Evolution of Seizure Precursors in Refractory Epilepsy

Sponsored by
University of Pennsylvania

Final Report
September 28, 2007

Principal Investigators:

Dr. George Vachtsevanos
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0250
(404) 894-6252
george.vachtsevanos@ece.gatech.edu

Feature Analysis of Phantom fMRI Data

Feature analysis of voxel time courses is a recent advance in the study of fMRI data. Because this work explores a novel method for analyzing areas of activation in fMRI data, rigorous verification and validation of the techniques are necessary to prove their utility. To that end, in addition to examining real data from human subjects with nebulous epileptogenic zones, phantom data—simulated data modeled to closely resemble actual patient data—with known areas of activation are examined for statistical validation of the methodology.

I. Phantom Data Generation

The first step in the process of generating the synthetic fMRI data was the creation of the baseline series, i.e., a set of low noise images without activations or motion artifacts. The same baseline dataset was used for all generated phantoms. The baseline series was formed by computing the mean label and control images from a scanned subject's ASL perfusion dataset. The images were motion corrected and spiking artifacts were removed prior to calculating their means. These mean images were then replicated 400 times each and interleaved to create a baseline set consisting of 800 alternating label and control images. The perfusion series—a set of images produced by subtracting each label image from the subsequent control image—had constant values in all of its voxel time courses. The voxel time course, a signal corresponding to a voxel's intensity as a function of time, is obtained by extracting the voxel intensity from the images at each point in time as illustrated in Figure 1. The baseline image series with constant time courses formed the basis for the synthetic datasets.

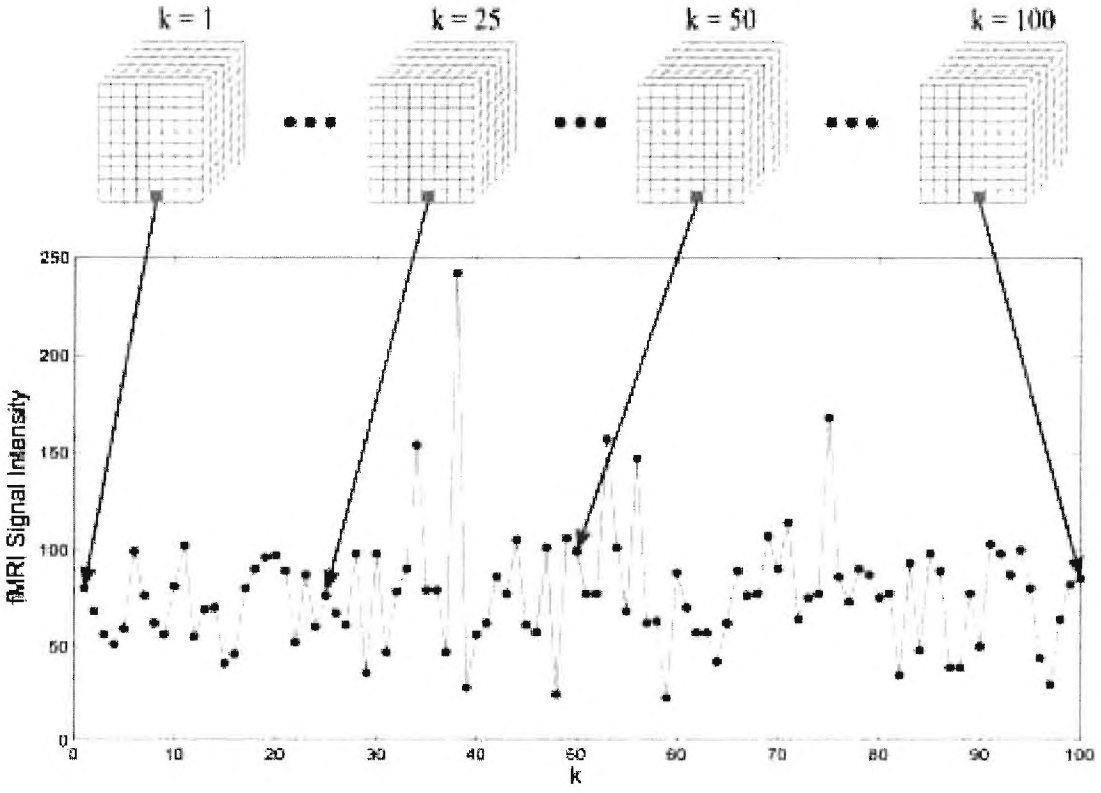

Figure 1 An illustration of the voxel time course extraction procedure.

An accurate model of the hemodynamic response function (HRF), i.e., the fMRI signal change resulting from a single brief stimulus in the brain, is necessary to produce realistic phantom fMRI datasets. The nature of the hemodynamic response to neural activity has been widely explored by researchers, but no one HRF model that fits all human subjects has been found because HRFs vary greatly from person to person and also vary within a subject from one brain region to another [1]. Differences in the shapes of the HRFs, time-to-onset following stimulus, maximum signal change, time-to-peak, and the full width at half maximum (FWHM) of the response have been reported in both BOLD and perfusion fMRI studies [2, 3]. A post-peak undershoot (a drop in the signal intensity below the baseline value) with variable degree and timing among subjects has also been observed in many empirical HRFs but is not always

present. The presence of an 'initial dip' in the fMRI signal due to an increase in oxygen consumption following stimulus and prior to the signal peak has been observed as well in some BOLD studies, but its existence is not universally accepted [4].

