Published in final edited form as:
Neuroimage. 2011 September 15; 58(2): 537–548. doi:10.1016/j.neuroimage.2011.06.043.

TWave: High-Order Analysis of Functional MRI

Michael Barnathan[a], **Vasileios Megalooikonomou**[a,b], **Christos Faloutsos**[c], **Scott Faro**[d], and **Feroze B. Mohamed**[d]

Michael Barnathan: mbarnath@temple.edu; Vasileios Megalooikonomou: vasilis@temple.edu; Christos Faloutsos: christos@cs.cmu.edu; Scott Faro: scott.faro@tuhs.temple.edu; Feroze B. Mohamed: feroze.mohamed@tuhs.temple.edu

[a]Data Engineering Laboratory, Center for Information Science and Technology, Temple University, Philadelphia, USA

[b]Department of Computer Engineering and Informatics, University of Patras, Patras, Greece

[c]School of Computer Science, Carnegie Mellon University, Pittsburgh, USA

[d]Department of Radiology, Temple University School of Medicine, Philadelphia, USA

Abstract

The traditional approach to functional image analysis models images as matrices of raw voxel intensity values. Although such a representation is widely utilized and heavily entrenched both within neuroimaging and in the wider data mining community, the strong interactions among space, time, and categorical modes such as subject and experimental task inherent in functional imaging yield a dataset with “high-order” structure, which matrix models are incapable of exploiting. Reasoning across all of these modes of data concurrently requires a high-order model capable of representing relationships between all modes of the data in tandem. We thus propose to model functional MRI data using *tensors*, which are high-order generalizations of matrices equivalent to multidimensional arrays or data cubes. However, several unique challenges exist in the high-order analysis of functional medical data: naïve tensor models are incapable of exploiting spatiotemporal locality patterns, standard tensor analysis techniques exhibit poor efficiency, and mixtures of numeric and categorical modes of data are very often present in neuroimaging experiments. Formulating the problem of image clustering as a form of Latent Semantic Analysis and using the WaveCluster algorithm as a baseline, we propose a comprehensive hybrid tensor and wavelet framework for clustering, concept discovery, and compression of functional medical images which successfully addresses these challenges. Our approach reduced runtime and dataset size on a 9.3 GB finger opposition motor task fMRI dataset by up to 98% while exhibiting improved spatiotemporal coherence relative to standard tensor, wavelet, and voxel-based approaches. Our clustering technique was capable of automatically differentiating between the frontal areas of the brain responsible for task-related habituation and the motor regions responsible for executing the motor task, in contrast to a widely used fMRI analysis program, SPM, which only detected the latter region. Furthermore, our approach discovered latent concepts suggestive of subject handedness nearly 100x faster than standard approaches. These results suggest that a high-order model is an integral component to accurate scalable functional neuroimaging.

Address correspondence to: Vasileios Megalooikonomou, Data Engineering Laboratory, Center for Information Science and Technology, Temple University, 1805 N. Broad St., Philadelphia, PA, (215)-204-5774 (tel.), (215)-204-5082 (fax), vasilis@temple.edu.

Keywords

Tensors; Wavelets; Clustering; Latent Semantic Analysis; Parallel Factor Analysis; fMRI

1. INTRODUCTION

The traditional approach to data representation utilizes a matrix structure, with observations in the rows and features in the columns. Although this model is appropriate for many datasets, it is not always a natural representation because it assumes the existence of a single target variable and lacks a means of modeling dependencies between other features. Additionally, such a structure assumes that observed variables are scalar quantities by definition. This assumption may not be valid in certain domains, such as diffusion tensor imaging, where higher-order features predominate, or in domains which have strong spatiotemporal components, such as functional MRI.

Traditionally, these problems have been solved by reducing the features to scalars and fitting the dataset to a matrix structure. However, as well as potentially losing information, this strategy also employs a questionable approach from a philosophical standpoint: attempting to fit the data to an imprecise model rather than attempting to accurately model the existing structure of the data. Finally, while it may be possible to model dependencies between features by repeating the methodology multiple times, each with a different target variable, this yields suboptimal performance and may not be computationally feasible when real-time performance is required or when the dataset is very large.

To address these issues, we propose to model such datasets using *tensors*, which are generalizations of matrices corresponding to *r*-dimensional arrays, where *r* is known as the *order* of the tensor. Using a combination of wavelet and tensor analysis tools, we propose a novel framework for summarization, clustering, concept discovery, and compression of high-order datasets, which we call TWave (Barnathan *et al.*, 2010). Applying our technique to analysis of a large real-world digital opposition fMRI dataset, we compare the performance of TWave against voxelwise, SVD-based, wavelet-only, and tensor-only techniques and demonstrate our TWave method achieves superior results and reduces computation time vs. competing methodologies such as Latent Semantic Analysis.

Our approach has several advantages:

- Compression of the data model through grid quantization and wavelet preprocessing (inherent in the WaveCluster algorithm).
- Exploitation of spatial neighborhoods and local patterns.
- Efficiency up to two orders of magnitude faster than naive tensor approaches.
- The ability to identify noncontiguous clusters based on patterns in the projected space.
- Naturally fuzzy clustering based on similarities to discovered concepts.
- The projected space may reveal latent dataset concepts (our method revealed information about subject handedness in our dataset).

2. BACKGROUND

2.1 Tensor Tools

Tensors are defined within the context of data mining as multidimensional arrays. The number of indices required to index the tensor is referred to as the *order* of the tensor, while each individual dimension is referred to as a *mode*. The number of elements defined on each mode is referred to as the mode's *dimensionality*. The dimensionality of a tensor is written in the same manner as the dimensionality of a matrix; for example, 20×50×10.

Tensors represent generalizations of scalars, vectors, and matrices, which are tensors of orders 0, 1, and 2, respectively. Tensors and the notion of order are illustrated in Table 1.

An important operation applicable to our analysis is the *tensor product*. This product generalizes not from matrix multiplication, but from the Kronecker product operation defined on matrices, which is given as follows:

Given an $m \times n$ matrix $\mathbf{A}$ and a $p \times q$ matrix $\mathbf{B}$, the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is defined by the following $mp \times nq$ block matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{1,1}\mathbf{B} & \cdots & a_{1,n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m,1}\mathbf{B} & \cdots & a_{m,n}\mathbf{B} \end{pmatrix}$$

The tensor product is similar, but the result is another tensor rather than a block matrix. Specifically, if given order r and s tensors $\mathcal{A}$ and $\mathcal{B}$, their tensor product $\mathcal{A} \otimes \mathcal{B}$ is a tensor of order $r + s$ defined as follows:

$$(\mathcal{A} \otimes \mathcal{B})_{i_1, i_2, \ldots, i_r, j_1, j_2, \ldots, j_s} = \mathcal{A}_{i_1, i_2, \ldots, i_r} * \mathcal{B}_{j_1, j_2, \ldots, j_s}$$

For example, the procedure of taking a tensor product is shown in Figure 1, with arrows representing the direction of multiplication:

The operation known as the **Khatri-Rao product** is useful in the computation of several tensor decompositions and is defined in terms of the Kronecker product. Let $\mathbf{A}$ be a $p \times n$ matrix and $\mathbf{B}$ be a $q \times n$ matrix. Their Khatri-Rao product $\mathbf{A} \odot \mathbf{B}$ is as follows:

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1, \mathbf{a}_2 \otimes \mathbf{b}_2, \ldots, \mathbf{a}_n \otimes \mathbf{b}_n]$$

Singular value decomposition (SVD) is a unique matrix factorization by which an $m \times n$ matrix is decomposed into two projection matrices and a core matrix, as follows:

$$\mathbf{A} = \mathbf{U} \times \sum \times \mathbf{V}^{\mathbf{T}}$$

where $\mathbf{A}$ is an $m \times n$ matrix, $\mathbf{U}$ is an $m \times r$ column-orthonormal projection matrix, $\mathbf{V}$ is an $n \times r$ column-orthonormal projection matrix, and $\mathbf{\Sigma}$ is a diagonal $r \times r$ *core matrix*, where r is the rank of the projection.

Singular value decomposition has a wide variety of applications: for example, truncation of the SVD coefficients provides an optimal low-rank approximation (i.e. minimizes the

Frobenius norm). This indicates a close relationship between Principal Component Analysis (PCA) and SVD.

SVD is also used to discover the rank of a matrix, find the pseudoinverse, and solve least squares minimization problems. Additionally, the solution to SVD may be used in an unsupervised summarization technique known as Latent Semantic Analysis (LSA) (Deerwester *et al*., 1999). In this technique, **A** is treated as a term-document matrix. Here, singular value decomposition automatically derives a user-specified number of latent *concepts* from the given terms which form a basis for the rows and columns of the matrix. The projection matrices **U** and **V** then contain term-to-concept and document-to-concept similarities, respectively. Thus, SVD can be used to provide simple yet powerful automatic data summarization. This technique may be naturally viewed as a form of co-clustering, in which the rows and columns of a matrix cluster to the same space. An alternative graphical interpretation exists, in which clusters represent shared "waypoints" through which edges pass between vertices. Use of the eigendecomposition or SVD is also common in a graphical context, where it is known as *spectral graph theory*; here a common technique is to cluster on the eigenvector corresponding to the second smallest eigenvalue of the Laplacian matrix, thereby partitioning vertices along edges which are likely to be minimal cuts. This technique is known as *Fiedler retrieval*. It is also possible to project new query vectors into the space defined by the SVD, known as *folding in*; this enables recommendation as the query projects to the same space as both the rows and columns and can be assessed using a distance metric.

