A New Generation of Brain-Computer Interfaces Driven by Discovery of Latent EEG-fMRI Linkages Using Tensor Decomposition

Gopikrishna Deshpande
AU MRI Research Center
AU Department of Electrical and Computer Engineering
AU Department of Psychology
Alabama Advanced Imaging Consortium

Rangaprakash Deshpande
Department of Psychiatry and Biobehavioral Sciences
UCLA

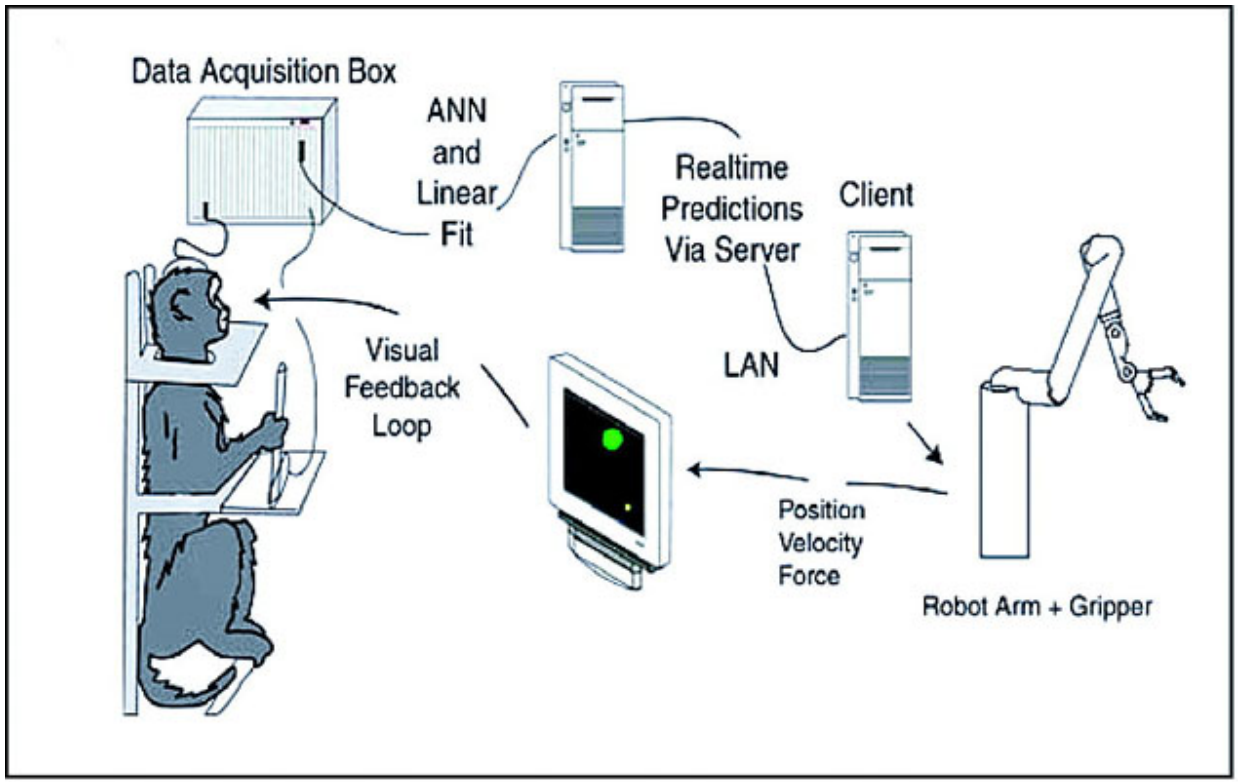
Luke Oeding
Department of Mathematics and Statistics
Auburn University

HYPOTHESIS AND THEORY ARTICLE
Frontiers in Neuroscience, 07 June 2017

Andrej Cichocki
Skolkovo Institute of Science and Technology (Skoltech), Moscow, Russia
Nicolaus Copernicus University (UMK), Torun, Poland
Systems Research Institute, Polish Academy of Science, Warsaw, Poland

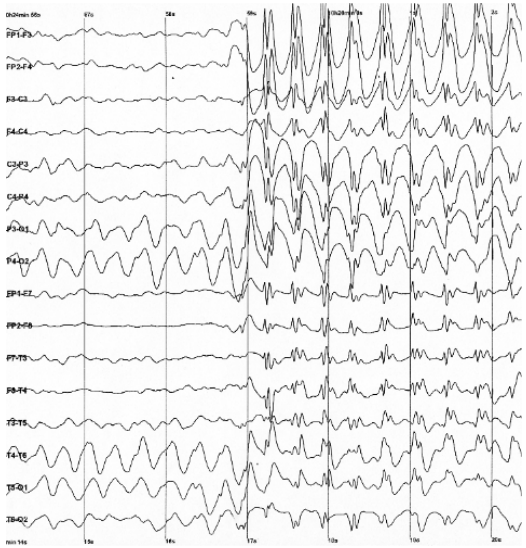
Xiaoping Hu
Professor of Biomedical Engineering
University of California, Riverside

