Constrained Principal Component Analysis for fMRI Data: relating fMRI signal to experimental conditions

Todd S. Woodward, Ph.D.

Assistant Professor
Department of Psychiatry, University of British Columbia
Vancouver, Canada

MSFHR Scholar
CIHR New Investigator
NARSAD Young Investigator

Constrained Principal Component Analysis (CPCA) is a method that was (and is continuing to be) developed by Yoshio Takane of McGill university.

Takane Y, Hunter MA. (2001): Constrained principal component analysis: A comprehensive theory. Applicable Algebra in Engineering, Communication and Computing 12:391-419.

Takane Y, Shibayama T. (1991): Principal component analysis with external information on both subjects and variables. Psychometrika 56(1):97-120.

Constrained Principal Component Analysis (CPCA) is a method that was (and is continuing to be) developed by Yoshio Takane of McGill university.

Takane Y, Hunter MA. (2001): Constrained principal component analysis: A comprehensive theory. Applicable Algebra in Engineering, Communication and Computing 12:391-419.

Takane Y, Shibayama T. (1991): Principal component analysis with external information on both subjects and variables. Psychometrika 56(1):97-120.

CPCA is a general method for structural analysis of multivariate data that combines regression analysis and principal component analysis into a unified framework.

Constrained Principal Component Analysis (CPCA) is a method that was (and is continuing to be) developed by Yoshio Takane of McGill university.

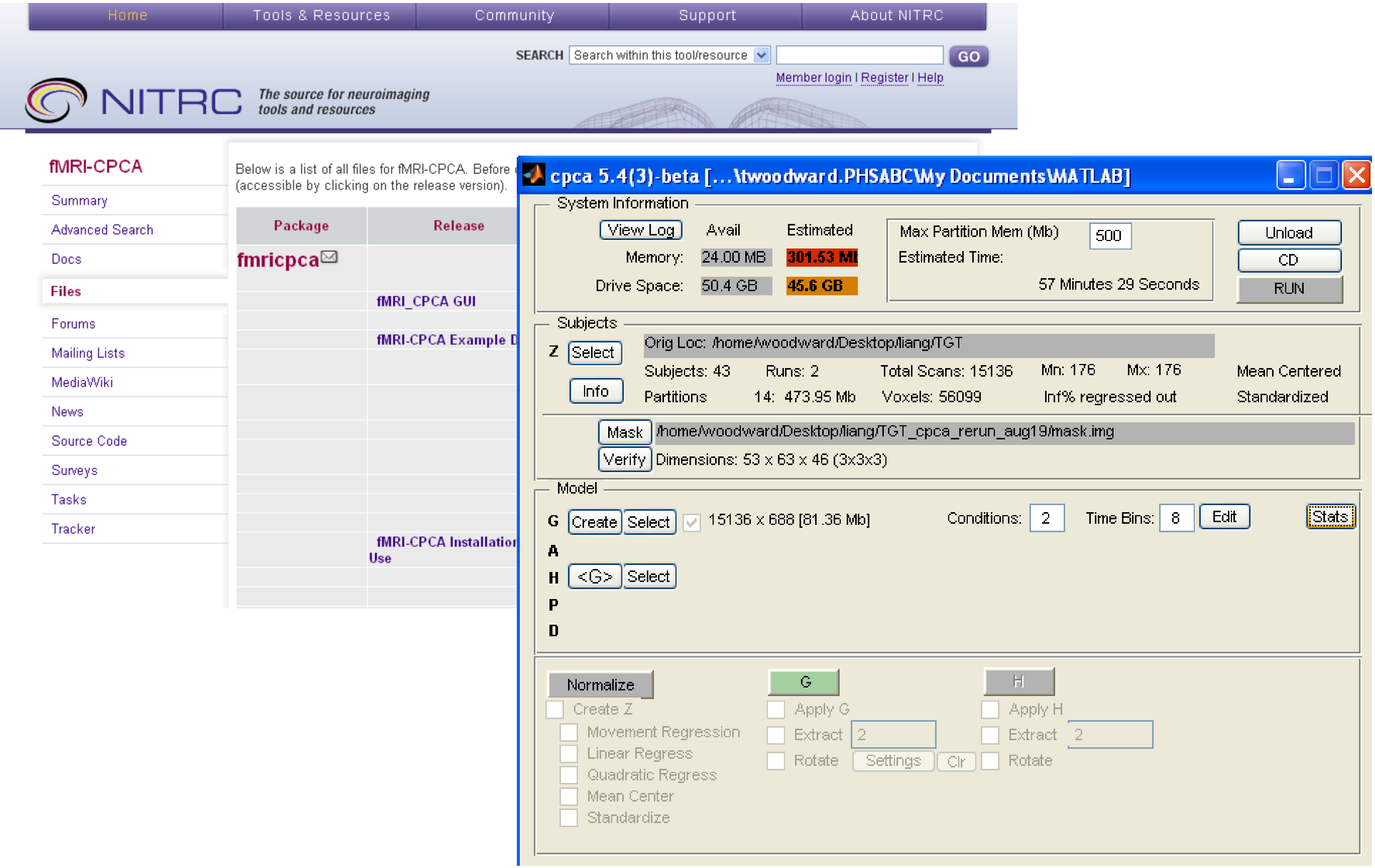
Takane Y, Hunter MA. (2001): Constrained principal component analysis: A comprehensive theory. Applicable Algebra in Engineering, Communication and Computing 12:391-419.

Takane Y, Shibayama T. (1991): Principal component analysis with external information on both subjects and variables. Psychometrika 56(1):97-120.

CPCA is a general method for structural analysis of multivariate data that combines regression analysis and principal component analysis into a unified framework.

Applied to functional magnetic resonance imaging (fMRI) data, CPCA can identify components directly relevant to experimental conditions of interest by integrating this information prior to computation of components.

fMRI-CPCA software is available for free download (MATLAB code)



- # Principal Component Analysis (PCA)

Table 1. Rotated Component Matrix for Signs and Symptoms of Psychotic Illness items (N = 68).

	Negative	Disorganized	Positive
Underactivity	**0.87**	-0.10	0.04
Flattened affect	**0.87**	-0.07	-0.23
Poverty of speech	**0.67**	-0.29	-0.40
Inappropriate affect	-0.27	**0.75**	-0.02
Disordered form of thought	0.00	**0.88**	0.08
Delusions	-0.19	0.17	**0.80**
Hallucinations	-0.07	-0.10	**0.82**

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. All loadings above .60 are written in bold ink.

- Components represent separable patterns of intercorrelation

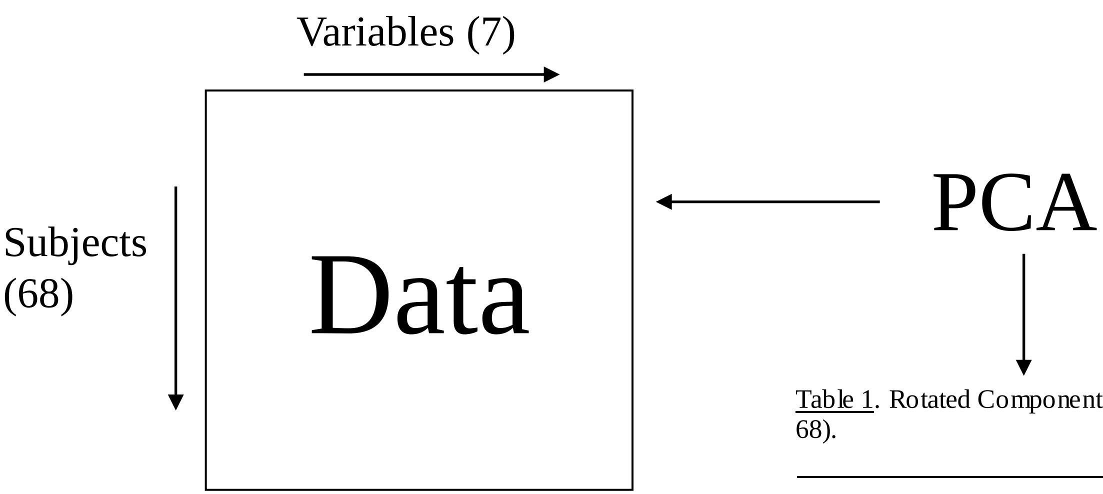


Table 1. Rotated Component Matrix for Signs and Symptoms of Psychotic Illness items (N = 68).

	Negative	Disorganized	Positive
Underactivity	**0.87**	-0.10	0.04
Flattened affect	**0.87**	-0.07	-0.23
Poverty of speech	**0.67**	-0.29	-0.40
Inappropriate affect	-0.27	**0.75**	-0.02
Disordered form of thought	0.00	**0.88**	0.08
Delusions	-0.19	0.17	**0.80**
Hallucinations	-0.07	-0.10	**0.82**

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. All loadings above .60 are written in bold ink.

