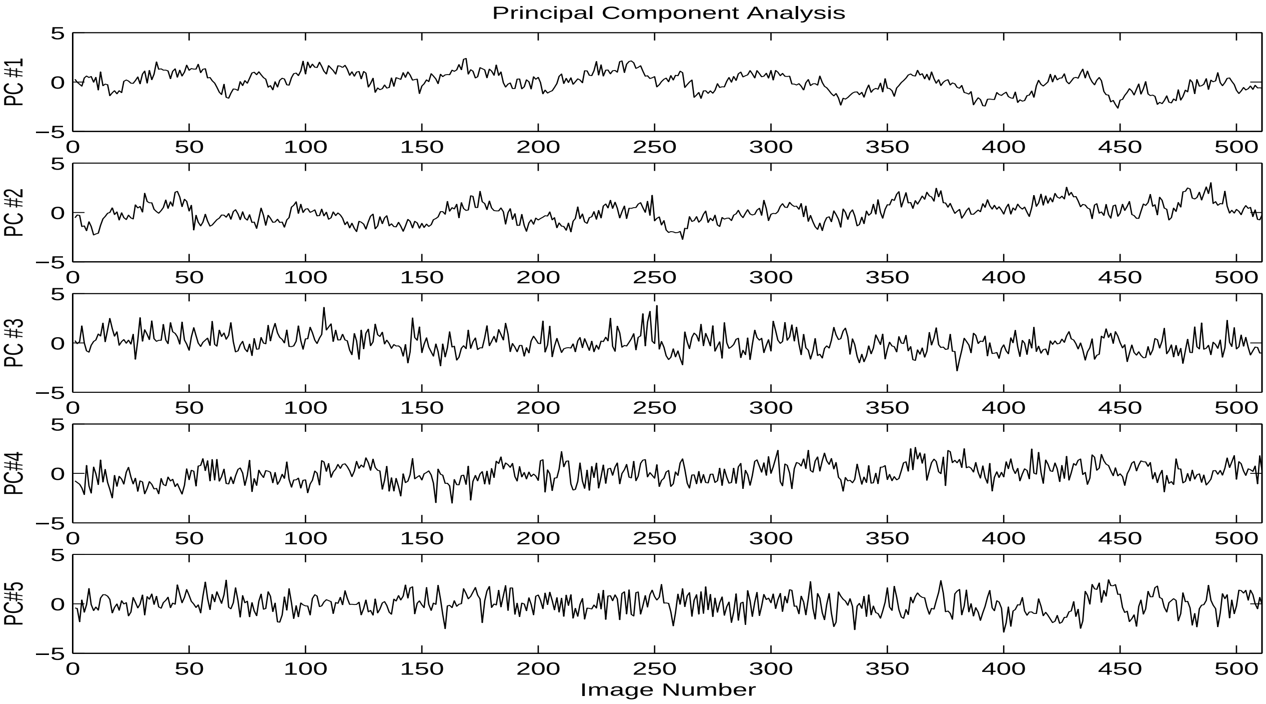

Figure 3. Principal component analysis of original time course signals.