Brain Computer Interface



https://en.wikipedia.org/wiki/File:Brain-computer_interface_(schematic).jpg

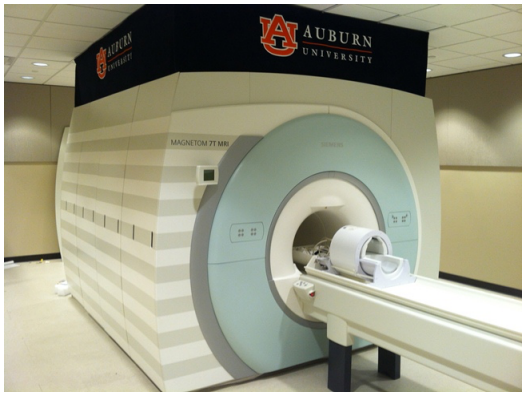
EEG Based BCI

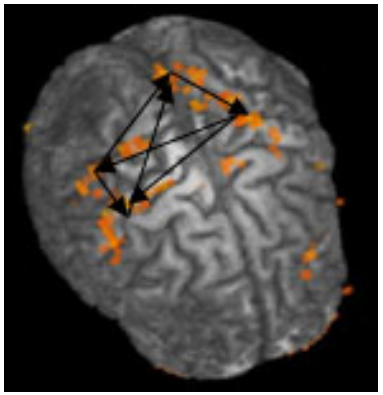


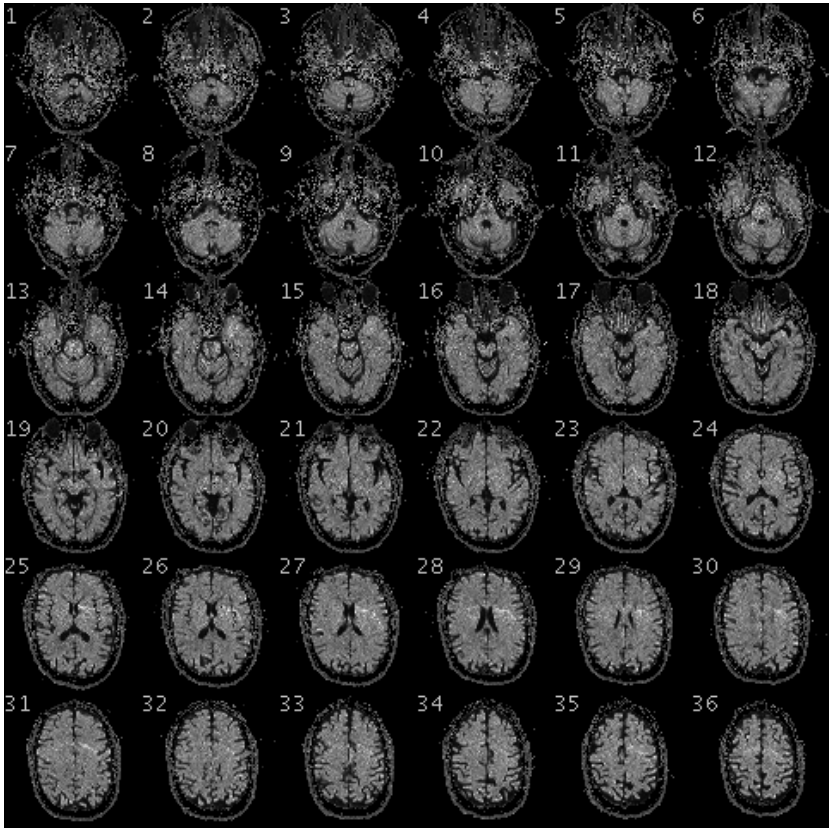
- + Non-invasive
- + High temporal resolution for real-time interaction
- + Inexpensive, light weight, and highly portable
- - Poor spatial specificity
- - EEG Signals in different channels are highly correlated, reducing ability to distinguish neurological processes.
- - Long training time required

Real-time fMRI Based BCI

- + High spatial specificity -- more accuracy
- - High cost
- - Non-portable
- - Low temporal resolution
- - Restrictive environment