The natural extension of singular value decomposition to tensors is known as *high-order singular value decomposition*, or HOSVD. This decomposition, in turn, is a special case of the *Tucker decomposition*, which is capable of concurrent data co-clustering across every mode of a tensor. Formally, let $\mathcal{A}$ be a tensor of order r; i.e. $\mathcal{A} \in \Re^{d_1 \times d_2 \times \ldots \times d_r}$. We may then define the Tucker decomposition as the following factorization into a *core tensor* and a product of r projection matrices:

$$\mathcal{A} = \mathcal{G} \times \mathbf{U}_1 \times \mathbf{U}_2 \times \ldots \times \mathbf{U}_r$$

Note that while either the core tensor must be diagonal or the projection matrices must be column-orthonormal, the Tucker decomposition does not guarantee that both conditions are simultaneously true. When the projection matrices are unitary, the factorization is called high-order singular value decomposition.

When used as a data summarization technique, the Tucker decomposition exhibits similar behavior to singular value decomposition. Specifically, the core tensor's elements represent the strengths of the discovered concepts (in terms of variance captured), while the projection matrices each represent the strength of the individual term-to-concept relationships on their corresponding modes.

Although the Tucker decomposition generalizes singular value decomposition, evaluating it requires computing the norm of $\mathcal{A}$'s covariance matrix. This can come at a memory cost of $\Omega(n^2)$, which, for large datasets such as ours, may be prohibitive.

Fortunately, another method exists that does not suffer from this problem, called both parallel factor analysis (PARAFAC) and canonical decomposition (CANDECOMP) due to simultaneous discovery of the method in 1970 by Harshman (1970) and Carrol and Chang (1970). In this paper, we will refer to the method as PARAFAC.

PARAFAC (Harshman, 1970) is a generalization of PCA and forms the basis of our tensor analysis approach. Given a user-specified number of concepts c, PARAFAC decomposes an

order-*r* tensor $\mathcal{A}$ into a columnwise sum of the tensor product of *r* projection matrices, denoted $\mathbf{U}^{(1)} \ldots \mathbf{U}^{(r)}$. Formally, we define the decomposition as follows:

$$\mathcal{A} = \sum_{i=1}^{c} \lambda_i \mathbf{U}_{1,i}^{(1)} \otimes \mathbf{U}_{1,i}^{(2)} \otimes \ldots \otimes \mathbf{U}_{1,i}^{(r)}$$

Where the **U** matrices represent projection matrices containing mode-to-concept similarities and λ represents a *c*-element scaling vector, in which each element represents the strength of a concept. The notation $\mathbf{U}_{1,i}$ refers to the entire *i*th column of **U**. The number of columns in each projection matrix will therefore be equal to the number of user-specified concepts and the number of rows in each individual projection matrix equal to the number of observations made on the corresponding mode of the tensor.

Both the Tucker and PARAFAC decompositions may be computed using alternating least squares (ALS) (Sands and Young, 1980). Though the computation of the tensor decompositions we use is outside of the scope of this paper, the procedure for computing PARAFAC using ALS is shown below:

1. Given an order- *r* tensor $\mathcal{A}$, declare a set of projection matrices $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, …, $\mathbf{U}^{(r)}$.
2. Let $i = 1$.
3. Holding all other matrices constant, solve the following equation for $\mathbf{U}^{(i)}$:

$$\mathbf{U}^{(i)} = \mathcal{A}^{(i)} \left(\sum_{j=1..r \wedge j \neq i} \mathbf{U}^{(j)} \right) \left(\prod_{j=1..r \wedge j \neq i} [\mathbf{U}^{(j)}]^T \mathbf{U}^{(j)} \right)^*$$

   Where ⊙ represents the *n*-ary Khatri-Rao product, * represents the Moore-Penrose pseudoinverse, and $\mathcal{A}^{(i)}$ represents $\mathcal{A}$ matricized (Kolda and Bader, 2007) on mode *i*.
4. Repeat for all from 1 to *r* until convergence is attained.

The resulting PARAFAC decomposition is illustrated in Figure 2.

2.2 Independent Component Analysis

While principal component analysis results in a projection that is uncorrelated, there is no guarantee of statistical independence. The orthogonal nature of the basis is often problematic in spatiotemporal scenarios such as neuroimaging. Independent component analysis (ICA) is a similar technique that assumes independence of the discovered components (Comon, 1994). As this is an important assumption in many applications, such as naïve Bayes classification, ICA can be applied to a wide variety of areas. Of course, ICA itself performs best when the underlying dataset meets this assumption of independence.

ICA is typically preceded by the preprocessing steps of centering (subtracting out the mean) and whitening (equalization of the power spectral density of the dataset, resulting in “white noise”). Though SVD is not used directly within the ICA algorithm, it is often used as a preprocessing step to make the problem of feature extraction easier for ICA to solve.

ICA, like PCA, is an example of a “blind signal separation” technique; that is, it can separate signals (or more generally, perform factor analysis) without preconceptions of the meaning of the signals’ components (or other prior domain knowledge). High-order ICA methods exist, with pioneering work in the field performed by (De Lathauwer *et al.*, 2000; Calhoun

and Pekar, 2000; Vasilescu and Terzopoulos, 2005), and have recently become widely accessible within the neuroimaging domain through implementations such as The Group ICA of fMRI Toolbox (Calhoun *et al.*, 2002) and MELODIC (Beckmann and Smith, 2004). We later compare a MELODIC analysis to our own clustering results.

2.3 Wavelets

Wavelets are spectral analysis tools capable of emphasizing patterns in scale space and exploiting spatiotemporal correlations between neighboring voxels. Wavelets have been applied to a wide variety of domains, including time-series analysis (Chan and Fu, 1999), musical analysis (Evangelista, 2001), and computer-aided diagnosis (Dinov *et al*., 2005). Among the properties of wavelets recently studied in functional MRI analysis are frequency-adaptive wavelet shrinkage for noise removal and compression (Dinov *et al.*, 2005; Wink *et al.*, 2008), multiresolution analysis (Desco *et al.*, 2001), reconstruction variance estimation (Aston *et al.*, 2005), and the use of wavelets with low-order dimensionality reduction techniques, such as PCA and ICA (Chen *et al.*, 2003; Roussos *et al.*, 2005; Gupta and Jacobson, 2006), where they are capable of introducing spatial information into typically neighborhood-insensitive dimensionality reduction approaches. Within the context of fMRI, (Ruttimann *et al.*, 1998) and (Turkheimer *et al.*, 1999) pioneered the use of wavelet decomposition as a method of statistical testing between images acquired under varying experimental conditions, yielding a significant increase in signal to noise ratio, which is in concordance with the results of our subject-based classification experiments. (Fadili and Bullmore, 2002) demonstrated the approximate optimality of wavelet-generalized least squares as unbiased linear estimators for solving least squares problems in the presence of long-memory noise, (Wink *et al.*, 2008) introduced the concept of wavelet shrinkage as a denoising technique, and (Van De Ville *et al.*, 2004) introduced the notion of wavelet-based statistical testing in the spatial domain, which we make use of.

Wavelet analysis permits decomposition into scaled and translated copies of a function known as the “mother wavelet”, denoted ψ. This carries the additional advantage of localization in space as well as frequency. Furthermore, it allows multiresolution analysis through variation of the scale parameter. These “daughter wavelets” can be obtained from the continuous wavelet transform, which is given by the following formula:

$$\gamma(s,t)=\frac{1}{\sqrt{|s|}}\int_{-\infty}^{\infty}\Psi\left(\frac{w-t}{s}\right)^{*}f(x)dx$$

Where * denotes the complex conjugate operation, *s* denotes the scaling parameter, *t* denotes the translation parameter, ψ denotes the mother wavelet, and f(x) denotes the original signal.

Fig. displays an example of a single functional MRI slice (with spatial clusters highlighted) and the linearized 4D multilevel wavelet decomposition of the volume of which this slice is a part.

However, many applications within the realm of machine learning and computing in general are discrete, in which case the continuous wavelet transform shown above is inappropriate. In such cases, we may use the *discrete wavelet transform*, which convolves a signal x with a low and high pass filter at each level (L[n] and H[n], respectively) and downsamples the signal by a factor of 2, as shown in Figure 4.

2.4 WaveCluster

WaveCluster (Sheikholeslami *et al*., 2000) is a grid and density-based clustering technique that is unique for performing clustering directly in wavelet space. Specifically, the WaveCluster algorithm:

- Quantizes data to a grid, using the counts of each grid cell in place of the original data.
- Applies a wavelet transformation using a hat-shaped wavelet (such as the (2,2) or (4,2) biorthogonal wavelets), retaining the approximation coefficients and emphasizing regions in which points cluster.
- Thresholds cells in the transformed space. Cells with values above a user-specified density threshold are considered “significant”.
- Applies a connected component algorithm to the significant cells to discover and label clusters.
- Maps the cells back to the original data using a lookup table built during quantization.

WaveCluster is optimally efficient [$O(n)$] as well as very fast in practice, can discover the boundaries of complex shapes, and natively supports multiresolution analysis. However, WaveCluster is primarily a spatial clustering method and is difficult to adapt to temporal clustering or other more complex clustering domains. Due to the use of a connected component algorithm, the algorithm will not operate on categorical modes or in other scenarios in which meaningful neighborhood relationships may not exist. Additionally, we observed a great deal of sensitivity to the cell significance threshold parameter in our experiments. More fundamentally, the WaveCluster algorithm is not capable of identifying clusters with large discontinuities. An illustration of the WaveCluster algorithm with an example threshold of 2 is shown in Figure 5.