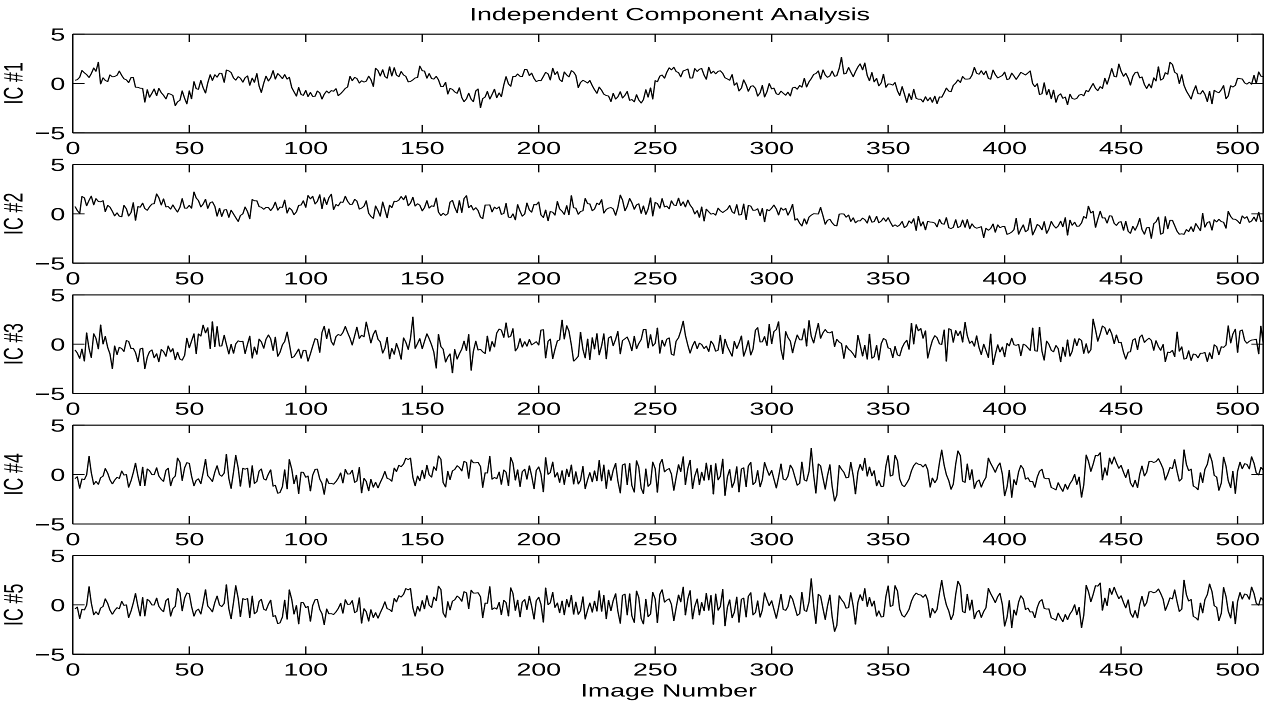

Figure 4. Independent component analysis of original time course signals.

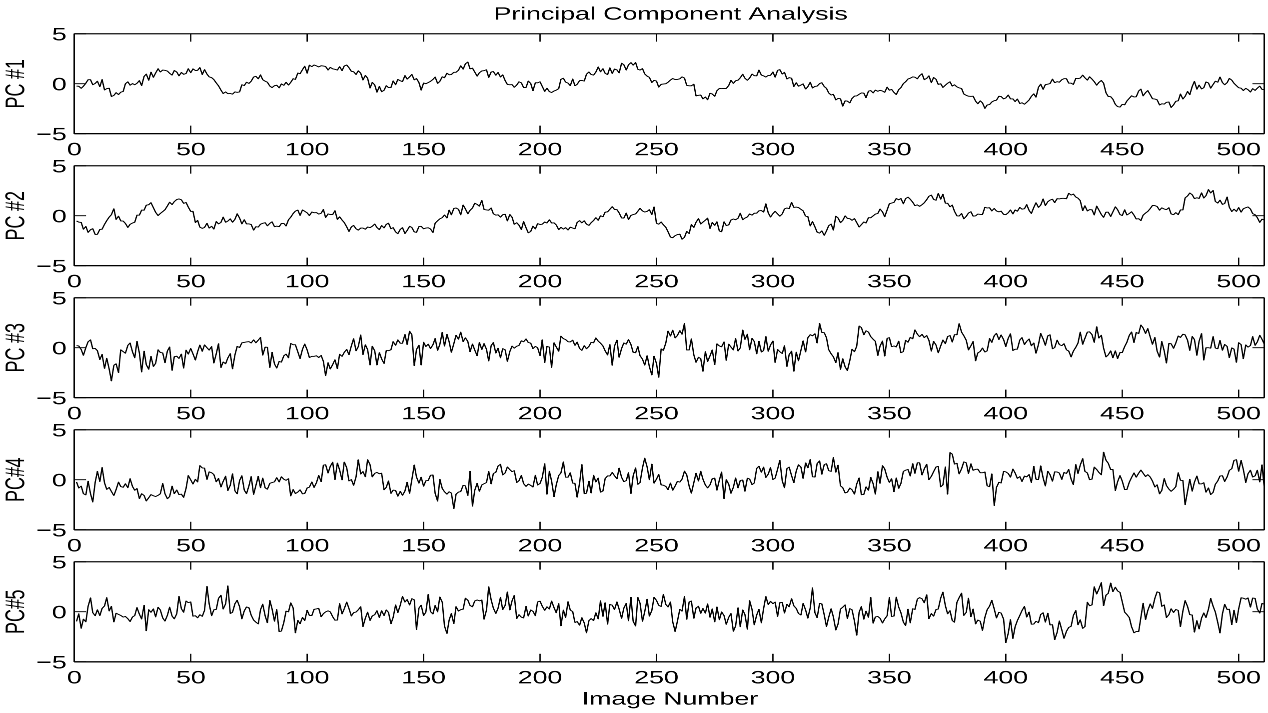

Figure 5. Principal component analysis of denoised time course signals.