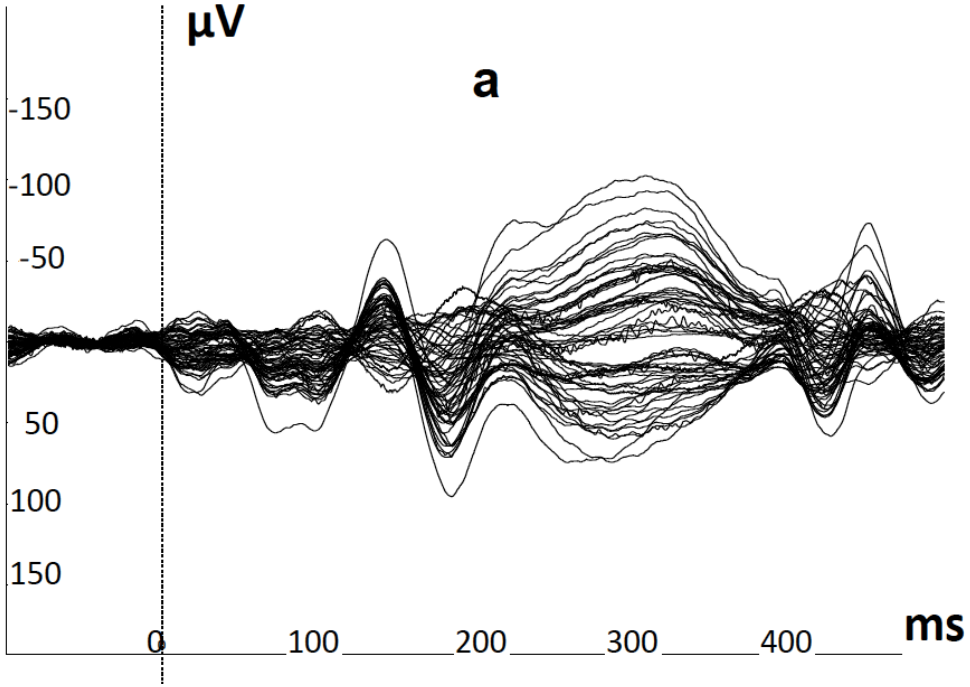




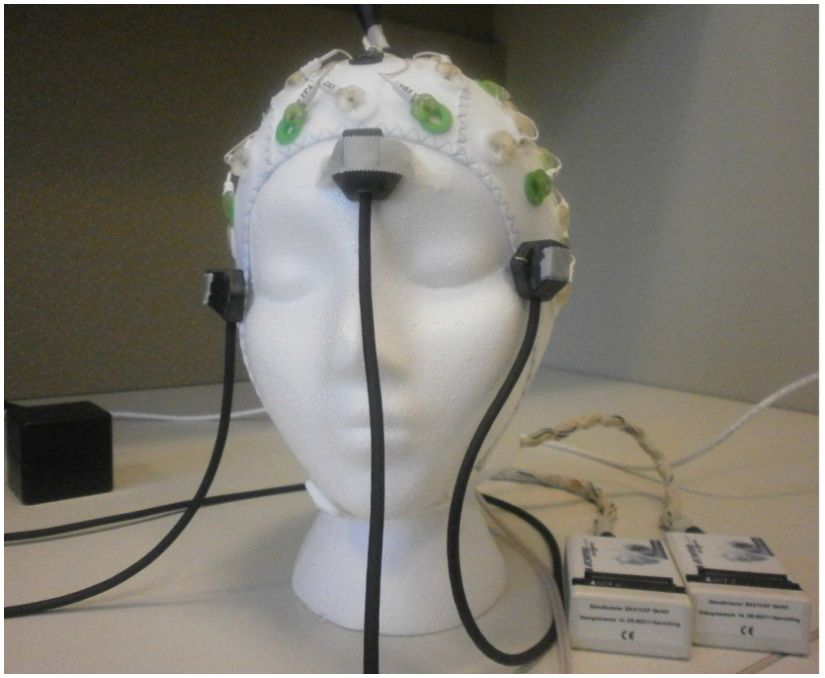
Simultaneous EEG – fMRI data acquisition

Fig.3 The stimulus grid used in the P300 based speller task

EEG Data



EEG data were epoched with respect to R peaks of EKG signal and averaged over trials.
The ballistocardiogram (BCG) artifact in the EEG signal obtained inside MR scanner is removed.



Electrode position measurements via Polhemus Fastrak 3D Digitizer System

Objective: Discovery of Latent Linkages between EEG and fMRI and improve BCI

- Hypothesis: latent linkages between EEG and fMRI can be exploited to estimate fMRI-like features from EEG data.
- This could allow an independently operated EEG-BCI to decode brain states in real time, with better accuracy and lower training time.
- Hypothesis: Features from a sub-set of subjects can be generalized to new subjects (for a homogeneous set of subjects).