2.5 Survey

Tensor decompositions were introduced by Hitchcock with his seminal paper (Hitchcock, 1927), but did not receive attention until the work of Tucker (Tucker, 1963) and Harshman (Harshman, 1970). Tensor decompositions have found many applications recently: Vasilescu and Terzopoulos, 2002 used Tucker decomposition in computer vision, Kolda and Bader, 2006 in web link analysis, and Savas, 2003 for handwritten digit recognition. Sun *et al*., 2006 presented a method that performs tensor decompositions in an incremental way and in (Sun *et al.*, 2008), a method that discovers spatio-temporal patterns. A comprehensive overview of the theory and the applications of tensor mining can be found in (Kolda and Bader, 2007).

Wavelets are widely used in signal processing (Mallat, 1999). They have an important advantage over the Fourier transform, which is the reason that we use them in our TWave method: they can provide information of when each frequency appears. Several different families of wavelets exist. We have evaluated Haar (Mallat, 1999), Daubechies (Daubechies, 1992), and CDF (Daubechies, 1992) wavelets in the context of our TWave methodology.

Thus, spikes are easily identified through the spectrum. Wavelets can be computed in a streaming way (Gilbert *et al*., 2003), giving us the opportunity to adapt our method to streaming scenarios as well.

3. MATERIALS AND METHODS

3.1 Experimental Setup

3.1.1 Datasets and Acquisition—We analyzed the TWave, pure tensor, pure wavelet, SVD, and voxelwise approaches on a high-order digital opposition fMRI dataset consisting of 11 subjects performing 4 simple motor tasks: left finger-to-thumb, left squeeze, right finger-to-thumb, and right squeeze (Khorrami and Faro, in press). We evaluated our experiments using leave-one-out cross-validation rather than partitioning into a separate test set due to the small number of subjects. Acquisition of the fMRI dataset took place using one scanner and one common set of acquisition parameters. Data was acquired from each subject over 120 time points, each 3 seconds long. The period of each task was 10 time points (or 30 seconds); this periodicity was easily visualized in the raw time series as well as via power spectral density.

Each volume after standard preprocessing (realignment, normalization, and smoothing, described in section 3.1.4) consisted of 79×95×69 voxels. Thus, the dataset was most easily represented as a 6^{th} order tensor of dimensionality 79×95×69×120×4×11, of which the first four modes were spatiotemporal and the remaining two (task and subject) were categorical. Our total dataset's size was thus 2,734,221,600 voxels, occupying approximately 9.3 GB of space.

3.1.2 fMRI Task—Nine right handed and two left handed healthy subjects participated in this study. The average subject age was 29.7 years. All subjects were assessed for any psychological or neurological problems and provided informed consents after listening to a detailed explanation of the scanning process. Handedness was defined using the inventory proposed by Coren, 1993.

The subjects were asked to perform dominant and non-dominant hand finger to thumb movement during the activation period. A block design fMRI experiment was performed and the subjects were presented with 6 blocks of activation and rest conditions. During each block (30 sec), 10 volumes of echo planar images (EPI) were acquired (thus the temporal resolution of our study was 3 seconds). All of the active conditions were interspersed with rest conditions of equal duration. The task lasted for six minutes. The subjects were instructed to start and stop the movements upon hearing the "START" and "STOP" commands. They were asked to move their hand at their own pace and keep their eyes closed during the scanning process. This task was chosen in this study since it produces robust and well understood fMRI activation patterns.

3.1.3 MRI Acquisition—Scanning was done using a 1.5 Tesla General Electric Imager. Initially T1-weighted axial structural images were acquired covering the entire brain parallel to the AC-PC line (Talairach and Tournoux, 1998). The structural imaging parameters were: TR = 700 msec, FOV = 22 cm, slices = 26, slice thickness = 5mm. Then fMRI images were acquired with an echo planar gradient echo (EPI) pulse sequence at the same locations as the structural images. The imaging parameters were: TR = 3 sec, TE=50ms, FOV = 22 cm, slices = 26, slice thickness = 5mm. The standard preprocessing methods of realignment, normalization, and smoothing were performed as described in (Khorrami *et al.*, 2011). Images are represented in radiological coordinates in this paper.

3.1.4 Image Preprocessing—Images were initially inspected to ensure freedom from substantial head movements and structural abnormalities. The post-acquisition preprocessing and statistical analyses were performed using SPM2 (Statistical Parametric Mapping, Welcome Department of Cognitive Neurology, University College of London, UK), running under the Matlab® (The Mathworks, Inc., Natick, MA) environment. Images were

converted from the DICOM format into the ANALYZE (AnalyzeDirect, Inc., Lenexa, KY) format adopted in the SPM package. A three dimensional automated image registration routine (six-parameter rigid body, sinc interpolation; second order adjustment for movement) was applied to the volumes to realign them with the first volume of the time series, which was used as a spatial reference. All functional and anatomical volumes were then transformed into the standard anatomical space using the T2 EPI template and the SPM normalization procedure (Friston *et al*., 1995). This procedure uses a sinc interpolation algorithm to account for brain size and position with a 12 parameter affine transformation, followed by a series of non-linear basis function transformations. Seven, eight, and seven nonlinear basis functions were applied in the x, y, and z directions, respectively, with 12 nonlinear iterations to correct for morphological differences between the template and given brain volume. Next, all volumes underwent spatial smoothing by convolution with a Gaussian kernel of 8 cubic mm full width at half maximum (FWHM) (Friston *et al*., 1995), to increase the signal-to-noise ratio (SNR). Per standard practice in neuroimaging, our kernel size was chosen to be approximately twice our original voxel size of 3.5 × 3.5 × 5 mm.

3.2 Analysis Methods

3.2.1 Overview—Our methodology makes use of both wavelets and tensors. Because spatiotemporal data tends to exhibit a high degree of spatial locality, the spatiotemporal modes of the dataset are first preprocessed using an *m*-dimensional discrete wavelet transform (obtained through cascading), where *m* is the number of spatiotemporal modes. We utilized the Daubechies-4 wavelet in all techniques except for clustering. Clustering was performed using the (2,2) and (4,4) biorthogonal wavelets, which emphasize dense regions of points while simultaneously suppressing sparse neighborhoods. We then linearized the wavelet coefficients to form a vector representing all spatiotemporal voxels in the dataset, reducing the order of the tensor by $m - 1$ and improving performance of the techniques. Without using wavelets, such a linearization technique would fail to capture spatial locality patterns and would destroy vital neighborhood information in the dataset. Furthermore, PARAFAC and other decompositions are fundamentally unaware of spatiotemporal neighborhood relations even in the absence of linearization (a property which allows them to be used on categorical data, but which makes them less useful for any sort of voxelwise image data). However, encoding the wavelet coefficients circumvents this issue, as each wavelet coefficient is the result of a convolution on its *m*-dimensional neighborhood. Thus no false neighborhoods are created when the dataset is linearized – all neighborhood information is already captured in the wavelet coefficients and PARAFAC will not discover more. The primary purpose of wavelet analysis in our framework is thus to make the metholodgy neighborhood-aware.

The remaining modes of the tensor are left in place after the spatiotemporal wavelet coefficients are linearized, resulting in an order $r - m + 1$ tensor. Per-subject mean fMRI volumes are subtracted from each image, as we found that task-specific deviations from the mean yielded improved separability on handedness. The columns were then centered to have means of 0, a standard preprocessing technique in matrix factorization, to avoid biasing the decomposition based on the columnwise voxel intensities of the image. PARAFAC was performed using alternating least squares and the resulting projection matrices (each representing similarities between a single mode of the tensor and a common set of latent concepts) were analyzed.

Because the concepts are held in common across all modes, so is the space onto which each mode is projected. This defines tensor-based concept discovery as a form of *co-clustering*, in which all modes of data are projected into the same space and associated with other modes

using distance metrics in the projected space. Thus the purpose of the tensor decomposition step is to integrate inter-modal reasoning and conceptual abstraction. The fewer the number of concepts, the greater the degree of abstraction but also the greater loss of dataset fidelity.

Affinities between modes of the decomposed tensor may thus be ranked directly against each other using a metric such as cosine or Tanimoto similarity – for instance, specific neighborhood patterns of activation associated with left-handedness may have strong links to left-handed subjects, corresponding to high cosine similarity in the concept space between wavelet coefficient-to-concept similarities on mode 1 and subject-to-concept similarities on mode 4.

This wavelet transformation and linearization step additionally allows us to (hard) threshold the discovered wavelet coefficients and store the results in a sparse matrix to achieve compression. This is known as wavelet shrinkage (Donoho *et al.*, 1995). We empirically demonstrate that it is possible to eliminate the majority of the coefficients in the dataset while preserving the meaning of the projection into the concept space.

Following calculation of the wavelet coefficients, the data is then converted to a tensor (which may also be stored in sparse form using the tensor toolbox). PARAFAC is then performed using alternating least squares and the resulting projection matrices are stored and analyzed. This method provides a general framework for further tensor and wavelet-based analysis, including data summarization, concept discovery, compression, clustering, and classification.

3.2.2 Other Methods—It is also possible to analyze data using wavelets and tensors alone, or by using neither preprocessing method (the voxelwise approach). Singular value decomposition run on the dataset in matrix representation additionally provides a benchmark for comparison of the tensor model and techniques.

Several low-order voxelwise and wavelet-based data summarization approaches exist, such as kernel density estimation, SVD, and data clustering. We chose to summarize the dataset in the voxelwise approach using *k*-means clustering and in the wavelet approach using WaveCluster (Sheikholeslami *et al.*, 2000).

