Analysis of Functional MRI Timeseries Data Using Signal Processing Techniques

Sea Chen

Department of Biomedical Engineering

Advisors: Dr. Charles A. Bouman and Dr. Mark J. Lowe

Overview

- Introduction
- Update: Supertemporal Resolution Analysis
  - Review
  - New simulations
  - New data
- Clustered Components Analysis
  - Motivation
  - Theory
  - Methods
  - Results
  - Conclusions

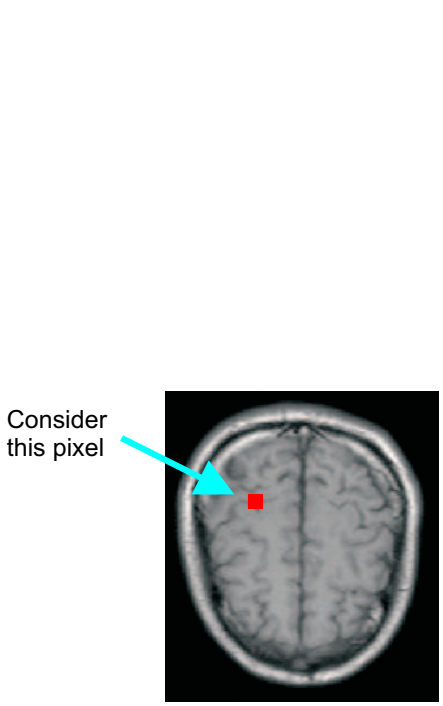
Goals

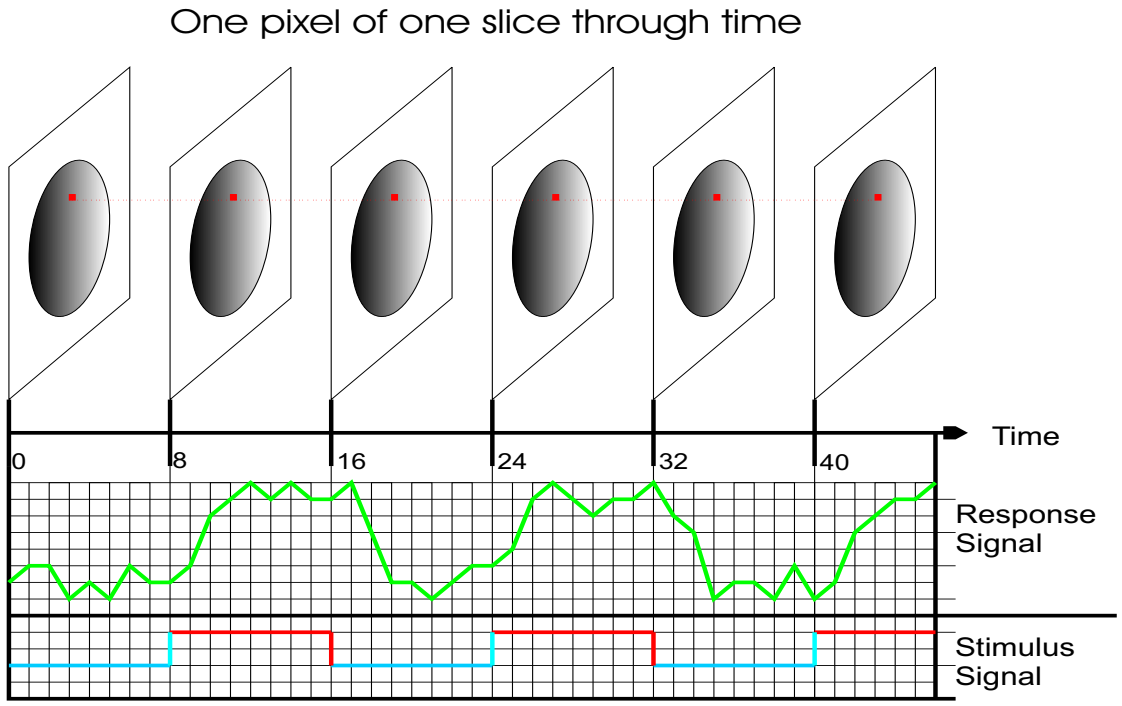
We would like to aid in the understanding of the blood-oxygenation-level-dependent (BOLD) contrast mechanisms used in functional magnetic resonance imaging (fMRI) through

- achieving a high signal-to-noise (SNR) estimate of the BOLD response.
- achieving a high temporal resolution estimate of the BOLD response.

fMRI: The basic idea

- Experimental paradigm designed to activate neuronal metabolism
- Changes in blood oxygenation during activation $\Rightarrow$ changes in physical parameters affecting MR signal
- Contrast produced by difference between active and control states
- Data set is volume of pixels repeated over time



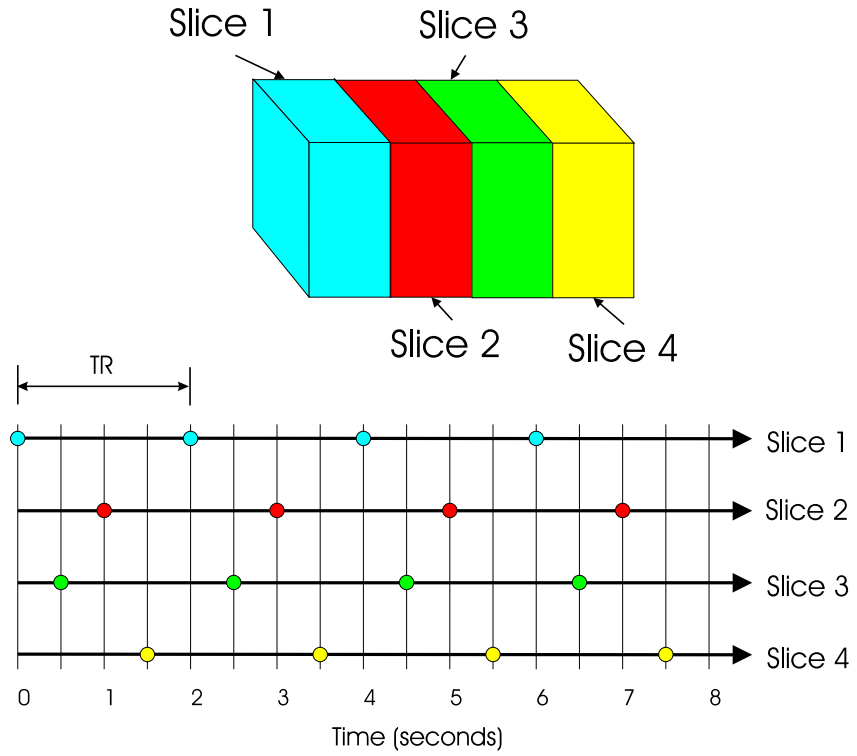


Supertemporal Resolution: Motivation

- Short TR
  - Better time resolution
  - Lower SNR due to saturation effects
  - BOLD signal is distorted by blood inflow effects
- Long TR
  - Poorer time resolution
  - BOLD effect more dominant in activation signal

Supertemporal Resolution: Review

- Assumption: Voxels exhibiting the same generating activation signal span different slices in a 2D acquisition
- Method exploits the timing characteristics of the 2D acquisition
- Bayesian prior used to implement temporal regularization



Supertemporal Resolution: Review

- MAP estimate for Supertemporal Resoution (STR)

$$\hat{e}_\sigma = \underset{e}{\operatorname{argmax}} \left\{ \frac{1}{2} \sum_{s=0}^{K-1} \frac{e^T \tilde{\mathbf{A}}_s^T \mathbf{R}_s \tilde{\mathbf{A}}_s e}{e^T \mathbf{A}_s{}^T \mathbf{A}_s e} - \frac{1}{p\sigma^p ||e||^p} \sum_{i=0}^{r-1} (e_i - e_{i+1})^p \right\}$$

where $\mathbf{R}_s = \sum_{n \in S} \tilde{y}_n \tilde{y}_n^T$

- Optimization performed using conjugate gradient
- Regularization parameter σ found by crossvalidation strategy