Strategies

- Obtain fMRI data with high temporal resolution:
  - Use multiband echo-planar imaging (M-EPI) [Feinberg, et al. 2010] to achieve whole brain coverage with sampling intervals (TR) as short as 200 ms.
  - View fMRI as convolution of HDF (Hemodynamic response function) and neuronal states. Use cubature Kalman filter based blind deconvolution of fMRI [Havlicek, et al. 2011] to recover driving neuronal state variables with higher effective temporal resolution.
- Obtain clean EEG data:
  - EEG signal sampled at 5000Hz to ensure accurate gradient artifact removal, then downsampled to 250Hz to make dataset more manageable.
- Use the complex Morlet wavelet [Teolis, 1998] to give a time-frequency representation of both EEG and fMRI for each trial.

Discover latent linkages between EEG and fMRI

- Simultaneous EEG/fMRI data collected using a P300 speller based paradigm.
- EEG modalities: trial–time–frequency–channel
  - 4ms updates, 63+1 channels, 4 trials
- fMRI modalities: trial–time–neuronal state–voxel
  - 200ms updates with whole brain coverage and 3mm voxels
- Apply Orthogonal Decomposition to each EEG and fMRI. [Zhou and Cichocki 2012]
- The first dimension of "trials" is the same for both tensors, permitting the application of HOPLS. This important property allows both EEG and fMRI to be sampled at different rates.
- It is not required to downsample EEG to fMRI's temporal resolution, as done by most researchers in the EEG-fMRI comparison literature (Goldman, et al. 2002) (Hinterberger, Veit, et al. 2005), which will lead to loss of vital temporal information.

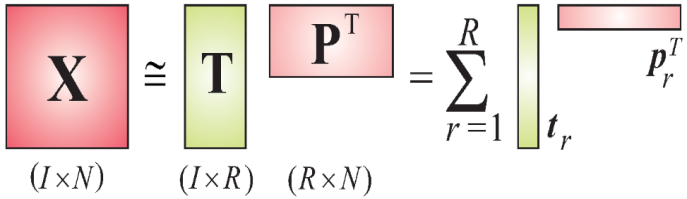
Discover latent linkages between EEG and fMRI

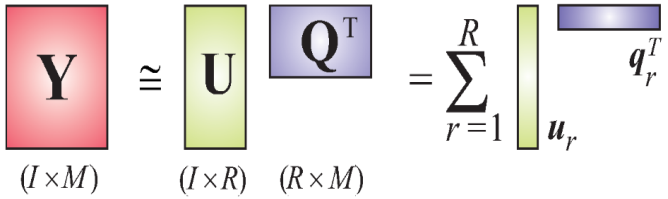
- Assumptions: EEG data is the independent variable *X, and* deconvolved fMRI (neuronal states) data is the dependent variable *Y*.
  - Reasonable assumption because the hemodynamic/metabolic activity is a secondary response to the electrical activity.
- Goal: Given *X* and *Y* over many trials, and assuming *F(X) = Y*, discover *F*.
  - Higher Order Multilinear Subspace Regression / Higher Order Partial Least Squares (HOPLS) [Q. Zhao, et al. 2011] to predict the dependent variable (deconvolved fMRI) from the independent variable (EEG).
  - HOPLS parameters (latent variables, core tensors and tensor loadings) are likely to provide information on latent EEG-fMRI relationships across the dimensions considered.

Partial Least Squares (PLS)

Partial least squares: Predicts a set of dependent variables *Y* from a set of independent variables *X*.
Attempts to explain as much as possible the covariance between X and Y.

PLS optimization objective is to maximize pairwise covariance of a set of latent variables by projecting both *X* and *Y* onto new subspaces.





$$X = TP^T + E = \sum_{r=1}^{R} t_r\, p_r^T + E$$

$$Y = UQ^T + F = \sum_{r=1}^{R} u_r\, q_r^T + F \qquad (3)$$

$T = [t_1, t_2, \cdots, t_R]$ and $U = [u_1, u_2, \cdots, u_R]$ are matrices of *R* extracted latent variables from *X* and *Y*, respectively.
U will have maximum covariance with *T* column-wise.
P and *Q* are latent vector subspace base loadings.
E and *F* are residuals.

The relation between *T* and *U* can be approximated as $U \approx TD$
where *D* is an *R*×*R* diagonal matrix of regression coefficients.

Least Squares (Undergraduate Linear Algebra)

- Given a linear transformation $P \rightarrow Q$ we want to simultaneously predict the subspaces $\mathbb{R}^m \subset P$ and $\mathbb{R}^n \subset Q$ so that the restricted map $\mathbb{R}^m \rightarrow \mathbb{R}^n$ gives a good approximation of the mapping.