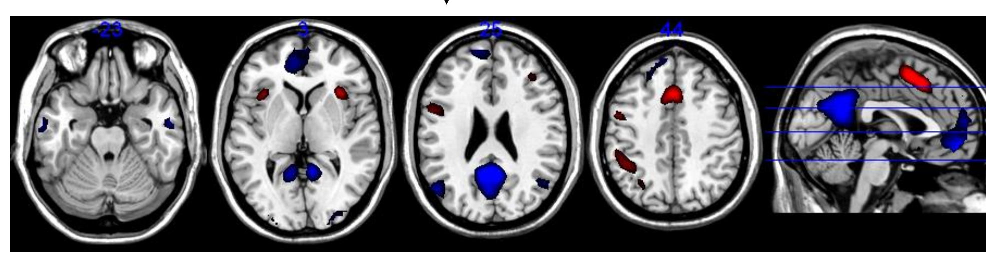
Principal Component Analysis (PCA) with fMRI

Voxels (~ = 25,000)

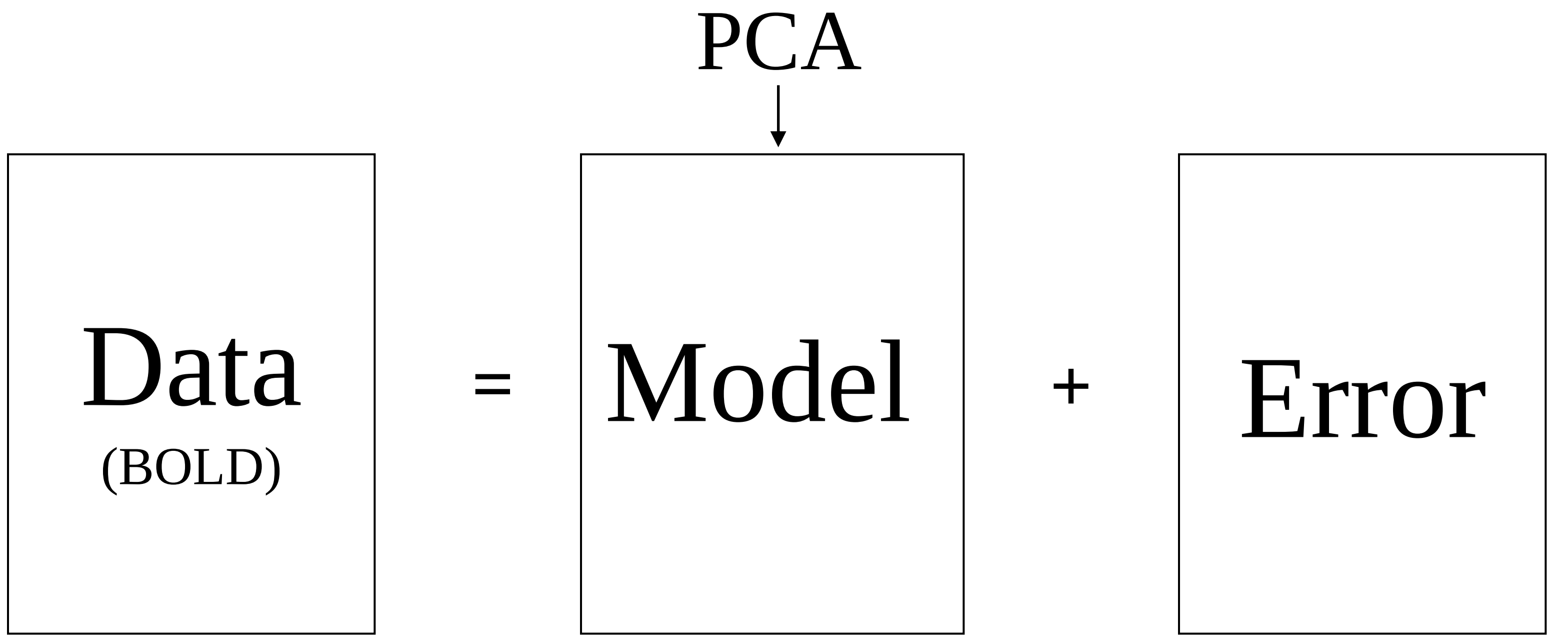
Scans
X
Subjects
(~=3000)

BOLD

PCA
(based on singular value decomposition [SVD])



Constrain the variance analyzed with PCA to that predictable from the timing of the conditions of interest (Model).



Model Types:

Finite Impulse Response Model (FIR model)

Hemodynamic Response Function Model (HRF model)

Constrained Principal Component Analysis (CPCA) – basic equations for FIR model

Regression-based constraints (external analysis)

$$ {}_{12840}Z_{23621} = {}_{12840}G_{960}C_{23621} + {}_{12840}E_{23621} $$

The columns in this subject-and-condition based Finite Impulse Response (FIR) model *G* matrix code eight peristimulus time points for each of four load conditions (2, 4, 6 and 8 letters) for each of 30 subjects, totaling 960 columns (8 × 4 × 30 = 960)

Data (BOLD)	=	Model (FIR model)	+	Error

Constrained Principal Component Analysis (CPCA) – basic equations

Regression-based constraints (external analysis)

$${}_{12840}Z_{23621} = {}_{12840}G_{960}C_{23621} + {}_{12840}E_{23621}$$

Singular Value Decomposition (internal analysis)

$${}_{12840}U_{3}D_{3}V'_{23621} \approx {}_{12840}G_{960}C_{23621}$$

Z

Data

(BOLD)

=

GC

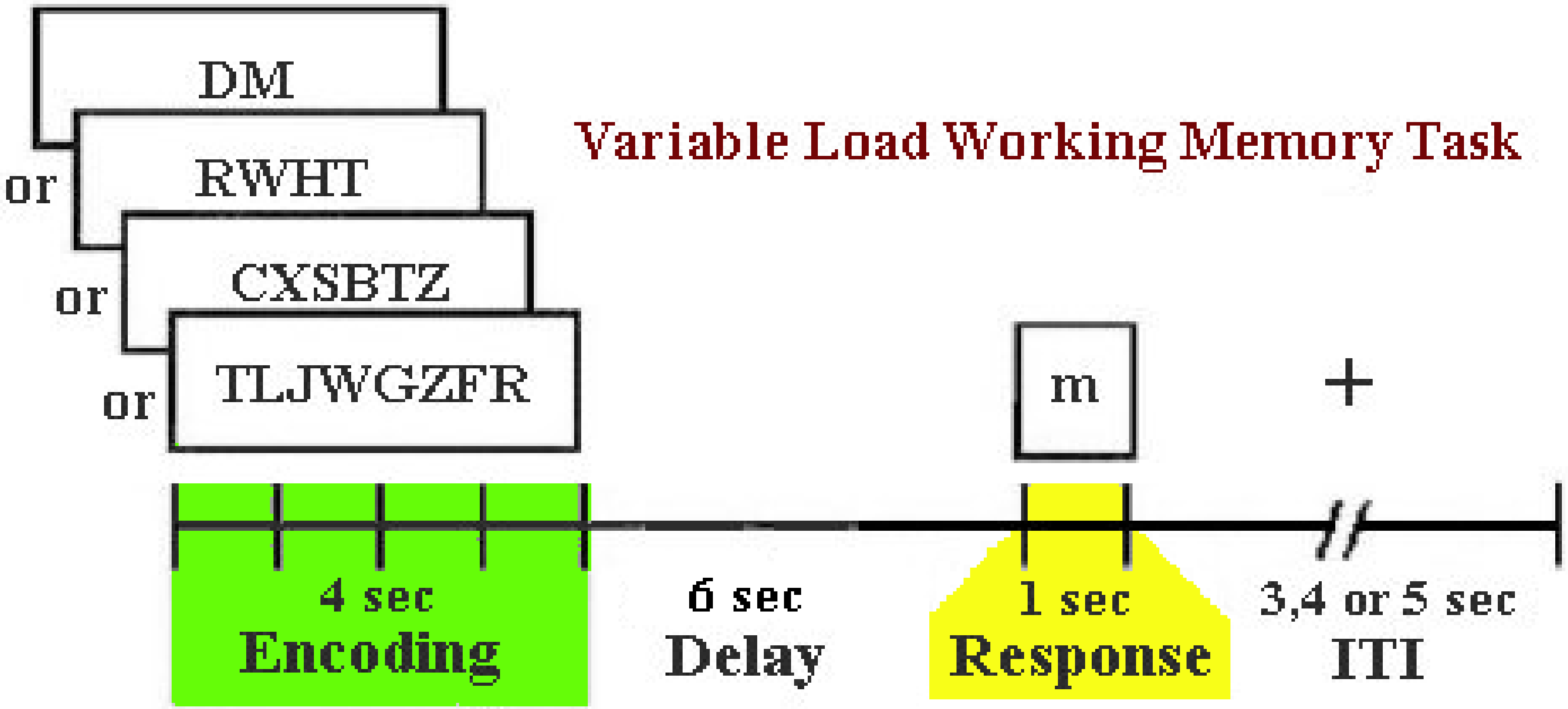
Model

(FIR model)

+

E

Error

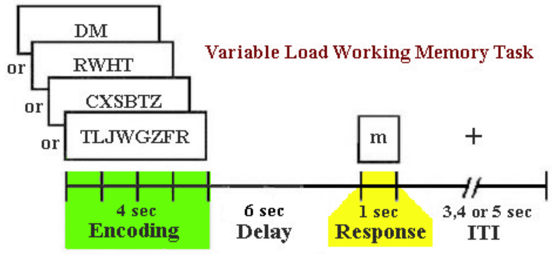


DM
or
RWHT
or
CXSBTZ
or
TLJWGZFR
Variable Load Working Memory Task
m
+
4 sec
Encoding
6 sec
Delay
1 sec
Response
3,4 or 5 sec
ITI

Constrained Principal Component Analysis (CPCA) – basic equations

Regression-based constraints (external analysis)

$${}_{12840}Z_{23621} = {}_{12840}G_{960}C_{23621} + {}_{12840}E_{23621}$$



The columns in this subject-and-condition based Finite Impulse Response (FIR) model *G* matrix code eight peristimulus time points for each of four load conditions (2, 4, 6 and 8 letters) for each of 30 subjects, totaling 960 columns (8 × 4 × 30 = 960)

FIR MODEL EXAMPLE (G):

Stimulus	Scan	S1_load2_scan1	S1_load2_scan2	S1_load2_scan3	S1_load2_scan4	S1_load2_scan5
	1	0	0	0	0	0
	2	0	0	0	0	0
DM	3	1	0	0	0	0
	4	0	1	0	0	0
	5	0	0	1	0	0
m	6	0	0	0	1	0
	7	0	0	0	0	1
	8	0	0	0	0	0

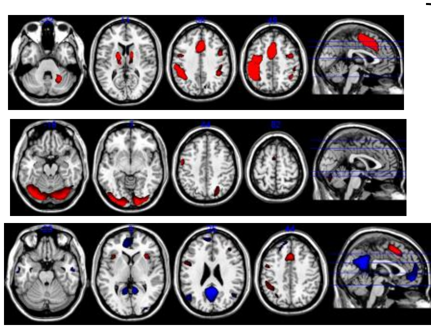
Constrained Principal Component Analysis (CPCA) – basic equations

Regression-based constraints (external analysis)

$$ {}_{12840}Z_{23621} = {}_{12840}G_{960}C_{23621} + {}_{12840}E_{23621} $$

The columns in this subject-and-condition based Finite Impulse Response (FIR) model G matrix code eight peristimulus time points for each of four load conditions (2, 4, 6 and 8 letters) for each of 30 subjects, totaling 960 columns ($8 \times 4 \times 30 = 960$)