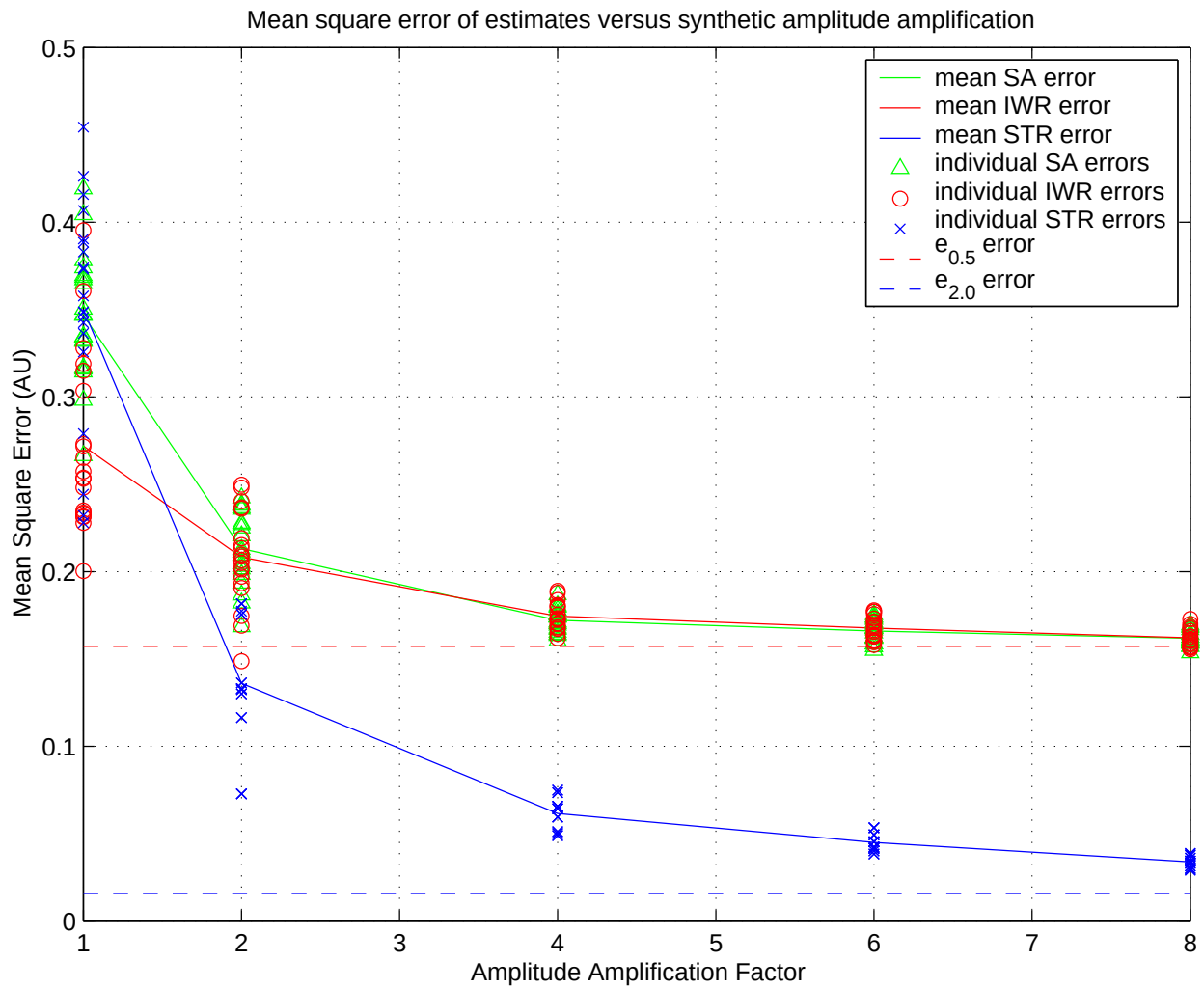
Supertemporal Resolution: Updates

- Reduction in computation time
  - Minor software revisions
  - New hardware
- New simulations
  - Introduced amplitude amplification factors κ (1x, 2x, 4x, 6x, 8x simulating increase in B_0-field)
  - Generated multiple (20 / κ) datasets
- New human visual system data
  - Three inch surface coil
  - Multiple runs (3 TR = 2.0s, 3 TR = 0.5) of 3.5 cycles

Performance comparison

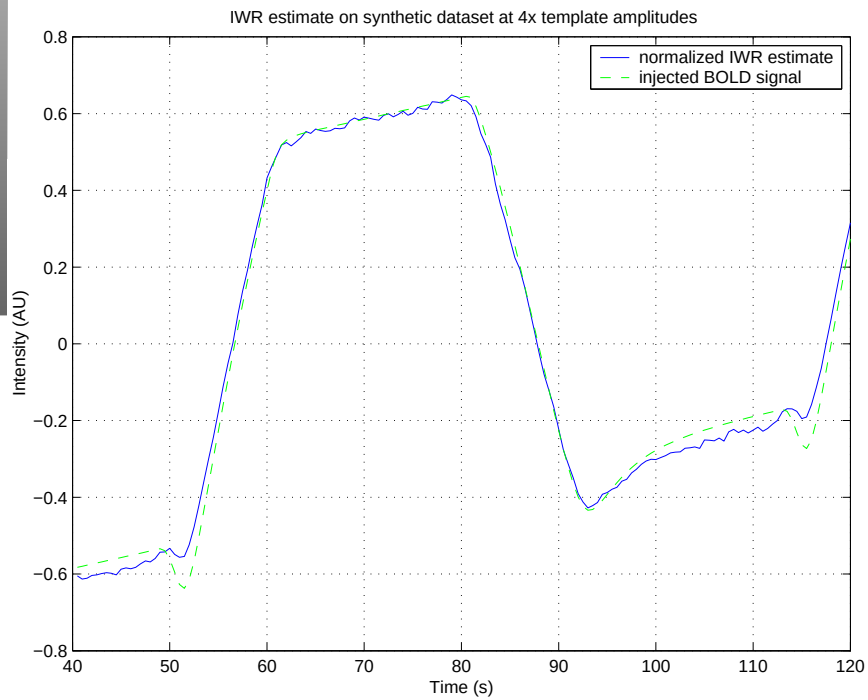
- Simple averaging (SA) method
  - Alignment of slices into timeframe of first slice
  - NO regularization
  - Closed form solution
  - 0.5 s time resolution estimate from TR = 0.5 s dataset
- Interpolation with regularization (IWR) method
  - Alignment of slices into timeframe of first slice
  - Regularization applied and chosen with crossvalidation
  - Numerical optimization with conjugate gradient
  - 0.5 s time resolution estimate from TR = 0.5 s dataset
- Supertemporal regularization (STR) method
  - Slice timing considered in data model
  - Regularization applied and chose with crossvalidation
  - Numerical optimization with conjugate gradient
  - 0.5 s time resolution estimate from TR = 2.0 s dataset

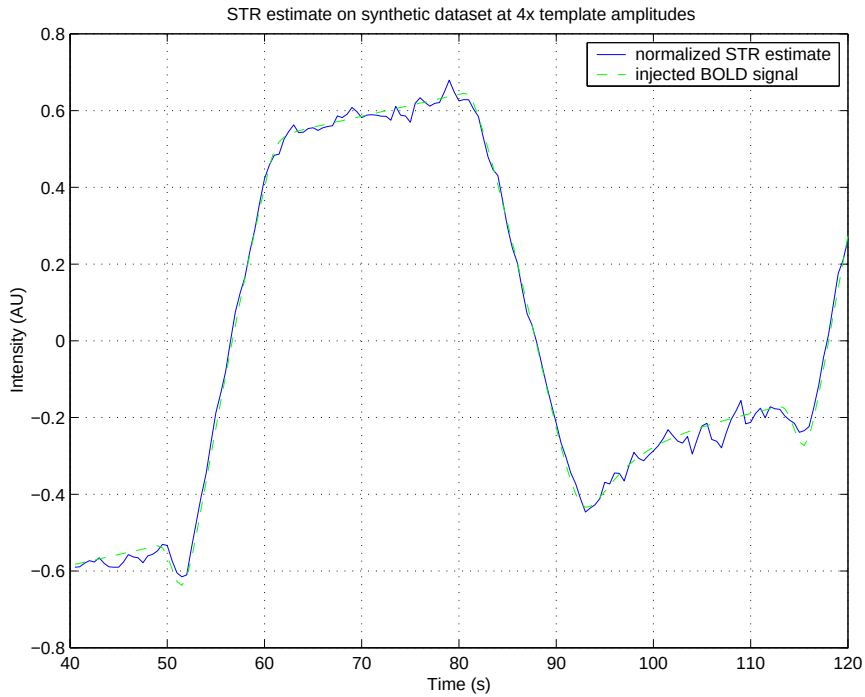
Simulation results: Performance



Mean square error of simulation results for different analysis methods plotted against amplitude amplification factor κ