In their work, Friston et al. found that the BOLD HRF can be modeled with a Poisson function [5], while Boynton et al. determined that a gamma function provided a more appropriate model [6]. The gamma function model has been shown to account for a much greater percentage of the fMRI evoked response to stimulus than the Poisson model and is therefore the more commonly employed model [1]. Although these studies focused on the hemodynamic response in BOLD signals, the same models can be applied to perfusion fMRI because the HRF shapes are similar for the two methodologies [2]. Because of the aforementioned reasons, a gamma function based model was exploited in this research to simulate the hemodynamic response to epileptic activity. Instead of applying the single gamma function model suggested by Boynton et al., the sum of two gamma functions model was chosen in order to incorporate a post-peak undershoot into the HRF [3]. The 'initial dip' is not included in this model. The sum of two gamma functions model is described by the following equation:

$$y(t) = A_1(t-\delta_1)^{h_1} e^{l_1(t-\delta_1)} + A_2(t-\delta_2)^{h_2} e^{l_2(t-\delta_2)} \quad (1)$$

where A_1 and A_2 control the peak and undershoot amplitudes, respectively, h_1 and h_2 affect the widths and timings of the peak and undershoot, l_1 and l_2 also affect the peak and undershoot widths, and δ_1 and δ_2 determine the time-to-onset of the HRF. In this work, the values for the parameters in Equation 1 were chosen based on empirically determined ranges from real subject HRF analyses reported in literature [1-3]. The shape of the HRF, the timings and numbers of simulated epileptic events, and the size and location of the epileptic focus were varied for each phantom dataset. The maximum percentage signal change for the simulated activations ranged

from 10% to 20% above the baseline value. Signal changes within this range have been reported in other perfusion data studies [2, 7]. The chosen HRF model is shown in Figure 2, and Table 1 summarizes the ranges of values used for some important HRF signal characteristics.

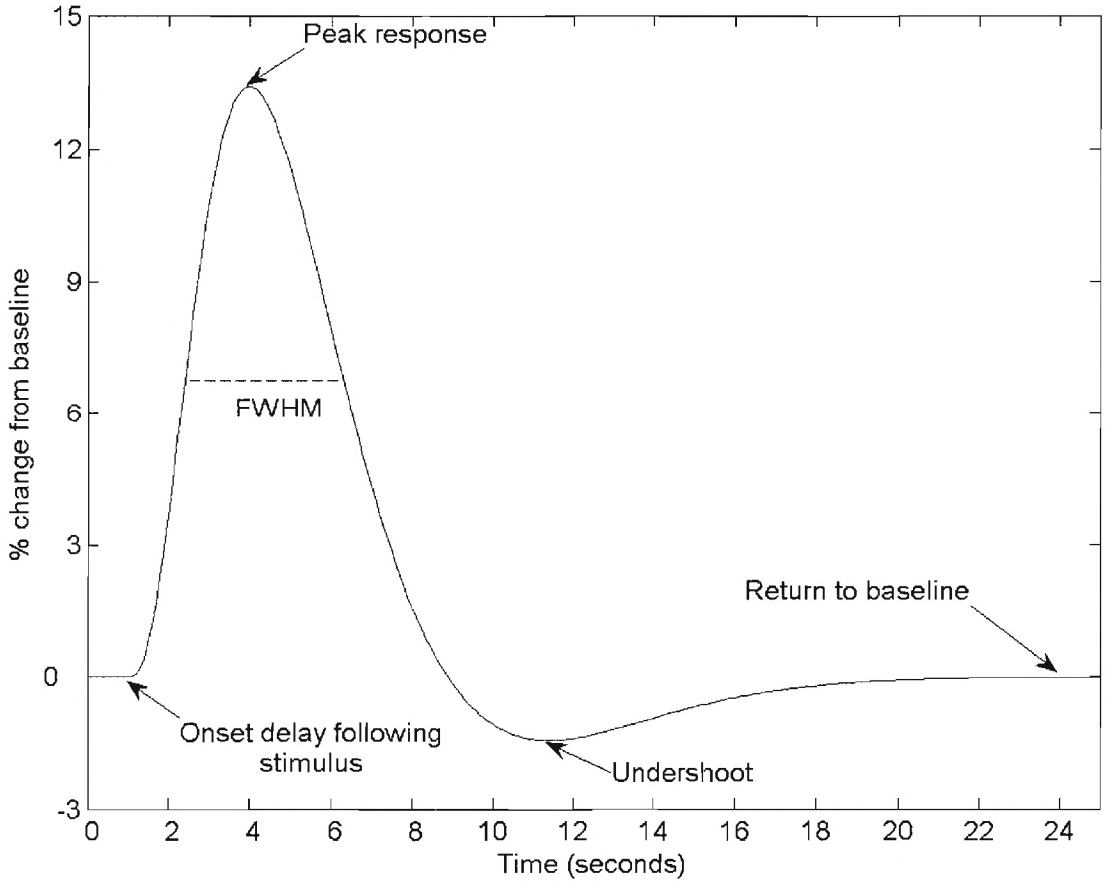

Figure 2 The sum of two gamma functions model of the HRF.

Table 1 Ranges of values for the simulated HRF characteristics.

Characteristic	**Values**
Percent maximum signal change	10%–20%
Time-to-peak response	4–5 seconds
Time-to-undershoot	9–10 seconds
FWHM	5–6 seconds
Ratio of peak response to undershoot	4–5

After adding simulated activations to the baseline dataset, motion deformations were applied to the images to model the effects of subject motion during scanning. Movement related fMRI signal changes can account for a majority of the signal variance in extreme cases and can result in false positive activations if not accounted for prior to data analysis [8]. Random head movements were simulated in this work to create phantoms with realistic motion characteristics. Task correlated-motion, which arises as a result of a subject's actions during a task activation study (e.g., a finger tapping exercise), is not present in resting functional data, so this type of motion was not added to the phantoms.