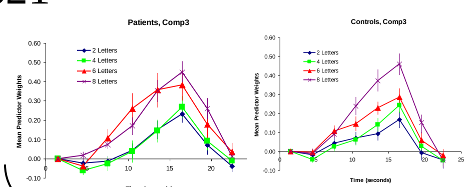
Singular Value Decomposition (internal analysis)



$$ {}_{12840}U_{3}D_{3}V'_{23621} \approx {}_{12840}G_{960}C_{23621} $$

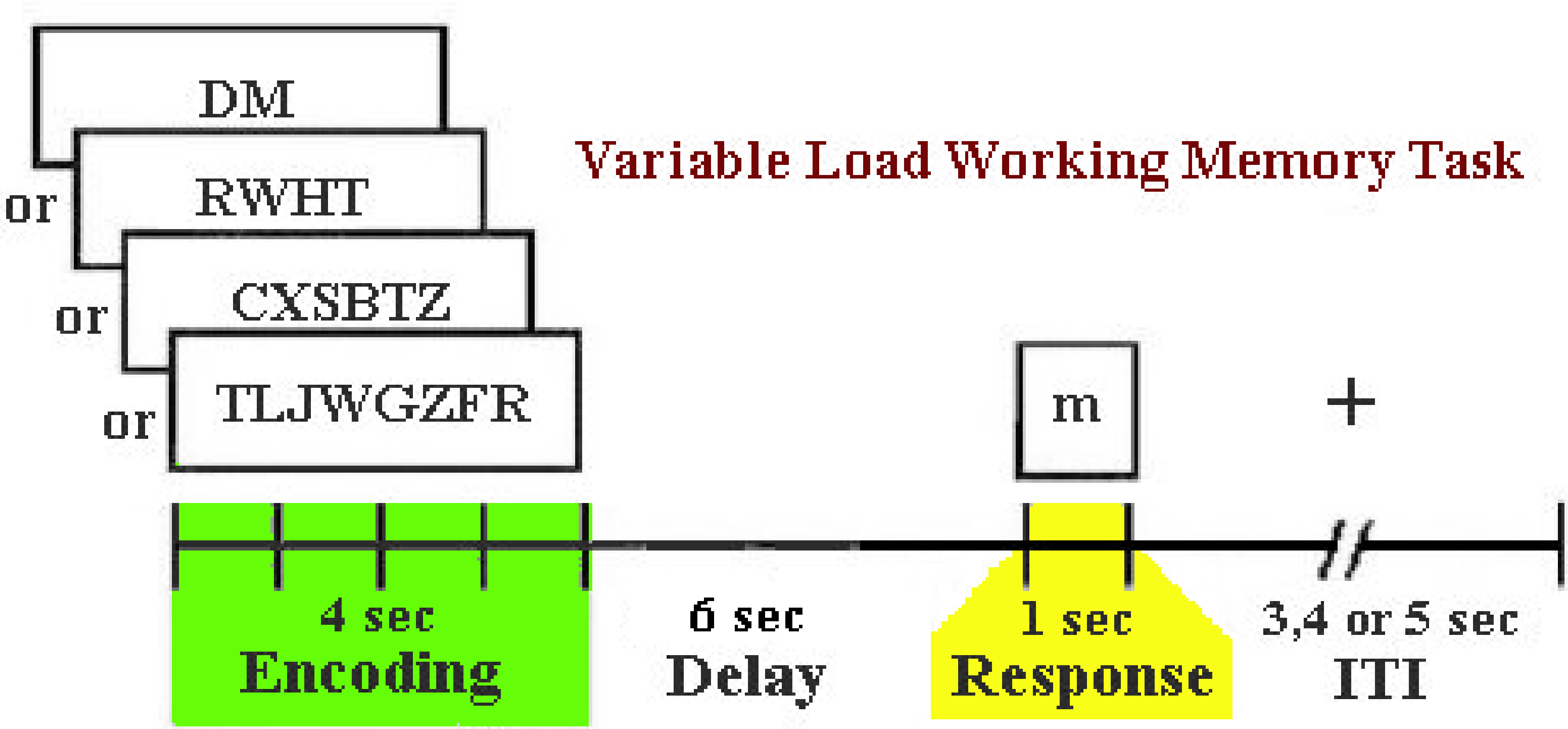
Predictor weights

$$ {}_{12840}U_{3} = {}_{12840}G_{960}P_{3} $$



Data From Dr. Elton Ngan, UBC

Cairo, T. A., Woodward, T. S. & Ngan, E.T.C. (2006). Decreased encoding efficiency in schizophrenia. Biological Psychiatry, 59(8), 740-746.



CPCA analysis of UBC WM task

Inspection of scree plot suggested a 3 component solution

Removal of linear and quadratic trends in pre-processing accounted for 32% of BOLD variance

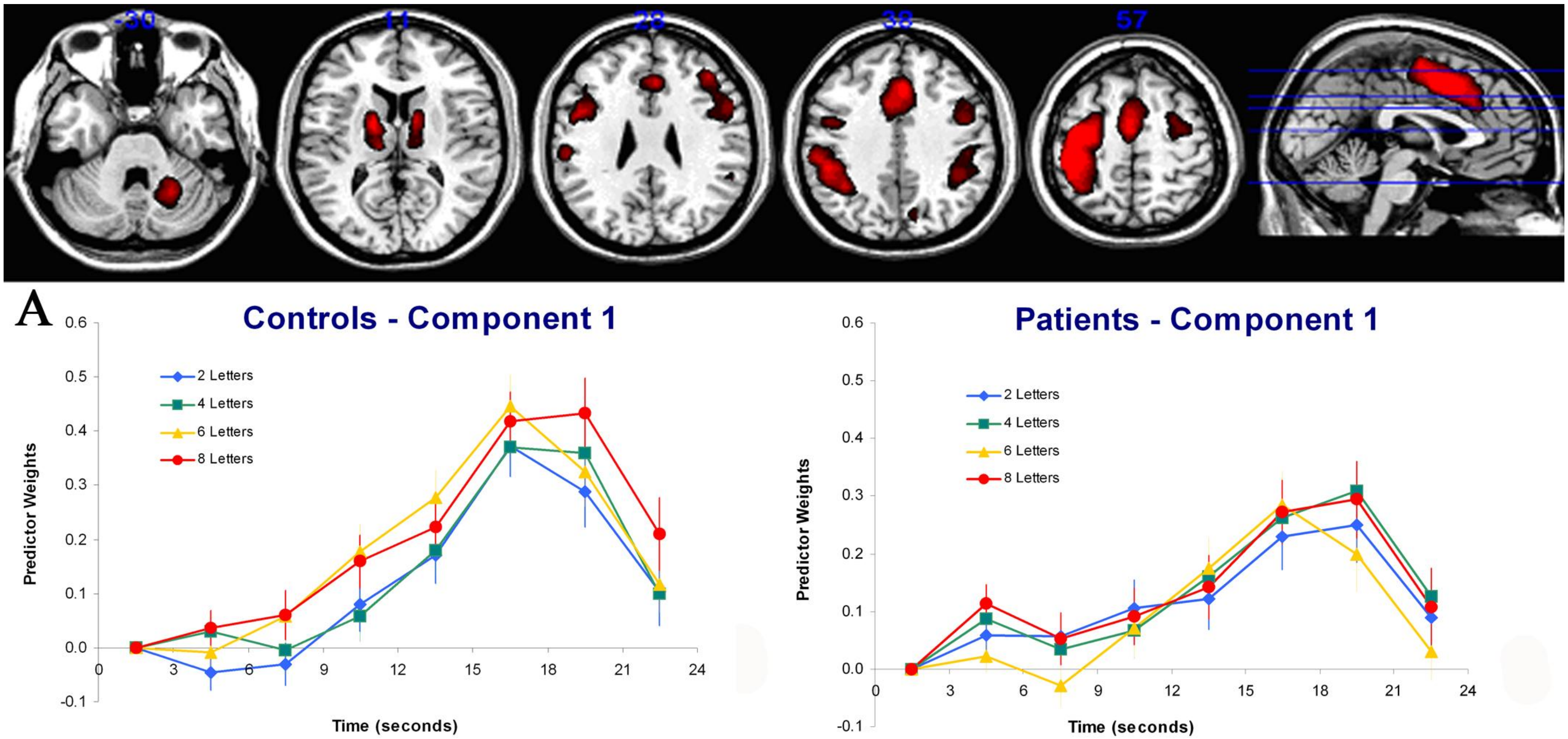
Of this, 13% of BOLD signal variance is predictable from stimulus presentation (ss(GC) / ss(Z))

Of this, 39% of BOLD signal variance can be explained by a 3 component solution (ss(UDV) / ss(GC))