- Given an underdetermined matrix equation $A\,\vec{x} = \vec{b}$, we can attempt to square the system and solve: $A^\top A\,\vec{x} = A^\top \vec{b}$
  - Perhaps use QR.
- The standard least-squares solution is $\hat{x} = (A^\top A)^{-1}\ A^\top \vec{b}$.
  - Project to a linear subspace and replace *A* with a full rank matrix using SVD.

Singular Value Decomposition

- The Singular Value Decomposition $A = U \Sigma V^\top$
  - The quasi-diagonal matrix of singular values $\sigma_1, \sigma_2, \ldots$ can be truncated to the largest *r* singular values to give the best rank *r* approximation.
  - The orthogonal matrices *U* and *V* (called *loadings*) give the embeddings of the respective subspaces on which *A* is best approximated to rank *r*.
  - The pseudoinverse of *A* is $A^\dagger = V \Sigma^\dagger U^\top$, where $\Sigma^\dagger = \mathrm{diag}(\sigma_1^{-1}, \sigma_2^{-1}, \ldots)$
  - The minimal norm solution to $A\vec{x} = \vec{b}$ is $\hat{x} = A^\dagger \vec{b} = \Sigma_{i=1}^{r} \frac{u_i^\top \vec{b}}{\sigma_i} v_i$.

SVD and PLS

- Given *m* data observations of *n* participants stored in a data matrix *X* (independent variables).
- Given *k* responses of the *n* participants stored in a data matrix Y (dependent variables).
- Find a linear function *F* that explains the maximum covariance between *X* and *Y*.

$$Y = XF + E$$

- Center and normalize both *X* and *Y*.
- Compute the Covariance Matrix $R = Y^{\top} X$
- Perform SVD: $\mathrm{R} = U\Sigma V^{\top}$ (compact form, iterative algorithm)
- The latent variables of *X* and *Y* are obtained by projections:

$$L_X = XV \quad L_Y = YU$$

  - *U* and *V* give the embeddings of the subspaces (the loadings of the variables)

Structured variables

- In the situation of EEG-fMRI data, even after a wavelet decomposition of the data, we still have extra structure in the dependent variables (fMRI) *Y* and in the independent variables *X.*
- EEG modalities: trial–time–frequency–channel
  - 4ms updates, 63+1 channels, 4 trials
  - So X could be 4 trials x 100 wavelets x 63 channels
- fMRI modalities: trial–time–neuronal state–voxel
  - 200ms updates with whole brain coverage and 3mm voxels
  - So Y could be 4 trials x 200 wavelets x 36,000 voxels
- Don't think of 100x63 as 6,300, don't think of 200x36,000 as 7.2×10^6
  - Unfolding leads to "small *p* large *n*" problem and a loss of information.

Modal Products for Tensors

- For $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ and $U \in \mathbb{R}^{J_n \times I_n}$ the n^{th} mode tensor-matrix product is

$$\mathcal{A} \times_n U \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \times \cdots \times I_N}$$

$$\mathcal{A} \times_n U := \left(\sum_{i_n \in I_n} a_{i_1, i_2, \ldots, i_n, \ldots i_N} u_{j_n, i_n} \right)$$

- This modal product generalizes the matrix product and vector outer product and replaces transpose.

If $A \in \mathbb{R}^{I_1 \times I_2}$ and $B \in \mathbb{R}^{I_1 \times J_2}$ then $A \times_1 B = B^\top A \in \mathbb{R}^{J_2 \times I_2}$

If $\vec{x} \in \mathbb{R}^{1 \times n}$ and $\vec{y} \in \mathbb{R}^{1 \times m}$ then $\vec{y}^\top \vec{x} \in \mathbb{R}^{m \times n}$

Matrix SVD using modal product notation

- SVD Theorem: Every complex $I_1 \times I_2$ matrix F has an expression

$$F = S \times_1 U^{(1)} \times_2 U^{(2)}$$

with

$U^{(1)}$ a unitary $I_1 \times I_1$ matrix

$U^{(2)}$ a unitary $I_2 \times I_2$ matrix

S pseudodiagonal $I_1 \times I_2$ matrix, $S = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_{\min\{I_1, I_2\}})$

The singular values are ordered: $\sigma_1 \geq \sigma_2 \geq \cdots, \geq \sigma_{\min\{I_1, I_2\}} \geq 0$

Tensor SVD (Orthogonal Tucker Decomposition)

[De Lauthawer 2005, Zhao-Cichocki 2013]

Theorem:

- Every $I_1 \times I_2 \times \cdots \times I_N$ array $\mathcal{A}$ can be written as a product:
$$\mathcal{A} = S \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_N U^{(N)}$$
- Each $U^{(n)}$ is a unitary $I_n \times I_n$ matrix.
- S is a $I_1 \times I_2 \times \cdots \times I_N$ complex tensor with slices having norm $\left\| S_{i_n=i} \right\| = \sigma_i^{(n)}$, the *n*-mode singular values of $\mathcal{A}$
- For each *n* the singular values are ordered $\sigma_1^{(n)} \geq \sigma_2^{(n)} \geq \cdots \geq \sigma_{I_n}^{(n)} \geq 0$
- The slices $S_{i_n=i}$ are *all-orthogonal*:
$$\left\langle S_{i_n=\alpha}, S_{i_n=\beta} \right\rangle = 0 \quad \forall \alpha \neq \beta \quad \forall n$$
- Compute the *n*-mode singular matrix $U^{(n)}$ and *n*-mode singular values by the matrix SVD of the *n*-th unfolding of size $I_n \times I_2 I_3 \cdots I_{n-1} I_{n+1} I_N$.
- S is computed by $S = \mathcal{A} \times_1 U^{(1)^*} \times_2 U^{(2)^*} \cdots \times_N U^{(N)^*}$

Tucker Decomposition for EEG--fMRI

- Take an $N_1 \times N_2 \times N_3 \times N_4$ tensor and express it as a (small) core tensor of size $L_1 \times L_2 \times L_3 \times L_4$ together with changes of bases (loadings) to put the core back into the larger tensor space.
- Let $\underline{X}$ and $\underline{Y}$ be tensors of EEG and deconvolved fMRI, respectively with modalities: trials, voxels/channels, time and frequency.
- Obtain new tensor subspaces via the Tucker model for each trial:
  - Approximate $\underline{X}$ with a sum of multilinear rank-$(1, L_2, L_3, L_4)$ terms
  - Approximate $\underline{Y}$ with a sum of multilinear rank-$(1, K_2, K_3, K_4)$ terms
- The core tensors model the underlying biophysics and are different for EEG and fMRI.
  - Perform HOSVD on the $L_2 \times L_3 \times L_4 \times K_2 \times K_3 \times K_4$ contraction $\underline{X} \times_1 \underline{Y}$

Higher order Partial Least Squares (HOPLS)

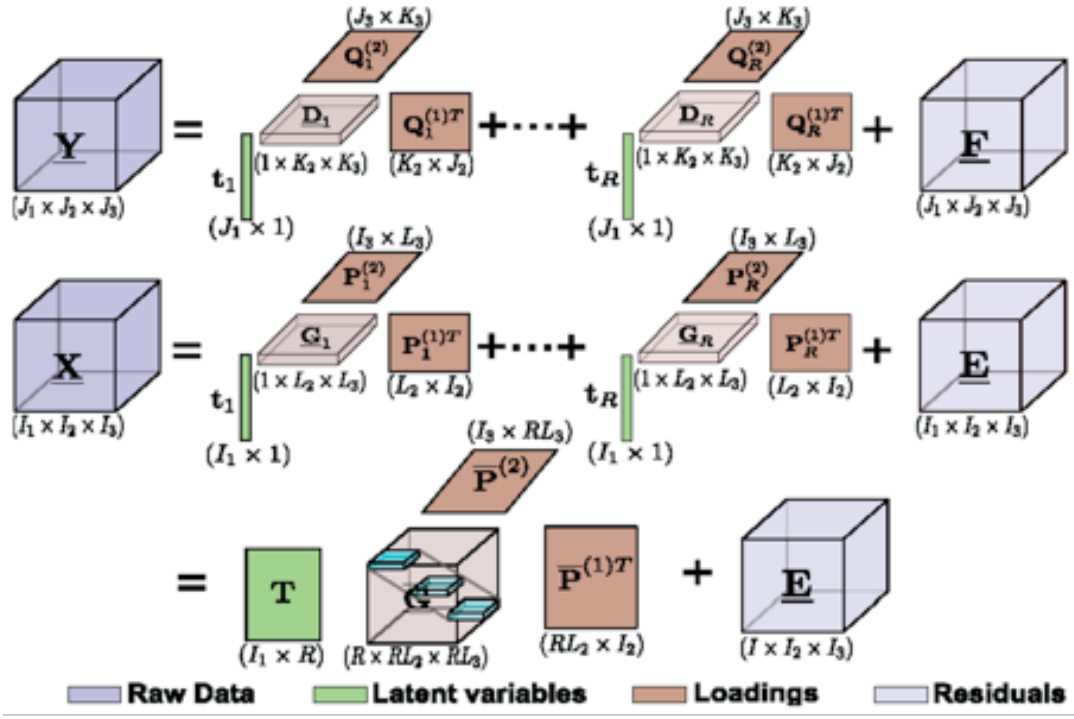


Fig.2 Schematic diagram of the HOPLS model

The HOPLS is expressed as

$$\underline{Y} = \sum_{r=1}^{R} \underline{D_r} \times_1 t_r \times_2 Q_r^{(2)} \times_3 Q_r^{(3)} \times_4 Q_r^{(4)} + \underline{F}$$