Motion in fMRI data is typically modeled as a six-parameter rigid body transformation that describes translations along the Cartesian coordinate system axes and rotations about these axes [9, 10]. Using this model, the n^{th} volume in the series, $f_n(\mathbf{x})$, can be related to the first volume by the following equation:

$$f_n(\mathbf{x}) = \alpha_n f_1(\mathbf{R}_n \mathbf{x} + \mathbf{t}_n) + e_n(\mathbf{x}) \qquad (2)$$

where $\mathbf{x}$ is the spatial position of the voxel, $\mathbf{R}_n$ is a 3×3 orthonormal matrix characterized by the three rotation parameters, $\mathbf{t}_n$ is a 3×1 vector comprised of the three translation parameters, α_n is a scaling factor to account for intensity differences between volumes, and $e_n(\mathbf{x})$ is noise [8]. The motion parameters for each of the eight real subject datasets were estimated using the statistical parametric mapping software package SPM2 [9, 11, 12] and used as the basis for the motion time courses in the phantom data. The motion parameter time courses for a control subject are shown in Figure 3.

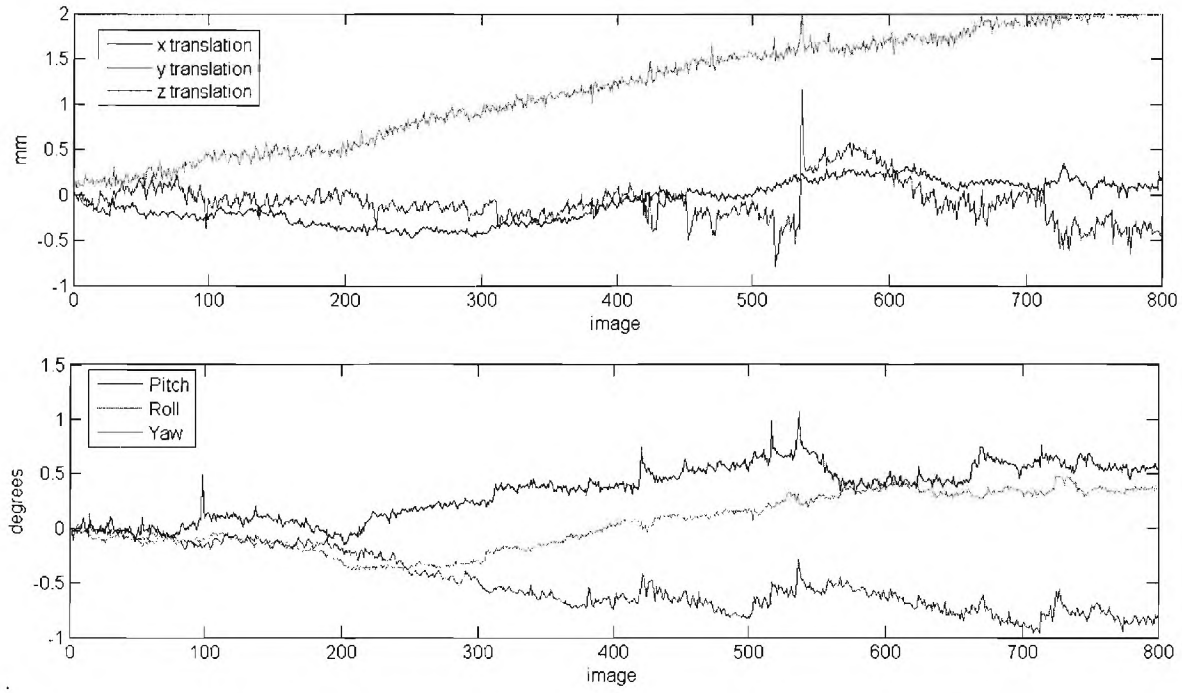

Figure 3 The translational (*top*) and rotational (*bottom*) motion parameters for a subject without epilepsy. Pitch, roll, and yaw represent the rotation about the x, y, and z axes, respectively.

The final consideration in the creation of the phantom datasets was the choice of noise models. The noise in fMRI is often assumed to be Gaussian, but studies have shown that the magnitudes in MR images actually assume a Rician distribution [13, 14]. At high SNRs, the Rician distribution approaches the Gaussian distribution, but the former model provides a closer approximation to the image noise at low SNRs. The Rician nature of the image noise is illustrated in Figure 4, which shows the histogram of the background voxel intensities for an ASL image from a control subject. To generate this histogram, the non-brain regions (background voxels) were manually segmented, and their intensity values were extracted. The Rician probability density function (PDF) provides a better overall fit to the data than the Gaussian PDF. The Rician PDF is a close match to the background noise histogram except at the point in the histogram where there is a brief excursion from the trend. Figure 5 shows the

histogram of image voxels corresponding to brain tissue regions. This higher SNR region has an intensity distribution that is almost Gaussian, as expected.

