Plumbing the Intertubes
A Hitchhiker's Guide to Information Filtering

satrajit ghosh satra@mit.edu

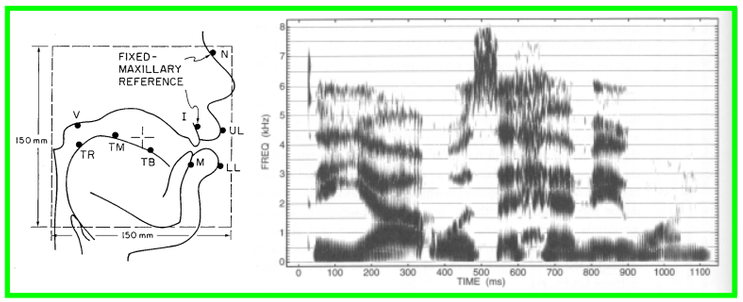
massachusetts institute of technology | harvard medical school

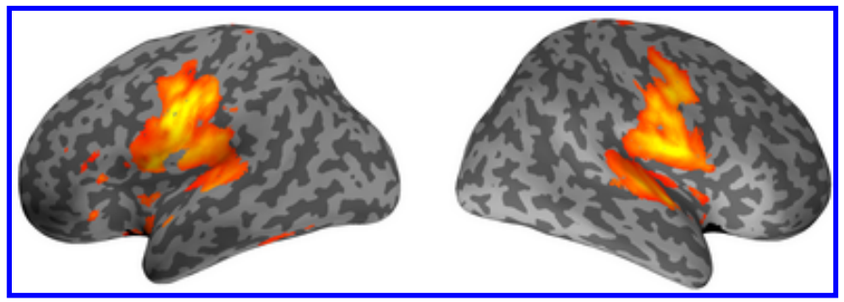
McGOVERN INSTITUTE FOR BRAIN RESEARCH AT MIT

HARVARD MEDICAL SCHOOL

SHBT Harvard University Speech and Hearing Bioscience and Technology

CC BY





$$M(t) = M(0) + \alpha_{ff} \int_0^t \dot{M}_{Feedforward}(t) g(t) dt + \alpha_{fb} \int_0^t \dot{M}_{Feedback}(t) g(t) dt$$

Behavior **Brain Structure and Function** **Computational Models**

Neuroinformatics

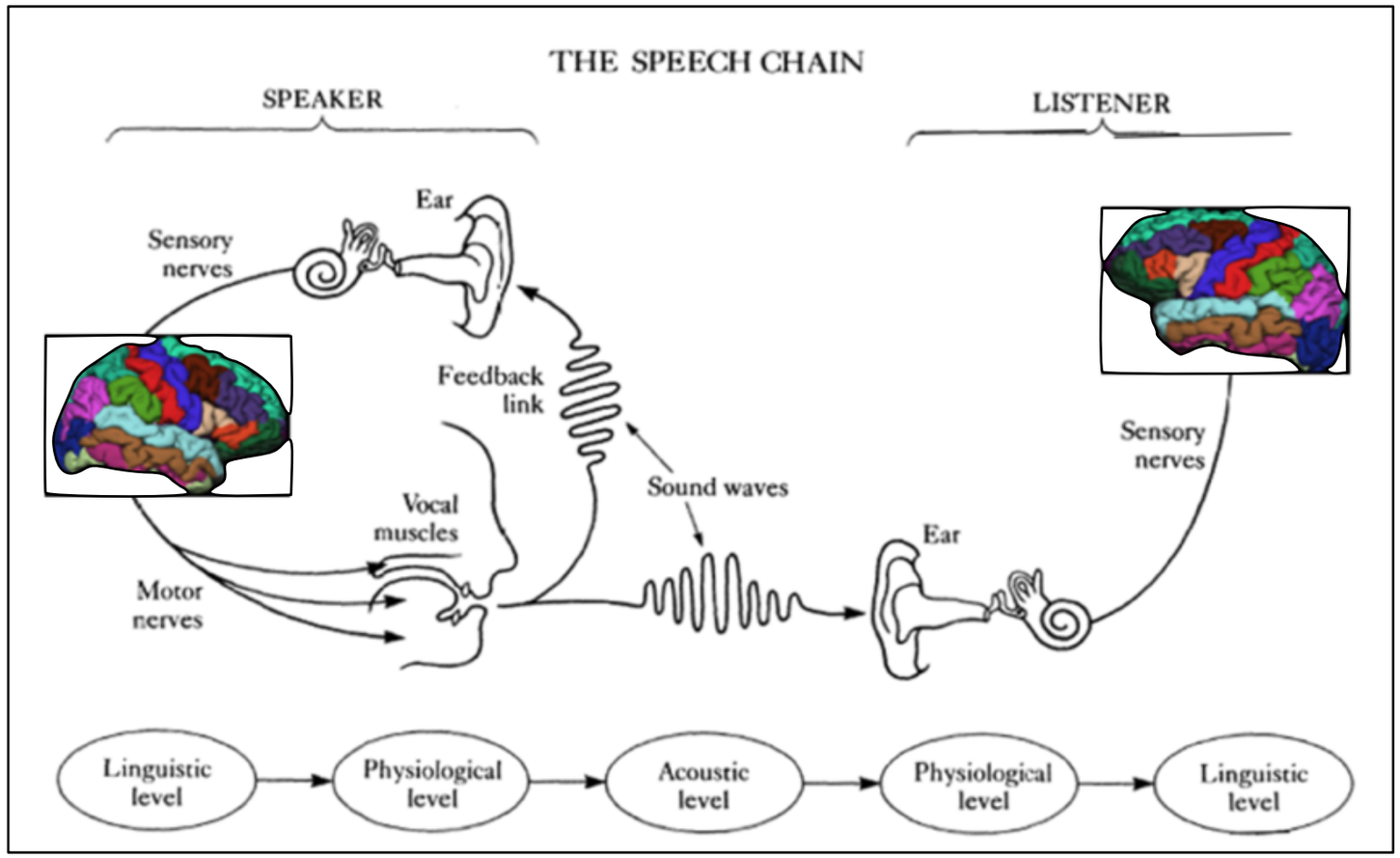
logo: John Salvatore

How I became a plumber

What types of plumbing we do

NeuroStars

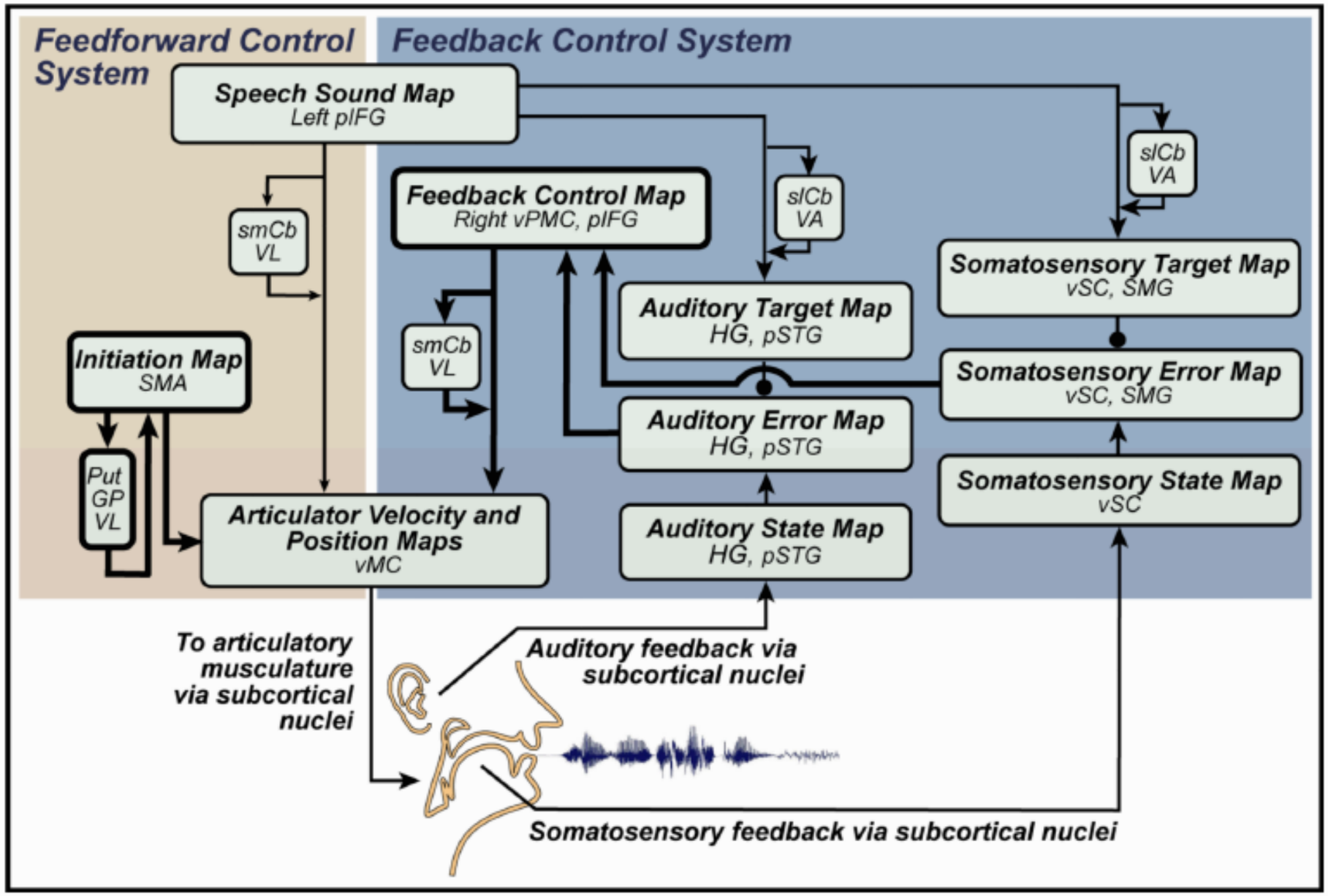
Spoken Communication



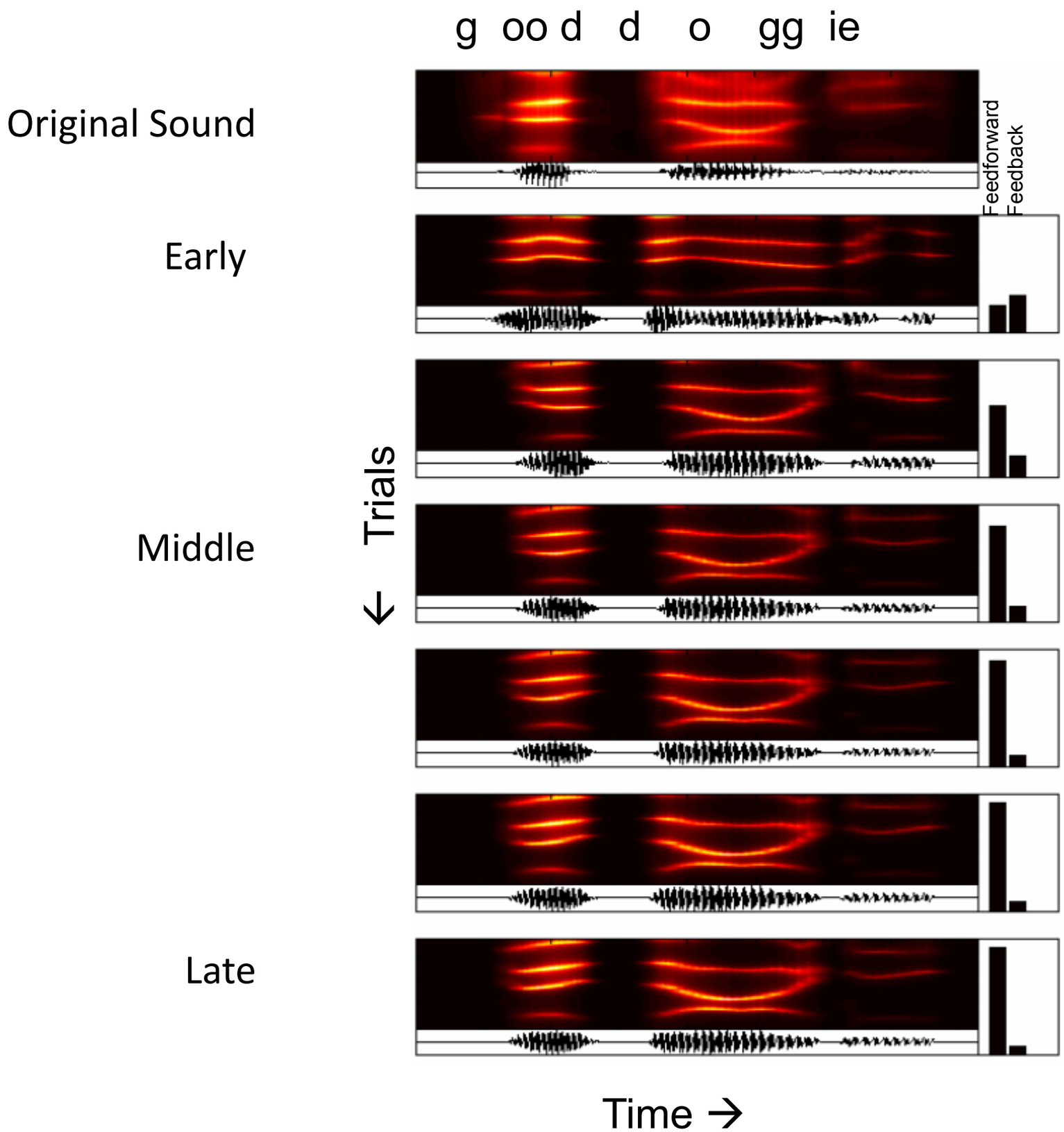
- Speaker to speaker
- Self-perception
- Context
- Environment
- Tone
- Diction
- Sensorimotor interactions
- Cranial nerves
- 100+ muscles