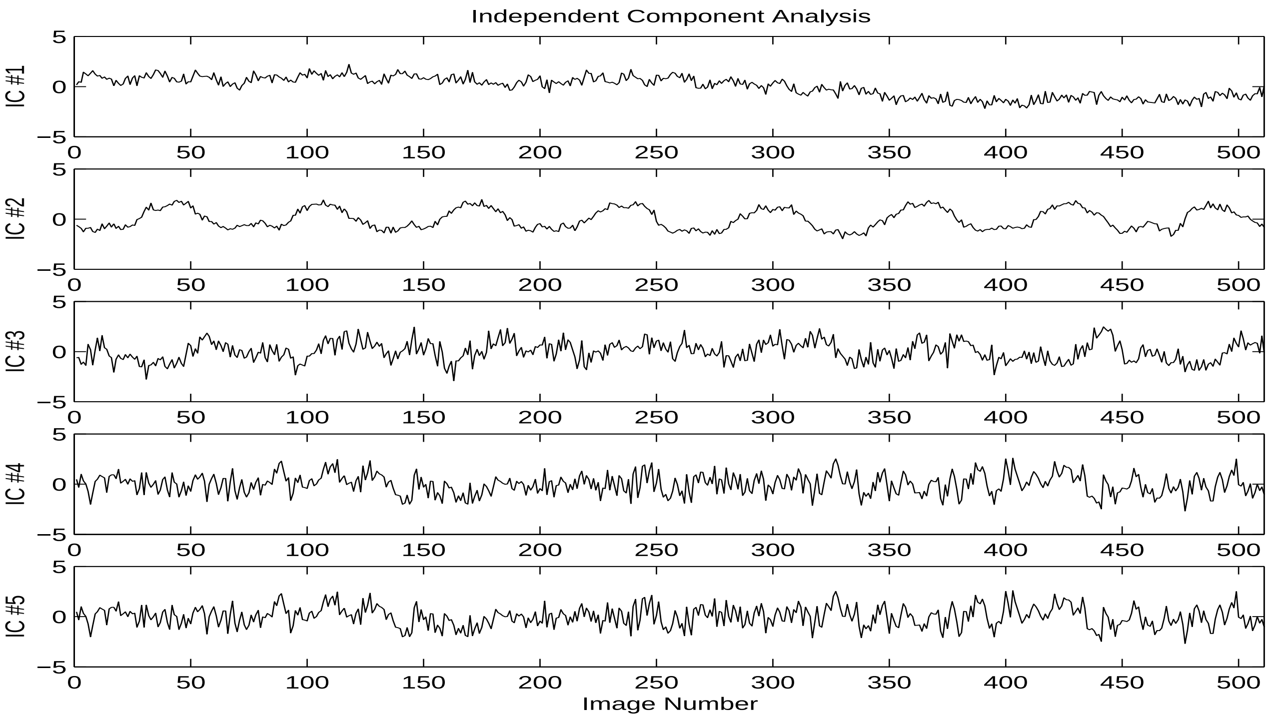

Figure 6. Independent component analysis of denoised time course signals.