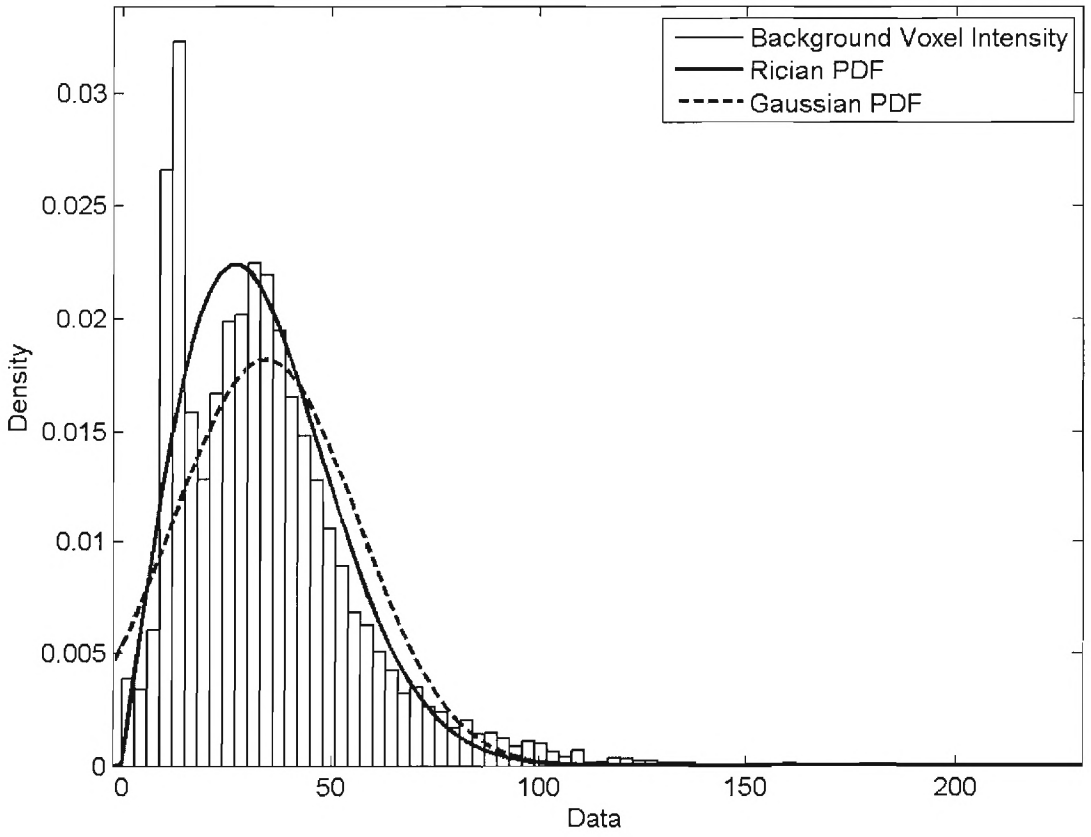

Figure 4 This figure demonstrates the close fit between the Rician PDF and the histogram of the background noise in an fMR image. The Rician PDF clearly provides a better fit than the Gaussian PDF over the entire intensity range.

The PDF of the noisy images is defined by the following equation:

$$p(r) = \frac{r}{\sigma^2} e^{\frac{-(r_0^2 + r^2)}{2\sigma^2}} I_0\left(\frac{r_0 \cdot r}{\sigma^2}\right) \tag{3}$$

where r_0 is the intensity of the underlying noise-free image, σ controls the scale of the noise, r_0/σ is the image SNR, and I_0 is the modified zeroth order Bessel function of the first kind. As

Equation 3 indicates, the Rician noise in these images depends on the noise-free image intensity. Because of this, Rician noise cannot be simply added to the images. The image intensities are made Rician distributed by adding complex Gaussian white noise to the images and taking the magnitude of the complex data to form real-valued, noisy images [14, 15].

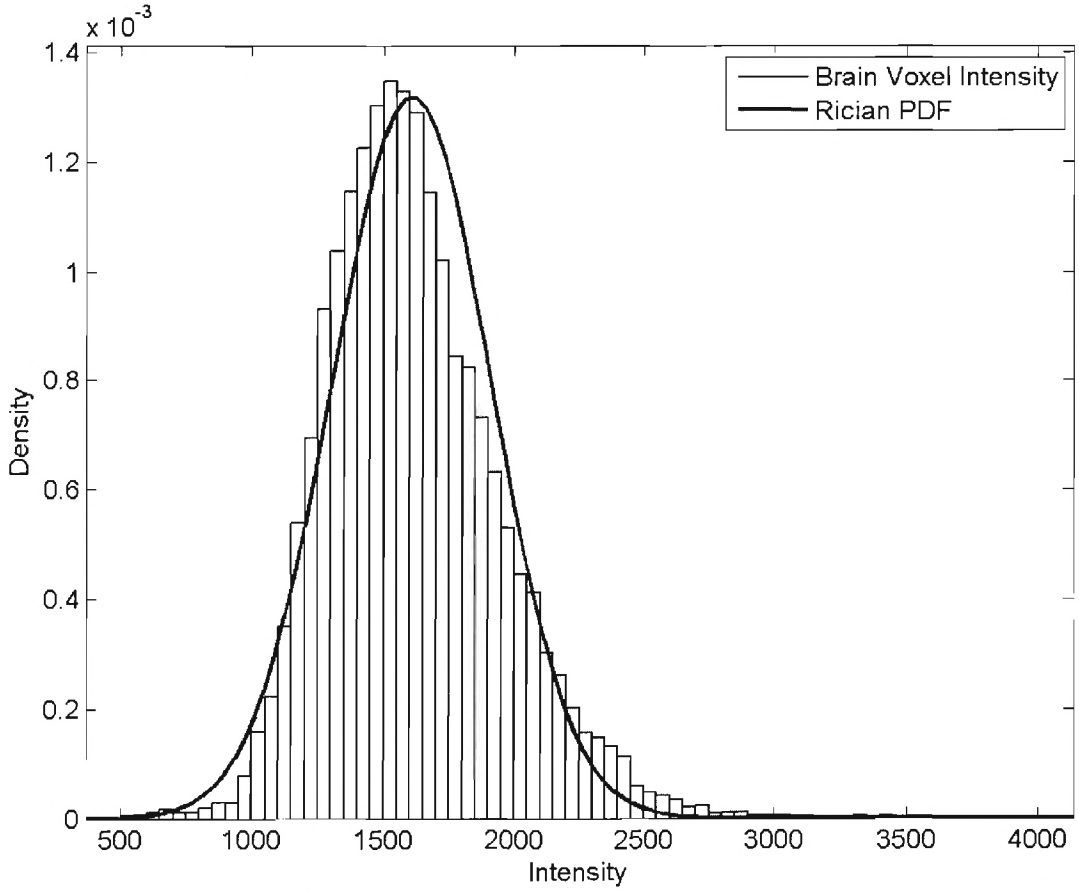

Figure 5 Histogram of the brain voxel intensity values with a Rician PDF overlaid. For this high SNR region of the image, the Rician distribution strongly resembles a Gaussian distribution.