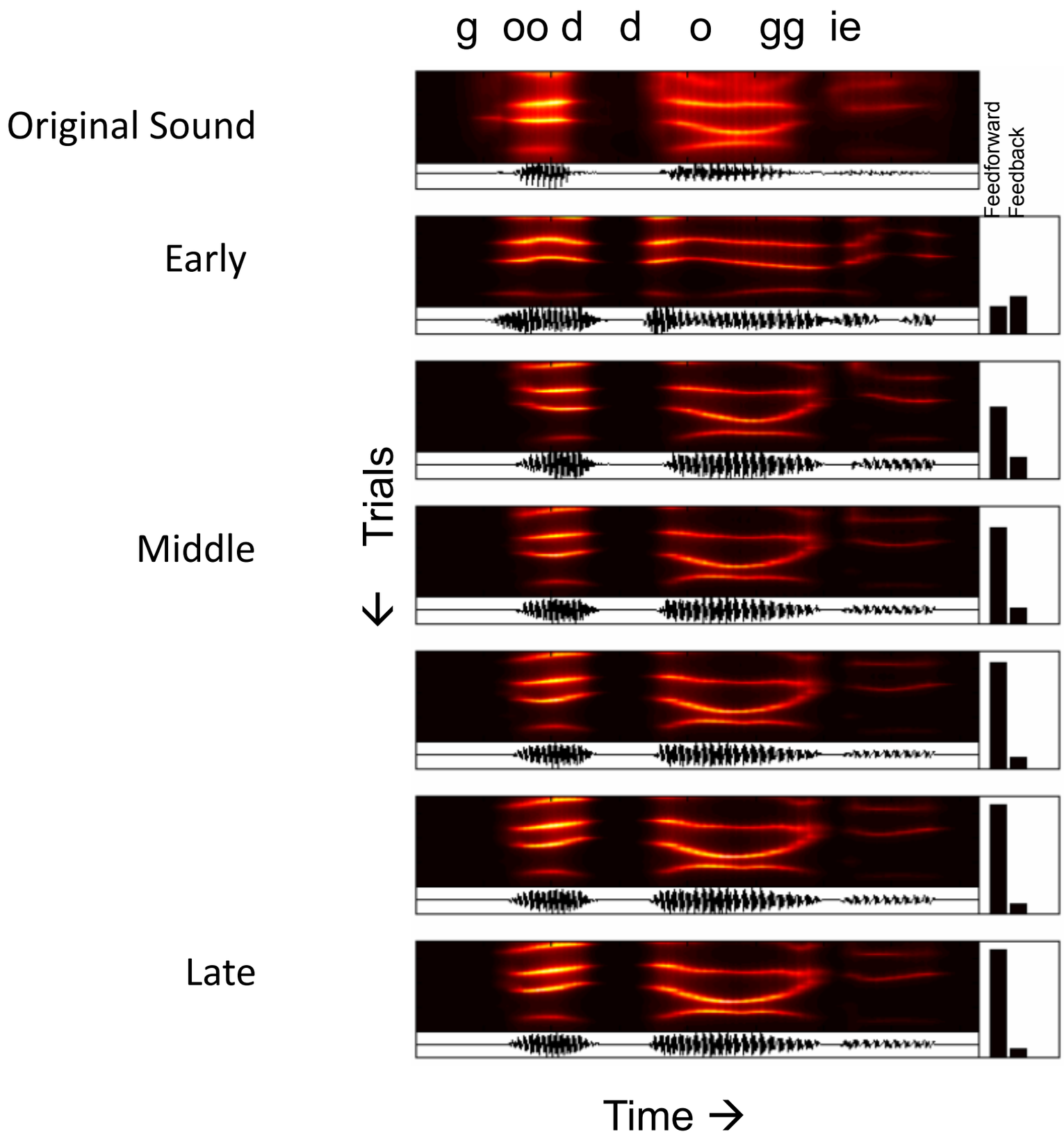
Speech engages numerous cognitive and physiological processes

Denis and Pinson

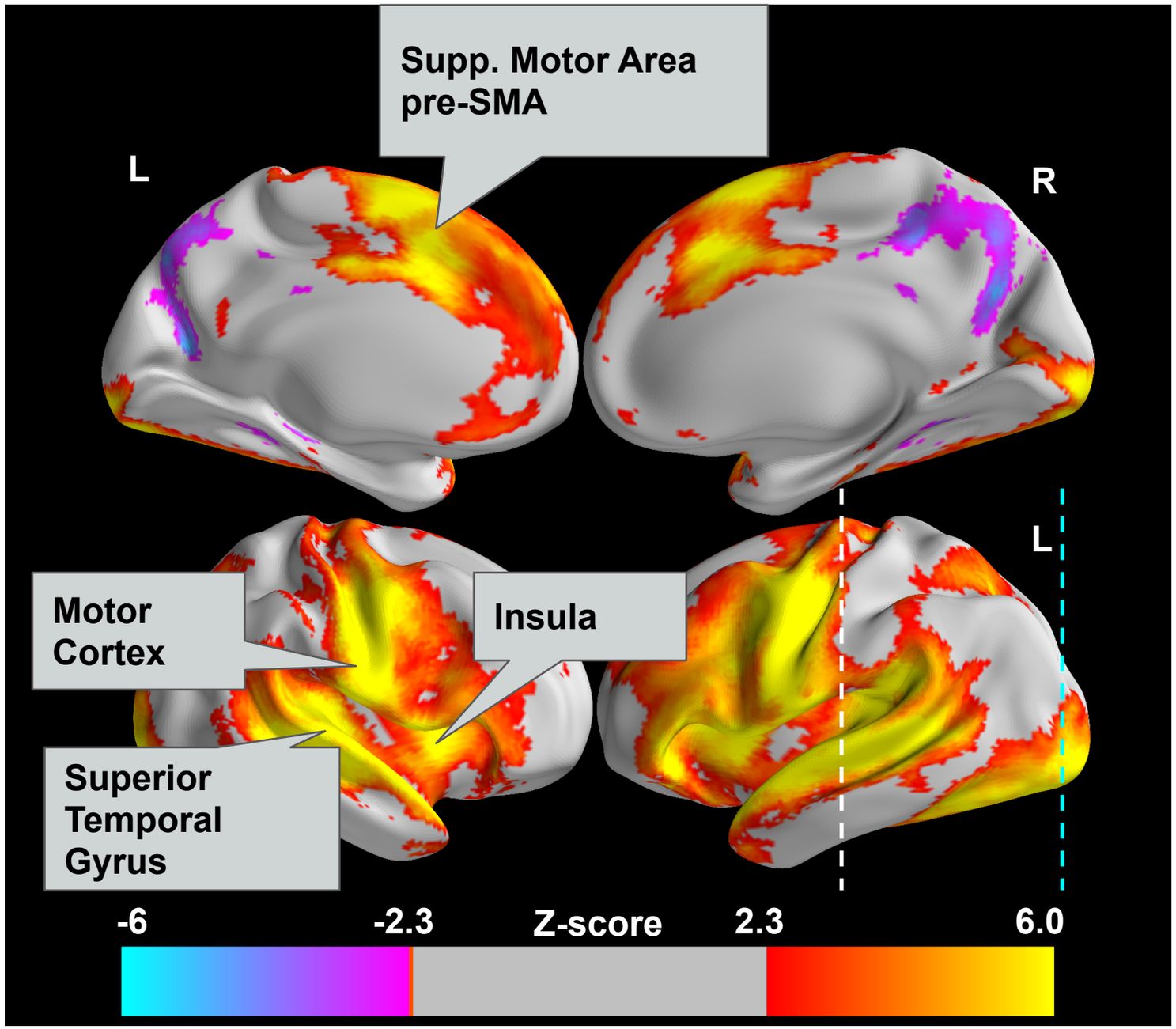


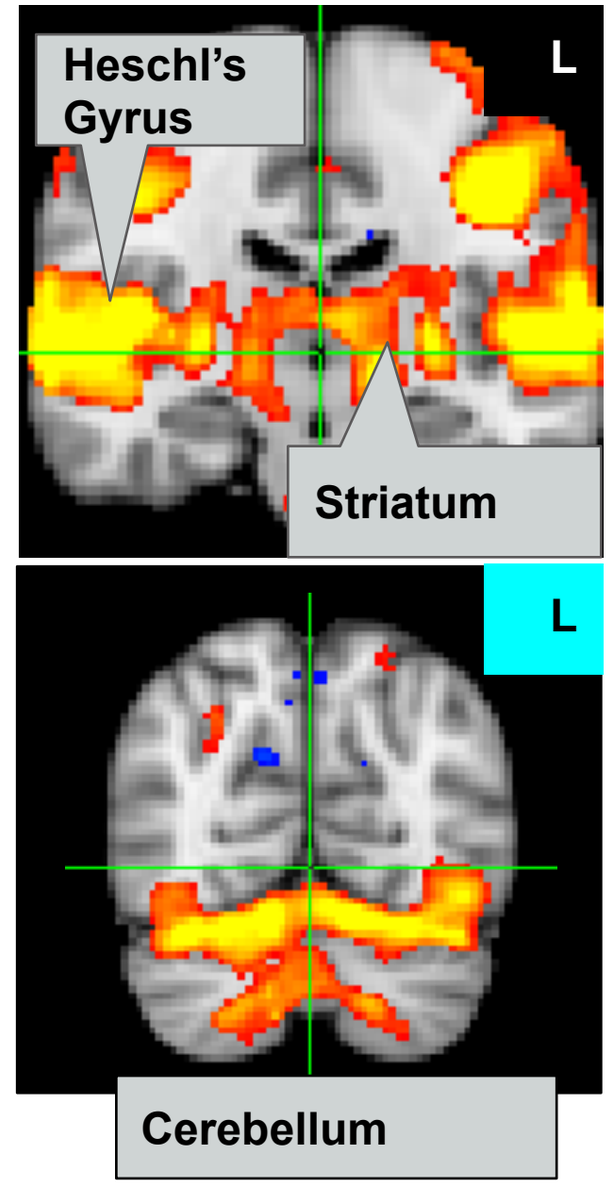
Guenther and Ghosh (2003); Guenther, Ghosh, Tourville (2006); Tourville, Guenther (2011)


g oo d d o gg ie
Original Sound
Early
Middle
Late
Trials
Feedforward
Feedback
Time


g oo d d o gg ie
Original Sound
Early
Middle
Late
Trials
Feedforward
Feedback
Time

A Robust Speech Network





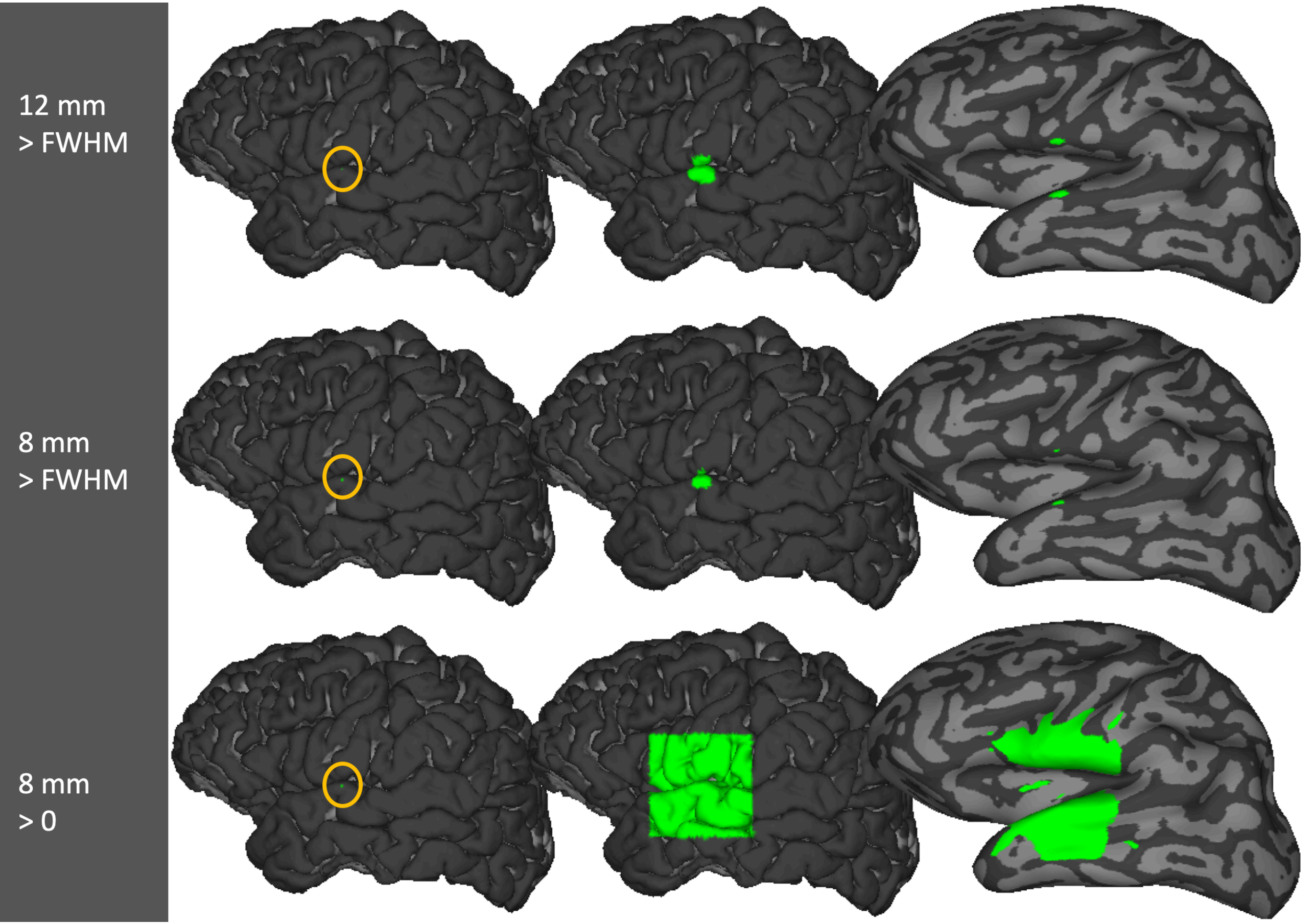
Gp: All Subjects
Subj: Speaking > Not Speaking

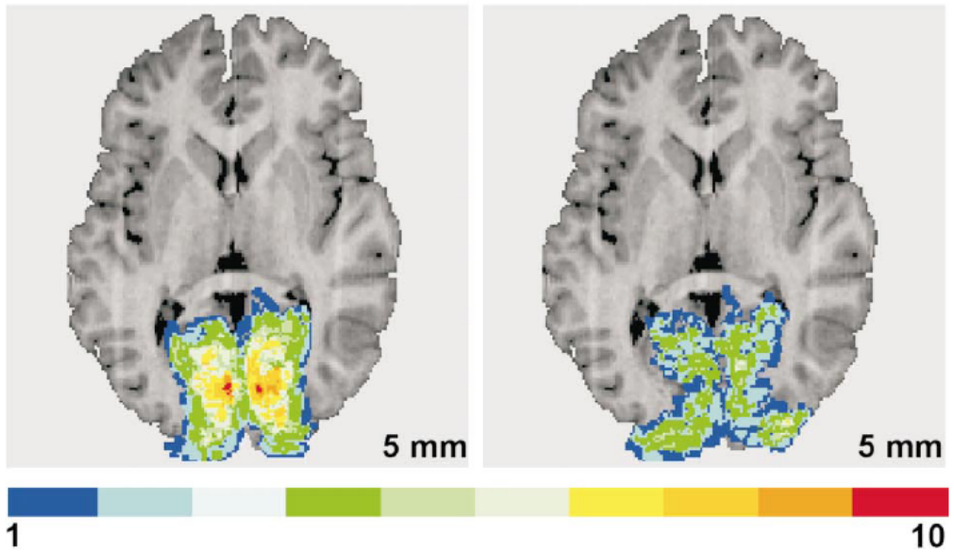
Voxel threshold: 2.3
Cluster p value: 0.05