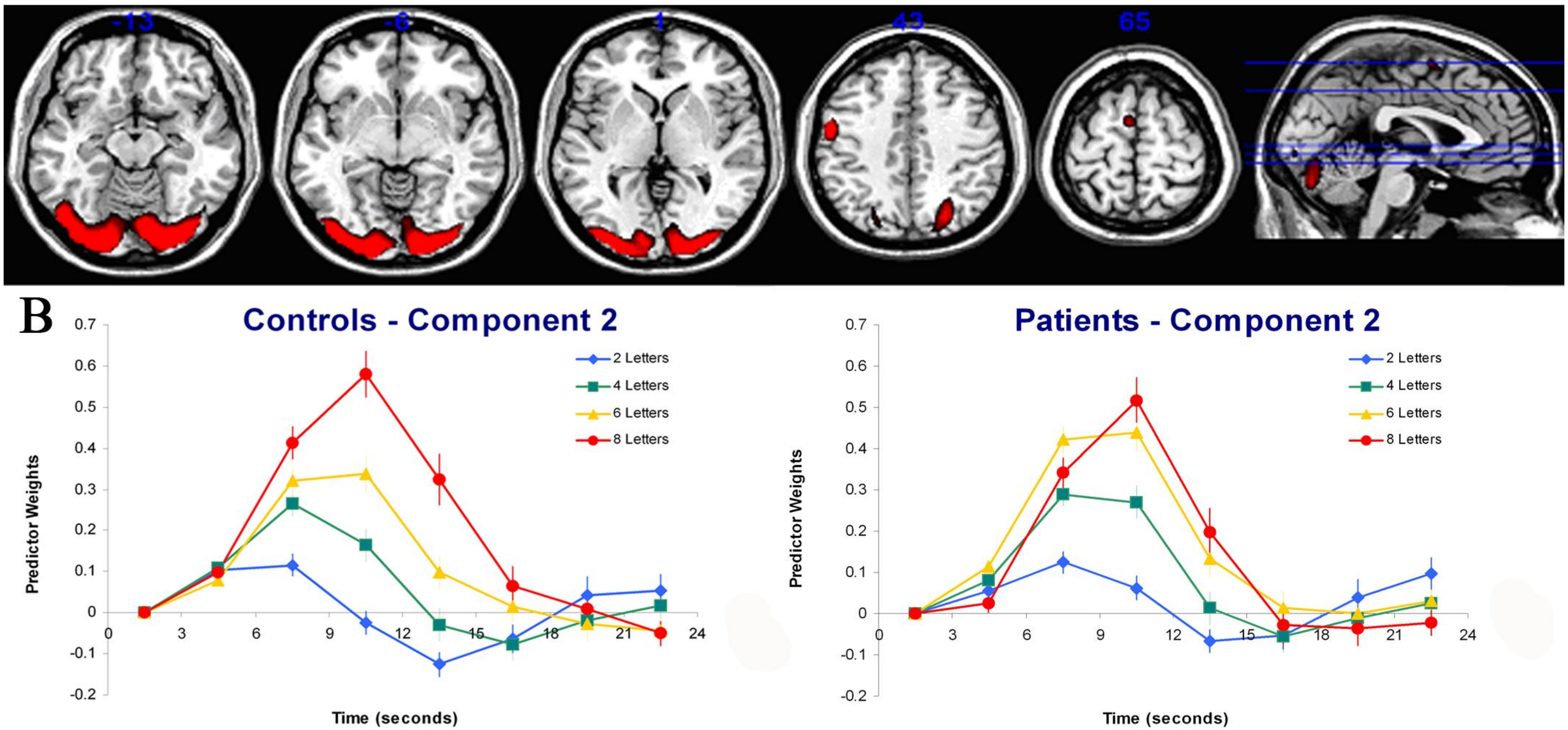
Components were rotated (orthogonal rotation)

These components all displayed HDR shapes in the predictor weights, and significant effects of peristimulus time using repeated measures ANOVA on the predictor weights

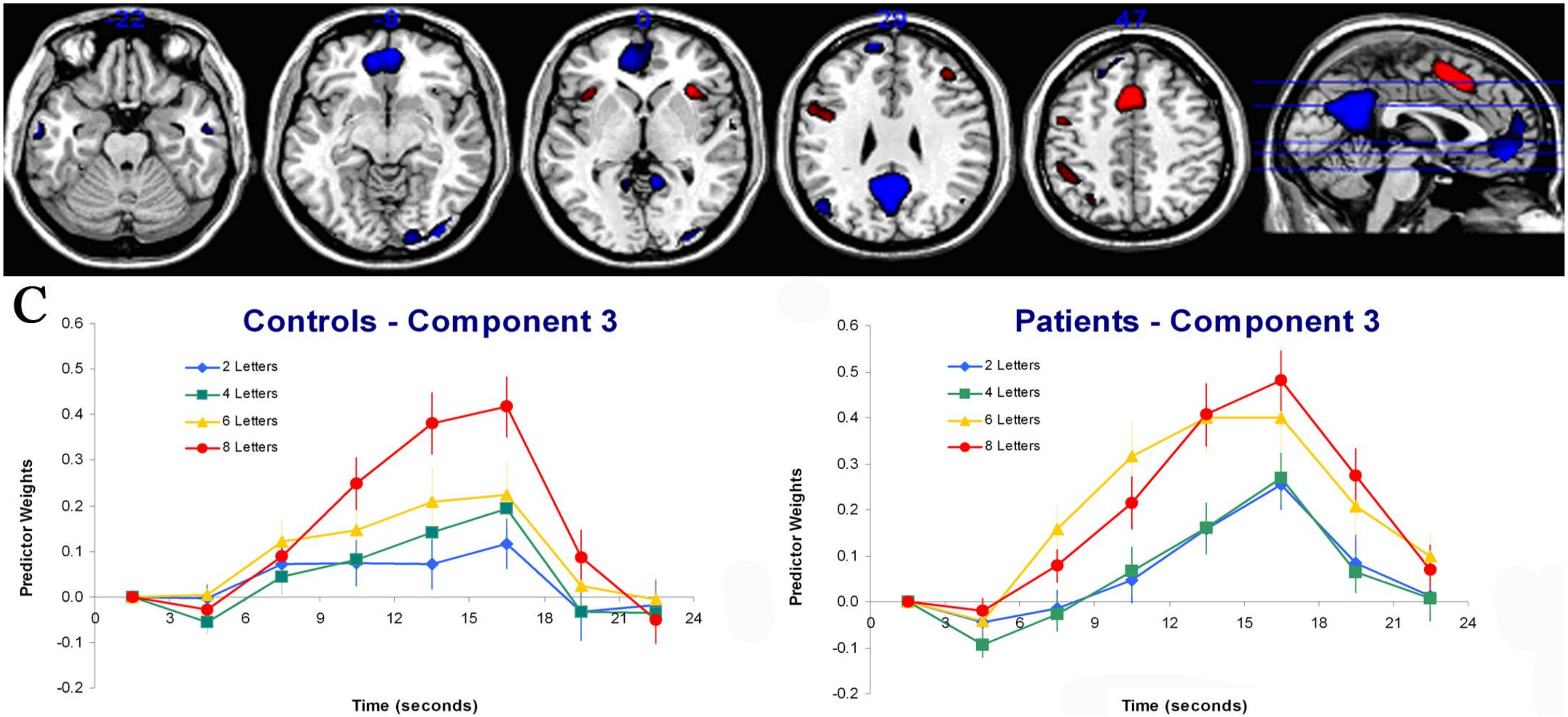
C1: Task positive network (20% variance)



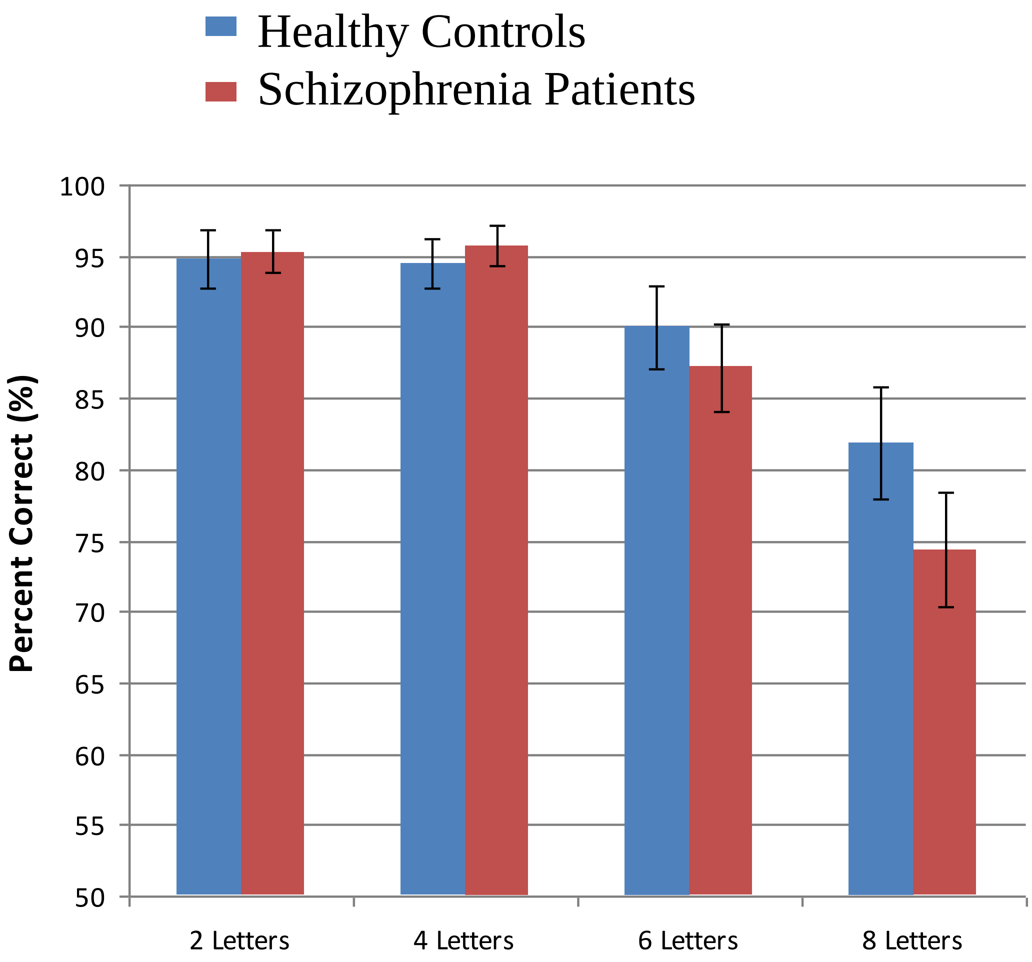
C2: Visual Cortex (14% variance)