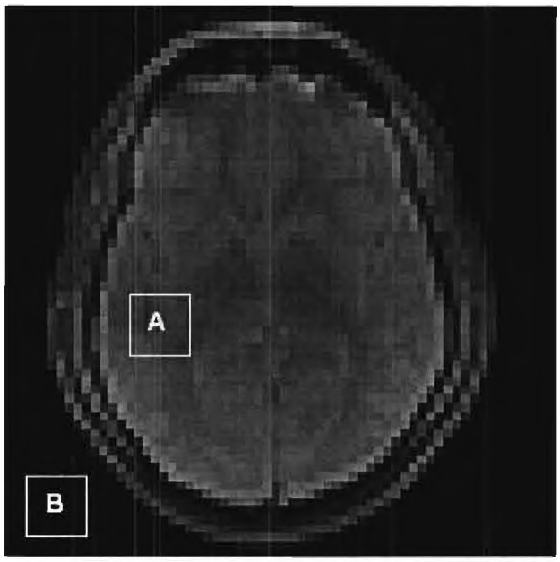

Figure 6 The regions of interest (ROIs) selected for estimating the image SNR as defined in Equation 4. The same ROIs were used for all subjects.

To generate the noisy images, first an appropriate SNR was chosen. The SNR values from the subject ASL scans were estimated to determine reasonable choices for phantom data SNRs. Applying the technique outlined by Pickens et al. in [16], the value of the SNR was computed for each acquired image, and a realistic range of SNR values was obtained. This procedure involved choosing a region of interest (ROI) inside the brain (ROI_A) and a background ROI without visible artifacts (ROI_B) as depicted in Figure 6. An estimate of the image SNR was then calculated using Equation 4. The computed SNR values for the data collected in this study ranged from approximately 29 dB to 49 dB with an average of 40 dB. After selecting an SNR, the value of the standard deviation of the background voxels necessary to yield the chosen SNR was computed using Equation 4, and the Rician distribution parameter σ was then found using Equation 5. The correction factor in the denominator of Equation 5 is needed due to the fact that the variance of the Rician distribution is $\sigma_R^2 = \sqrt{\sigma^2(2-\pi/2)}$ when the image intensity is zero (as

it would be in the noise-free background voxels). σ is the standard deviation of the Gaussian noise used to generate the Rician distribution, not the distribution itself. The final step of the procedure was to generate the noisy images using Equation 6, where **x** is the voxel spatial location and n_1 and n_2 are the real and imaginary components, respectively, of complex Gaussian white noise with zero mean and variance σ^2.

$$SNR = 20 \cdot \log_{10}\left(\frac{\text{mean}(ROI_A)}{\text{std}(ROI_B)}\right) \tag{4}$$

$$\sigma = \frac{\text{std}(ROI_B)}{\sqrt{2-\pi/2}} \tag{5}$$

$$r(\mathbf{x}) = \sqrt{(r_0(\mathbf{x}) + n_1(\mathbf{x}))^2 + n_2^2(\mathbf{x})}, \quad n_1, n_2 \sim N(0, \sigma^2), \text{ i.i.d.} \tag{6}$$

The following is a summary of the procedure implemented for generating realistic fMRI phantom data:

1. The baseline data were created from a control subject dataset after the images were motion corrected and spike artifacts were removed.
2. The HRF model given in Equation 1 was added to a subset of the voxels to simulate epileptic activity in the brain.
3. Motion artifacts were added to the data using the relationship in Equation 2 in order to simulate the effects of patient movement during scanning.
4. Finally, the image intensities were modified using Equation 6 to model fMRI data with Rician noise characteristics.

II. Feature Extraction

Features are measures of signal attributes that are used to compress raw data into smaller, more informative sets of data for classification, pattern recognition, and machine learning applications. For large datasets, analyzing the raw signals themselves can be computationally expensive and even impractical. By examining features of signals, the most important characteristics of these signals can be gleaned and some of the irrelevant information can be discarded. Feature extraction is therefore a key step in any classification problem. In fact, when a highly representative set of features is selected, the choice of a classifier becomes almost immaterial [17].

The features described below were chosen based on their proven usefulness in the study of other physiologic signals and in signal processing in general [18-22]. Features were selected from the time, statistics, information theory, and frequency domains to reveal some of the most important signal characteristics. In the following equations, x represents the signal samples, N is the length of the signal, p is the signal histogram, P_{xx} is the power spectral density (PSD) of x, and $X(f)$ is the Fourier transform of x.

- *Energy*: measures the average instantaneous energy in the signal.

$$\frac{1}{N}\sum_i x_i^2$$

- *Curve Length*: sum of the lengths of the vertical line segments between samples. It provides measures of both time and frequency characteristics.

$$\sum_i |x_i - x_{i-1}|$$

- *Nonlinear Energy*: provides information about amplitude and frequency content of a signal. It is also known as Teager's energy [20].

$$\frac{1}{N}\sum_i \left(x_i^2 - x_{i+1}\cdot x_{i-1}\right)$$

- *Katz Fractal Dimension*: measures the number of elemental units comprising a curve. It is useful for detecting transients in a signal [22, 23].

$$\frac{\log_{10}(N-1)}{\log_{10}\left(\frac{\max_i\left(\sqrt{(x_i - x_1)^2 + i^2}\right)}{\sum_i \sqrt{(x_i - x_{i-1})^2 + 1}}\right) + \log_{10}(N-1)}$$

- *Mean*: average (arithmetic mean) of the signal amplitudes.

$$\mu = \frac{1}{N}\sum_i x_i$$

- *Variance*: measure of the dispersion of the amplitude values.

$$\sigma^2 = \frac{1}{N}\sum_i (x_i - \mu)^2$$

- *Skewness*: measure of the asymmetry of the data distribution.

$$\frac{1}{N}\sum_i \left(\frac{x_i - \mu}{\sigma}\right)^3$$

- *Kurtosis*: measure of how prone to outliers the signal distribution is.

$$\frac{1}{N}\sum_i \left(\frac{x_i - \mu}{\sigma}\right)^4 - 3$$

- *Interquartile Range*: measures statistical dispersion in the signal amplitudes. It is difference between the 75th and 25th percentile samples in x.