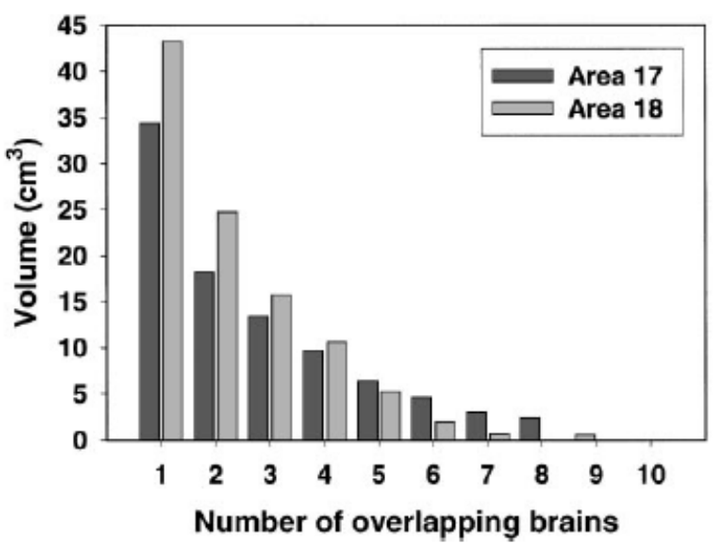
Smoothing

Registration

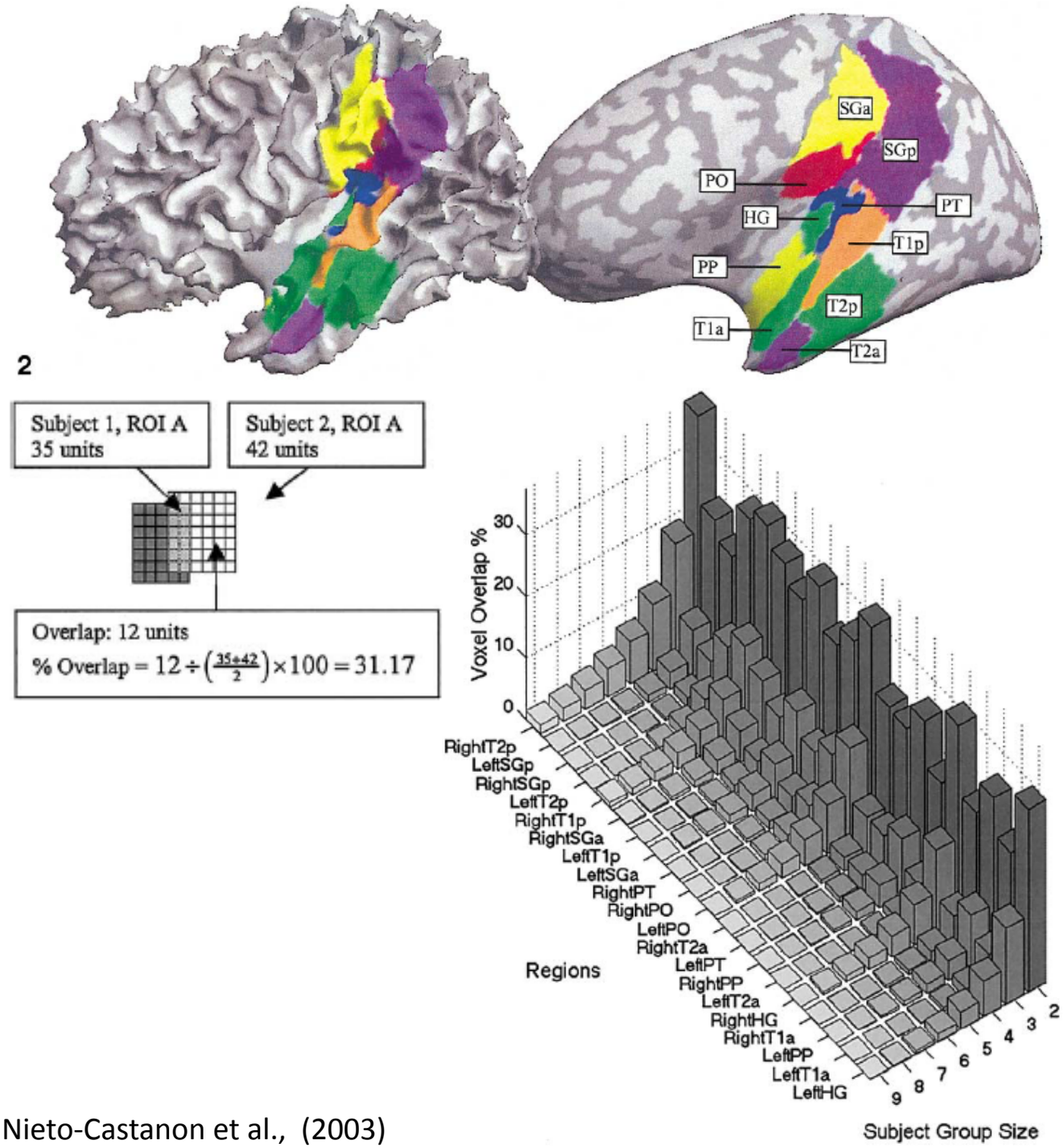
GUIs


12 mm
> FWHM
8 mm
> FWHM
8 mm
> 0





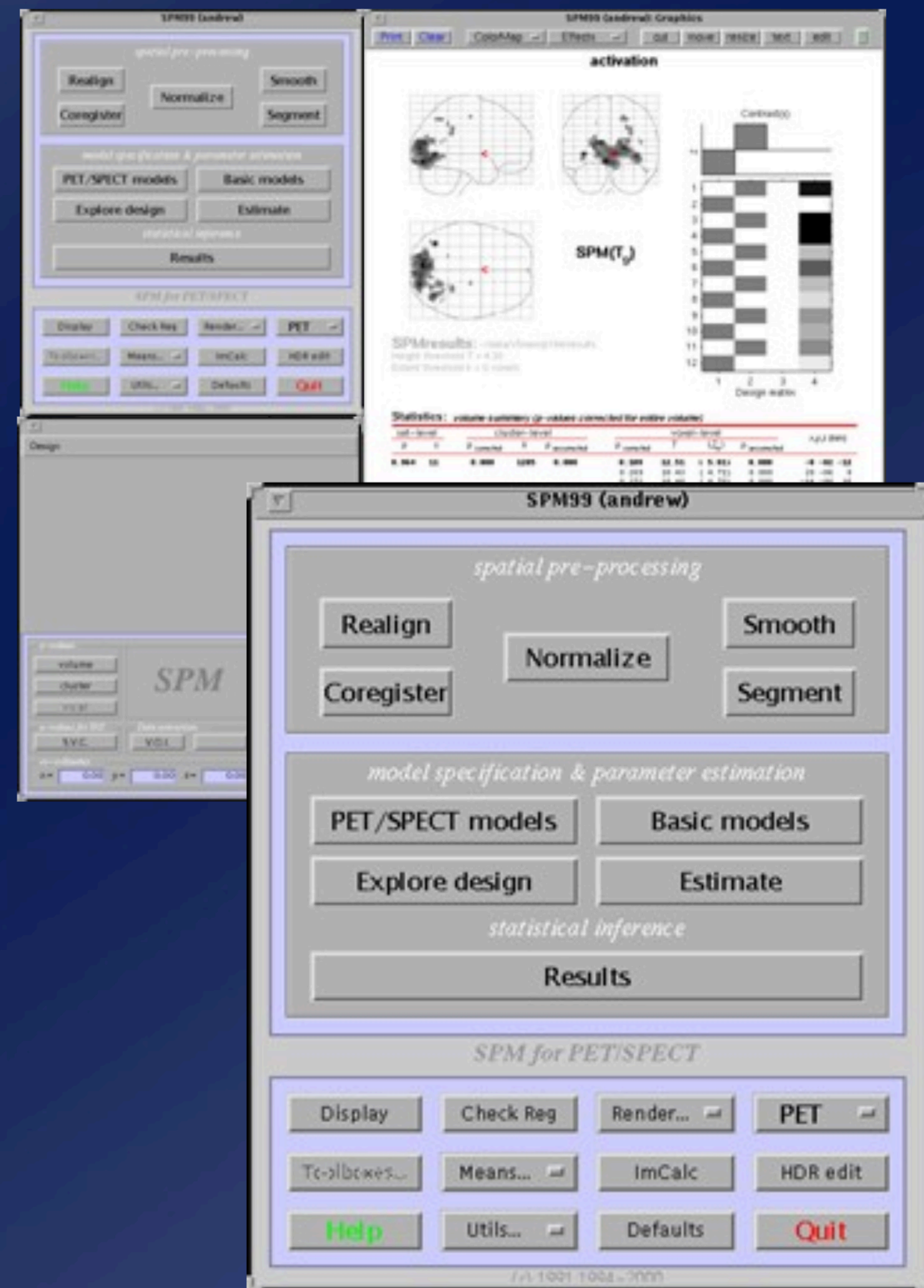
Amunts et al., (2000)

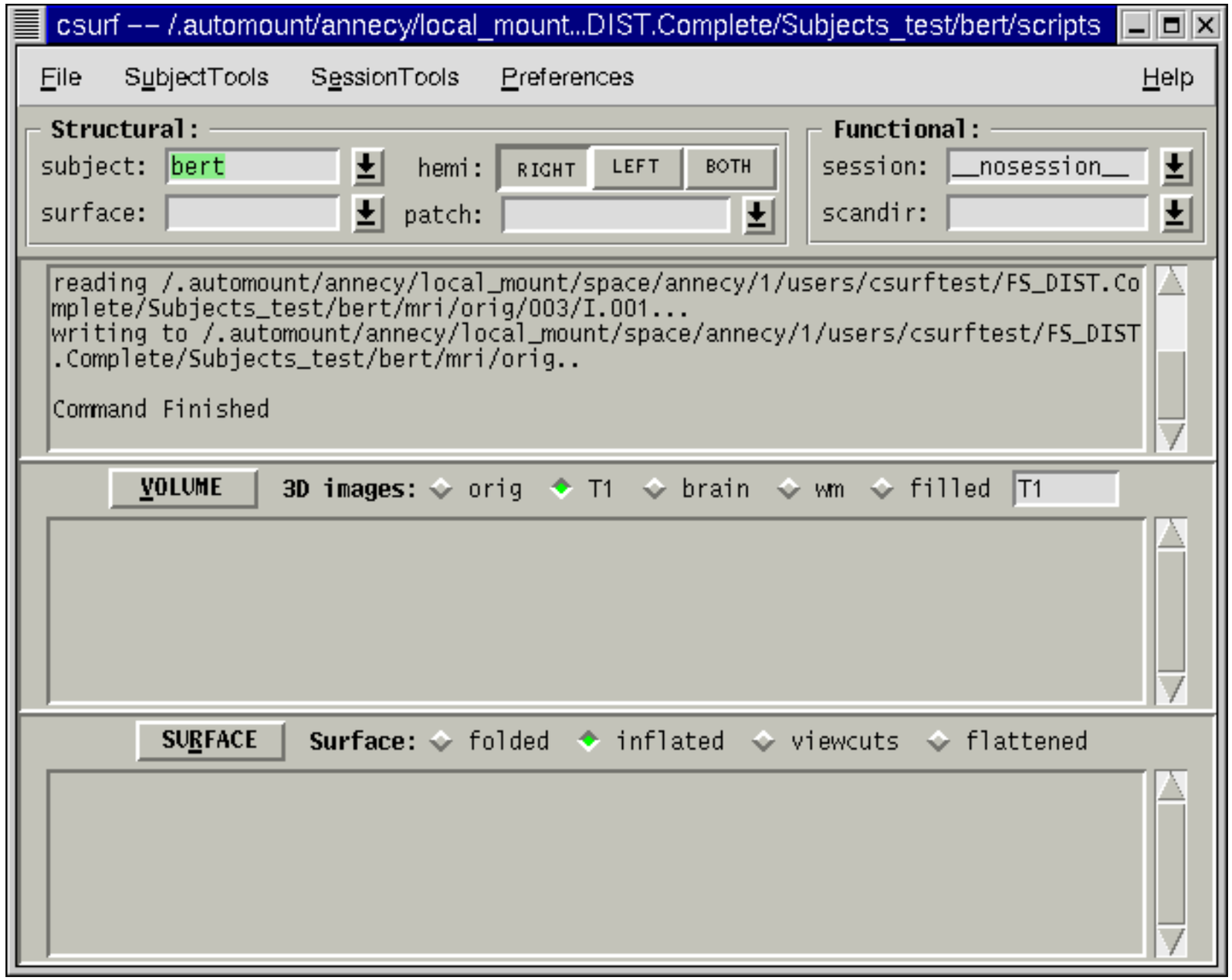


Nieto-Castanon et al., (2003)

SPM features…

- **Spatial**
  - realignment, spatial normalisation, segmentation, coregistration, spatial smoothing
- **Statistical**
  - voxel by voxel statistical analysis
  - general linear model, generalised for temporal autocorrelation
  - random effects analyses
  - multiple comparisons: Corrected *p*-values from random field theory
  - plotting & results interrogation
- **Utilities**
  - image display, CheckReg, rendering, brain extraction, adjusted means, image algebra




csurf -- /.automount/annecy/local_mount...DIST.Complete/Subjects_test/bert/scripts
File
SubjectTools
SessionTools
Preferences
Help
Structural:
subject: bert
hemi: RIGHT LEFT BOTH
surface:
patch:
Functional:
session: __nosession__
scandir:
reading /.automount/annecy/local_mount/space/annecy/1/users/csurftest/FS_DIST.Complete/Subjects_test/bert/mri/orig/003/I.001...
writing to /.automount/annecy/local_mount/space/annecy/1/users/csurftest/FS_DIST.Complete/Subjects_test/bert/mri/orig..
Command Finished
VOLUME
3D images: orig T1 brain wm filled T1
SURFACE
Surface: folded inflated viewcuts flattened

csurf -- /.automount/annecy/local_mount...DIST.Complete/Subjects_test/bert/scripts

File SubjectTools SessionTools Preferences Help

Structural:

subject: bert hemi: RIGHT LEFT BOTH

surface: patch:

Functional:

session: __nosession__

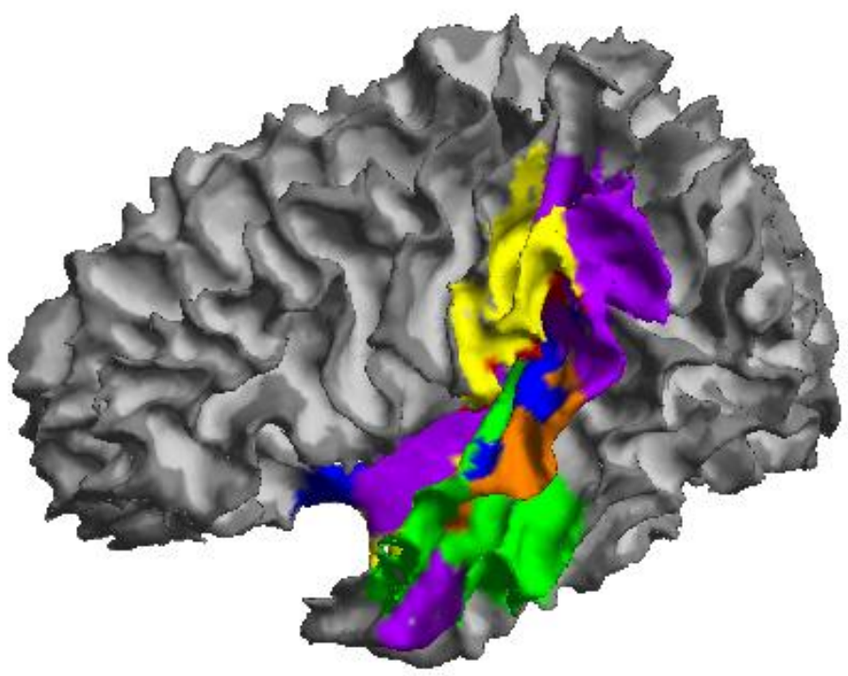
scandir:

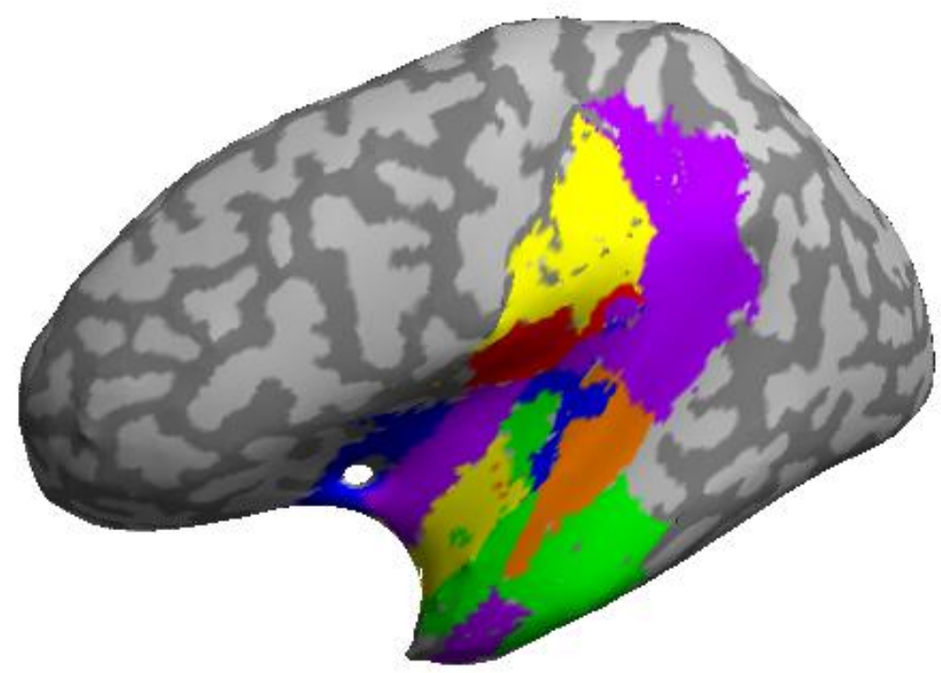
```
reading /.automount/annecy/local_mount/space/annecy/1/users/csurftest/FS_DIST.Co
mplete/Subjects_test/bert/mri/orig/003/I.001...
writing to /.automount/annecy/local_mount/space/annecy/1/users/csurftest/FS_DIST
.Complete/Subjects_test/bert/mri/orig..

Command Finished
```

VOLUME **3D images:** orig T1 brain wm filled T1

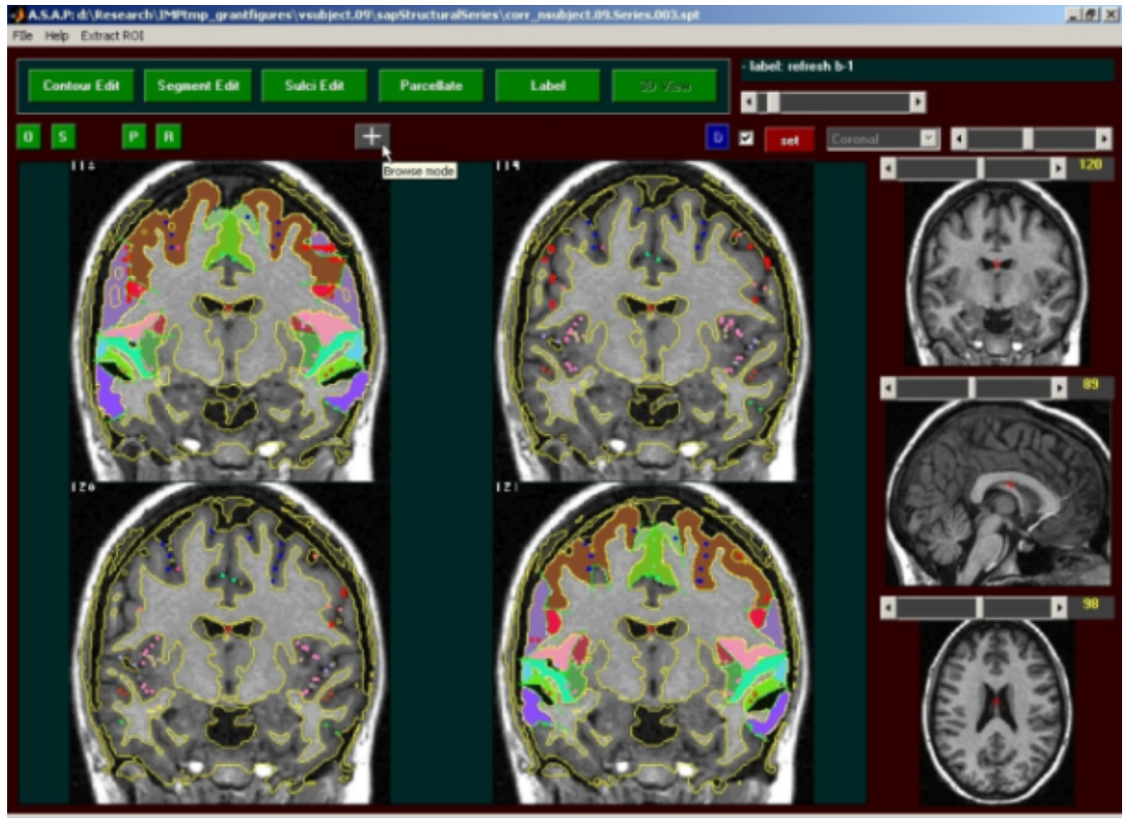
SURFACE **Surface:** folded inflated viewcuts flattened

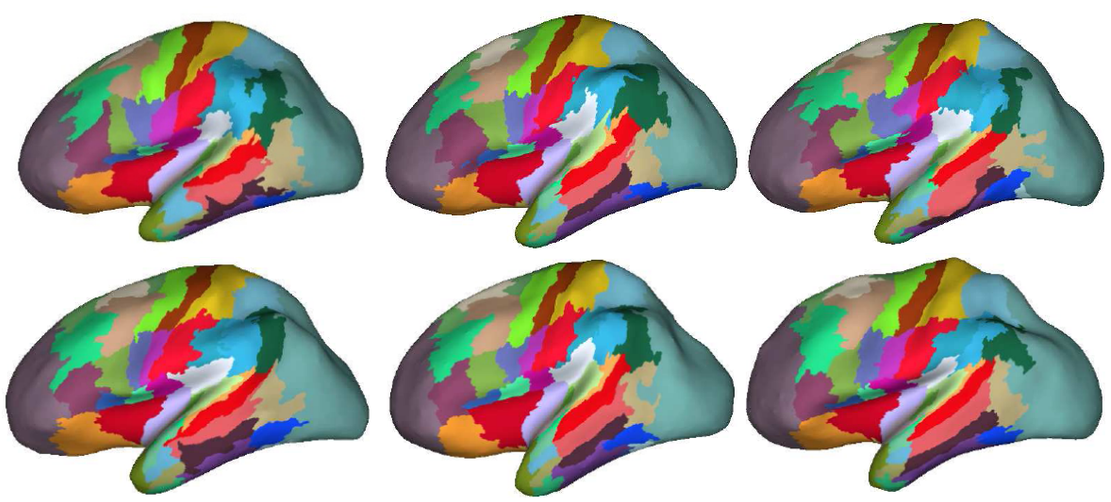




SpeechLab Toolkit
version 1.0

19 September 2003



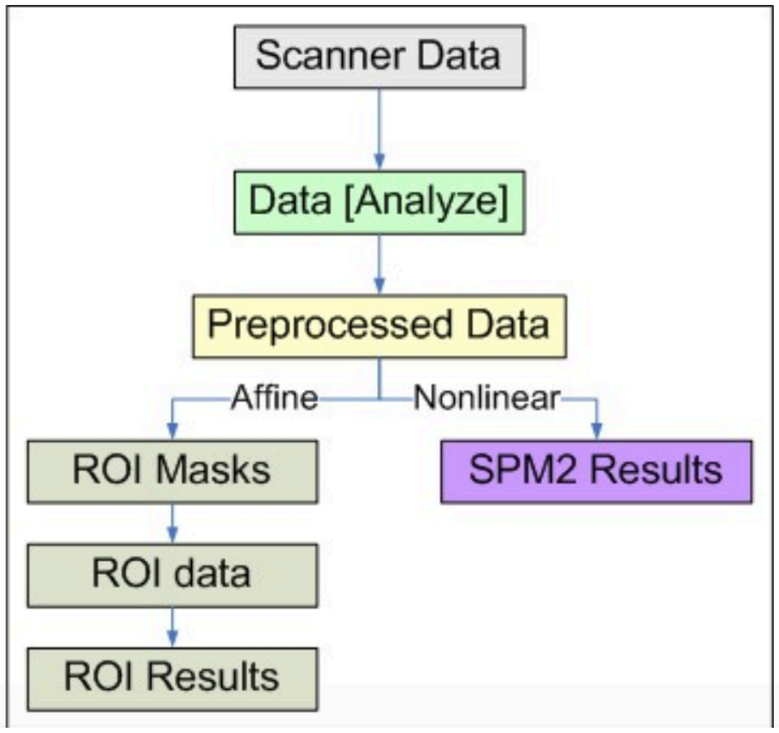


Components

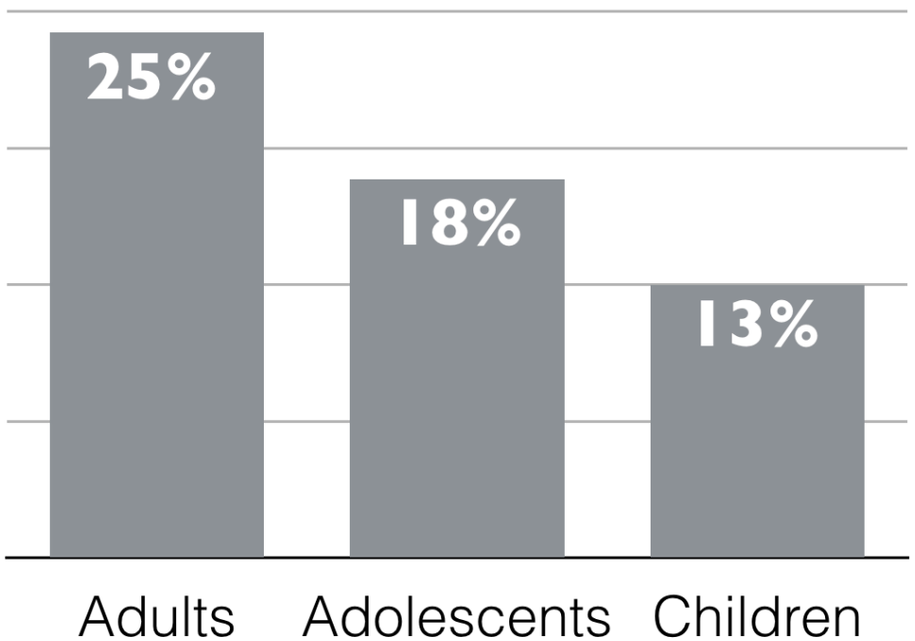
The **SLT** consists of several different components: some more general than others. The current composition is described in brief below.

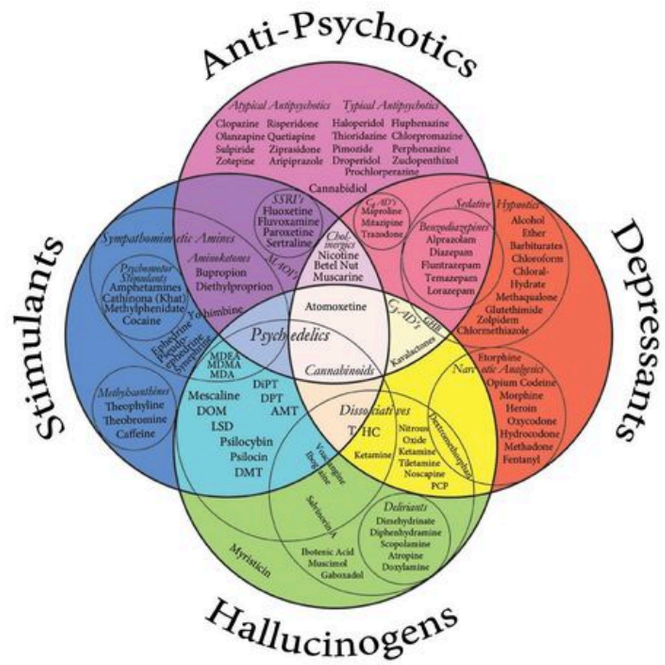
1) **ROI Analysis Toolbox 2 (RAT2):** This is the main component of the package and includes:
   - ✓ An SPM2 batch processing system
   - ✓ ROI analyses scripts (spm_ROI)
   - ✓ Figure creation scripts for both SPM2 and spm_ROI results
   - ✓ Preprocessing scripts for ASAP
   - ✓ FreeSurfer interaction tools (UNIX only)
2) **A Segmentation and Parcellation Program (ASAP):** A graphical interface for creating ROI masks for brain images. This tool is a Matlab based clone of the program CardViews.
3) **SurfTools:** A set of computational geometry tools to work with manifold triangular meshes. In addition it provides the means for including VTK functionality in Matlab.
4) **FSTools:** A set of Matlab scripts provided by the authors of FreeSurfer for accessing and manipulating different FreeSurfer file formats.