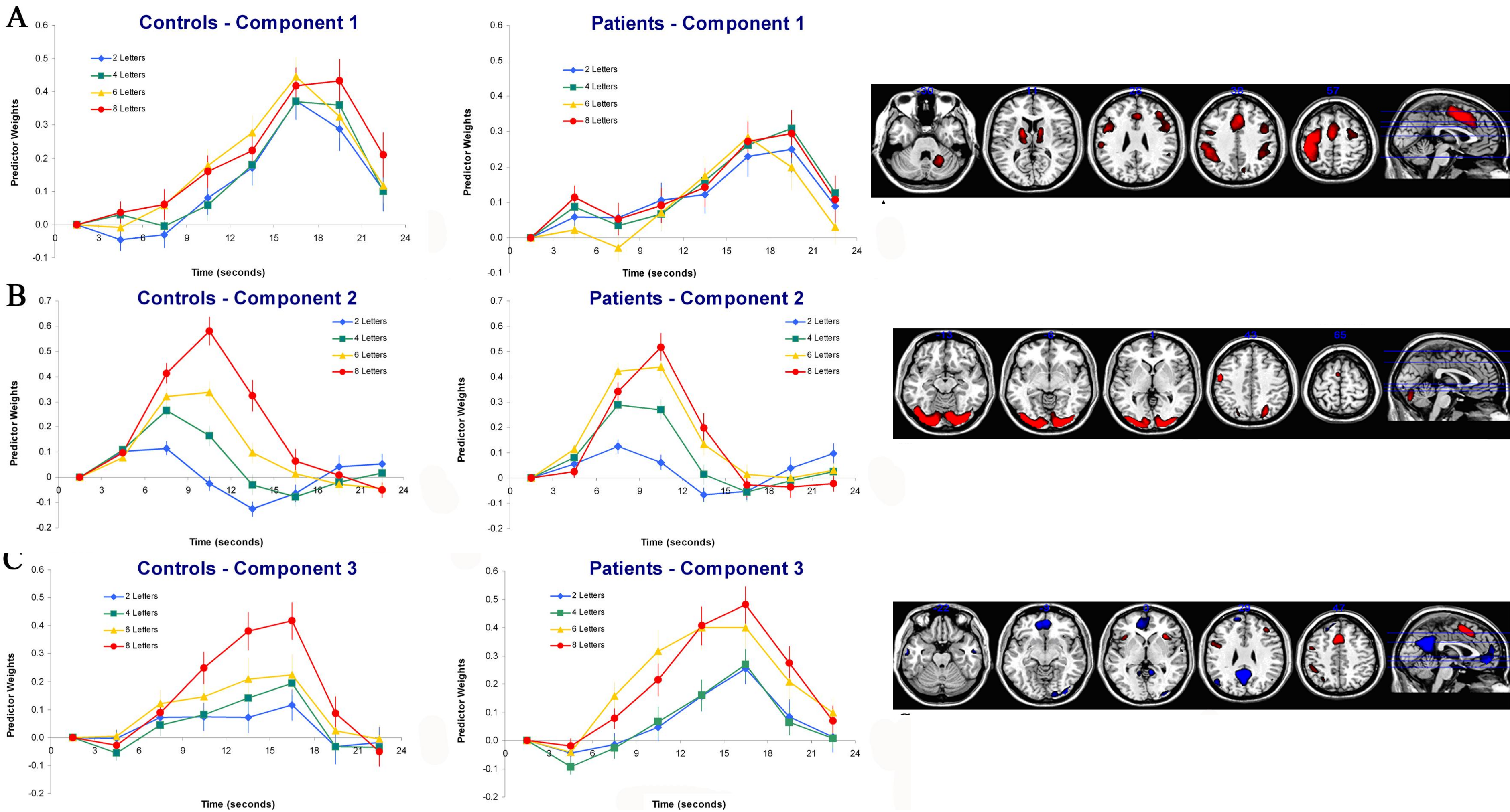
C3: Task Negative/Positive (6% variance)

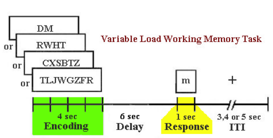


WM Performance



Inefficient Task Positive/Negative Networks – Patients must activate more to achieve equivalent performance. If not possible to activate more, performance deficits.

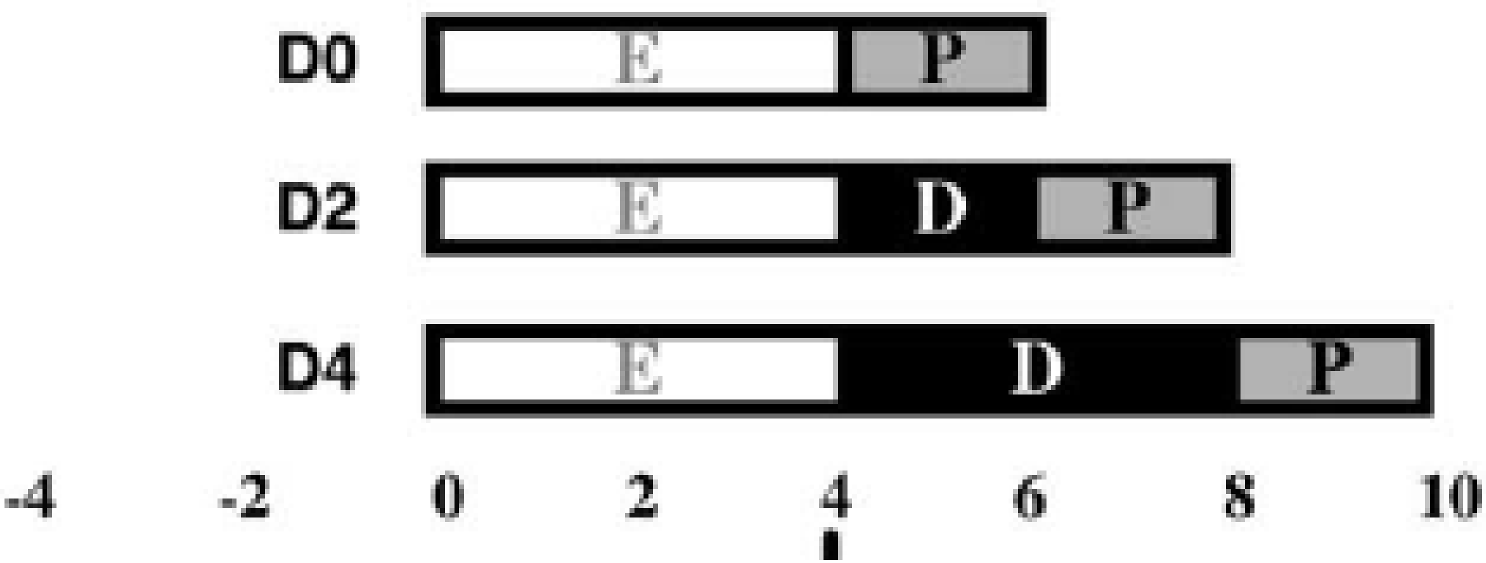

A
Controls - Component 1
Patients - Component 1
B
Controls - Component 2
Patients - Component 2
C
Controls - Component 3
Patients - Component 3
2 Letters
4 Letters
6 Letters
8 Letters
Predictor Weights
Time (seconds)


DM
RWHT
CXSBTZ
TLJWGZFR
or
Variable Load Working Memory Task
m
+
4 sec
Encoding
6 sec
Delay
1 sec
Response
3,4 or 5 sec
ITI

Data From Dr. Dara Manoach, Harvard

Manoach DS, Greve DN, Lindgren KA, Dale AM (2003). Identifying regional activity associated with temporally separated components of working memory using event-related functional MRI NeuroImage, 20(3):1670-1684

No load manipulation but delay period manipulation



CPCA analysis of Harvard WM task

Inspection of scree plot suggested a 3 component solution

Removal of linear and quadratic trends in pre-processing accounted for 16% of BOLD variance

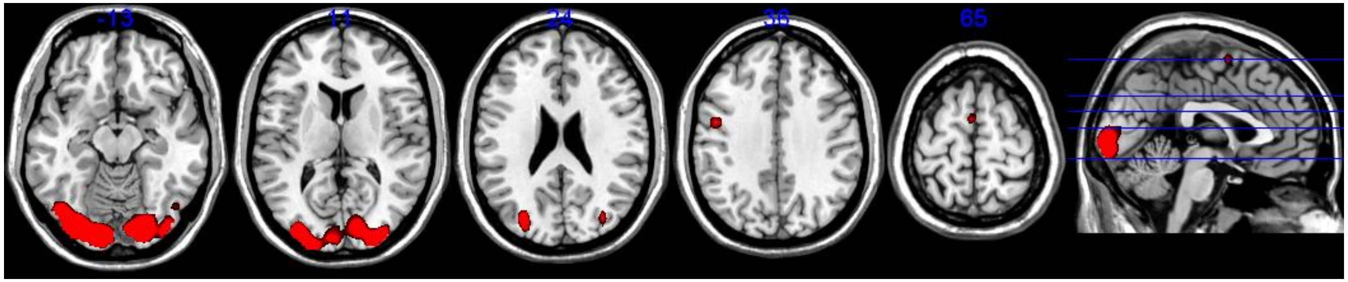
Of this, 9% of BOLD signal variance is predictable from stimulus presentation (ss(GC) / ss(Z))

Of this, 53% of BOLD signal variance can be explained by a 3 component solution (ss(UDV) / ss(GC))

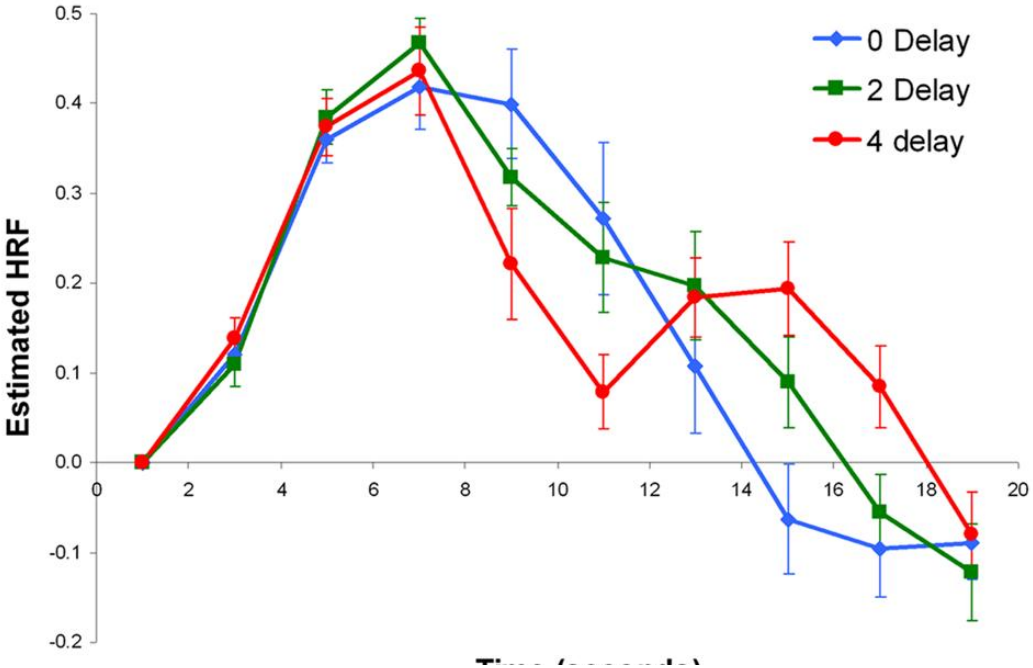
Components were rotated (orthogonal rotation)