- *Spectral Entropy*: measure of the randomness in the frequency spectrum of a signal.

$$-\sum_i P_{xx} \times \log_2(P_{xx})$$

- *Shannon Entropy*: measure of the randomness in the time signal.

$$-\sum_i p_i \times \log_2(p_i)$$

- *Renyi Entropy*: measure of the randomness in the time signal.

$$\frac{1}{1-q}\log_2 \sum_i p_i^q$$

- *Complexity*: measures disorder in a sequence using the procedure described in [24].
- *Median Frequency*: frequency at which equal amounts of the spectral power lie and below.

$$f_{med} \ni \sum_{f=0}^{f_{med}} |X(f)|^2 = \frac{1}{2}\sum_{f=f_{med}}^{\infty} |X(f)|^2$$

- *Mean Frequency*: frequency that is the centroid of the PSD.

$$\frac{\sum_f f \cdot |X(f)|^2}{\sum_f |X(f)|^2}$$

- *Peak Frequency*: frequency at which the PSD reaches its maximum value.

$$f_{\max} = \arg\max_f |X(f)|$$

III. Feature Selection

Seemingly, having more information about the samples from the two distributions to be classified would improve the overall classification accuracy; however, this is not always true. Correlations in the features used in classifying data and noise in these features can increase the probability of misclassifications. For this reason, feature selection techniques were exploited in order to determine the best combinations of features for accurately distinguishing epileptogenic and non-epileptogenic voxels.

The features presented in Section II formed the basis for the library of signal attributes investigated for discernment of epileptogenic and non-epileptogenic brain regions. The aim of feature selection was to find a feature subset of size $d < D$ (where D is total number of features in the library) that had the greatest ability to distinguish between the two classes of voxels. Because the goal was to maximize the classification performance, the classification accuracy was selected as an appropriate objective function for evaluating the feature subsets. The four feature selection approaches employed in this work include:

- Sequential forward selection (SFS)

  SFS is a greedy search algorithm that determines the best set of features for extraction by starting from an empty set and sequentially adding a single feature to the subset if it increases the value of the chosen objective function.

- Sequential backward selection (SBS)

  SBS is similar to SFS but works in the opposite direction. The search initializes with the full set of features and removes a single feature that improves or minimally worsens the objective function to obtain the best subset of features.

- Sequential floating forward selection (SFFS)

  SFFS addresses the tendency of SFS and SBS to gravitate toward local minima due to the inability to re-evaluate the usefulness of features that were previously added or discarded. SFFS performs a dynamically chosen number of forward selection steps followed by backward selection as long the objective function increases. The process then repeats until the desired number of features has been selected.

- Branch and bound (B&B)

B&B is an exponential search algorithm with the ability to find "optimal" solutions under certain assumptions and conditions [25]. The algorithm creates a tree structure such as the one shown in Figure 7. Then, based on the assumption that the superset always has a higher objective function value than its subsets—which is not always true—the algorithm is able to eliminate certain branches of the tree as possible best solutions without evaluating them. In this way, the optimal solution can be found without an exhaustive search of the feature space.

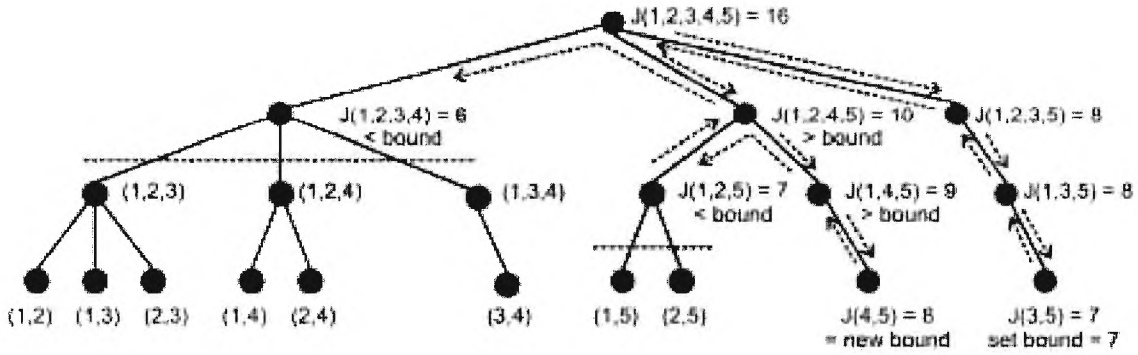

Figure 7 Example of a branch and bound search tree for $D = 5$ and $d = 2$. The dashed black lines represent the points at which branches can be pruned in order to reduce the computational burden without loss of optimality, and the arrows represent the order of tree traversal starting from the right.