Simulation results: Examples





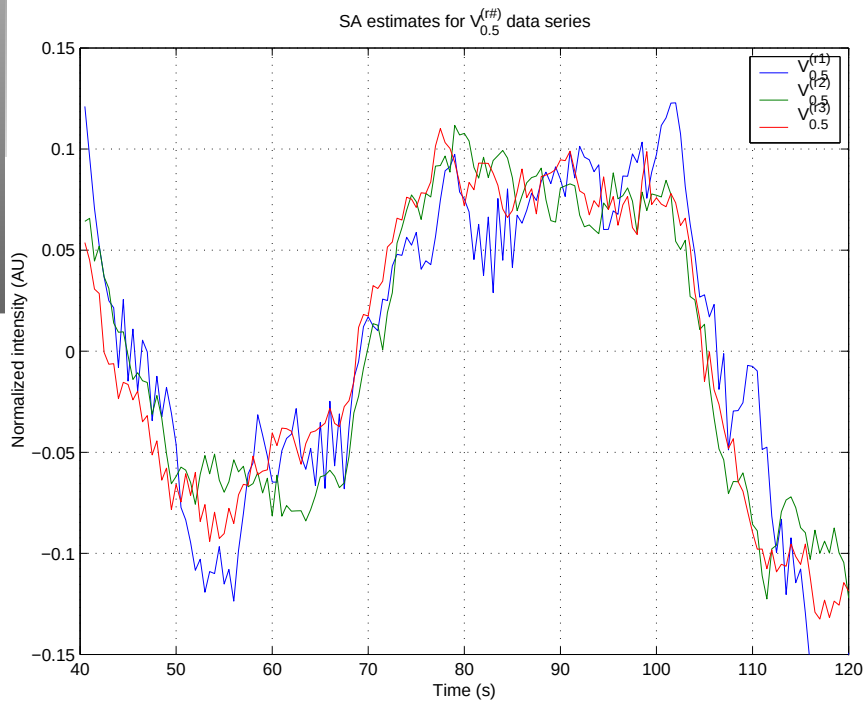
IWR estimate on TR=0.5s data

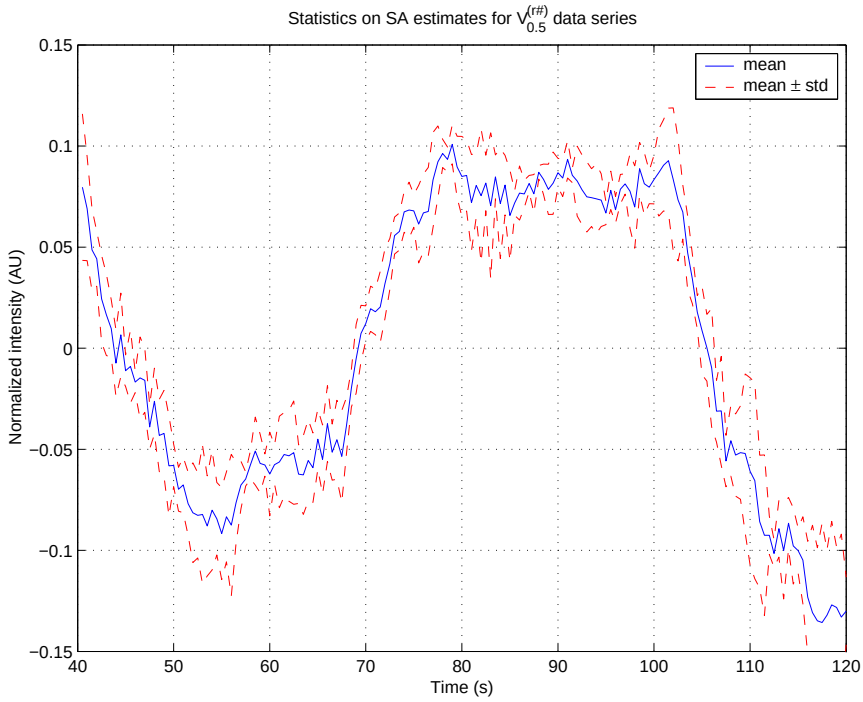
STR estimate on TR=2.0s data

Examples of 0.5 second estimates at $\kappa = 4$

Human data results:

Simple averaging method



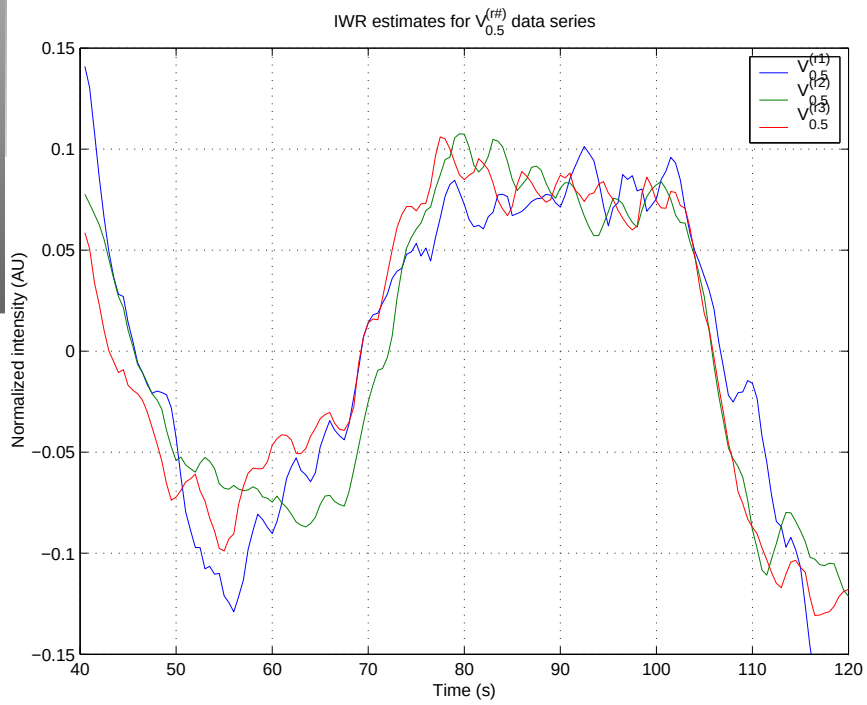


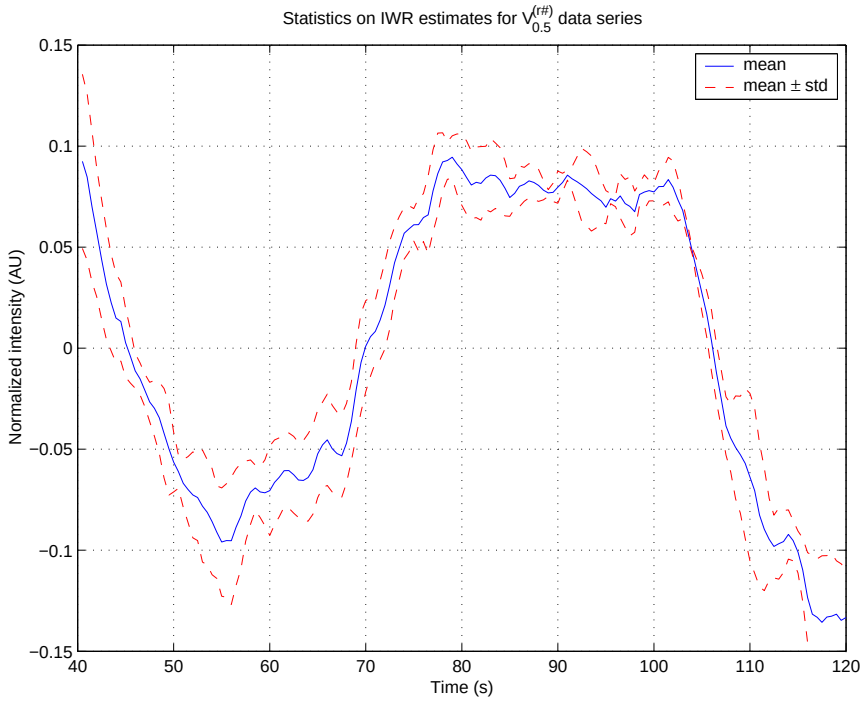
SA estimates

Mean and std. dev. of SA estimates

Simple averaging estimates on the TR=0.5 second dataset (3 experiments)