3.2.3 Summarization—Data summarization was performed by direct inspection of the term-to-concept similarities discovered by PARAFAC (stored in the projection matrices $\mathbf{U}^{(1)}$ … $\mathbf{U}^{(r)}$). Both the tensor and TWave methods perform automatic data summarization through computation of term-to-concept similarities on each mode. Additionally, concept strengths themselves are represented by the corresponding elements of the λ vector of PARAFAC or the core tensor of the Tucker decomposition. Consequently, no postprocessing is necessary.

3.2.4 Concept Discovery—Concept discovery is the higher-order analogue to traditional latent variable methods in matrices, and as such is performed using high-order projections, such as PARAFAC, HOSVD, or Tucker. In TWave, these patterns are extracted using PARAFAC following wavelet decomposition, normalization, and recentering (i.e. subtracting the mean of the dataset). As shown later in Figure 6, patterns are often directly present in the projected space and further analysis following decomposition is often unnecessary. Each projection matrix $\mathbf{U}^{(i)}$ automatically discovers patterns along the *i*th mode of the dataset. However, as these methods are unsupervised, one key component in concept discovery (as with traditional PCA) is interpretation of the derived concepts.

Spatiotemporal concept discovery is also possible using clustering. Tensor and TWave methods utilize a different approach to perform clustering: using the wavelet/voxel-to-concept similarities discovered by PARAFAC, TWave assigned each voxel/wavelet coefficient to the cluster corresponding to its strongest concept; i.e., the one to which it is most similar – information directly present within the decomposed matrices. In this way, PARAFAC represents an automatic fuzzy clustering on potentially noncontiguous voxels. This method has the innate advantage of modeling voxels that belong to multiple clusters by degree of membership in each. It is also capable of discovering voxels that do not belong to any clusters (those with very low similarities to all discovered concepts); these may represent outliers. Alone, this has the disadvantage of failing to capture spatial locality patterns. However, encoding the wavelet coefficients, as in the TWave method, avoids this issue, as the wavelet coefficients capture neighborhood information by nature. The clustered wavelet coefficients may be mapped back to the original volume using an inverse wavelet transformation.

3.2.5 TWaveCluster—We derive a clustering method for high-order data based on the WaveCluster algorithm (Sheikholeslami *et a*l., 2000), a grid and density-based method that utilizes quantization and wavelet transformation to identify significant cells, defined as cells whose count exceeds a user-specified density threshold. Unlike most clustering methods which operate on high-order data, our method operates on the high-order structure directly, without unfolding or linearization. “Hat-shaped” wavelets such as the (2,2) biorthogonal wavelet or the Cohen-Daubechies-Feauveau wavelet are used to emphasize the center of clustered regions while suppressing boundary information (Sheikholeslami *et a*l., 2000).

Following wavelet transformation, the approximation (LL) subband of the wavelet transformed space is used as the new grid and connected components are discovered among significant cells using one of several algorithms available in the literature. Each connected component is assigned a unique cluster label. The resulting quantized grid cells are then mapped back to the points that quantize to them and the points are assigned the cluster labels of the cells. This method has several advantages, including an innate potential for multiresolution analysis, robustness against noise, and an optimal runtime of $O(n)$ Moreover, both the quantization and analysis are easily parallelizable, offering the potential of an additional speedup.

As we have demonstrated that wavelets and tensors are complementary analysis tools in our development of the TWave methodology, WaveCluster represents a logical starting point for our own clustering methodology. As a grid-based algorithm which counts or weights points falling in each grid cell, the WaveCluster algorithm represents a spatiotemporally constrained version of the LSA problem, thus matrix methods are particularly applicable. We further extend WaveCluster to operate on high-order data modeled using tensors and sum the points in each cell by a weight rather than simply counting them to allow the algorithm to operate on real-valued rather than strictly binary data. We call our extended approach TWaveCluster. The primary distinction in our approach is the use of tensor decompositions such as Tucker or higher-order SDD in place of a connected component-finding algorithm. This carries a number of advantages: first, the concept discovery properties of these decompositions automatically provide a parallel clustering on all modes of the dataset, with concepts representing clusters. Different modes of the tensor (e.g. subject, task, time point, or spatial location) are thus directly comparable in the projected space, and each receive cluster similarity weights. Thus it is possible to map clusters to meanings such as “activity 30 seconds into a task in the right frontal region of left handed subjects” – an analysis which is impossible in traditional low-order clustering methods. Furthermore, these methods are capable of automatically identifying spatially distant cluster components based on similarity to discovered concepts. These methods also provide a

naturally fuzzy clustering, due to the presence of concept-to-mode similarities, and automatically provide a useful measure of clustering quality: the proportion of the dataset variance captured by the cluster, directly represented by the strength of the concept corresponding to that cluster within the core tensor of the decomposition ($\mathcal{G}$ for Tucker, λ for PARAFAC). As the concepts are identical across all modes, discovered clusters may even extend spatially along multiple modes of the tensor, highlighting potential associations between as well as within modes.

The first few steps of our algorithm remain identical to WaveCluster:

- Quantize data to a grid, using the count of each grid cell in place of the original data.
- Apply a wavelet transformation using a hat-shaped wavelet (such as the (2,2) or (4,2) biorthogonal wavelet), retaining the approximation coefficients.
- Threshold cells in the transformed space. Cells with values above a user specified density threshold are considered “significant”.

However, the remaining steps in our algorithm differ:

- Model significant cells as a tensor $\mathcal{X} \in \Re^{d_1 \times d_2 \times \ldots \times d_r}$.
- For a user-specified k, run a k-concept PARAFAC analysis on $\mathcal{X}$:

$$\mathcal{X} = \sum_{i=1}^{k} \lambda_i \mathbf{U}_{1,i}^{(1)} \otimes \mathbf{U}_{1,i}^{(2)} \otimes \ldots \otimes \mathbf{U}_{1,i}^{(r)}.$$

- For each c from 1 to k, recompose a tensor using only column c of each projection matrix. Omit λ_c: we are interested in concept similarities, not strengths. The resulting tensor $\mathcal{X}_c$ contains voxel similarities to concept *c*:
  $\mathcal{X}_c = \mathbf{U}_{1,c}^{(1)} \otimes \mathbf{U}_{1,c}^{(2)} \otimes \ldots \otimes \mathbf{U}_{1,c}^{(r)}$
- Assign every voxel the cluster label of the concept to which it is most similar:
  $(\forall x \in \mathcal{X}) \mathcal{L}_x = \arg \max_{1<c<k} (\mathcal{X}_c)_x$
- Threshold: take only the *s*% most similar voxels to each cluster concept, where *s* is a parameter chosen by the user. The higher the value of *s*, the fewer the number of voxels returned but the stronger their association with the underlying concept.

Our approach exhibits a number of advantages, including the ability to create a fuzzy clustering (where each voxel’s degree of membership in cluster *c* is determined by its similarity to concept *c* in the decomposed tensor), the ability to cluster noncontiguous voxels, and even the ability to discover clusters that extend across modes of the tensor. Our approach also has the advantages of efficiency, which is comparable to WaveCluster, and simple cluster validation, as the terms in the λ vector of the PARAFAC decomposition automatically represent the variance captured by each cluster!

4. RESULTS AND ANALYSIS

The methodologies described in the previous sections were applied to fMRI datasets acquired during dominant and non-dominant hand finger to thumb movement. Results of each method are enumerated below:

4.1 Discovered Concepts

Though the number of left-handed subjects in our dataset is small, we discovered strong separability between left and right handed subjects in a 2-concept PARAFAC analysis, which persisted for TWave but not for SVD. This pattern was made even more explicit when subtracting the subject means from each subject's set of images, suggesting that the task residuals discriminate better between left and right handed subjects than task activations biased by subjects' means (however, it should be noted that variance was much greater between subjects than among tasks performed by the same subject in this dataset). The results of TWave using the Daubechies-4 wavelet and mean subtraction are shown in Figure 6. A larger dataset with a more complex task (e.g. a cognitive task) would be beneficial to support these findings; however, they represent an important step towards discovering clinically significant neuroimaging concepts using high-order approaches.

4.2 Classification

k-nearest neighbor classification experiments were run, but due to the extremely long per-fold running times of naïve PARAFAC and Tucker decopositions and the difficulty of folding new observations into an existing high-order decomposition, cross-validated classification accuracies are not reported for the high-order techniques. Among low-order techniques, subject classification accuracies on the fMRI dataset reached 52% for voxelwise analysis and 98% for wavelet-based classification (using the Daubechies-4 wavelet). The Haar wavelet was also used with the wavelet and TWave classification approaches, but only yielded 82% classification accuracy in both contexts.

We were able to hard threshold up to the weakest 98% of wavelet coefficients (i.e. set these coefficients to zero) without any loss of subject or task classification accuracy, greatly improving time and space costs while preserving the discriminative power of the classifier. This held true in both the low and high-order cases. Further compression is possible in the decomposed tensor through truncation of weak concept similarities; however, this would necessarily take place after the tensor decomposition is performed and thus would not significantly affect the memory cost or speed of the analysis.

Task classification was more difficult because the intra-subject between-task variance (σ^2 = 179.29) was much less than the between-subject variance (σ^2 = 9066.85). Initial results yielded only 2% accuracy for voxelwise analysis and 27% accuracy for wavelet-based analysis. However, by subtracting the voxelwise mean of each subject across all tasks, we were able to improve classification accuracy to 34% for the voxelwise approach and 68% for the wavelet approach. From a medical standpoint, this suggests that residual task activations between scans of the same subject have higher discriminative power in motor task classification than overall activations with subject means included.