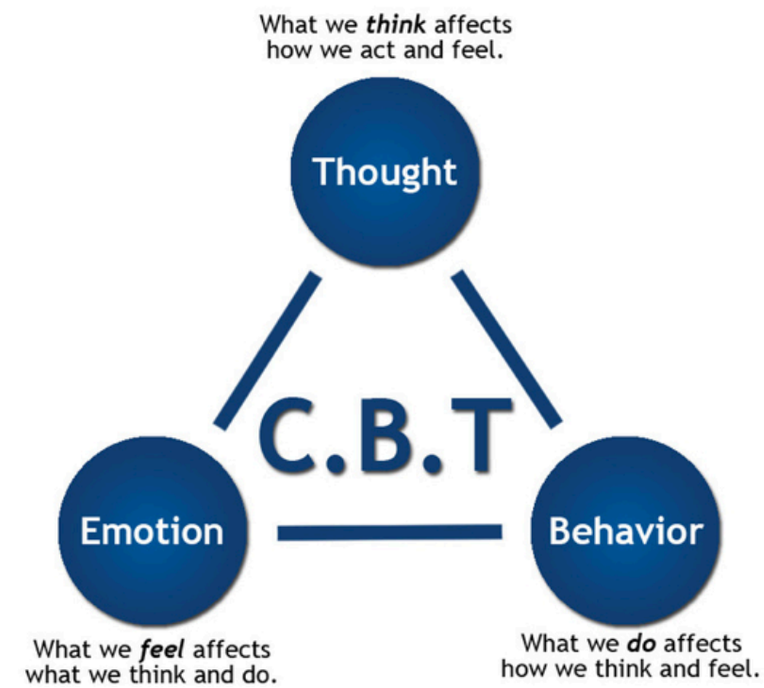
Operational flowchart

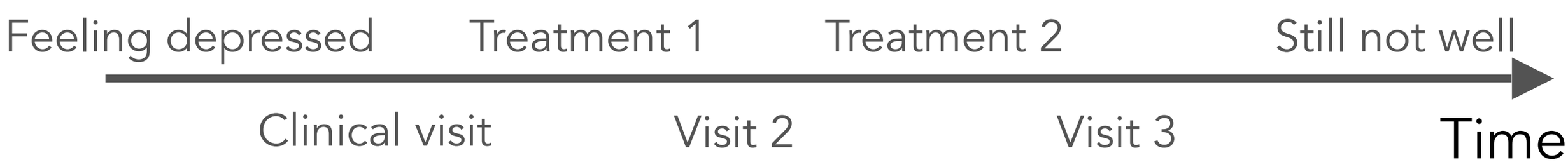


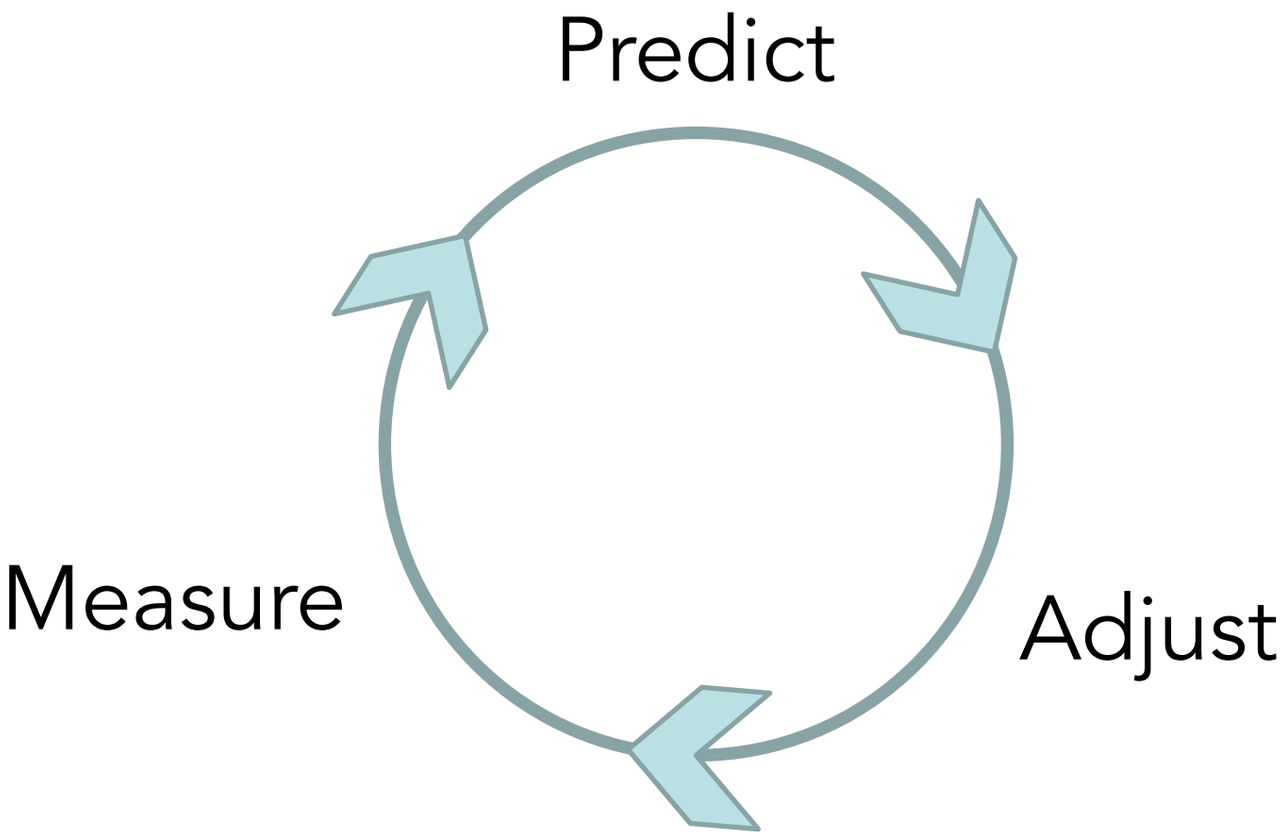
https://github.com/satra/SLT


25%
18%
13%
Adults
Adolescents
Children


Anti-Psychotics
Stimulants
Depressants
Hallucinogens
Atypical Antipsychotics
Typical Antipsychotics
Clopazine
Olanzapine
Sulpiride
Zotepine
Risperidone
Quetiapine
Ziprasidone
Aripiprazole
Haloperidol
Thioridazine
Pimozide
Droperidol
Fluphenazine
Chlorpromazine
Perphenazine
Zuclopenthixol
Prochlorperazine
Cannabidiol
SSRI's
Fluoxetine
Fluvoxamine
Paroxetine
Sertraline
Nicotine
Betel Nut
Muscarine
Atomoxetine
Psychedelics
Cannabinoids
Sympathomimetic Amines
Amphetamines
Cathinona (Khat)
Methylphenidate
Cocaine
Aminoketones
Bupropion
Diethylproprion
Yohimbine
Methylxanthines
Theophyline
Theobromine
Caffeine
MDE
MDMA
MDA
DiPT
DPT
AMT
Mescaline
DOM
LSD
Psilocybin
Psilocin
DMT
Dissociatives
THC
Ketamine
Nitrous Oxide
Ketamine
Tiletamine
Noscapine
PCP
Deliriants
Dimehydrinate
Diphenhydramine
Scopolamine
Atropine
Doxylamine
Ibotenic Acid
Muscimol
Gabonadol
Myristicin
Sedative Hypnotics
Alcohol
Ether
Barbiturates
Chloroform
Chloral-Hydrate
Methaqualone
Glutethimide
Zolpidem
Chlormethiazole
Benzodiazepines
Alprazolam
Diazepam
Flunitrazepam
Temazepam
Lorazepam
Narcotic Analgesics
Etorphine
Opium Codeine
Morphine
Heroin
Oxycodone
Hydrocodone
Methadone
Fentanyl


What we *think* affects how we act and feel.
Thought
C.B.T
Emotion
Behavior
What we *feel* affects what we think and do.
What we *do* affects how we think and feel.


Feeling depressed
Treatment 1
Treatment 2
Still not well
Clinical visit
Visit 2
Visit 3
Time