$$\underline{X} = \sum_{r=1}^{R} \underline{G_r} \times_1 t_r \times_2 P_r^{(2)} \times_3 P_r^{(3)} \times_4 P_r^{(4)} + \underline{E}$$

where R is the number of latent vectors,

t_r is the r^{th} latent vector,

$P_r^{(n)}$ and $Q_r^{(m)}$ are loading matrices corresponding to latent vector t_r on mode-n and mode-m, respectively,

$\underline{G_r}$ and $\underline{D_r}$ are core tensors,

and $\times_k$ is the product in the k^{th} mode.

Compute the t_r as the leading left singular vector of an unfolding, deflate, and repeat.

Feasibility Study

- We performed the simultaneous EEG/fMRI experiment and EEG-only BCI using the P300 speller paradigm in 4 right-handed male subjects (mean age: 21.5 years) with no history of neurological or other illness.
- fMRI data were acquired using the M-EPI sequence (TR=200 ms, multiband factor=8, 3 mm isotropic voxels, full coverage) and deconvolved using the cubature Kalman filter approach.
- The analyses were carried out on a high performance computer with Intel® Core™ i7-3820 (Quad Core, 10MB Cache) overclocked up to 4.1GHz processor with a top of the line NVidia GPU Quadro Plex 7000.
- We obtained significantly high correlation using both the full and the significant HOPLS models, with the latter providing better accuracy with run times under a second.