These components all displayed HDR shapes in the predictor weights, and significant effects of peristimulus time using repeated measures ANOVA on the predictor weights

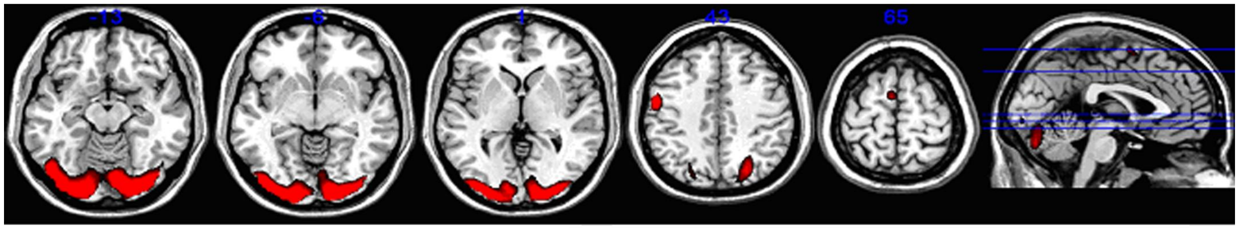
C1: Visual Cortex (24% variance)



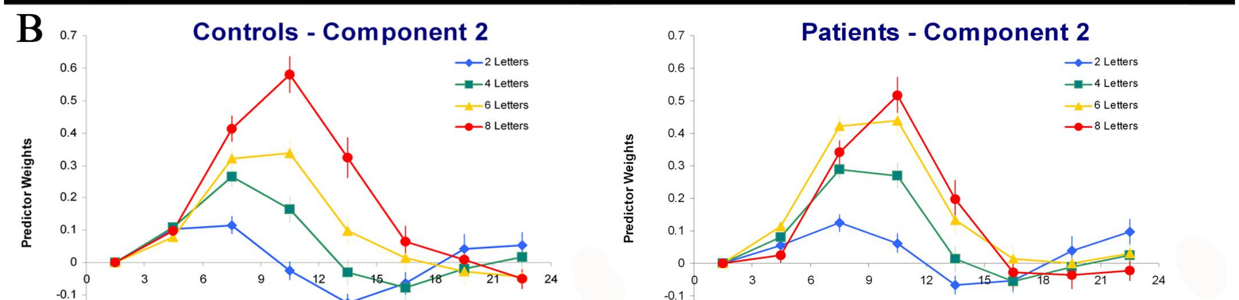
A

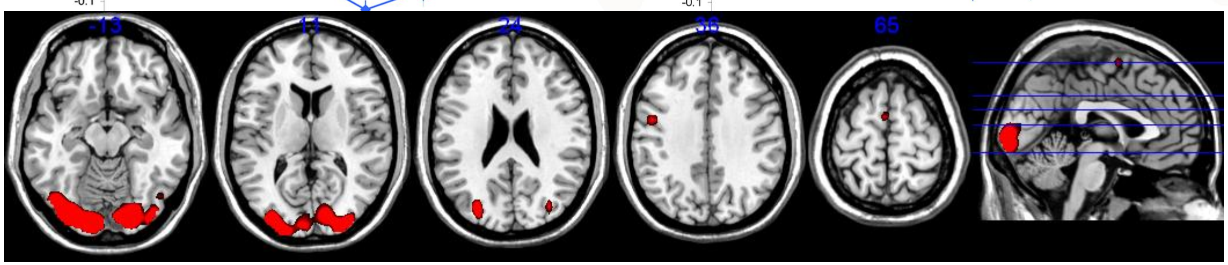


Visual Cortex

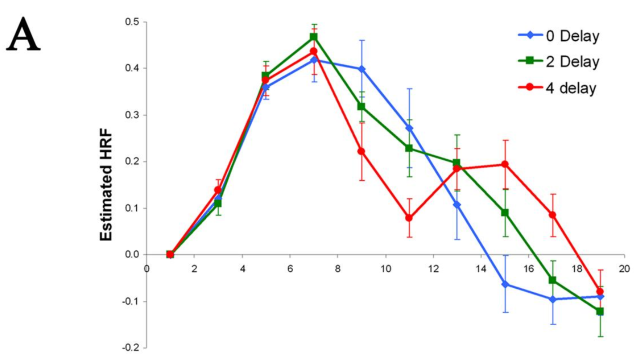


14% variance
UBC study

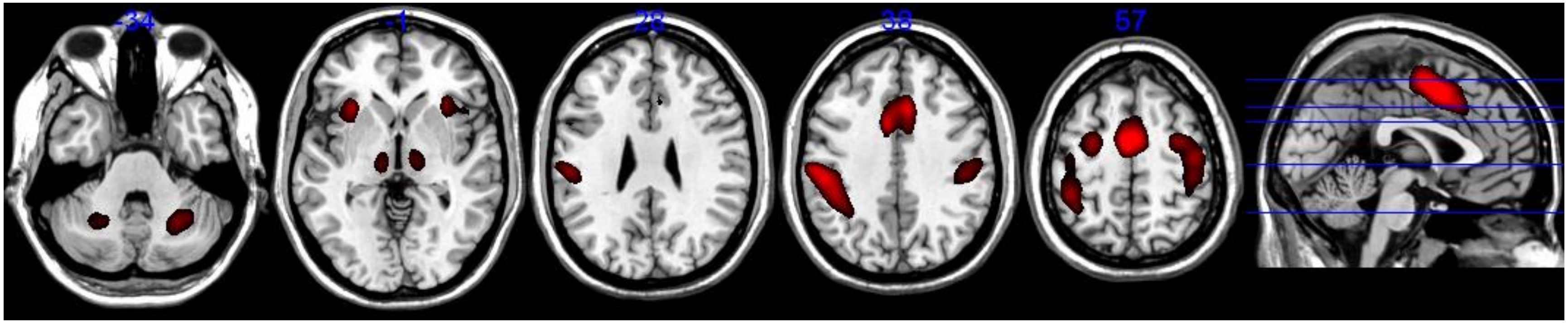




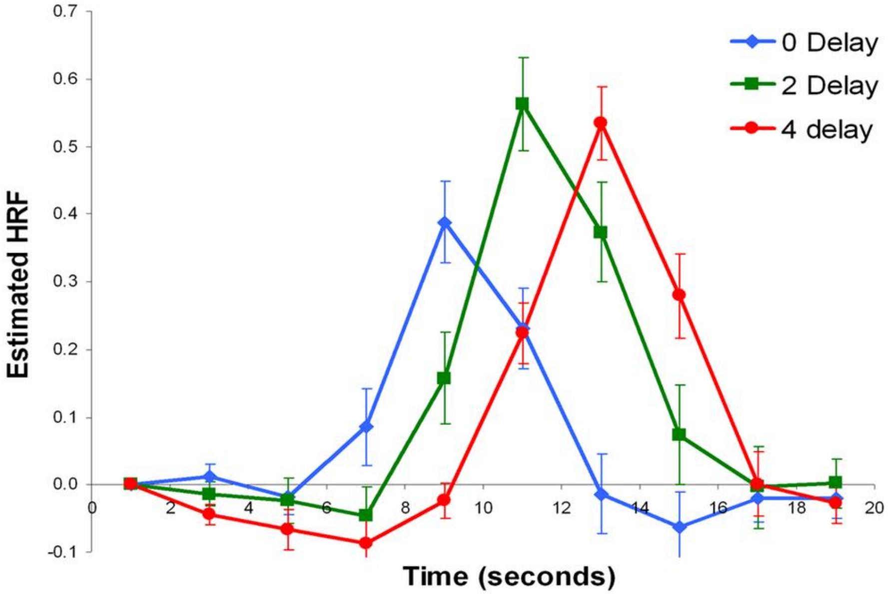
24% variance
Harvard study



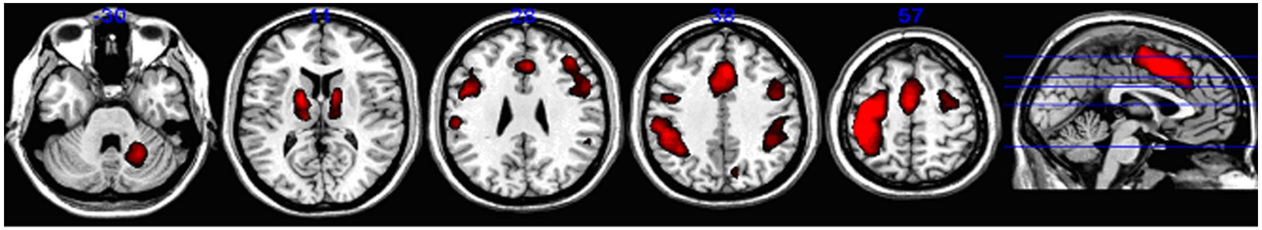
Responding network (16% variance)