4.3 Clustering and TWaveCluster

We analyzed two subjects on all four spatiotemporal modes of the fMRI tensor using the *k*-means (*k*=4) and TWaveCluster (k=5, density threshold=85th percentile) approaches. Average running times for each individual volume were 53 seconds and 23 seconds, respectively. Discovered clusters using traditional approaches are shown in Figure 7, while TWaveCluster results are shown in Figure 8. A demarcation can be seen between the frontal and temporal regions of the brain in the TWaveCluster results; this distinction is less clear in *k*-means.

By contrast, SPM fails to recognize the frontal cluster at all at $\alpha = 0.05$, indicating that our method is discovering medically useful data which the leading technique in the field is failing to pick up. Additionally, MELODIC, though it recognizes the motor cluster

distinctly, considers the noisy regions at the corners of the volume highly significant (as opposed to TWaveCluster, which projected them into the least significant component) while deemphasizing the frontal task-related regions, which are more interesting from a clinical standpoint. As a native high-order technique (with a number of bundled preprocessing steps), MELODIC's runtime on an individual volume was orders of magnitude higher than TWaveCluster's, at 1175 seconds per volume vs. 53 seconds. By contrast, a PARAFAC analysis using alternating least squares on the raw tensor data completed in 785 seconds. Though this may be partially due to the NIFTI imaging format used by MELODIC, which inflated the 181 MB raw volume to 474 MB, the order of magnitude is consistent with the performance of pure high-order techniques in our other experiments. SPM and MELODIC results are shown in Figure 9.

The clusters discovered by TWaveCluster show a greater degree of symmetry and homogeneity than the *k*-means clusters. Moreover, the most significant voxels in the first two clusters spatially localize in the frontal and motor regions of the brain, involved with planning and executing the motor task, respectively. As is natural for PARAFAC, clusters are ranked by salience. Corresponding values in the λ vector were 28,812, 18,033, 15,958, and 11,366 for the four discovered clusters. Unlike partition-based algorithms such as *k*-means, the existing clusters will not significantly change as new clusters are added (and will not change at all if a globally optimal algorithm is used to compute the PARAFAC decomposition).

Based on the variance captured by the clustering and on visual inspection, we found that $k=4$ yielded good results on our dataset. Following clustering, we grouped each voxel's time series by cluster and generated an overall time series for the cluster by taking the mean of these series. The individual voxel and mean cluster time series representations are shown in Figures 10 and 11, respectively. The mean cluster time series may be used for anatomical segmentation and discovery of functional patterns, such as frontal lobe task-related motor learning.

4.4 Speed and Summary of Results

Classification runtime per fold was assessed for the voxelwise, wavelet-based, and TWave approaches on a dual-processor 2.2 GHz Opteron system with 4 GB of memory. The SVD and pure tensor approaches were measured on an 8 processor (16 core) 2.6 GHz Opteron supercomputer with 128 GB of memory. Despite running on a much more powerful system, the PARAFAC and SVD approaches still took significantly longer to complete than other approaches, as shown in Table 2.

Pure high-order methods were typically dominated by the computational cost of factorization, while low-order methods and TWave tended to be dominated by the lesser cost of wavelet extraction, or, in the voxelwise case, by classification, as shown in Table 3. This is expected, as each step of the methodology reduces the feature set size.

5. CONCLUSION

From these results, we may conclude that the combination of wavelets and tensors tools in the analysis of fMRI motor task datasets yields better performance in space, time, and accuracy than the voxelwise approach, as well as either wavelets or tensors alone. Specifically, a combined approach achieves benefits including the sensitivity to locality that wavelets provide and the ability to reason across modes using tensor techniques, but without the prohibitive space and time costs of using tensors constructed on raw voxels for analysis. Additionally, tensor tools provide powerful automatic data summarization techniques, as demonstrated through discovery of left-handed subjects in our dataset. Of medical

significance, our techniques were able to distinguish between left and right handed subjects and discovered a frontal cluster which exhibited motor habituation in time and which was not detected by SPM. Potential avenues for future research include use of different wavelet functions, extension of our methods to streaming and sparse tensor data, and applications to high-order datasets in other fields. We believe our method demonstrates that tensors provide an appropriate model for datasets with interactive or nonscalar features.

Acknowledgments

We would like to thank Charalampos Tsourakakis for assistance with the literature review and experiments.

This work was supported in part by the National Science Foundation under grants IIS-0237921, IIS-0705215, IIS-0705359, and the National Institutes of Health under grant R01 MH68066-05 funded by the National Institute of Mental Health, the National Institute of Neurological Disorders and Stroke, and the National Institute on Aging. All agencies specifically disclaim responsibility for any analyses, interpretations, or conclusions.

References

Aston JAD, Gunn RN, Brett M, Hinz R, Turkheimer FE. Wavelet Analysis of PET and fMRI Data. IEEE Transactions on Nuclear Science. 2001; 21:1–7.

Barnathan, M.; Megalooikonomou, V.; Faloutsos, C.; Faro, S.; Mohamed, F. TWave: A High-Order Framework for Classification, Clustering, and Concept Discovery. http://denlab.temple.edu/tool&data.htm

Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans on Medical Imaging. 2004; 23(2):137–152.

Calhoun JPekar V. Where and Where Are Components Independent? On the Applicability of Spatial and Temporal ICA to Functional MRI Data. NeuroImage. 2000; 11:682.

Calhoun VD, Adali T, Pearlson GD, van Zijl PC, Pekar JJ. Independent Component Analysis of FMRI Data in the Complex Domain. Magnetic Resonance and Medicine. 2002; 48:180–192.

Carroll JD, Chang J. Analysis of Individual Differences in Multidimensional Scaling via an n-way Generalization of ‘Eckart-Young’ Decomposition. Psychometrika. 1970; 35(3):283–319.

Chan, K.; Fu, AW. Efficient Time Series Matching by Wavelets. Proc. 15th Int’l Conf. on Data Eng; Sydney, Australia. 1999. p. 126-133.

Chen, CL.; Wu, TH.; Wu, YT.; Hsieh, JC.; Huang, YH.; Lee, JS. Proc. 1st International EMBS Conference on Neural Engineering; 2003. p. 557-560.

Comon P. Independent Component Analysis: A New Concept? Signal Processing. 1994; 36(3):287–314.

Coren S. Measurement of Handedness Via Self-report: the Relationship Between Brief and Extended Inventories. Percept Mot Skills. 1993; 76(3/1):1035–1042. [PubMed: 8321574]

Daubechies, I. Society for Industrial and Applied Mathematics. Philadelphia: Capital City Press; 1992. Ten Lectures on Wavelets.

Deerwester S, Dumais ST, et al. Indexing by latent semantic analysis. J Am Soc Inf Sci. 1999; 41(6): 391–407.

De Lathauwer L, De Moor B, Vandewalle J. A Multilinear Singular Value Decomposition. SIAM Journal on Matrix Analysis and Applications. 2000; 21(4):1253–1278.

Desco M, Hernandez JA, Santos A, Brammer M. Multiresolution analysis in fMRI: Sensitivity and specificity in the detection of brain activation. Human Brain Mapping. 2001; 14:16–27. [PubMed: 11500987]

Dinov ID, Boscardin JW, et al. A Wavelet-based Statistical Analysis of fMRI Data. Neuroinformatics. 2005; 3(4):319–342. [PubMed: 16284415]

Donoho D, Johnstone IM, Kerkyacharian G, Picard D. Wavelet Shrinkage: Asymptopia. Journal of the Royal Statistical Society. 1995:371–394.

Evangelista G. Flexible Wavelets for Music Signal Processing. Journal of New Music Research. 2001; 30(1):13–22.

Fadili MJ, Bullmore ET. Wavelet-Generalized Least Squares: A New BLU Estimator of Linear Regression Models with 1/f Errors. Neuroimage. 2002; 15(1):217–232. [PubMed: 11771991]

Friston KJ, Holmes AP, et al. Statistical parametric maps in functional imaging: A general linear approach. Hum Brain Mapp. 1995; 2(4):189–210.

Gilbert AC, Kotidis Y, et al. One-pass Wavelet Decompositions of Data Streams. IEEE Trans Knowl Data Eng. 2003; 15(3):541–554.

Gupta, MR.; Jacobson, NP. Wavelet Principal Component Analysis and its Application to Hyperspectral Images. IEEE International Conference on Image Processing; Atlanta, GA. 2006; 2006. p. 1585-1588.

Harshman R. Foundations of the PARAFAC Procedure: Models and Conditions for an 'Explanatory' Multimodal Factor Analysis. UCLA Working Papers in Phonetics. 1970:1–84.

Hitchcock FL. The Expression of a Tensor or a Polyadic as a Sum of Products. J Math Phys. 1927; 6(1):164–189.

Khorrami MS, Faro SH, et al. Functional MRI of sensory motor cortex: Comparison between finger to thumb and hand squeeze tasks. J Neuroimaging. 2011 in press.

Kolda, TG.; Bader, BW. The TOPHITS Model for Higher-order Web Link Analysis. Workshop on Link Analysis, Counterterrorism and Security; 2006.

Kolda, TG.; Bader, BW. Technical Report. Sandia National Laboratories; 2007. Tensor Decomposition and Applications.

Kolda TG, Sun J. Scalable Tensor Decompositions for Multi-aspect Data Mining. Proc IEEE Int'l Conf Data Min. 2008:363–372.

Mallat, S. Wavelet Analysis & Its Applications. 2. Academic Press; 1999. A Wavelet Tour of Signal Processing.

Roussos E, Roberts S, Daubechies I. Variational Bayesian Learning for Wavelet Independent Component Analysis. Bayesian Inference and Maximum Entropy Methods in Science and Engineering. 2005; 803:274–281.