Human data results: Interpolation with regularization method



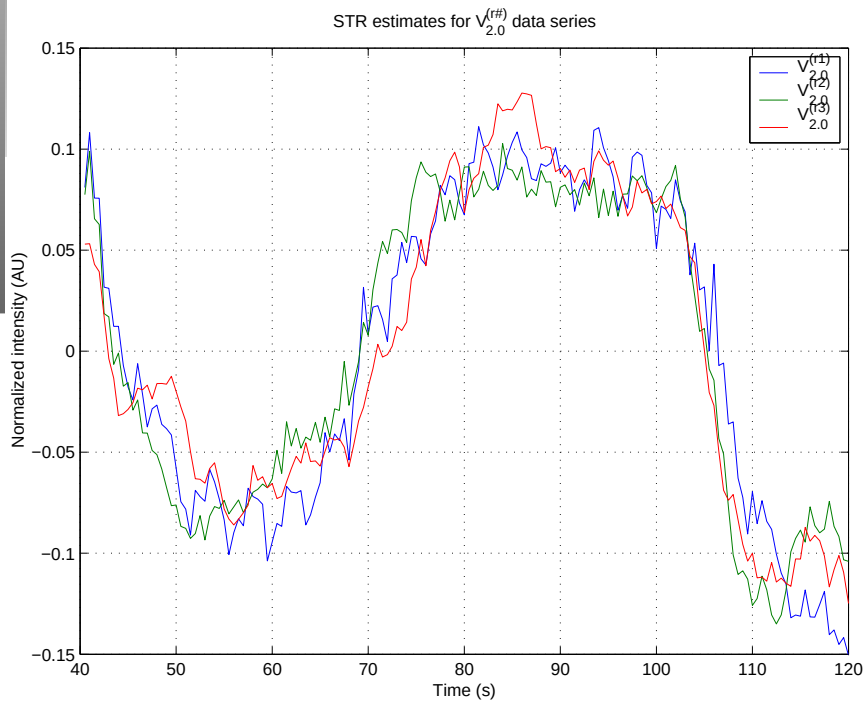


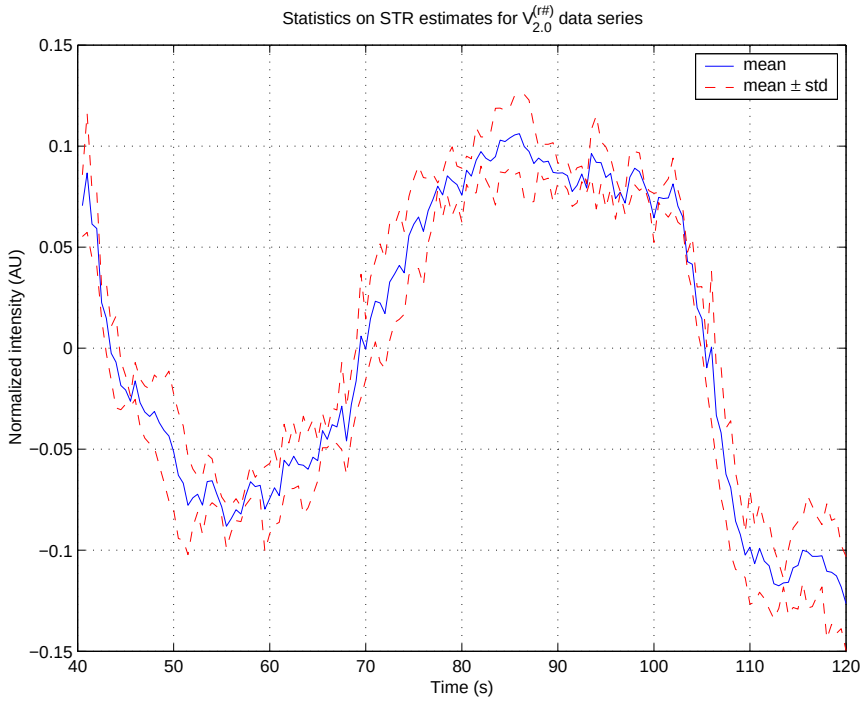
IWR estimates

Mean and std. dev. of IWR estimates

Interpolation with regularization estimates on the TR=0.5 second dataset (3 experiments)

Human data results: Supertemporal Resolution method





STR estimates

Mean and std. dev. of STR estimates

Supertemporal resolution estimates on the TR=2.0 second dataset (3 experiments)

Discussion and conclusions

- Simulated data
  - Initially at low SNR, STR performs worse than IWR because small features masked by noise
  - At increasing SNR, STR performs better than IWR/SA as inherent physical advantange becomes apparent
- In human data, STR estimates qualitatively different from IWR/SA estimates
- Conclusion: STR may be a valuable tool in characterizing small features in the BOLD signal at higher static field strengths or higher SNR

Clustered components analysis: Objectives

Hypothesis: Activation by specific functional tasks $\Rightarrow$ Distinct neural responses in different parts of the brain

Therefore, we propose the following goals:

- Design and run fMRI experiment activating visual, auditory, and motor cortices.
- Estimate number of distinct neural responses (# of classes/clusters)
- Extract an estimate for each response
- Determine voxel memberships

Existing approaches to signal estimation

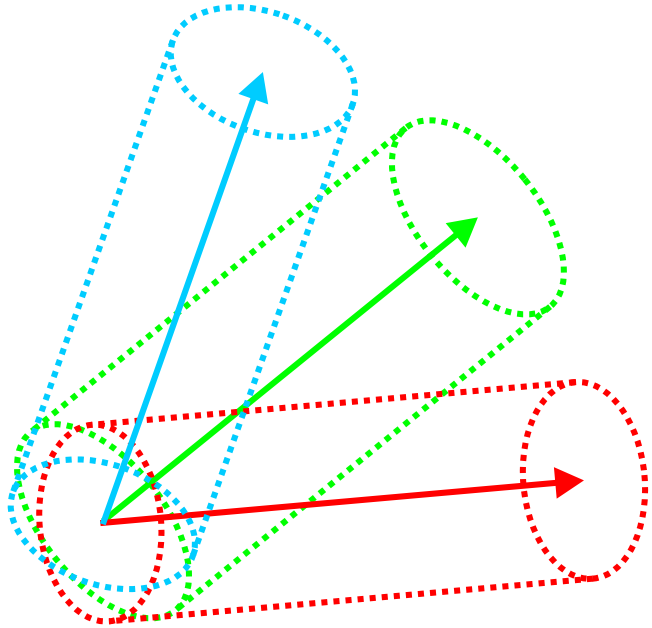
- Principle component analysis (PCA)
  - Extracts orthogonal signals
  - Disadvantage: Signals not usually orthogonal
- Independent component analysis (ICA)
  - Extracts spatially independent signals
  - Disadvantage: Signals may not be independent
- Conventional Clustering
  - Groups signal vectors as spheres about a mean
  - Disadvantage: Signals may not form spherical clusters
- **General Comment**: None of these methods start with an explicit model of the data. All go about estimating the distinct signals in an ad hoc way.

Analysis framework

- Dimensionality Reduction
  - Signal subspace is orthogonal to noise subspace
  - Noise can be accurately modeled in fMRI
  - Separate signal subspace (dim M) from noise subspace (dim $N - M$)
- Clustered Components Analysis
  - Useful information is in shape of signal, amplitude unimportant
  - Component direction is found instead of mean
  - Amplitude can vary in cluster so long as shape preserved
  - Clusters found in cylinders instead of spheres

Interpretation of Clustered Components Analysis

- Because amplitude of the voxel signal is not important, the method clusters around component directions, not component means.
- This means the clusters can be thought of as cylinders rather than the traditional spheres.



Dimensionality reduction: Harmonic decomposition

Data model for harmonic decomposition

$$\mathbf{D} = \mathbf{A}\hat{\mathbf{\Theta}} + \epsilon$$

- $\mathbf{D}$: $P \times N$ detrended voxel timecourse matrix (P=# of timepoints, N=# of voxels)
- $\mathbf{A}$: $P \times L$ matrix of sampled sines and cosines (L harmonic components)
- $\hat{\mathbf{\Theta}}$: $L \times N$ harmonic image
- ϵ: $P \times N$ maxtrix of residuals from the least squares fit

Dimensionality reduction: Signal subspace estimation

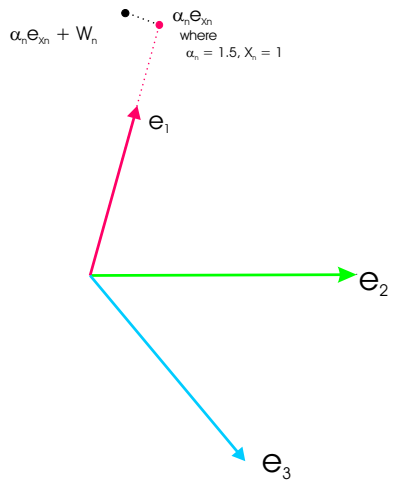
- Signal + noise covariance: $\hat{R}_s{}^n = \frac{1}{N}\hat{\Theta}\hat{\Theta}^t$
- Noise covariance: $\hat{R}_n = \text{trace}\{\epsilon\epsilon^t\}(\mathbf{A}^t\mathbf{A})^{-1}/(N(P-L-2))$
- Signal covariance: $\hat{R}_s = \hat{R}_s{}^n - \hat{R}_n$
- Eigen decomposition $\hat{R}_s = U_s\Lambda_s U_s^t$
- Only the columns of the $L \times L$ eigenvector matrix U_s corresponding to the M positive eigenvalues of Λ_s are retained, yielding the $L \times M$ modified eigenvector matrix $\hat{U}_s$.
- $M \times N$ reduced dimensionality feature vector matrix: $Y = W\hat{U}_s^t\hat{\Theta}$ (W is a whitening vector matrix derived from ϵ)

Data Model for Clustered Component Analysis

- Assumptions
  - Only shape of the response important
  - Amplitude is **NOT** important
  - Noise independent in space and time (time-independence can be relaxed)
- Our Model
  - Y_n is M-dimensional column vector representation of the n^{th} timecourse

$$Y_n = \alpha_n e_{x_n} + \omega_n$$

  - α_n is the unknown scalar amplitude for pixel n,
  - $E = [e_1, \cdots, e_K]$ are the K component directions
  - $1 \leq x_n \leq K$ is class of the pixel
  - ω_n is a Gaussian noise vector