PROGNOSIS

DIAGNOSIS

TREATMENT SELECTION

TREATMENT DEVELOPMENT

MONITORING

"how to commit suicide"
Search term

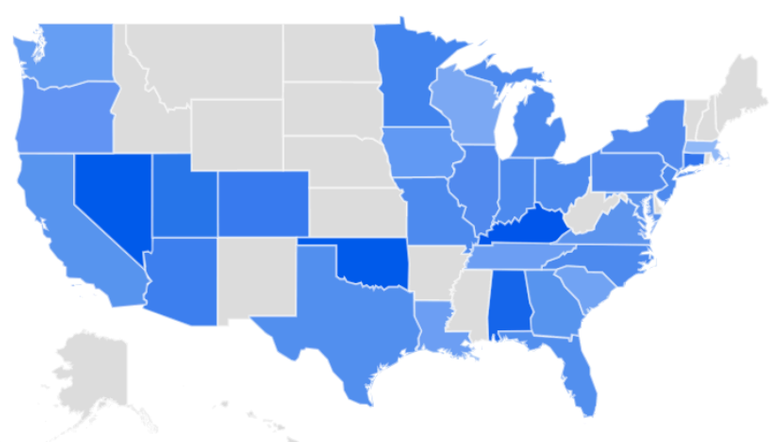
+ Compare

United States | Past 7 days | All categories | Web Search

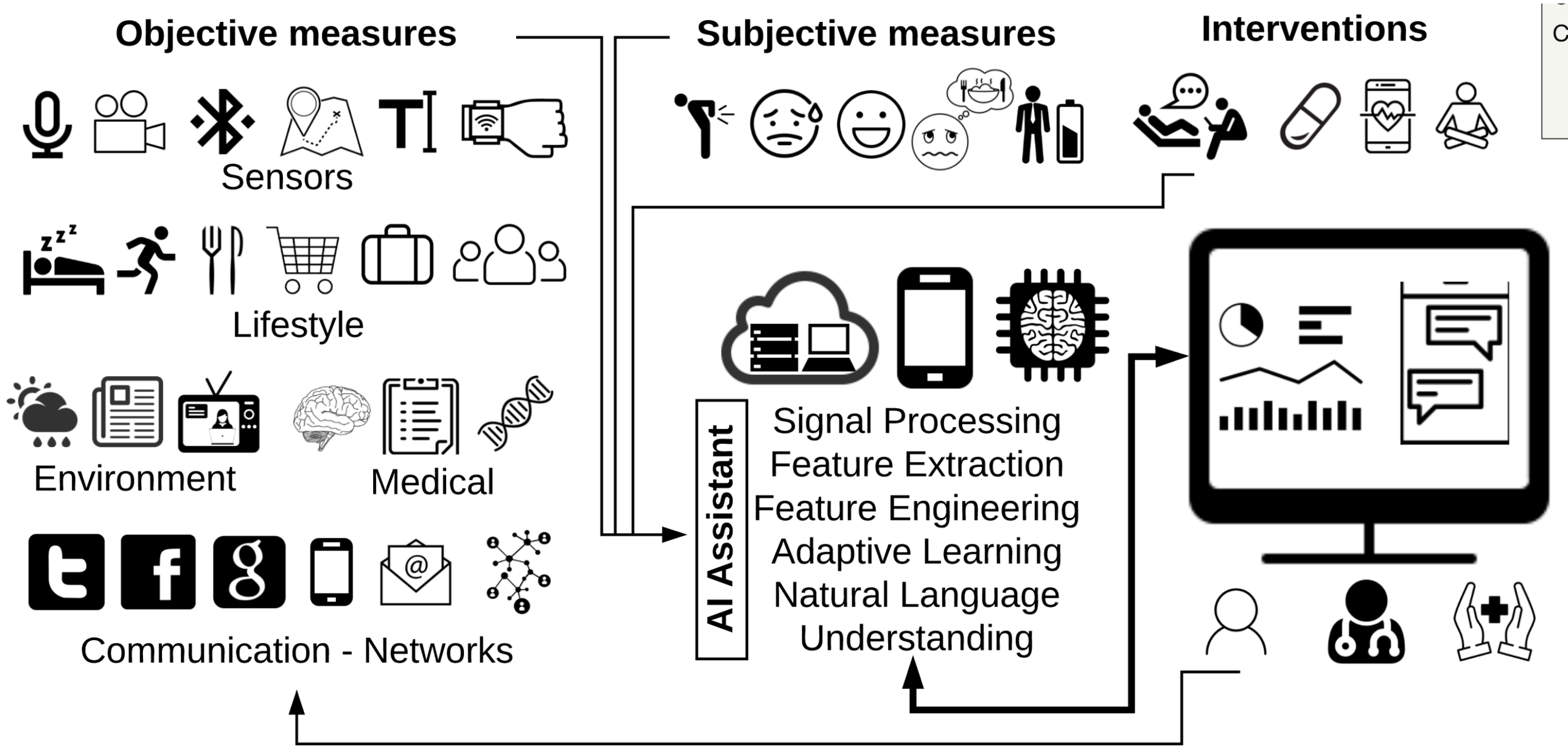
Interest over time

Interest by sub-region

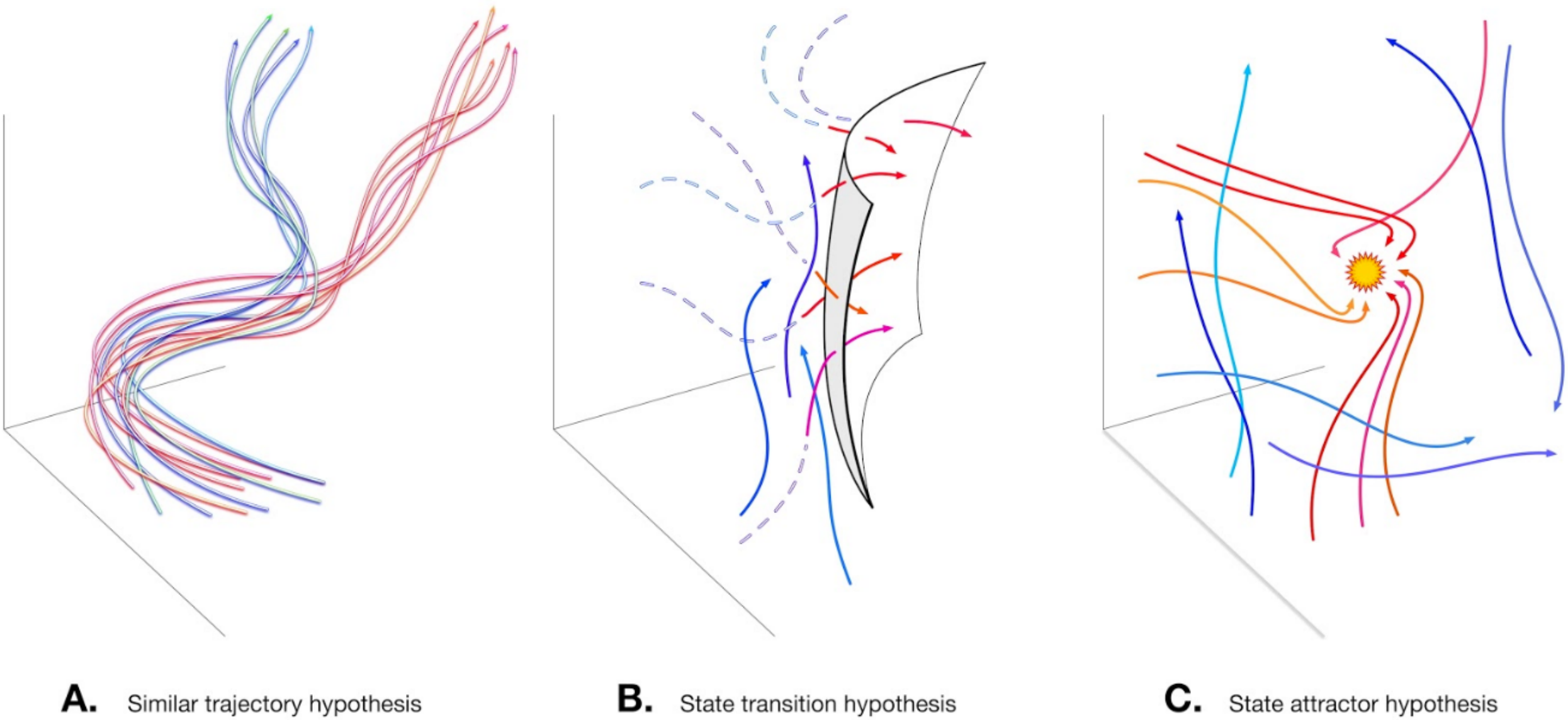
Sub-region



1	Kentucky	100
2	Oklahoma	97
3	Nevada	97
4	Alabama	88

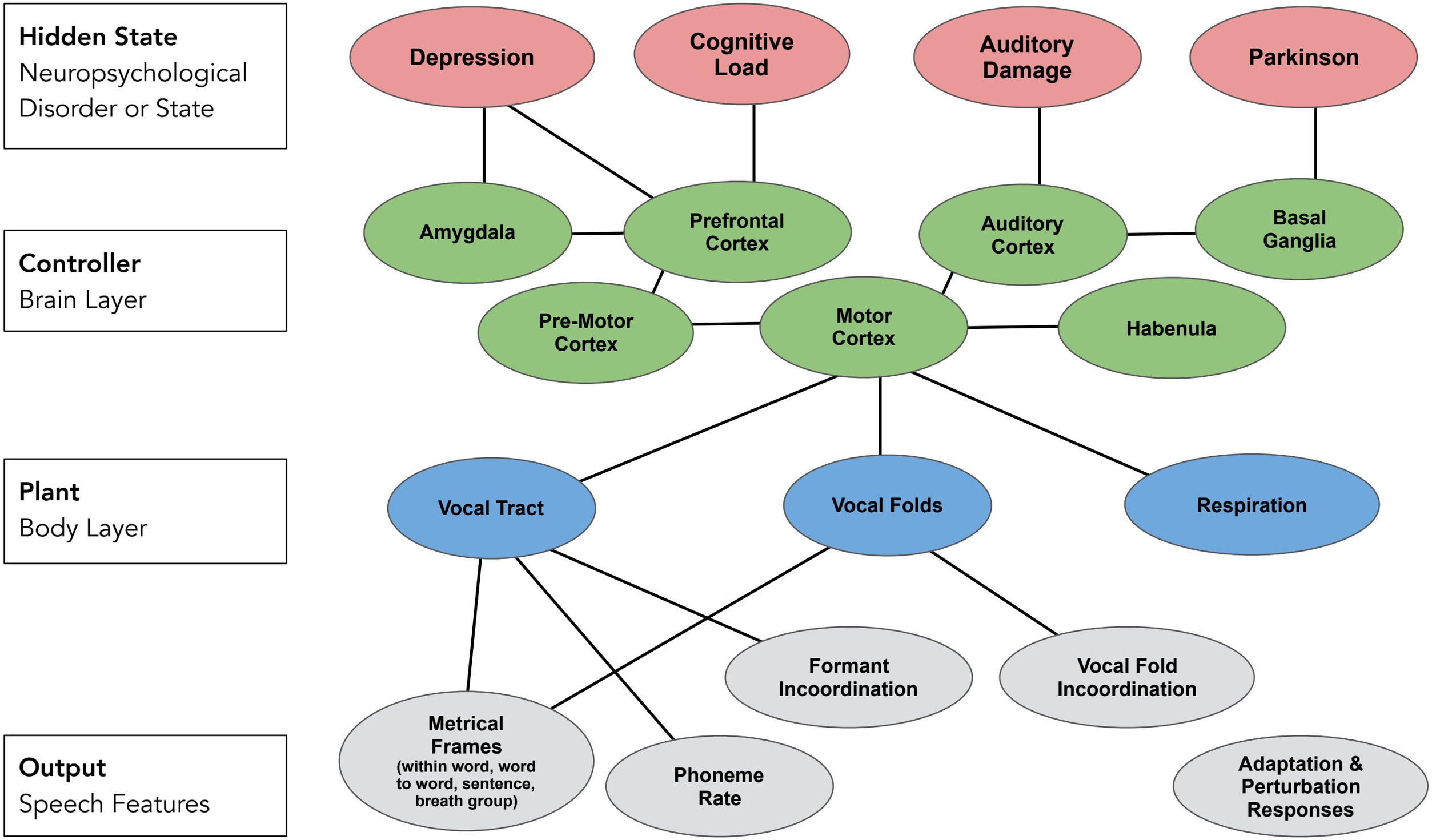

Objective measures
Sensors
Lifestyle
Environment
Medical
Communication - Networks
Subjective measures
Interventions
AI Assistant
Signal Processing
Feature Extraction
Feature Engineering
Adaptive Learning
Natural Language
Understanding

High-dimensional State Space for Health



Klein and Ghosh (2015)

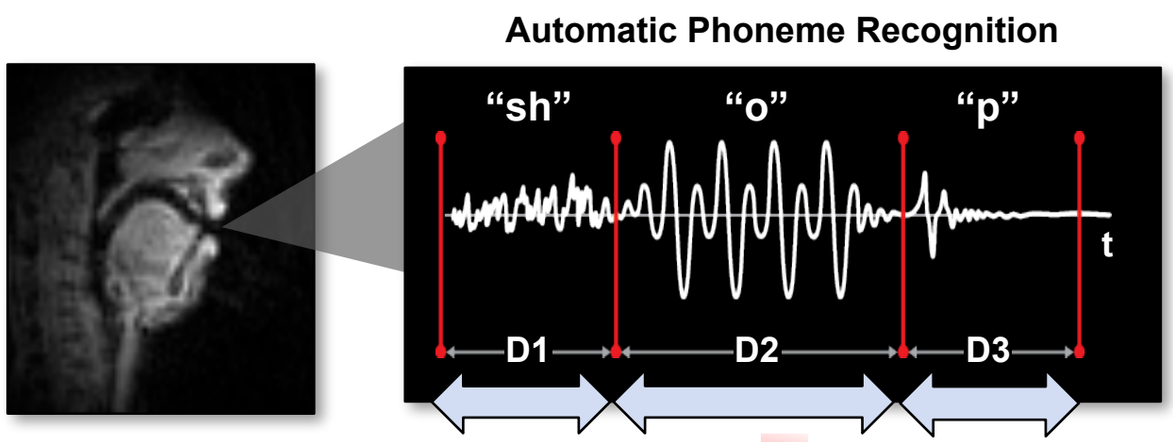
The Hypothetical Graph

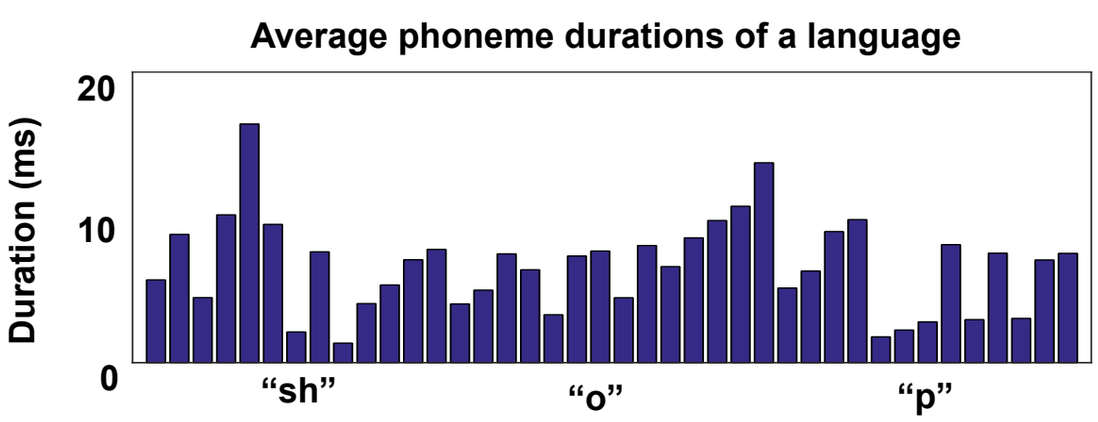


** Not all brain regions and connections shown for clarity*

MIT LL Novel Coordination features

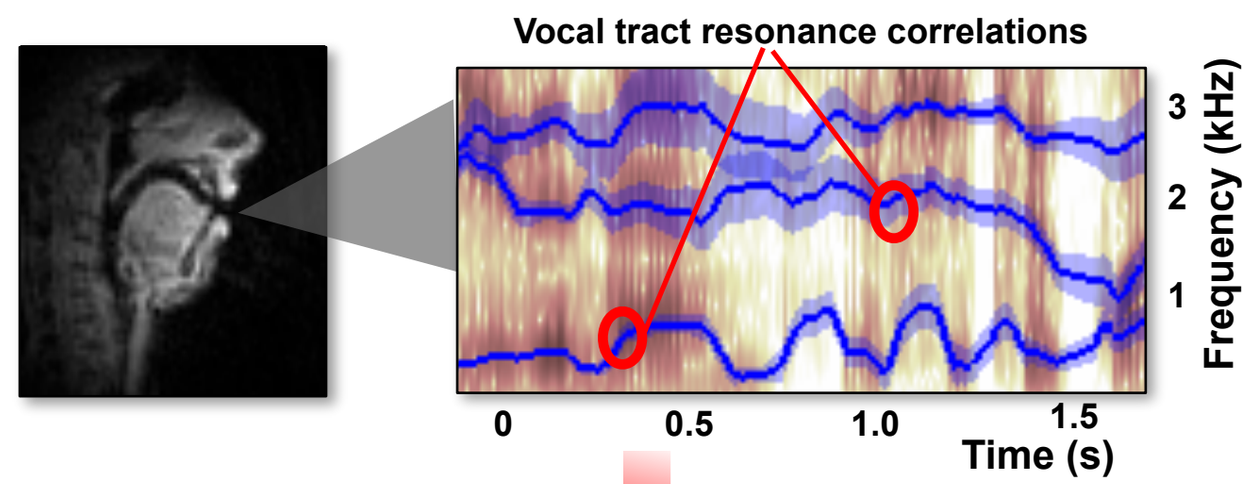
Changing of speaking rate at phoneme level

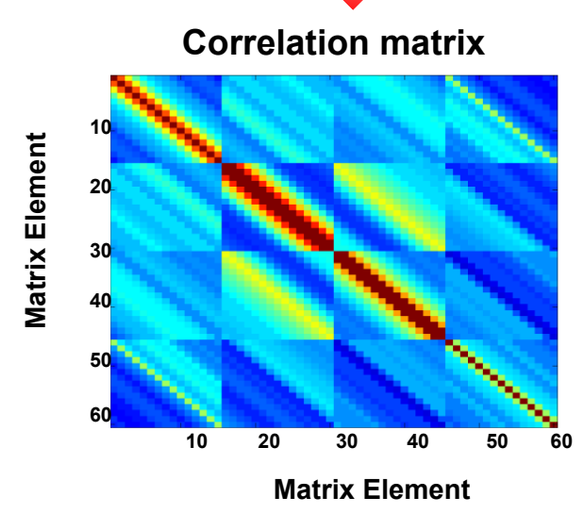




Average durations to quantify decrease *or* increase in rate

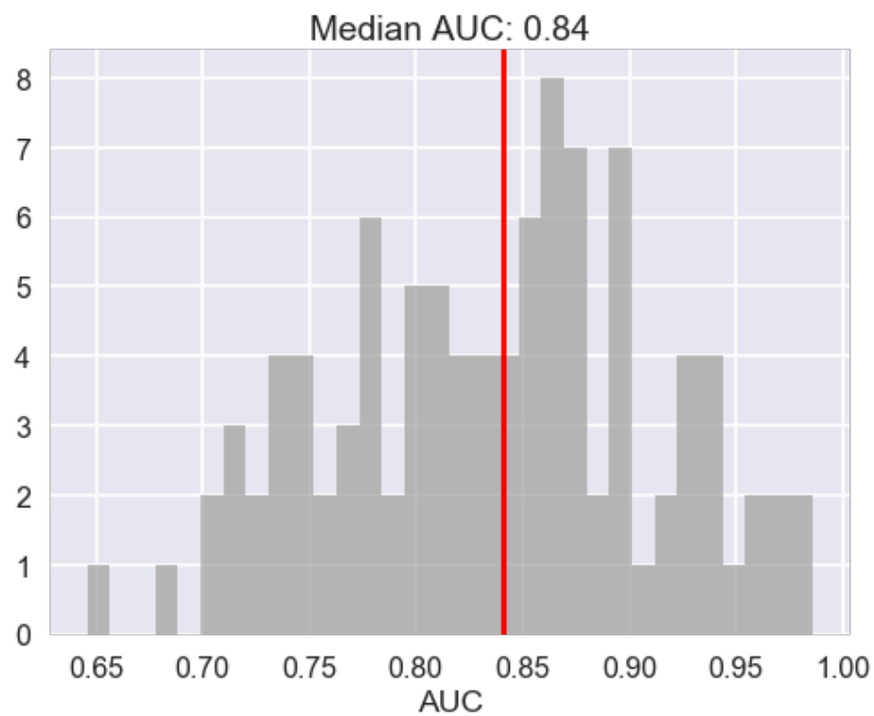
Loss of motor coordination in speech articulation