Ruttimann UE, Unser M, Rawlings RR, Rio D, Ramsey NF, Mattay VS, Hommer DW, Frank JA, Weinberger DR. Statistical Analysis of Functional MRI Data in the Wavelet Domain. IEEE Transactions on Medical Imaging. 1998; 17(2):142–154. [PubMed: 9688147]

Sands R, Young FW. Component Models for Three-way Data: An Alternating Least Squares Algorithm with Optimal Scaling Features. Psychometrika. 1980; 45(1):39–67.

Savas, B. Master's Thesis. Sweden: Linkoping University; 2003. Analyses and Tests of Handwritten Digit Recognition Algorithms.

Sheikholeslami G, Chatterjee S, et al. WaveCluster: a wavelet-based clustering approach for spatial data in very large databases. Int'l J Very Large Data Bases. 2000; 8(3–4):289–304.

Sun J, Tao D, et al. Beyond Streams and Graphs: Dynamic Tensor Analysis. Proc 12th ACM SIGKDD Int'l Conf Knowl Disc Data Min. 2006:374–383.

Sun J, Tsourakakis CE, et al. Two Heads Better than One: Pattern Discovery in Time-evolving Multi-aspect Data. Data Min Knowl Disc. 2008; 17(1):111–128.

Talairach, J.; Tournoux, P. Co-planar Stereotaxic Atlas of The Human Brain: 3-dimensional Proportional System: An Approach to Cerebral Mapping. New York: Thieme; 1998.

Tucker, LR. Implications of Factor Analysis of Three-way Matrices for Measurement of Change. In: Harris, CW., editor. Problems in Measuring Change. University of Wisconsin Press; 1963.

Turkheimer F, Brett M, Visvikis D, Cunningham VJ. Multiresolution Analysis of Emission Tomography Images in the Wavelet Domain. Journal of Cerebral Blood Flow and Metabolism. 1999; 19:1189–1208. [PubMed: 10566965]

Van De Ville D, Blu T, Unser M. Integrated Wavelet Processing and Spatial Statistical Testing of fMRI Data. Neuroimage. 2004; 23(4):1472–1485. [PubMed: 15589111]

Vasilescu, MAO.; Terzopoulos, D. Proc 7th European Conf Comp Vision, Lecture Notes in Computer Science. Vol. 2350. Springer; 2002. Multilinear Analysis of Image Ensembles: Tensor-Faces; p. 447-460.

Vasilescu, MAO.; Terzopoulos, D. Multilinear Independent Components Analysis. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2005; Washington, DC.

Wink AM, Hoogduin H, Roerdink JBTM. Data-driven haemodynamic response function extraction using Fourier-wavelet regularised deconvolution. BMC Medical Imaging. 2008; 8:7. [PubMed: 18402674]

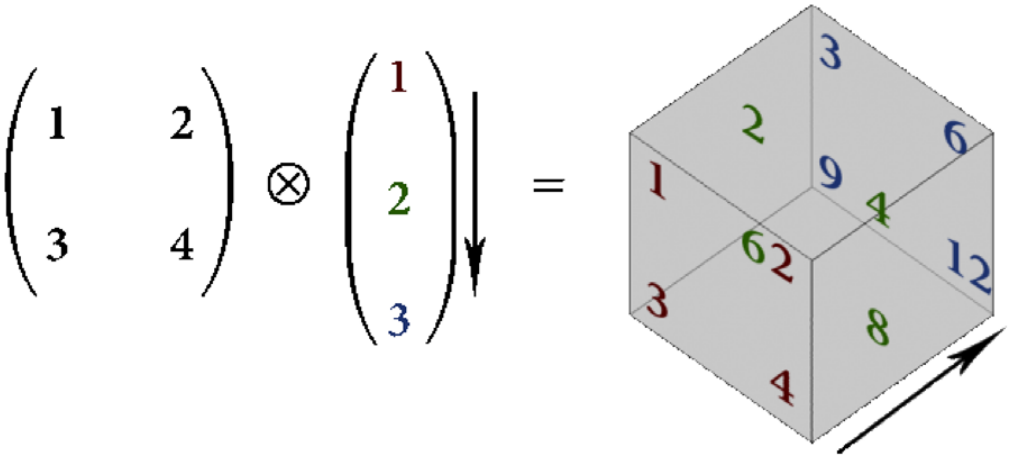


Fig. 1.
Graphical example of a tensor product.

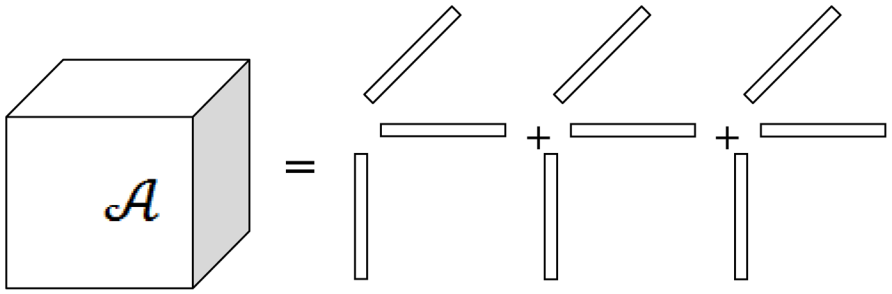


Fig. 2.
Illustration of a third-order PARAFAC decomposition, in which a tensor is decomposed into a sum of outer products.

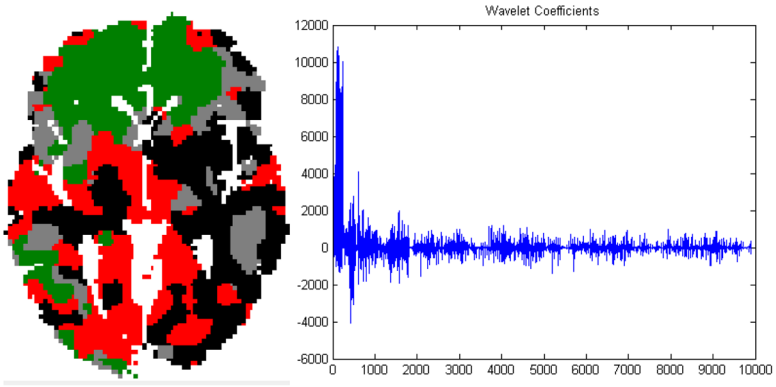


Fig. 3.
Wavelet transformation of a volume and linearization to a feature vector.

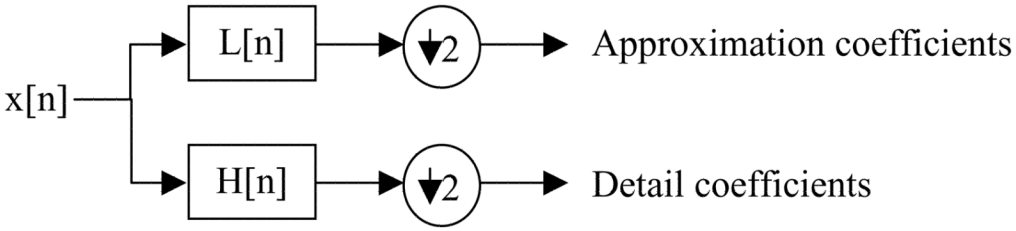


Fig. 4.
A single level wavelet decomposition represented as a filter. With certain wavelets, a perfect reconstruction is possible despite downsampling.