B

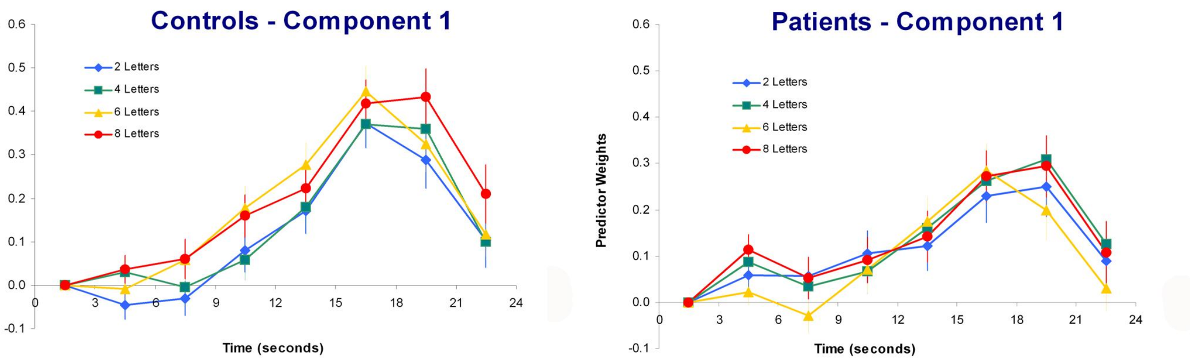


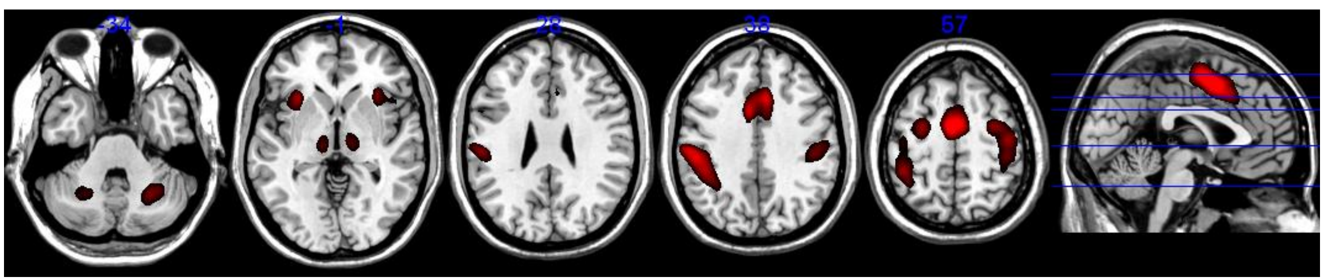
Responding network



24% variance
UBC study

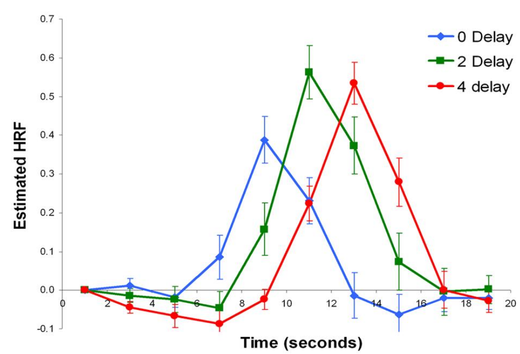
A



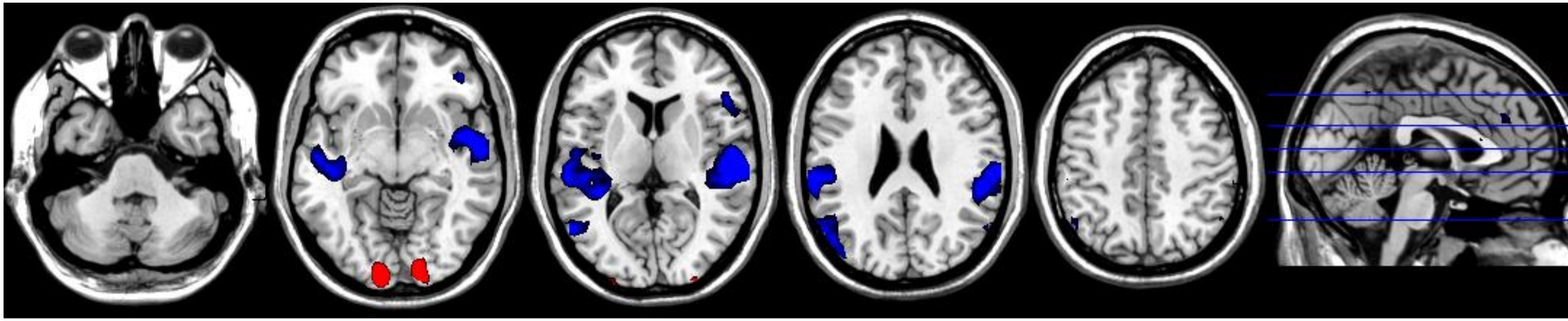


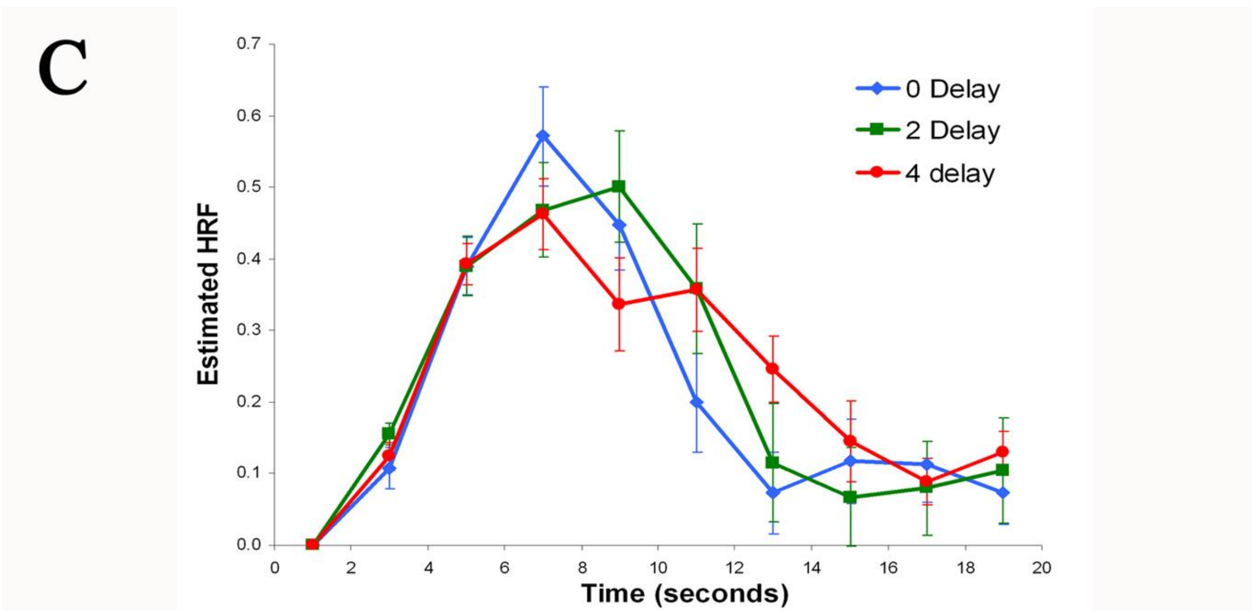
16% variance
Harvard study

B

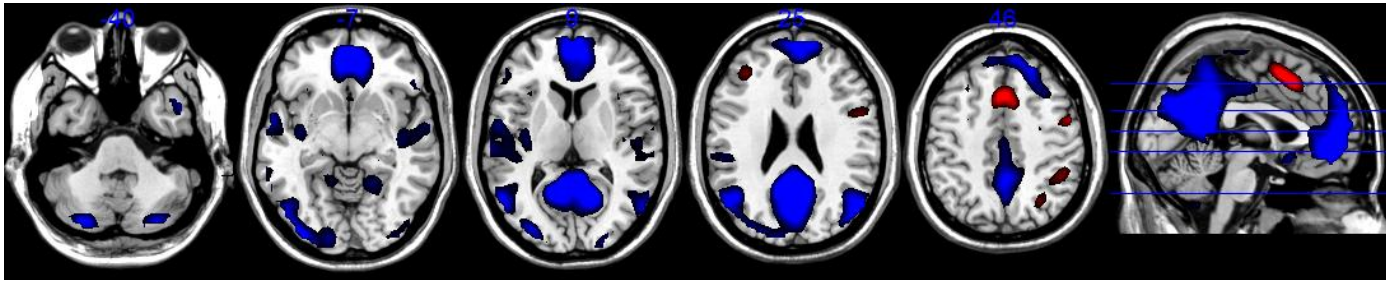


Working Memory (12% variance)



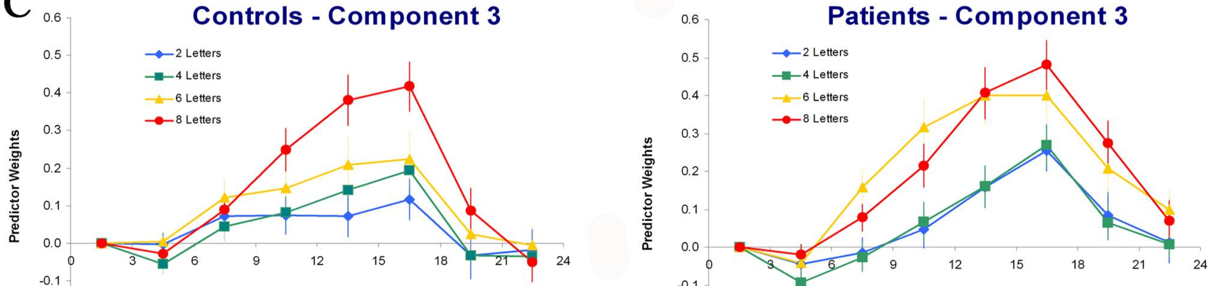


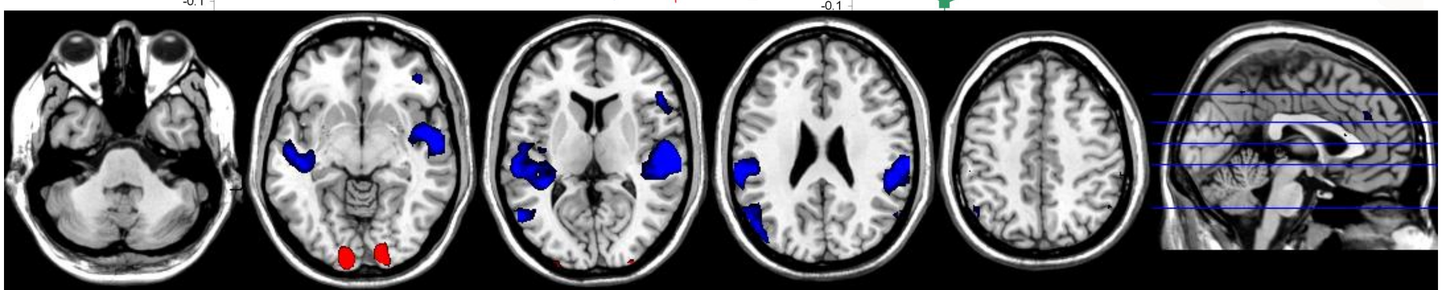
Working Memory



6% variance
UBC study

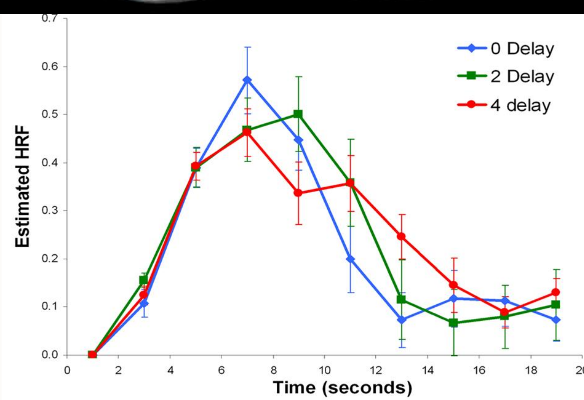
C





12% variance
Harvard study

C



UBC Scanner (1.5 Tesla GE)