Clustered components approach

- Goal: Minimize minimum description length (MDL) criterion

  $$\text{MDL} = -\text{loglikelihood} + \frac{1}{2}(\#\text{ of parameters})\log(\#\text{ of datapoints})$$

- Unknown model parameters
  - K is the model order (number of clusters)
  - α_n is the amplitude of each pixel
  - $E_K = [e_1, \cdots, e_K]$ is the set of distinct neural responses
  - $\Pi_K = [\pi_1, \cdots, \pi_K]$ are the prior probabilities for each class
- Use maximum likelihood (ML) estimate $\hat{\alpha}_n$ implicitly
- Find ML estimates $\hat{E}_K$ and $\hat{\Pi}_K$ using the Expectation-Maximization (EM) algorithm for each model order K
- Estimate model order $\hat{K}$ by cluster merging and minimizing the MDL criterion

Voxel likelihood function

- Likelihood for each voxel

$$p_{y_n|x_n}(y_n|k, E_K, \alpha_n) = \frac{1}{(2\pi)^{(M-1)/2}} \exp\left\{-\frac{1}{2}(y_n - \alpha_n e_k)^t (y_n - \alpha_n e_k)\right\}$$

- ML estimate of the scalar amplitude

$$\hat{\alpha}_n = {y_n}^t e_k$$

- Voxel log-likelihood

$$\log p_{y_n|x_n}(y_n|k, E_K, \hat{\alpha}_n) = -\frac{M-1}{2}\log 2\pi - \frac{1}{2}\left(y_n^t y_n - {e_k}^t y_n {y_n}^t e_k\right)$$

Maximum likelihood estimate

- Log-likelihood of the entire dataset

$$\log p_y(y|K, E_K, \Pi_K, \hat{\alpha}) = \sum_{n=1}^{N} \log \left(\sum_{k=1}^{K} p_{y_n|x_n}(y_n|k, E_K, \hat{\alpha}_n) \pi_k \right)$$

$$= \sum_{n=1}^{N} \log \left[\sum_{k=1}^{K} \left(\frac{1}{(2\pi)^{(M-1)/2}} \exp \left\{ -\frac{1}{2} \left(y_n^t y_n - {e_k}^t y_n {y_n}^t e_k \right) \right\} \pi_k \right) \right]$$

- ML estimate of the parameters

$$(\hat{E}_K, \hat{\Pi}_K) = \arg\max_{E_K, \Pi_K} \log p_y(y|K, E_K, \Pi_K, \hat{\alpha})$$

Expectation-maximization equations

- **Posterior probability**

$$p_{x_n|y_n}(k|y_n, K, E_K, \Pi_K, \hat{\alpha}_n) = \frac{p_{y_n|x_n}(y_n|k, E_K, \hat{\alpha}_n)\pi_k}{\sum_{l=1}^{K} p_{y_n|x_n}(y_n|k, E_K, \hat{\alpha}_n)\pi_l}$$

- **E-step**

$$\bar{N}_{k|K}^{(i)} = \sum_{n=1}^{N} p_{x_n|y_n}(k|y_n, {E_K}^{(i)}, {\Pi_K}^{(i)}, \hat{\alpha}_n)$$

$$\bar{R}_{k|K}^{(i)} = \frac{1}{\bar{N}_{k|K}^{(i)}} \sum_{n=1}^{N} y_n y_n^t \, p_{x_n|y_n}(k|y_n, {E_K}^{(i)}, {\Pi_K}^{(i)}, \hat{\alpha}_n)$$

- **M-step**

$$e_k^{(i+1)} = e_{max}\{\bar{R}_{k|K}^{(i)}\}$$

$$\pi_k^{(i+1)} = \bar{N}_{k|K}^{(i)}/N$$

Order Estimation through Cluster Merging

1. Start with large number of clusters (K) and initialize
2. Run EM algorithm to convergence
3. Choose the two clusters that minimize the distance function (which is the upper bound on the change in MDL)

$$d(l,m) = \sigma_{max}(\bar{R}_{l|K}) + \sigma_{max}(\bar{R}_{m|K}) - \sigma_{max}(\bar{R}_{l|K} + \bar{R}_{m|K})$$

4. Merge clusters using

$$\begin{aligned} \pi_{(\hat{l},\hat{m})} &= \pi_{\hat{l}} + \pi_{\hat{m}} \\ e_{(\hat{l},\hat{m})} &= e_{max}\{\bar{R}_{\hat{l}|K} + \bar{R}_{\hat{m}|K}\} \end{aligned}$$

5. Decrement K and initialize next iteration with new clusters
6. Repeat 2 through 5 until K = 1
7. Choose number of components minimizing the MDL criterion

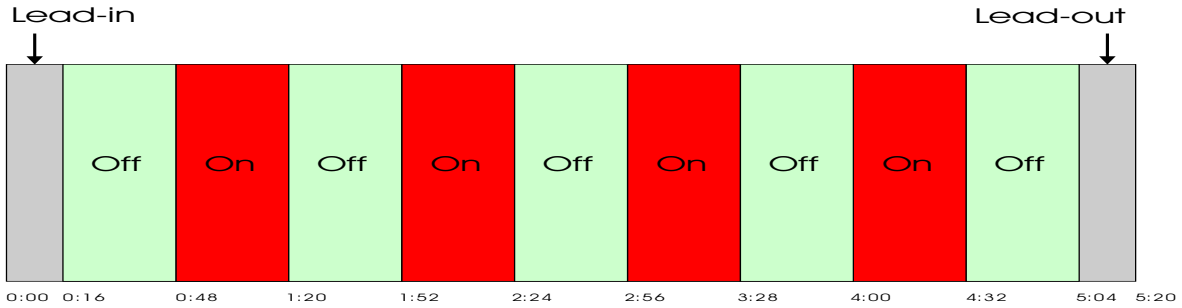
$$MDL(K, E_K, \Pi_K) = -\log p_y(y|K, E_K, \Pi_K, \hat{\alpha}) + \frac{1}{2} KM \log(NM)$$

Synthetic fMRI Images

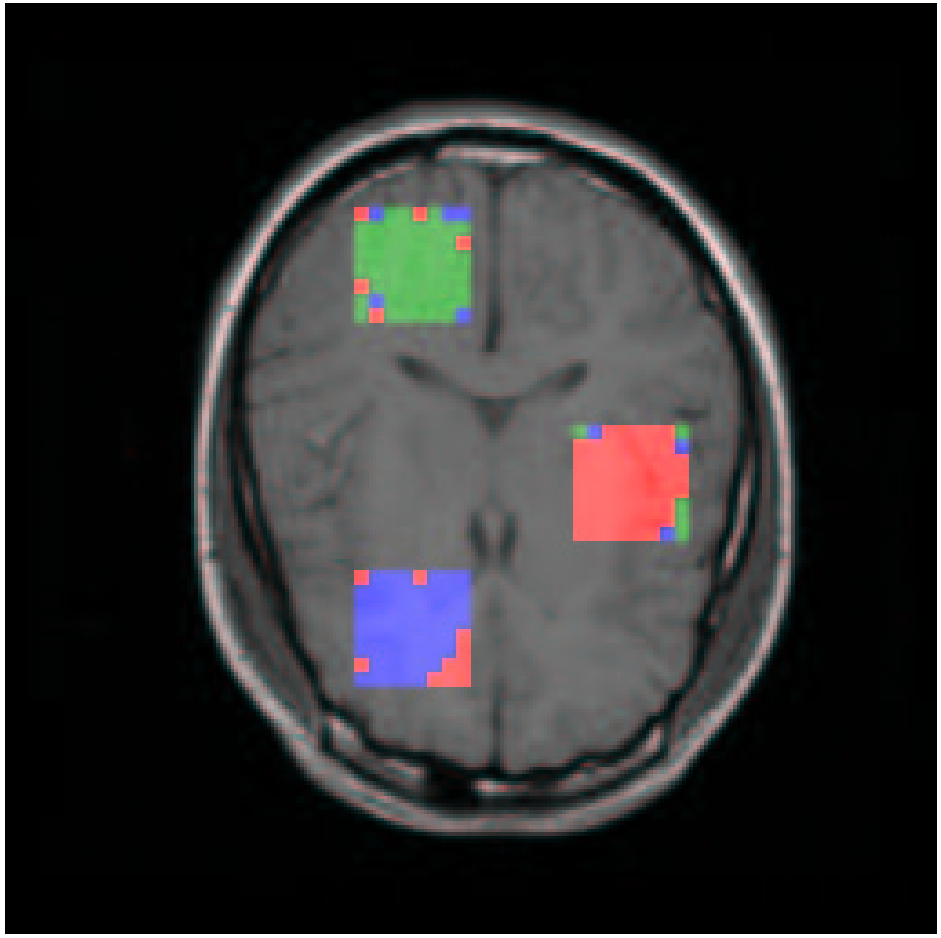
- Synthetic data
  - Baseline control images created at each sample point
  - During periods of activation, 3 different realistic signals with varying amplitudes were injected
  - Gaussian white noise added to each voxel at each timepoint time
- Verification and comparison using different analysis methods applied before and after signal subspace estimation (SSE)
  - PCA, using 3 components corresponding to the 3 largest variances
  - Spatial ICA constrained to yield 3 components
  - Spatial ICA unconstrained, using 3 best components
  - Fuzzy C-means (FCM) clustering constrained to yield 3 clusters
  - CCA