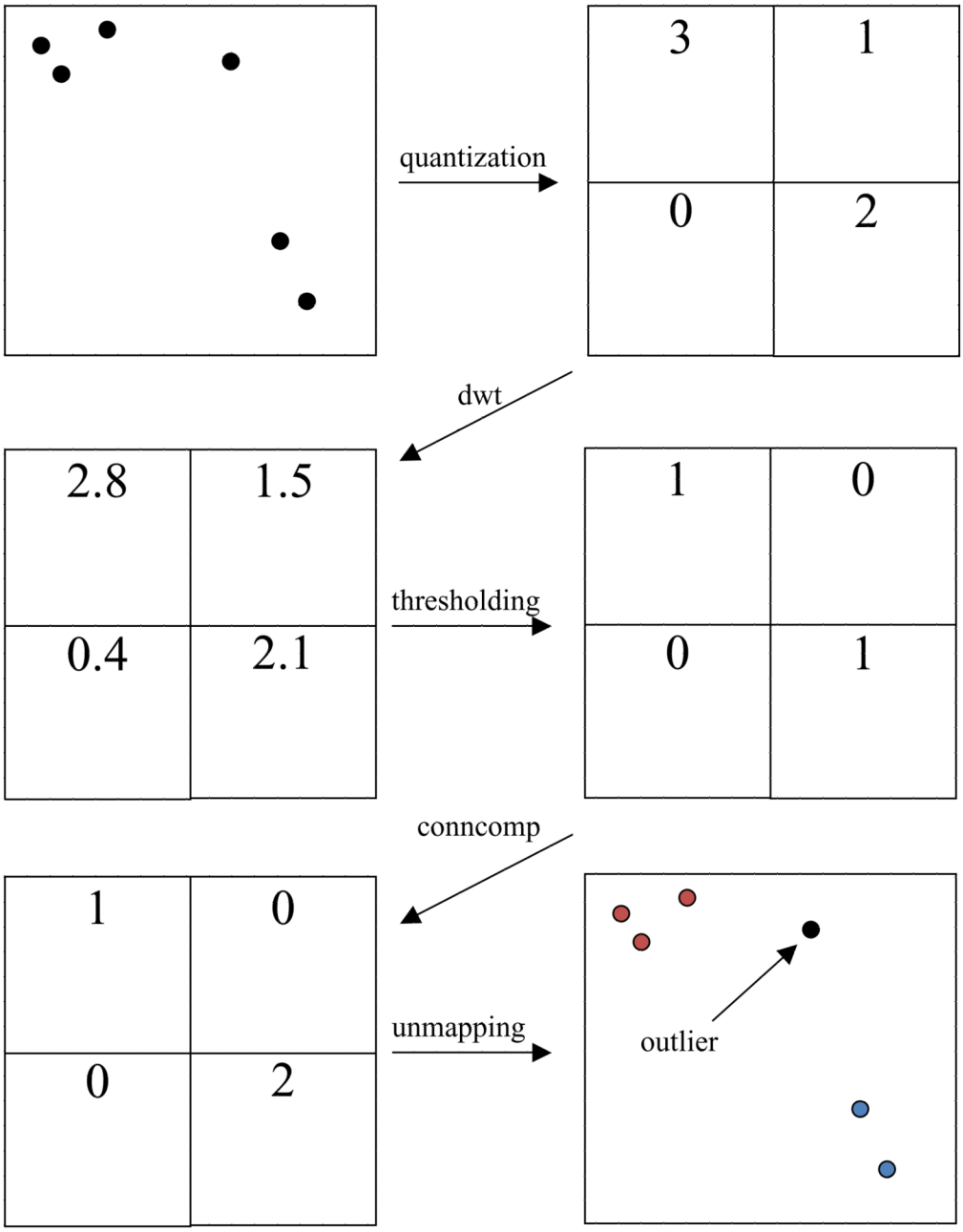


Fig. 5.
Step-by-step illustration of the WaveCluster algorithm. The cell corresponding to the isolated black dot falls below the significance threshold and is labeled an outlier.

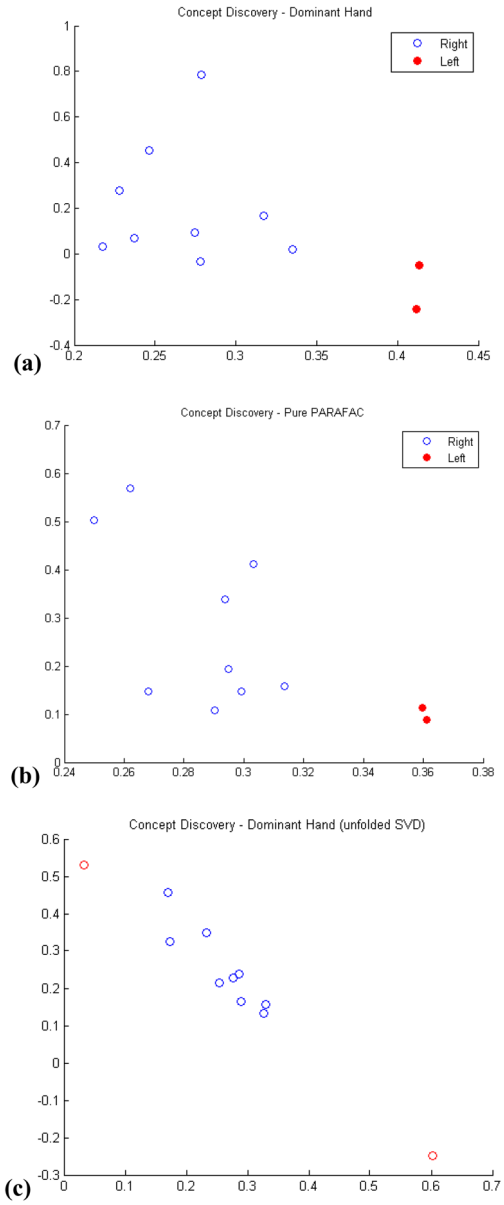


Fig. 6.

The two concepts discovered by (a) TWave, (b) PARAFAC, and (c) SVD on the subject mode. The outliers in the TWave and PARAFAC projections correspond to the two left handed subjects in the dataset, while there is no linear separation of these points when using SVD. Note the similarity of the TWave and PARAFAC spaces despite 98% compression of the dataset and a 100x difference in runtime.

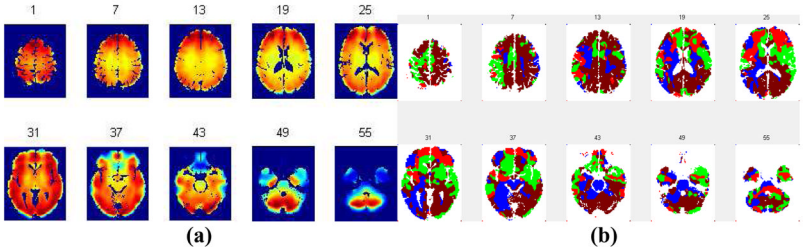
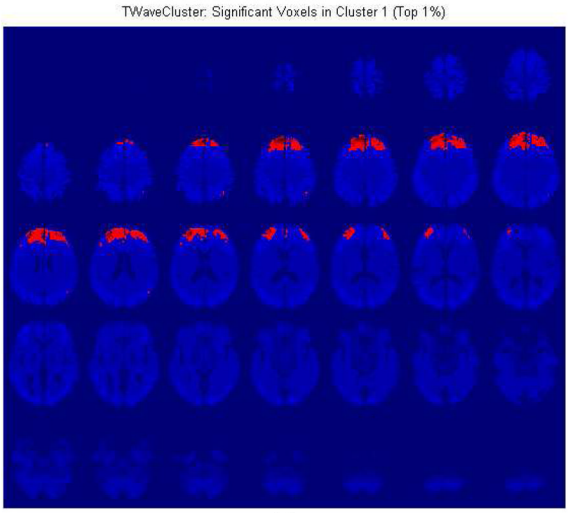
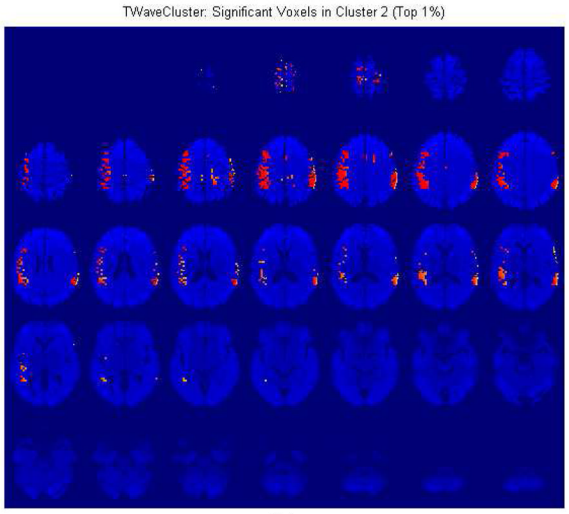


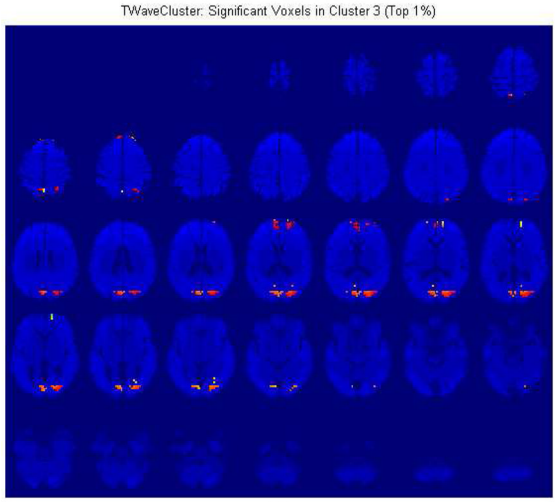
Fig. 7.
(a) Spatial activation patterns in a right-handed subject performing a left finger-to-thumb task and (b) the clusters discovered by k-means.



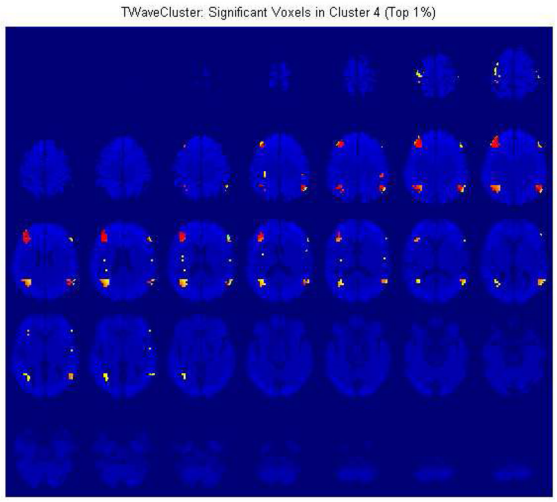
(a)



(b)