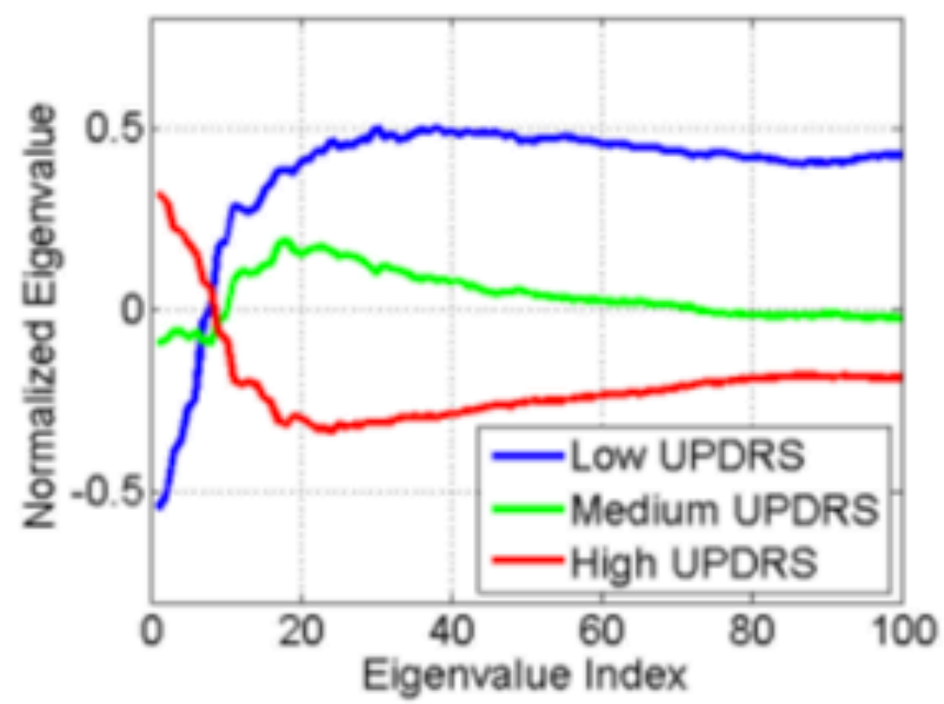


Eigenvalues to quantify coordination

Vocal Fold Paralysis

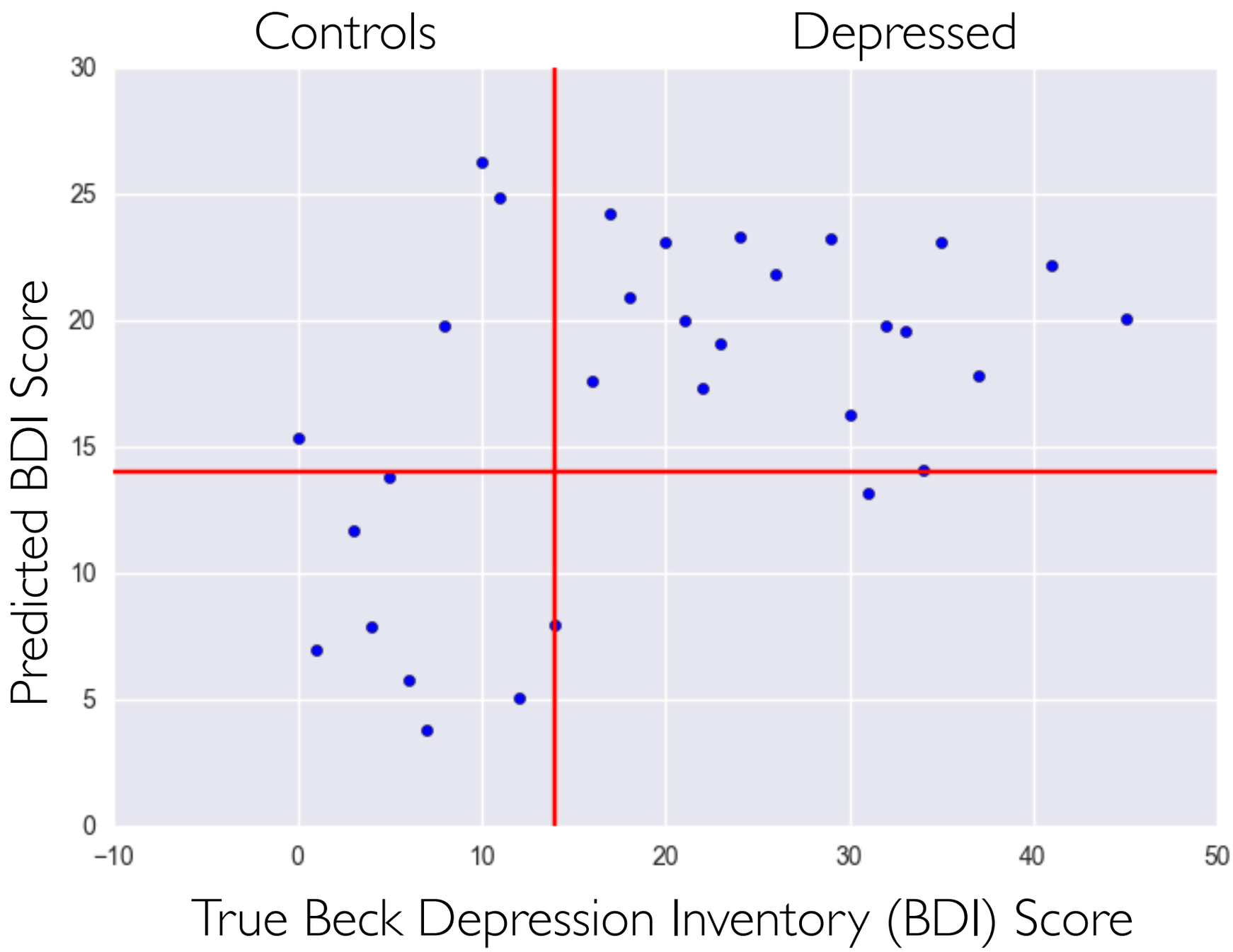


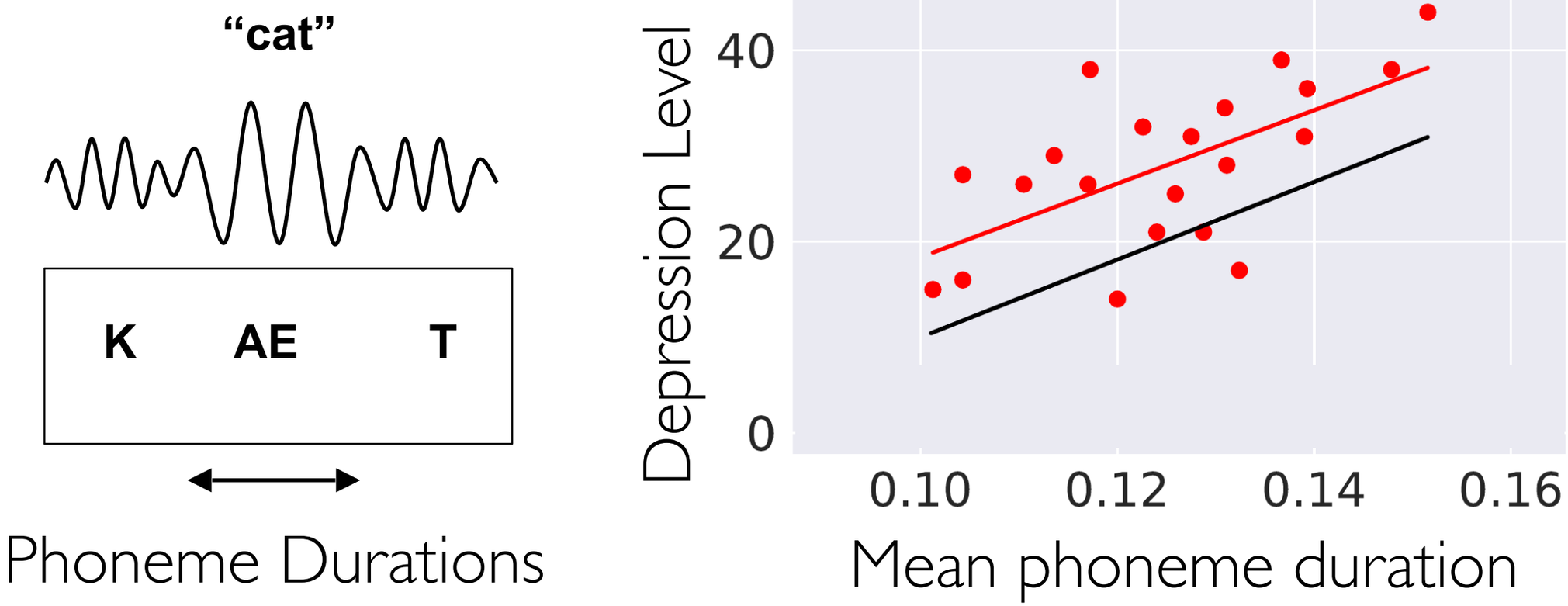
Parkinson severity



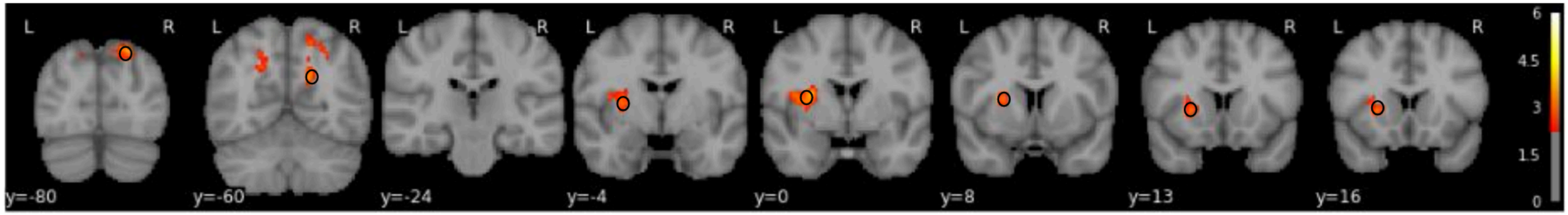
Williamson, Quatieri, Helfer, Perricone, Ghosh, Ciccarelli, Mehta (2015)

Predicting BDI levels from a sustained vowel



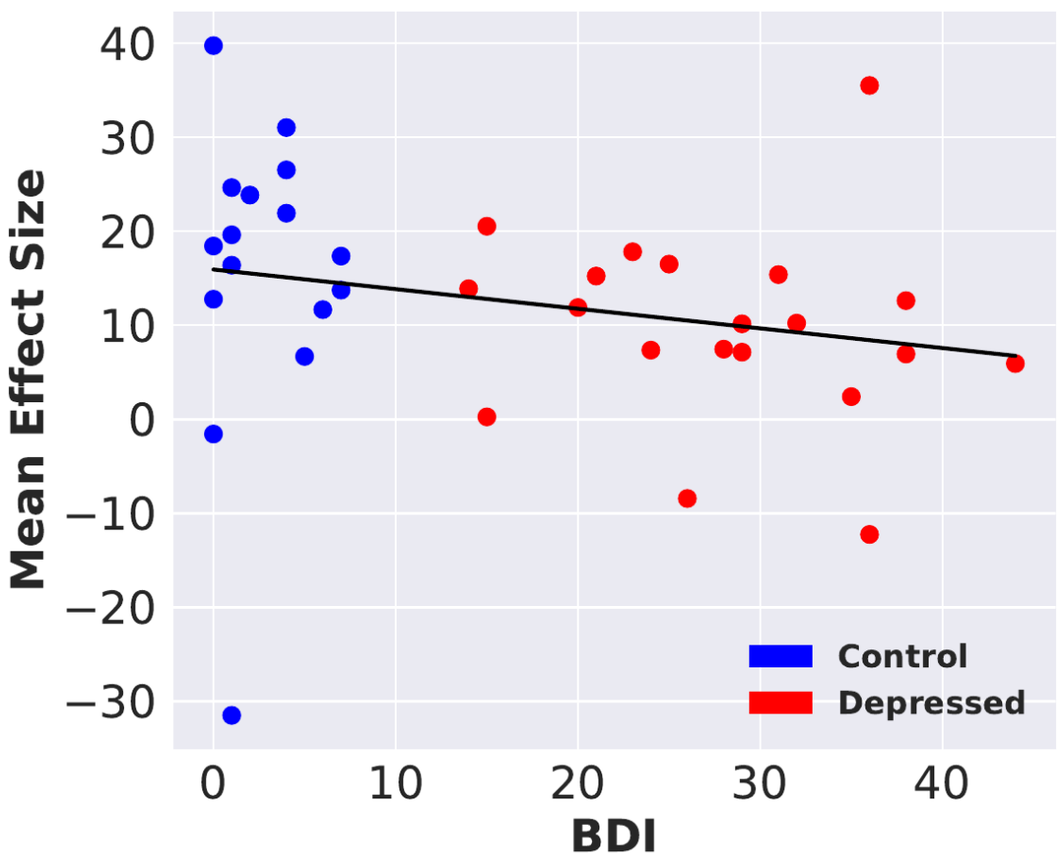

"cat"
K
AE
T
Phoneme Durations
Depression Level
40
20
0
0.10
0.12
0.14
0.16
Mean phoneme duration

Putamen activity during speaking is negatively correlated with BDI



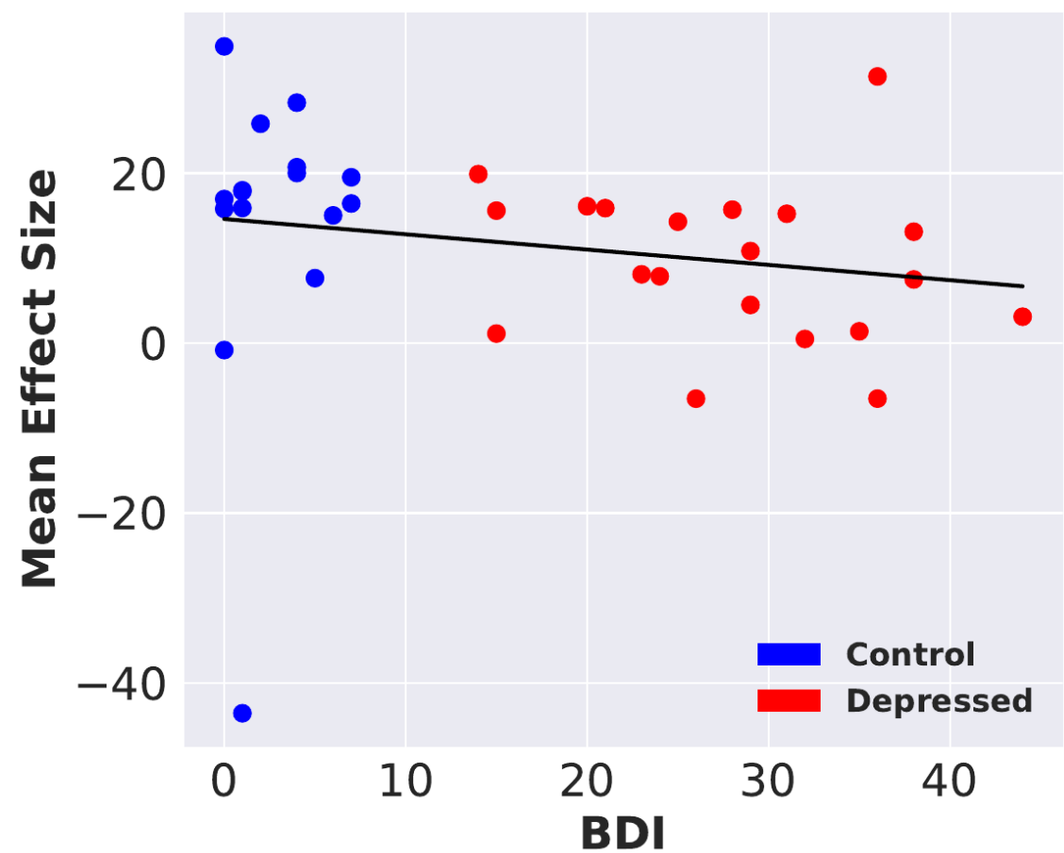
Lateral Superior Occipital Cortex | **Precuneus** | **Putamen** | **Central Opercular Cortex** | **Putamen** | **Putamen** | **Putamen and Insular Cortex**