Paradigm Design

- For our dimensionality reduction, activation must be periodic
  - Block activation scheme
  - 1 cycle = 32 seconds control (rest state), 32 seconds
  - 1 scan = 16 seconds lead in - 4.5 paradigm cycles - 16 seconds lead out (only use samples during paradigm)
  - 0.5 Hz sample rate (TR = 2 seconds)
- To illustrate the power of the clustering method, many different types of functional cortex must be activated
  - Visual: Flashing 8Hz checkerboard
  - Auditory: Forward vs. Backward speech (backwards is the control)
  - Motor: Self paced finger tapping

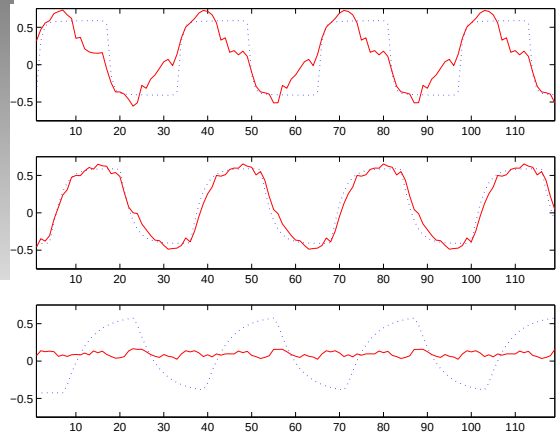


Simulated data: Hard classifications

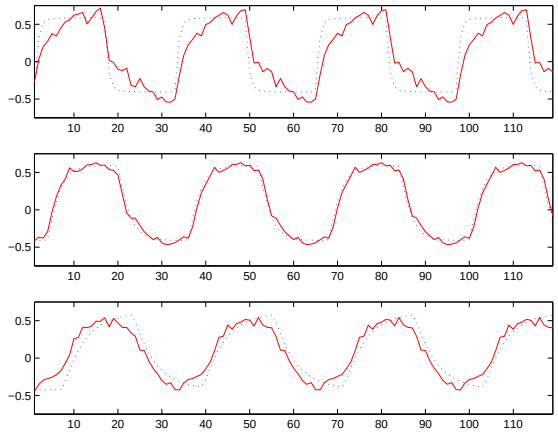


Results for CCA applied to synthetic data

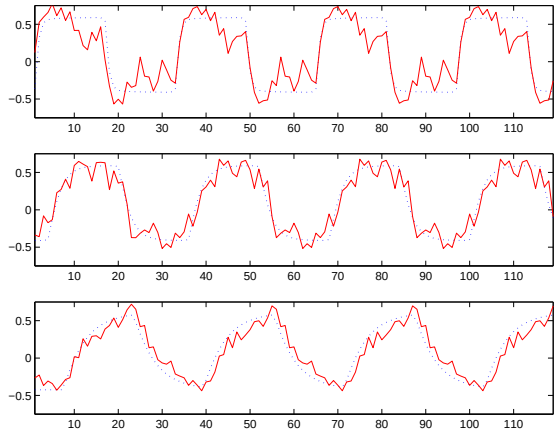
Simulated data: Qualitative results



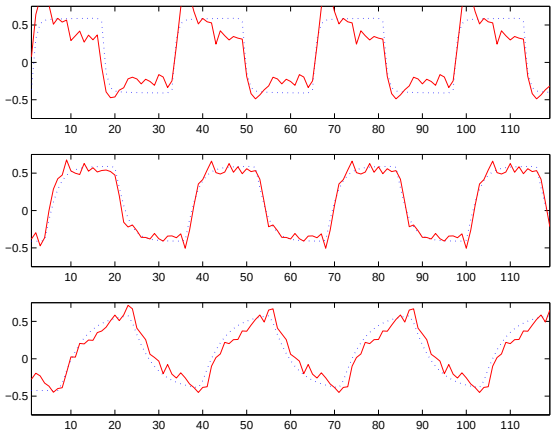
PCA



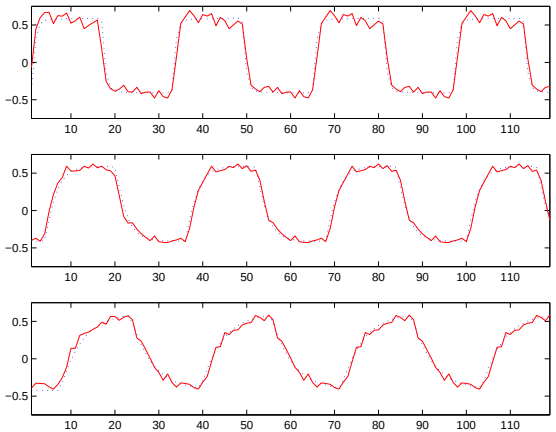
FCM



constrained ICA



unconstrained ICA



CCA

Estimates after application of SSE

Simulated data: Quantitative results

Mean squared error for analyses on synthetic data before and after signal subspace estimation (SSE)

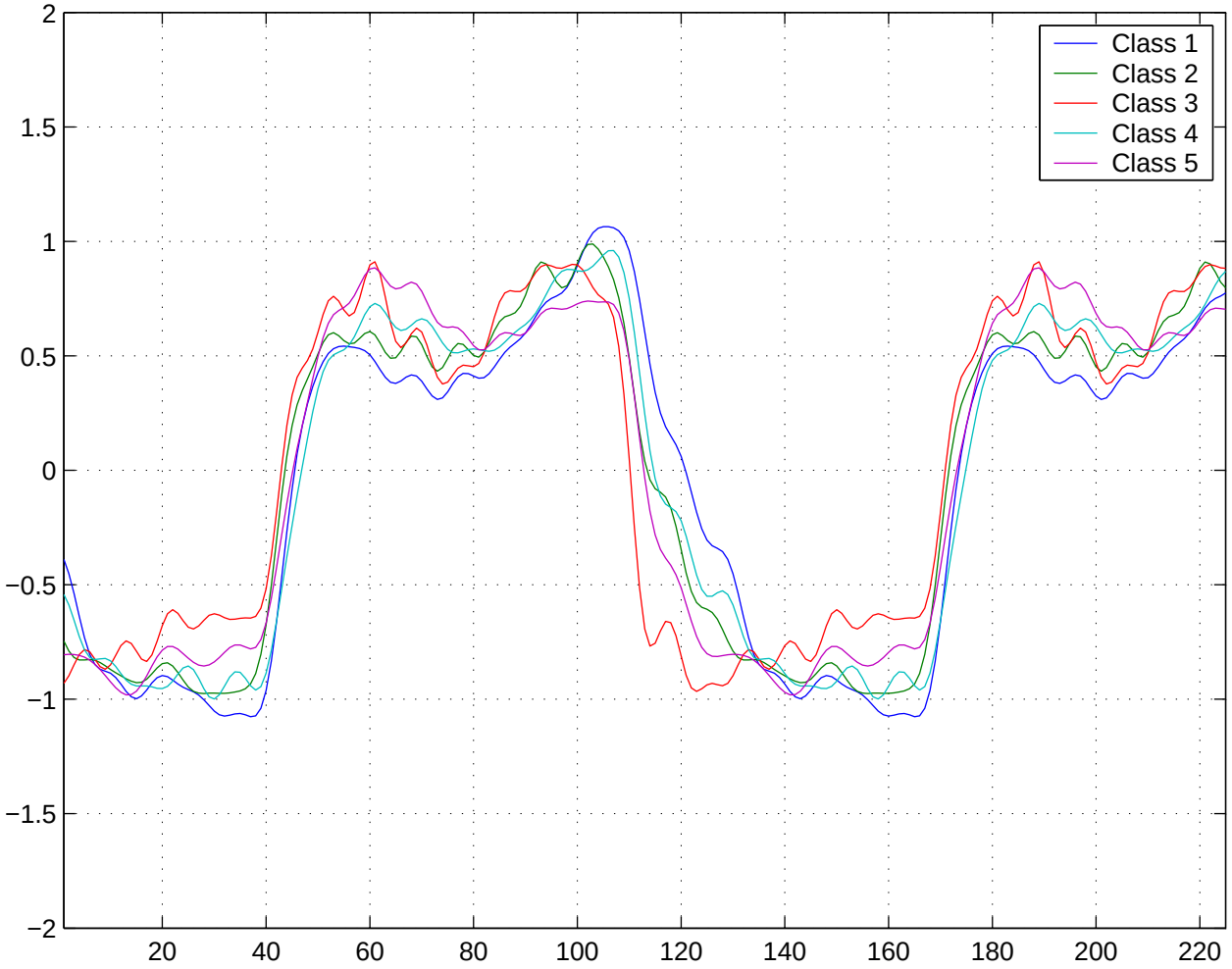
	PCA	FCM	ICA (c)	ICA (u)	CCA
Before SSE	5.73×10^{-4}	7.37×10^{-4}	1.02×10^{-3}	2.49×10^{-4}	3.49×10^{-5}
After SSE	5.80×10^{-4}	1.95×10^{-4}	2.21×10^{-4}	1.63×10^{-4}	3.09×10^{-5}