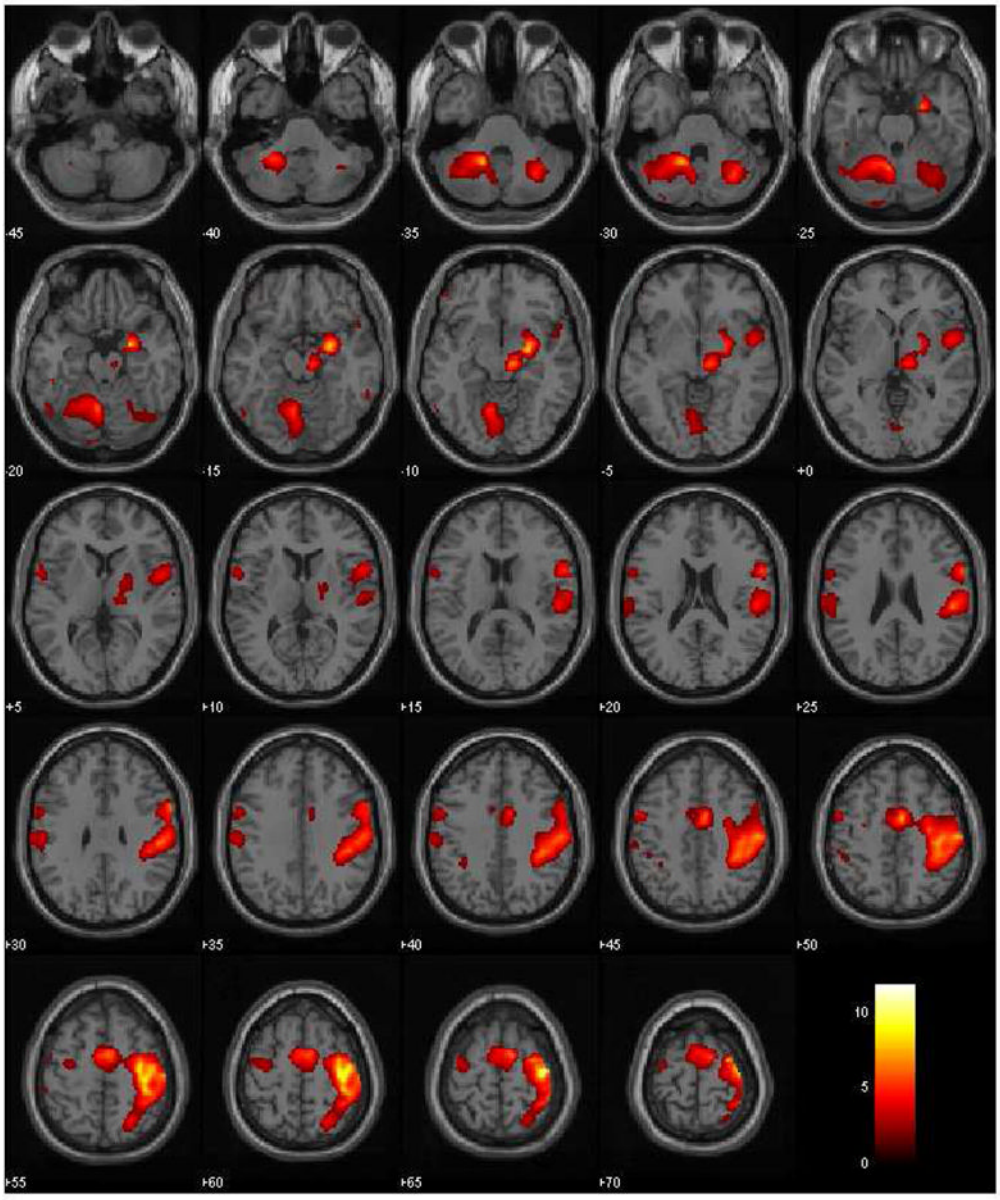
(c)



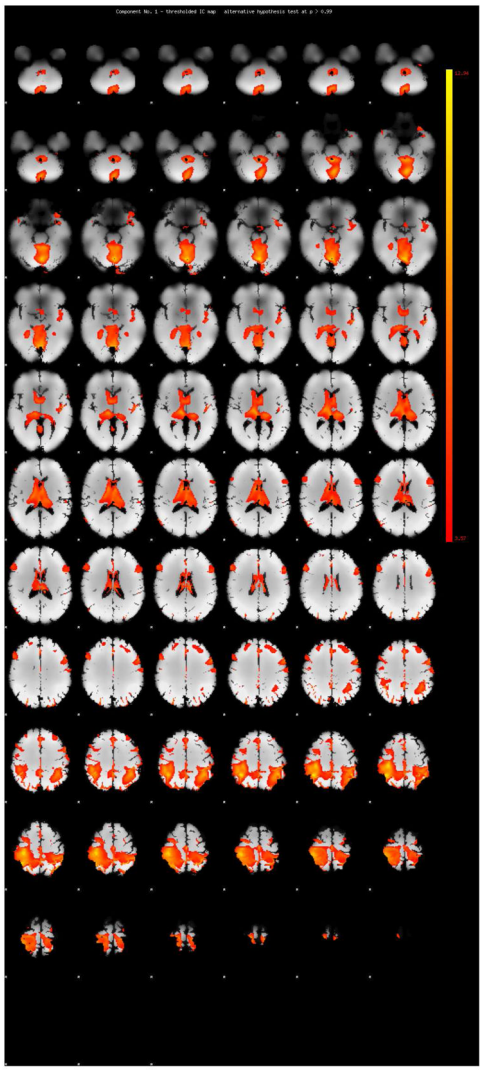
(d)

Fig. 8.

TWaveCluster results thresholded at the top 1% most significant voxels per cluster and overlaid on the full dataset. The two most salient clusters (a) and (b) spatially correspond to the frontal lobe and the motor strip of the brain, respectively. The third cluster (c) centralizes in the caudate region of the brain and the least salient cluster (d) appears to represent noise, as is common with matrix and tensor factorizations of sufficient rank.



(a)



(b)

Fig. 9.

(a) SPM on the same subject fails to discover significant frontal activation but does discover the activation in the motor strip, while (b) MELODIC overly emphasizes the noisy regions at the corners of the image but also discovers the motor activation (note that there is a difference in coordinate orientation).

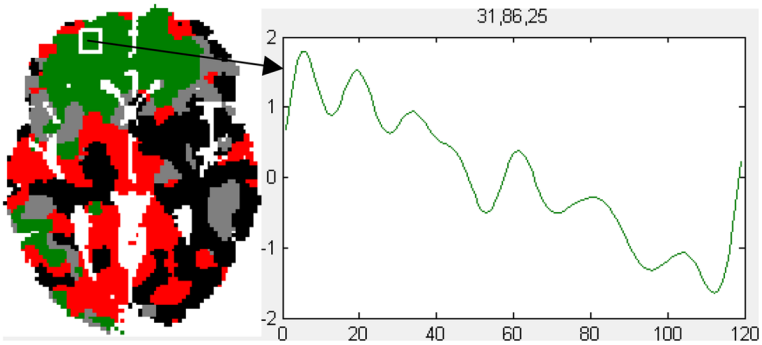


Fig. 10.
The spatiotemporal nature of our dataset: each voxel represents a time series.

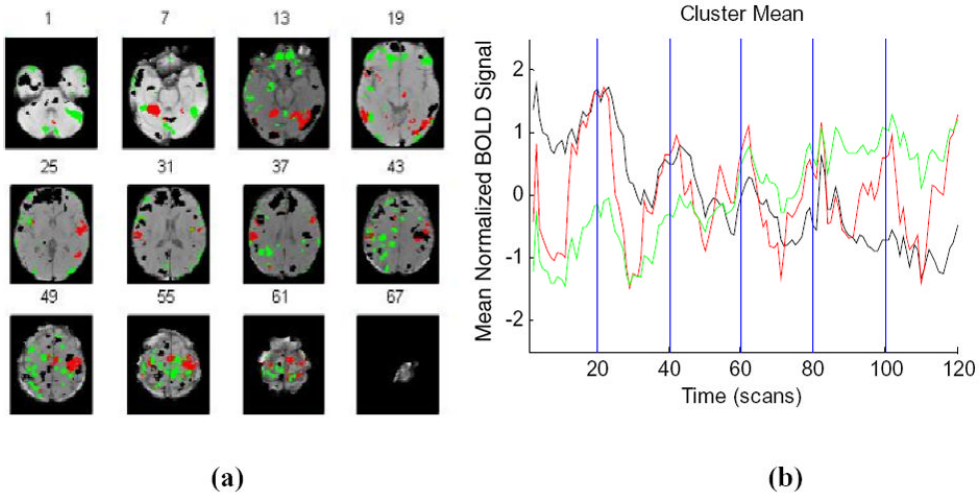


Fig. 11.

Voxelwise spatiotemporal clustering, shown in (a) space and (b) time: each voxel is represented as a time series and is clustered using the k-means algorithm. Colors represent clusters and are consistently used between (a) and (b); the black cluster corresponds to frontal lobe activation. The mean time series for the black cluster exhibits a decreasing trend due to motor learning with repetition of the task.

Table 1

Scalars, Vectors, Matrices, Tensors, and their orders.

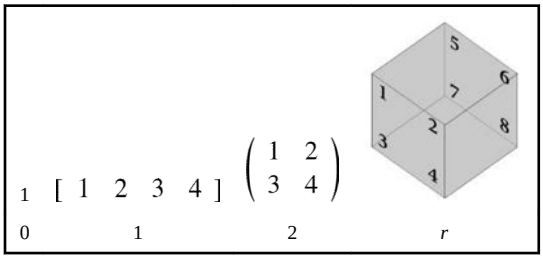


Table 2

High-order fMRI total-dataset runtimes and ability to automatically identify left-handed subjects for each method that we analyzed.

	Voxels	**Wavelets**	**SVD**	**PARAFAC**	**TWave**
Runtime	95 min	112 min	**3 days**	**8 days**	133 min
Lefties?	No	No	No	Yes	Yes

Table 3

Runtime breakdown during a second run by stage of analysis.

	Voxels	Wavelets	SVD	PARAFAC	TWave
Preprocessing	1%	0%	0%	0%	0%
Wavelet Transform	N/A	91%	N/A	N/A	87%
Shrinkage	N/A	3%	N/A	N/A	2%
Decomposition	N/A	N/A	100%	100%	9%
Classification	99%	6%	0%	0%	1%