linear + quadratic trends: 32% of BOLD variance

13% of remaining predictable from stimulus presentation

39% of remaining explained by a 3 component solution

Harvard Scanner (3 Tesla Siemens Allegra)

linear + quadratic trends: 16% of BOLD variance

9% of remaining predictable from stimulus presentation

53% of remaining explained by a 3 component solution

Constrained Principal Component Analysis (CPCA) - equations

$$Z = GC + E$$

where

$$C = (G'G)^{-1} G'Z$$

Singular Value Decomposition

$$UDV' = GC$$

Predictor weights

$$U = G^{*}P$$

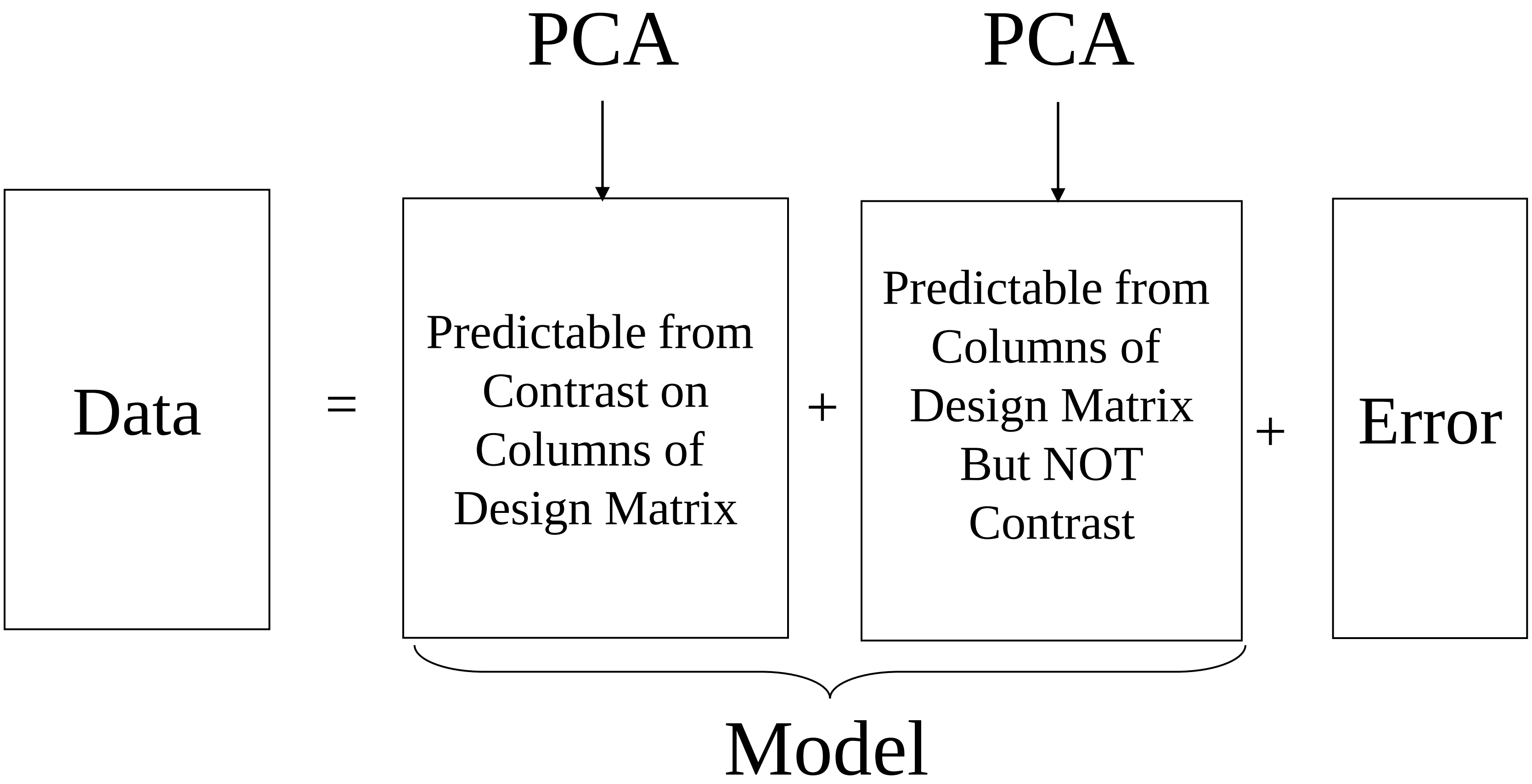
Using Contrasts of conditions
Employ the A matrix

$$C = AW + E^{*}$$

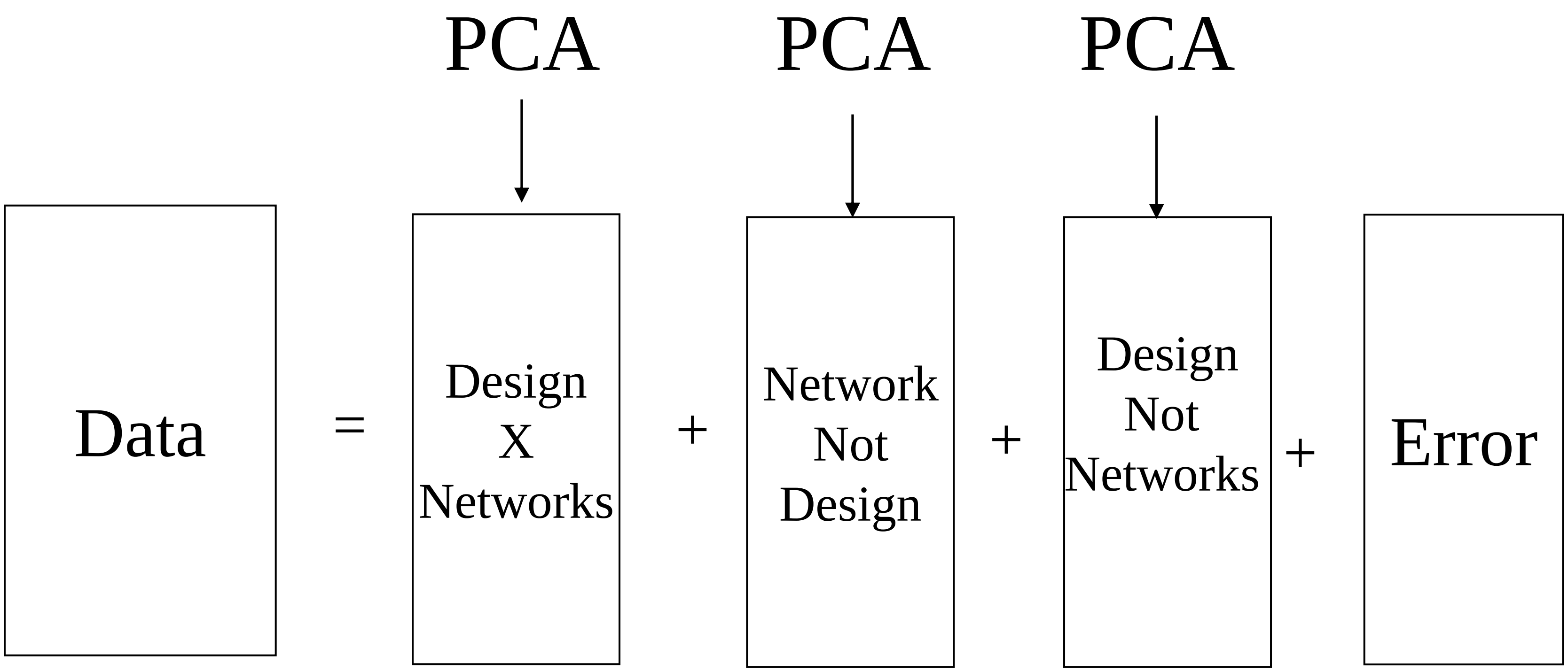
$$Z = G(AW + E^{*}) + E$$

$$Z = GAW + GE^{*} + E$$

Constrained Principal Component Analysis (CPCA) – Other Applications



Add Information about Hypothesized Functional Networks



Available Extensions of CPCA

$$Z = GMH' + BH' + GC + E$$

$$Z = GAMH' + G{\sim}AMH' + BH' + GAC + G{\sim}AC + E$$

$$Z = GMHQ' + GMH{\sim}Q' + BHQ' + BH{\sim}Q' + GC + E$$

$$Z = GAMHQ' + GAMH{\sim}Q' + G{\sim}AMHQ' + G{\sim}AMH{\sim}Q' + BHQ' + BH{\sim}Q' + GAC + G{\sim}AC + E$$

Acknowledgements

CPCA

Dr. Yoshio Takane

McGill University

fMRI

Dr. Elton Ngan

University of British Columbia

Dr. Dara Manoach

Harvard University

CNoS

Cognitive Neuroscience of Schizophrenia Lab

Ph.D. student

Jennifer Whitman

Masters students

Paul Metzak

Katie Lavigne

Programmer

John Paiement

fMRI - CPCA Funding

•NIH Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC) Administrative Supplements for Neuroimaging Informatics Software Enhancement to **Brad Postle** 2RO1 MH064498-05A2

•Canadian Institutes for Health Research (CIHR)

•Michael Smith Foundation for Health Research (MSFHR)

•NARSAD

•Natural Sciences and Engineering Research Council of Canada (NSERC)