Number of voxels classified correctly on synthetic data before and after signal subspace estimation (SSE) out of 192 total voxels

	PCA	ICA (c)	ICA (u)	FCM	CCA
Before SSE	61	113	38	95	167
After SSE	111	162	77	151	169

Human data:

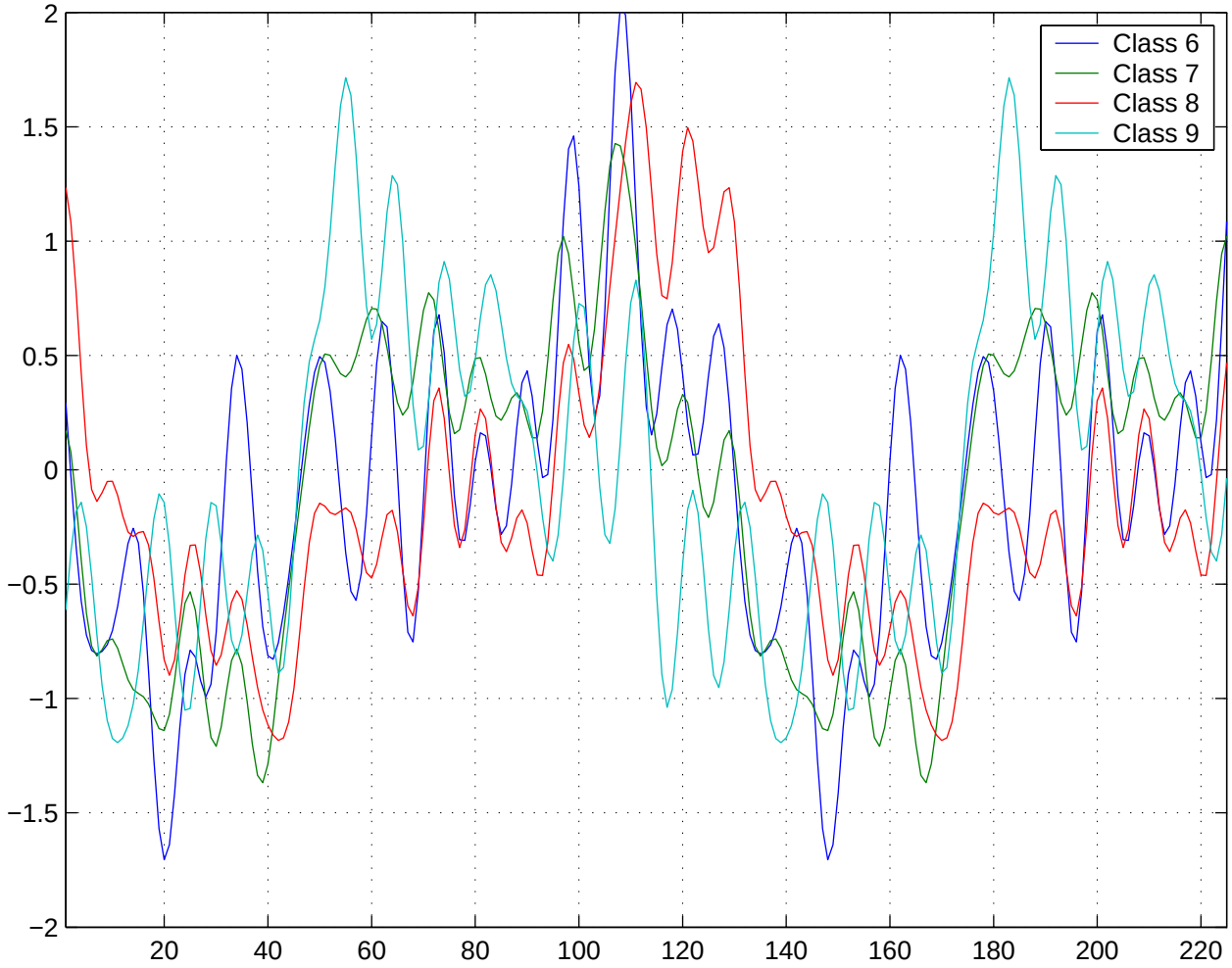
Timesequence realizations



First 5 clusters

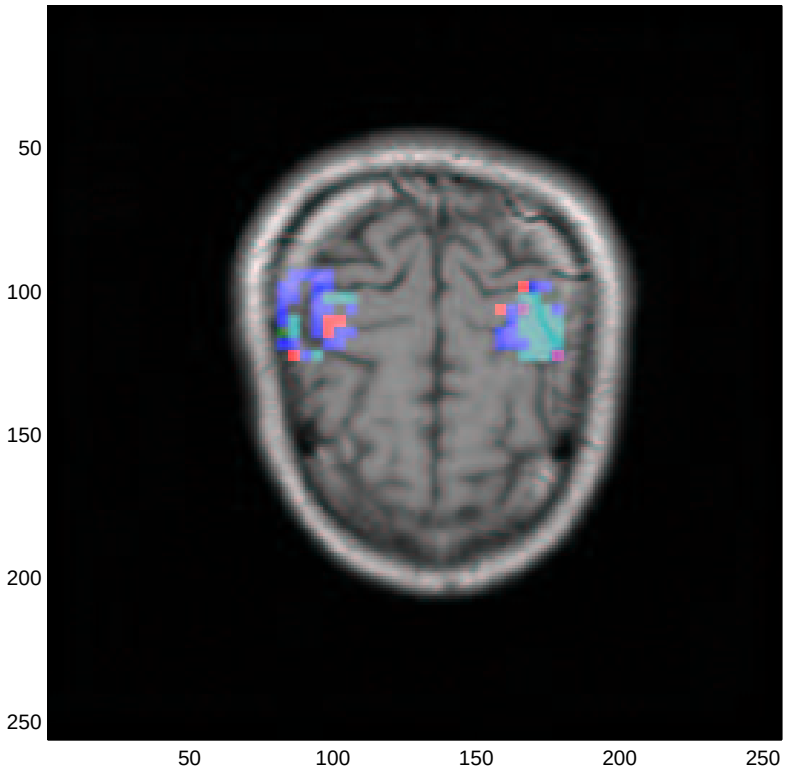
Human data:

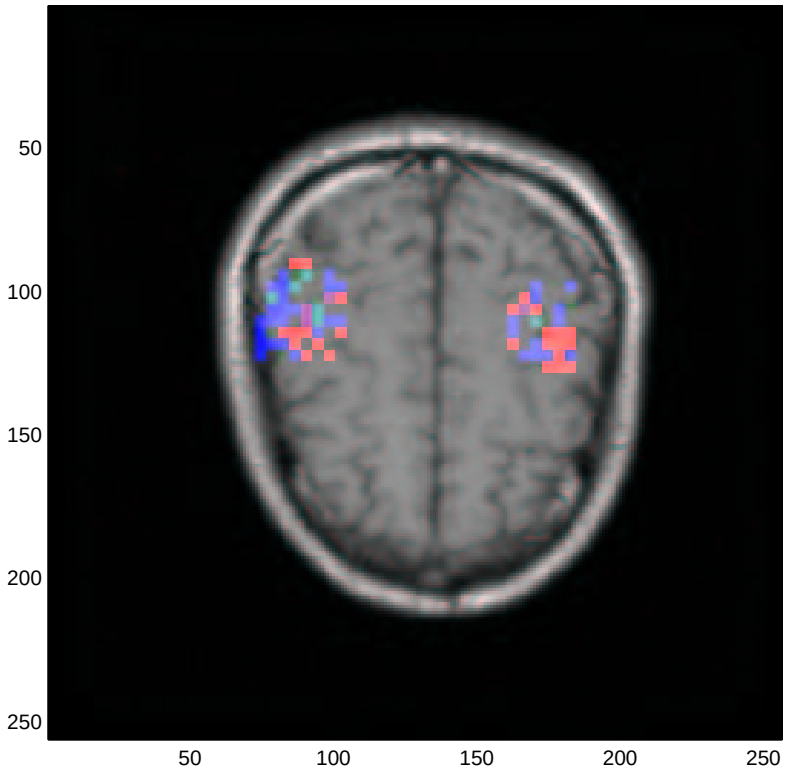
Timesequence realizations



Clusters 6-9

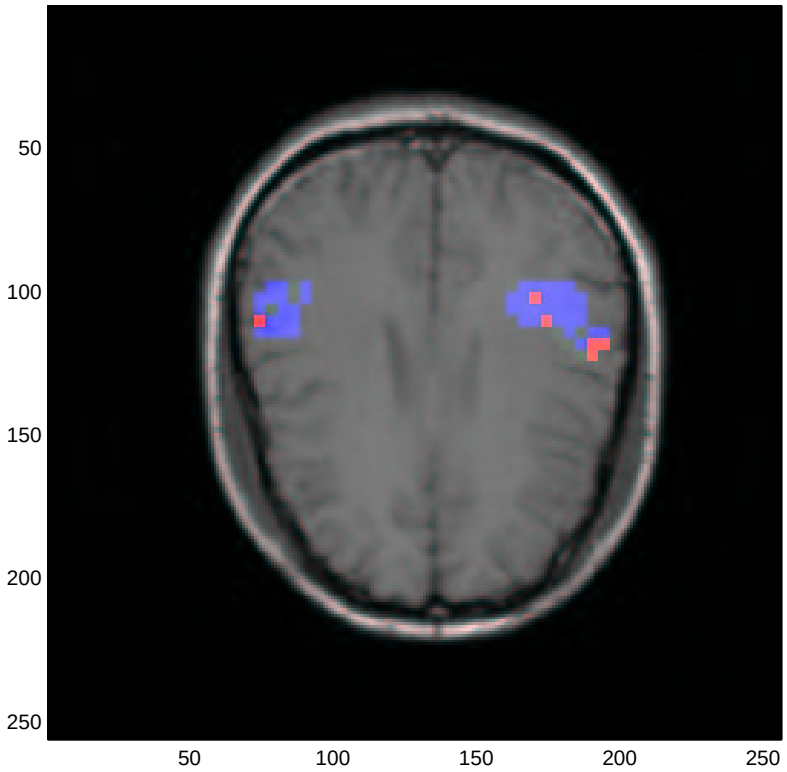
Human data: Hard classifications

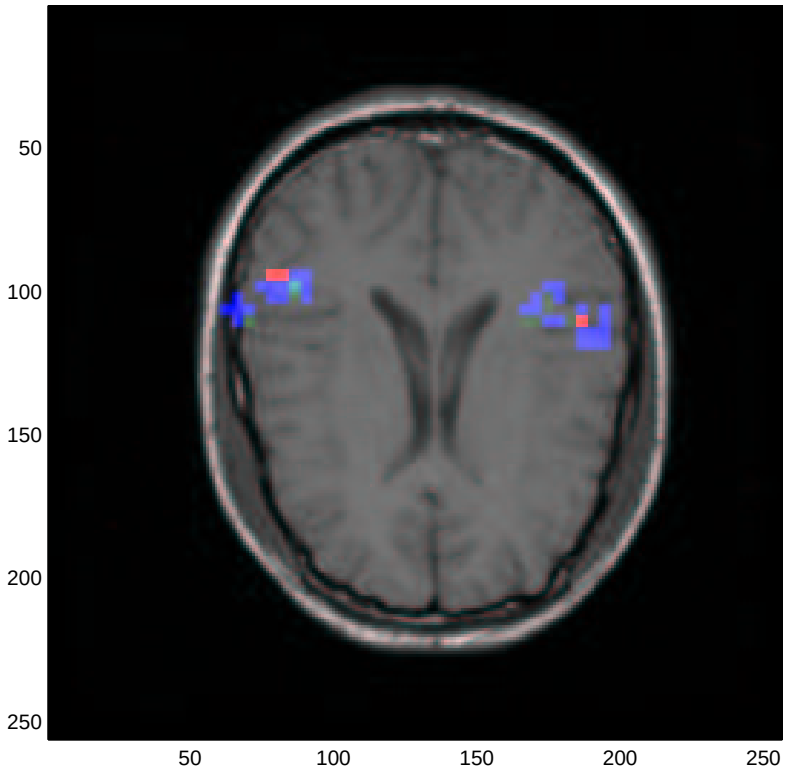




Motor cortex (first 5 clusters): (L) Upper slice, (R) Lower slice

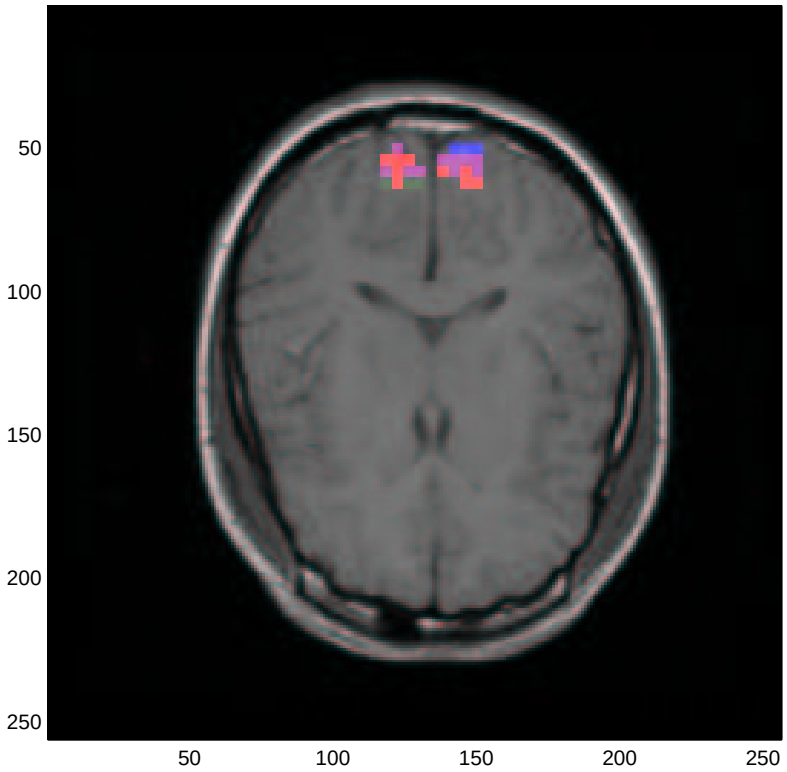
Human data: Hard classifications

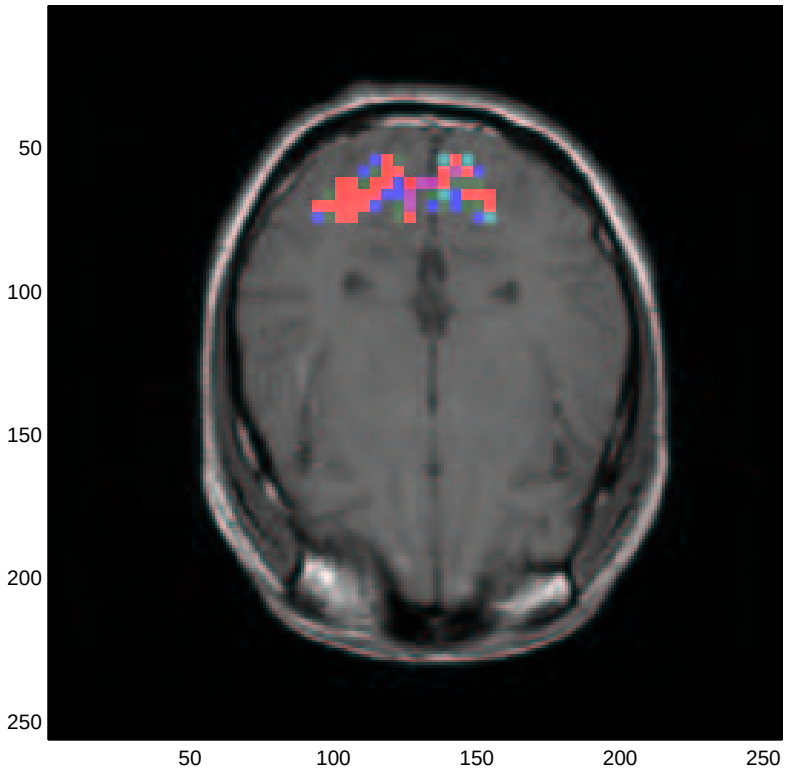




Auditory cortex (first 5 clusters): (L) Upper slice, (R) Lower slice

Human data: Hard classifications





Visual cortex (first 5 clusters): (L) Upper slice, (R) Lower slice

Conclusions

- Clustered component analysis is a new method of extracting signals where only shape, not amplitude, is important
- CCA has been shown to perform well on simulated data
- The experimental results show the following:
  - The distinct neuronal signals do not correlate strongly with the known functional cortices
  - The clusters tend to lie along sulcal-gyral boundaries, possibly correlated with vasculature
- CCA can be used with dimensionality reduction strategies other than the ones used in our experiments
- CCA may also be adapted for use with applications other than fMRI

Acknowledgements

- Major Professors: Dr. Mark J. Lowe and Dr. Charles A. Bouman
- Committee Members: Dr. Peter C. Doerschuk and Dr. Edward J. Delp
- Department of Biomedical Engineering and Division of Imaging Sciences