IV. Results and Discussion

Each feature selection algorithm in Section III was executed for varying feature subset sizes with the objective function, feature vectors, and classification vectors as the algorithm inputs. Since the features were to ultimately be used in a pattern detection system, the classification accuracy defined by Equation 7 was taken as the objective function for each feature selection method. A *k*-nearest-neighbor (*k*-NN) rule was used as the classifier in the system

where five neighbors (k = 5) was chosen. In the following expression for the classification accuracy, which is the number of correctly classified events normalized by the total number of events,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

where the values TP, TN, FP, and FN are numbers of true positives, true negatives, false positives, and false negatives, respectively. Positives indicate epileptic activity, and negatives denote non-epileptic activity. For each set of data, using the classification accuracy as an objective function necessitates separating the feature vectors and classification vector into training and testing sets for the k-NN. The counts for TP, TN, FP, and FN were determined after a testing set was classified and compared to the corresponding a priori classification vector of the testing set.

In the training set, a balanced proportion of values from the epileptic (1) and non-epileptic (0) classes was used because training a classifier with a balanced set provides a "guaranteed" measure of future classification accuracy, whereas training an unbalanced classifier may prove much better or much worse if the testing set were unbalanced. A 95% confidence interval for the accuracy was approximated using Equation 8 after thirty Monte Carlo simulations were run. Each simulation randomly drew a different training and testing data set, which was used for each selection method, before the accuracy was calculated and recorded. In Equation 8, $\bar{x}$, s, and n are the mean, standard deviation, and number of the recorded accuracies, respectively.

$$CI_{95\%} = \bar{x} \pm 2 \cdot \frac{s}{\sqrt{n}} \tag{8}$$

After running the Monte Carlo simulations, the best feature set of any size was determined and then used on a second testing set for classifying epileptic and non-epileptic voxels in each phantom dataset. The testing set was generated by randomly selecting voxels from each of the know classes, i.e. epileptic and non-epileptic. An ANOVA analysis was then performed to determine whether or not the choice of feature selection algorithm significantly affected the performance of the selected feature set and the ability of the chosen features to accurately classify the voxels. Figures 8 shows box plots of the classification accuracy versus the feature selection method for each of the fifteen phantom datasets. It is clear from these plots that no one method outperforms the others in all cases, but there are certain methods that substantially outperform others in certain cases. For instance, SBS provides greater accuracy than the other methods in subject P1, but for subject P11, all methods have relatively even performance. No matter which technique is chosen, all methods provide high accuracy in classifying epileptic and non-epileptic voxels.

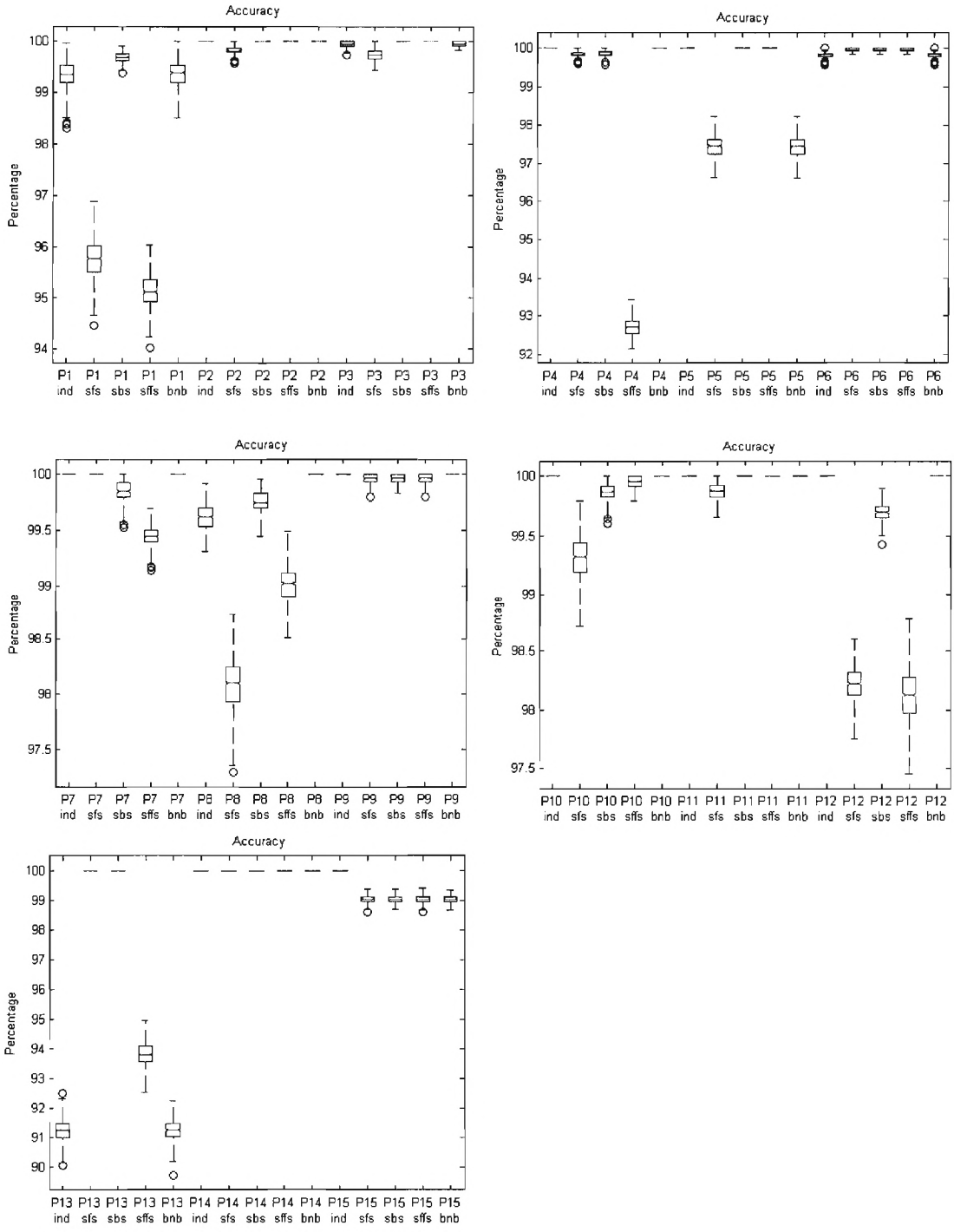

Figure 8 This figures shows that the accuracy differed statistically in some subjects depending on the feature selection method. In all cases ind = individual feature selection, sfs = sequential forward selection, sbs = sequential backward selection, sffs = sequential floating forward selection, and bnb = branch & bound.