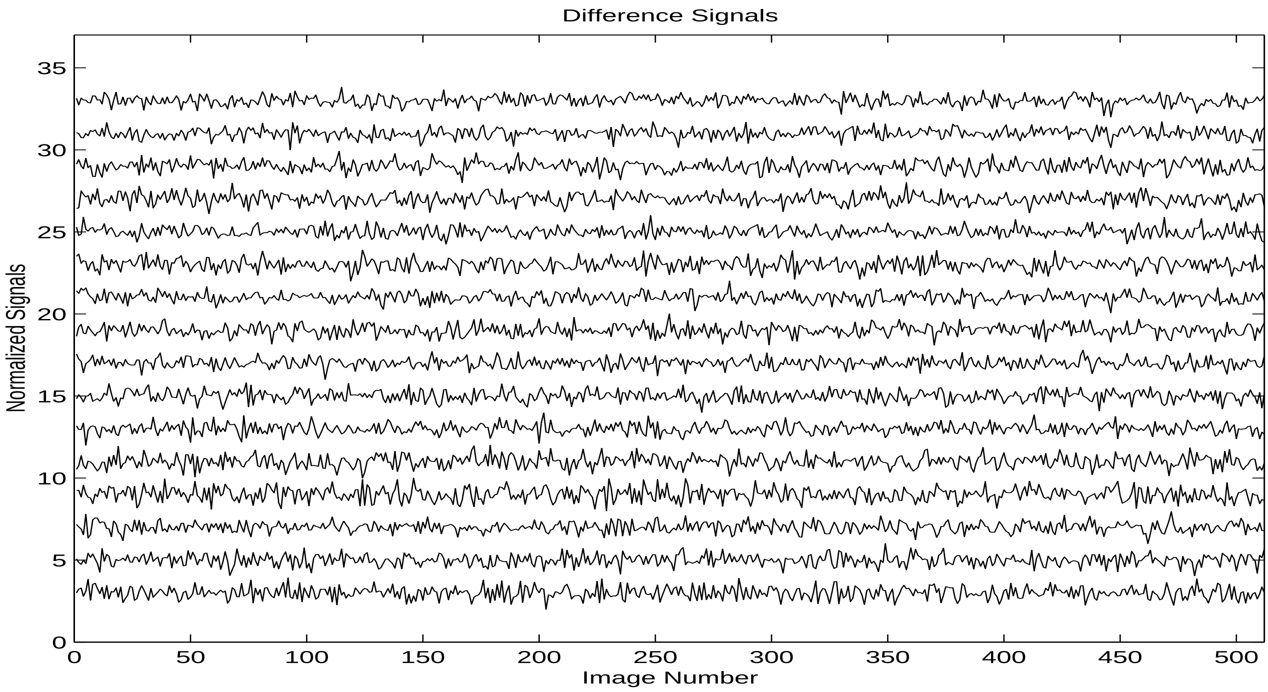

Figure 7. Difference signals between original and denoised time course signals.

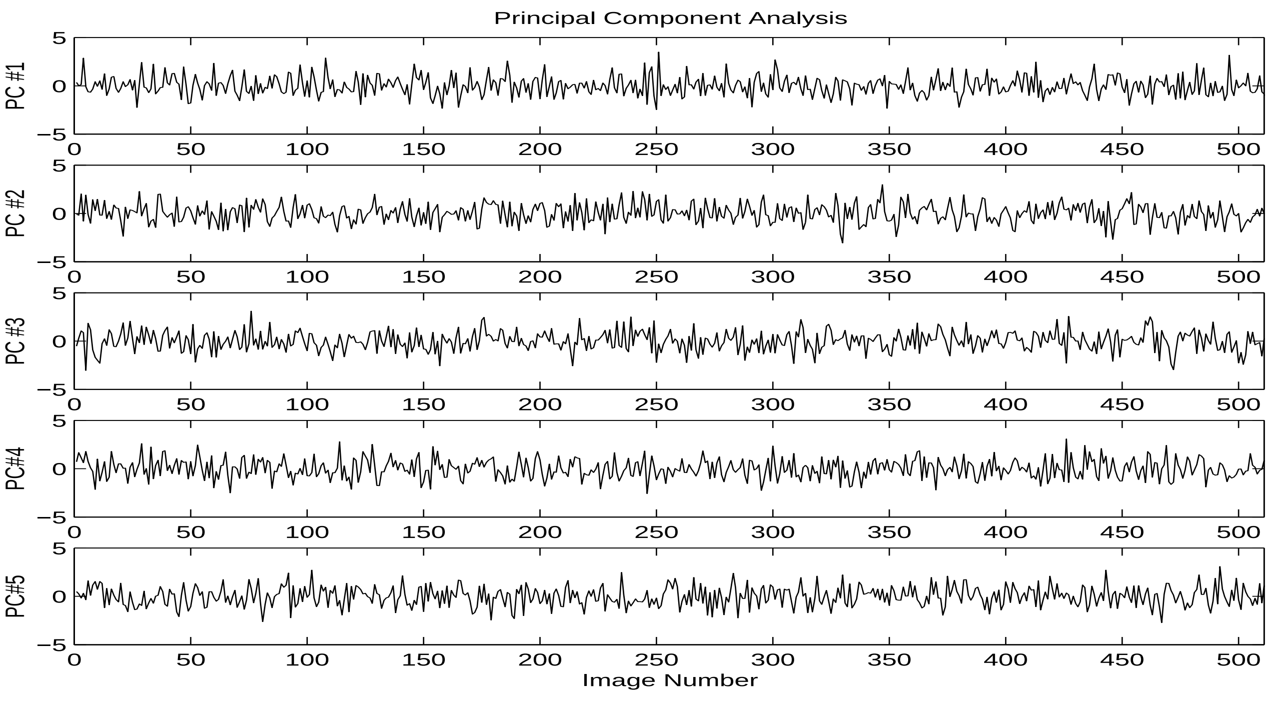

Figure 8. Principal component analysis of difference signals.