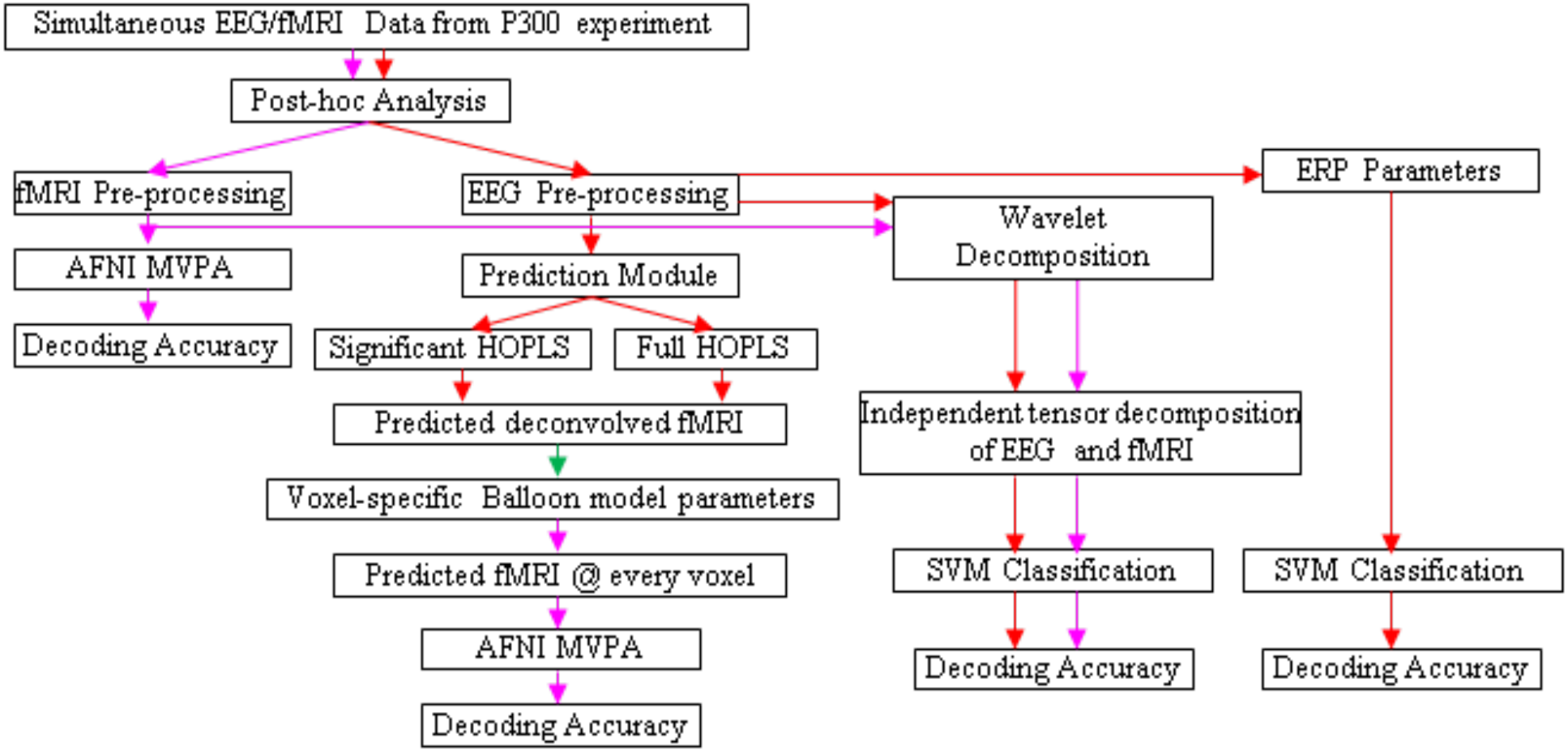


Fig.4 *Schematic for letter decoding from post-hoc analysis of simultaneous EEG/fMRI data. Arrow legend: red: EEG, magenta: fMRI, green: deconvolved fMRI*

Preliminary Results

Table.2 Prediction of deconvolved fMRI from simultaneously acquired EEG using offline analysis

Offline analysis of Simultaneous EEG/fMRI Expt	Full HOPLS forward model	Significant HOPLS forward model
Correlation between deconvolved fMRI data and that predicted from EEG	0.76 ± 0.17	0.84 ± 0.13
Approximate run time for 'prediction module' in sec	1.4	0.8

Table.3 Letter decoding accuracy from post-hoc analysis of simultaneous EEG/fMRI data

Off line analysis of simultaneous EEG/fMRI Expt		Original fMRI MVPA	fMRI predicted with significant HOPLS + MVPA	fMRI predicted with full HOPLS + MVPA	SVM based on EEG tensors (from sequential model)	SVM based on fMRI tensors(from sequential model)	SVM based on ERP amplitude and latency
Letter decoding accuracy	1 trial block	0.97±0.03	0.94 ± 0.04	0.93±0.05	0.84±0.10	0.86±0.12	0.68 ± 0.17
	8 trial blocks	1	1	1	0.98±0.02	0.98±0.02	0.84 ± 0.11
Run time per letter decoded (sec)		0.9	1.8	2.4	0.13	0.24	0.08

Preliminary results

Table.4 Letter decoding accuracy from real-time analysis of EEG data using predicted fMRI (from significant HOPLS) as features for MVPA

Online analysis of EEG-only BCI data		Parameters from same subject's EEG/fMRI run	Parameters from random prior subject's EEG/fMRI run	Parameters learned from all prior subjects' EEG/fMRI run		
				Subject-2	Subject-3	Subject-4
Letter decoding accuracy from fMRI predicted with significant HOPLS + MVPA	1 trial block	0.93 ± 0.04	0.87 ± 0.11	0.86	0.91	0.93
	8 trial blocks	1	0.94 ± 0.04	0.93	0.93	0.95

In spite of these encouraging results, we stress the fact that they are derived from a small, homogeneous sample of 4 subjects. We need to do more trials to demonstrate more broad generalizability.

(Extra Slide) Higher Order Partial Least Squares

- The subspace transformation is optimized using the following objective function, yielding the common latent variable t_r instead of 2 latent variables.
- $min_{\{P^{(n)},Q^{(n)}\}} \left\| \underline{X} - [\underline{G};\ t, P^{(2)}, P^{(3)}, P^{(4)}] \right\|^2 + \left\| \underline{Y} - [\underline{D};\ t, Q^{(2)}, Q^{(3)}, Q^{(4)}] \right\|^2$

$$\text{such that } \left\{ P^{(n)T} P^{(n)} \right\} = I_{L_n} \quad and \quad \left\{ Q^{(m)T} Q^{(m)} \right\} = I_{K_m}$$

- Simultaneous optimization of subspace representations and latent variable t_r. Solutions can be obtained by Multilinear Singular Value Decomposition (MSVD) (see [Q. Zhao, et al. 2011])
- Minimizing the errors is equivalent to maximizing the norms $\|G\|$ and $\|D\|$ simultaneously (accounting for the common latent variable). To do this we maximize the product $\|G\|^2 \cdot \|D\|^2$.
- Compute the latent variables t_r as leading left singular vectors and then deflate. Repeat until you reach the error bounds you want.