REFERENCES

[1] G. K. Aguirre, E. Zarahn, and M. D'Esposito, "The Variability of Human, BOLD Hemodynamic Responses," *NeuroImage*, vol. 8, pp. 360-369, 1998.

[2] H.-L. Liu, Y. Pu, L. D. Nickerson, Y. Liu, P. T. Fox, and J.-H. Gao, "Comparison of the temporal response in perfusion and BOLD-based event-related functional MRI," *Magnetic Resonance in Medicine*, vol. 43, pp. 768-772, 2000.

[3] D. A. Handwerker, J. M. Ollinger, and M. D'Esposito, "Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses," *NeuroImage*, vol. 21, pp. 1639-1651, 2004.

[4] J. A. Detre and J. Wang, "Technical aspects and utility of fMRI using BOLD and ASL," *Clinical Neurophysiology*, vol. 113, pp. 621-634, 2002.

[5] K. J. Friston, P. Jezzard, and R. Turner, "Analysis of functional MRI time-series," *Human Brain Mapping*, vol. 1, pp. 153-171, 1994.

[6] G. M. Boynton, S. A. Engel, G. H. Glover, and D. J. Heeger, "Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1," *Journal of Neuroscience*, vol. 16, pp. 4207-4221, 1996.

[7] J. Kim, J. Whyte, J. Wang, H. Rao, K. Z. Tang, and J. A. Detre, "Continuous ASL perfusion fMRI investigation of higher cognition: Quantification of tonic CBF changes during sustained attention and working memory tasks," *NeuroImage*, vol. In Press, Corrected Proof, 2005.

[8] B. A. Ardekani, A. H. Bachman, and J. A. Helpern, "A quantitative comparison of motion detection algorithms in fMRI," *Magnetic Resonance Imaging*, vol. 19, pp. 959-963, 2001.

[9] K. J. Friston, J. Ashburner, C. D. Frith, J. B. Poline, J. D. Heather, and R. S. J. Frackowiak, "Spatial registration and normalization of images," *Human Brain Mapping*, vol. 3, pp. 165-189, 1995.

[10] K. J. Friston, S. Williams, R. Howard, R. S. J. Frackowiak, and R. Turner, "Movement-related effects in fMRI time-series," *Magnetic Resonance in Medicine*, vol. 35, pp. 346-355, 1996.

[11] K. J. Friston, A. P. Holmes, K. J. Worsley, J. P. Poline, C. D. Frith, and R. S. J. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapping*, vol. 2, pp. 189-210, 1994.

[12] R. S. J. Frackowiak, K. J. Friston, C. Frith, R. Dolan, C. J. Price, S. Zeki, J. Ashburner, and W. D. Penny, *Human Brain Function*, 2nd ed: Academic Press, 2003.

[13] H. Gudbjartsson and S. Patz, "The rician distribution of noisy MRI data," *Magnetic Resonance in Medicine*, vol. 34, pp. 910-914, 1995.

[14] A. M. Wink and J. B. T. M. Roerdink, "BOLD Noise Assumptions in fMRI," *International Journal of Biomedical Imaging*, vol. 2006, Article ID 12014, pp. 1-11, 2006.

[15] R. D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *Image Processing, IEEE Transactions on*, vol. 8, pp. 1408-1419, 1999.

[16] D. R. Pickens, Y. Li, V. L. Morgan, and B. M. Dawant, "Development of computer-generated phantoms for FMRI software evaluation," *Magnetic Resonance Imaging*, vol. 23, pp. 653-663, 2005.

[17] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2001.

[18] R. Esteller, J. Echauz, T. Tcheng, B. Litt, and B. Pless, "Line length: An efficient feature for seizure onset detection," in *Proceedings of the 23th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2001*, pp. 1707-1710, 2001.

[19] R. Esteller, G. Vachtsevanos, J. Echauz, T. Henry, P. Pennell, C. Epstein, R. Bakay, C. Bowen, and B. Litt, "Fractal dimension characterizes seizure onset in epileptic patients," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing 1999*, vol. 4, pp. 2343 – 2346, 1999.

[20] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echauz, and B. Litt, "Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients," *Biomedical Engineering, IEEE Transactions on*, vol. 50, pp. 603-615, 2003.

[21] E. N. Bruce, *Biomedical Signal Processing and Signal Modeling*. New York: Wiley, 2001.

[22] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt, "A comparison of waveform fractal dimension algorithms," *Circuits and Systems I: Fundamental Theory and Applications,*

IEEE Transactions on [see also Circuits and Systems I: Regular Papers, IEEE Transactions on], vol. 48, pp. 177-183, 2001.

[23] M. J. Katz, "Fractals and the analysis of waveforms," *Computers in Biology and Medicine*, vol. 18, pp. 145-156, 1988.

[24] X.-S. Zhang, Y.-S. Zhu, N. V. Thakor, and Z.-Z. Wang, "Detecting ventricular tachycardia and fibrillation by complexity measure," *IEEE Transactions on Biomedical Engineering*, vol. 46, pp. 548-555, 1999.

[25] P. Somol, P. Pudil, and J. Kittler, "Fast branch & bound algorithms for optimal feature selection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, pp. 900-912, 2004.