Left Putamen



$R_{spearman}$ = - 0.38

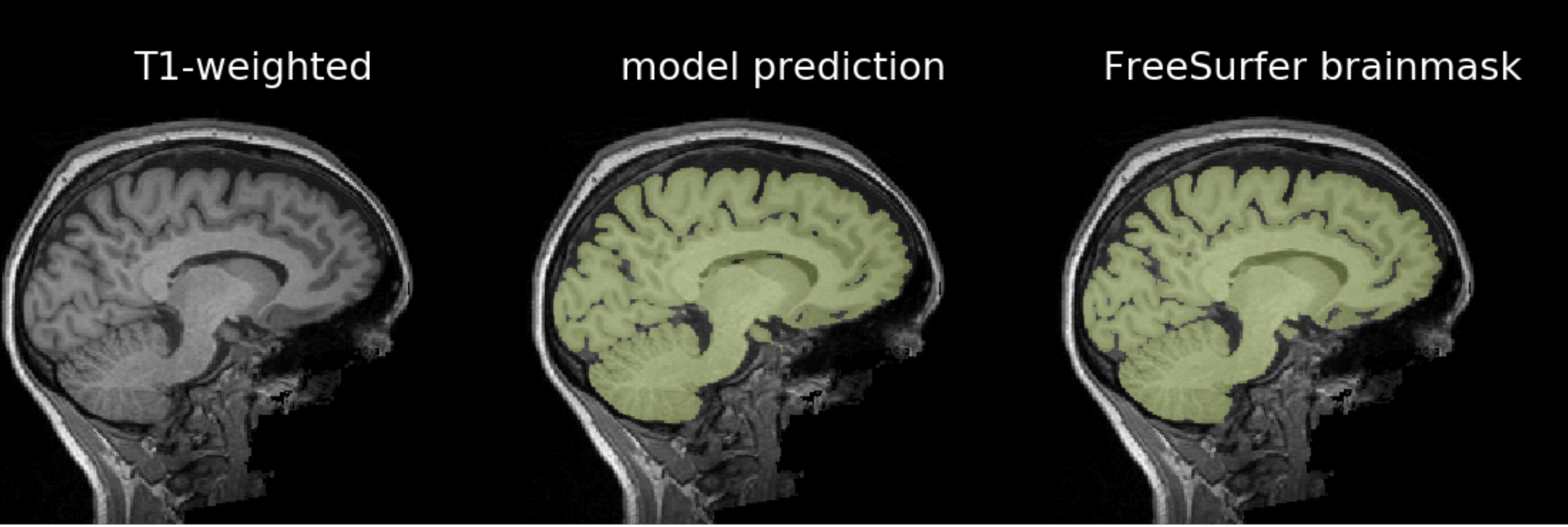
Right Putamen



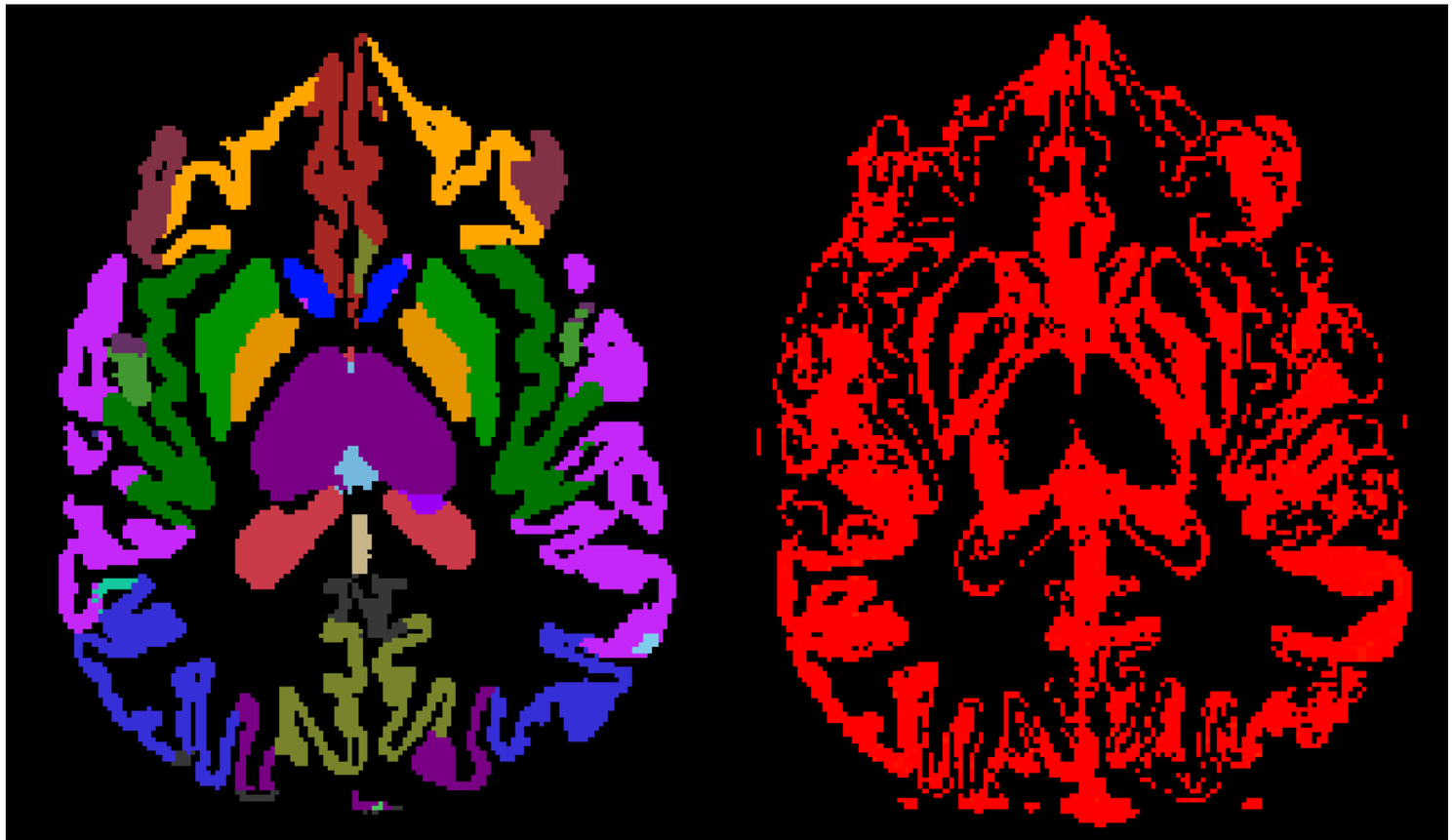
$R_{spearman}$ = - 0.43

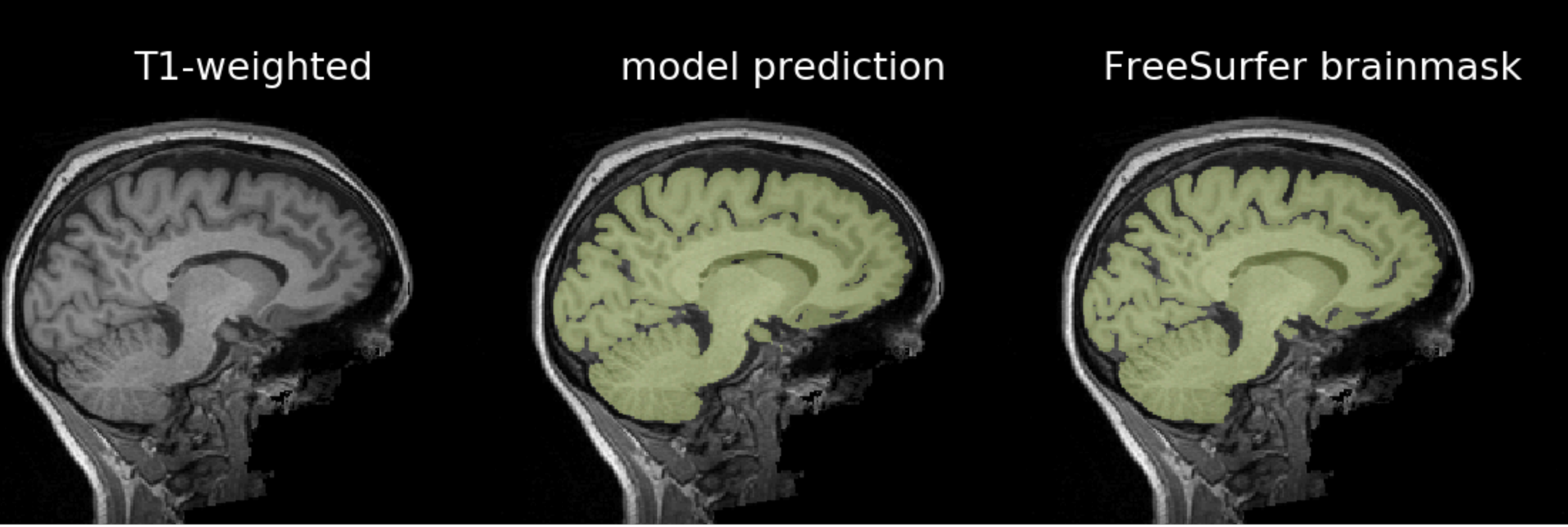
Karras et al. (2019)

https://tinyurl.com/2d-brain-gan

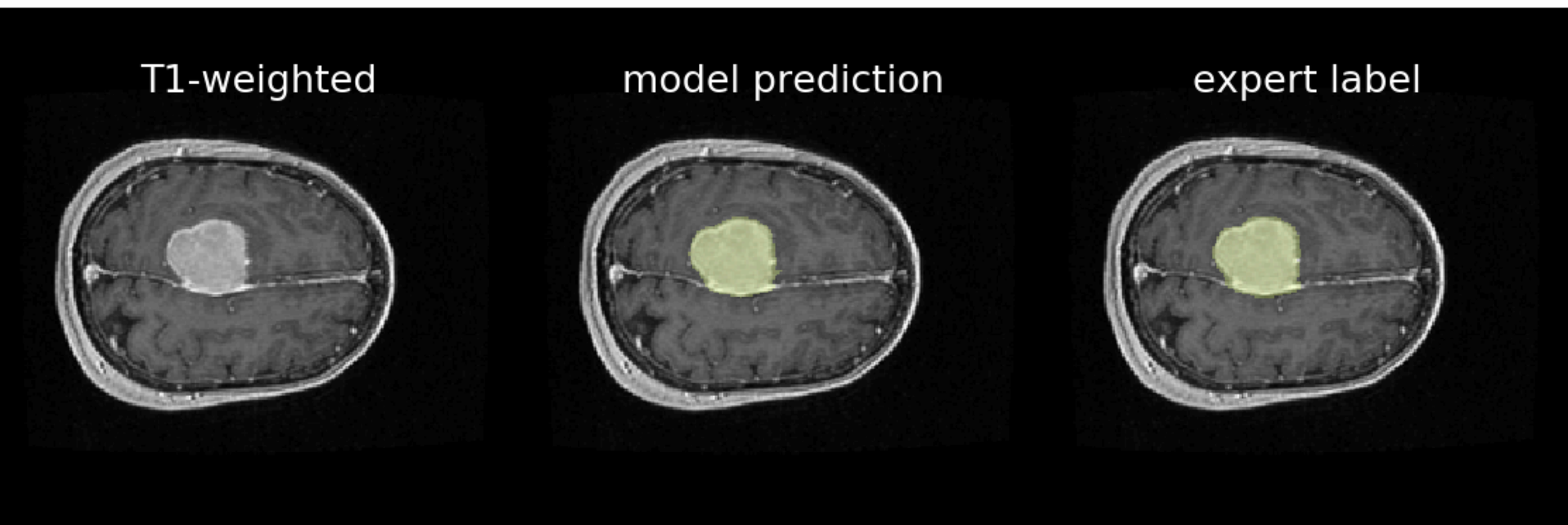


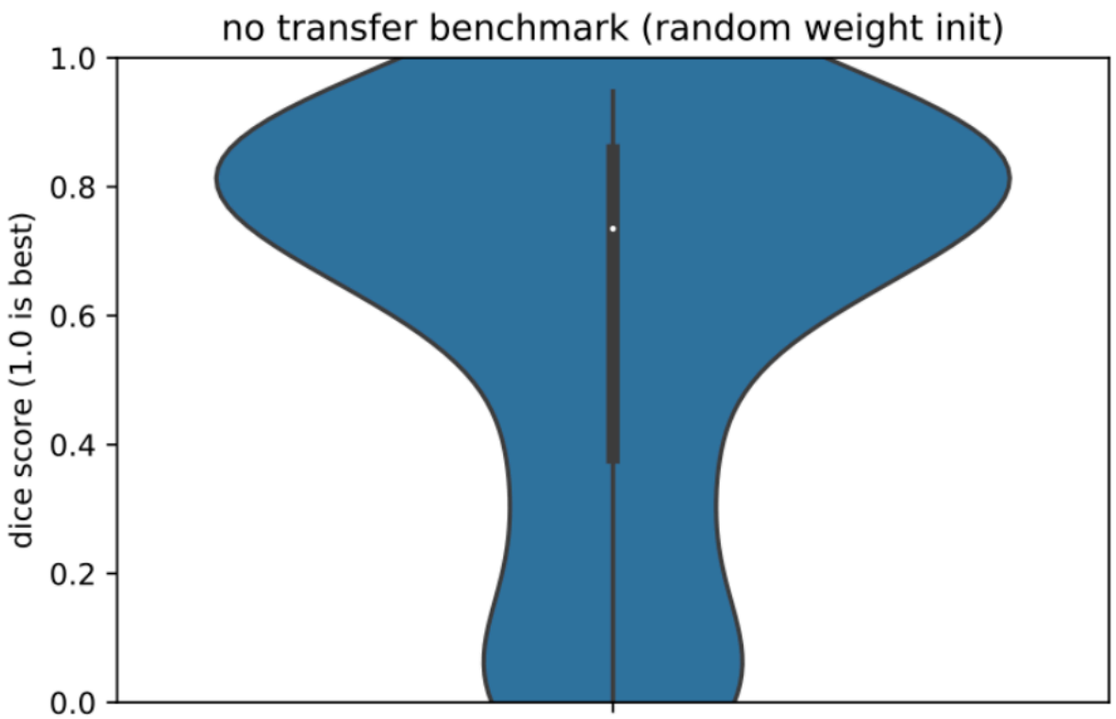
https://github.com/neuronets/nobrainer

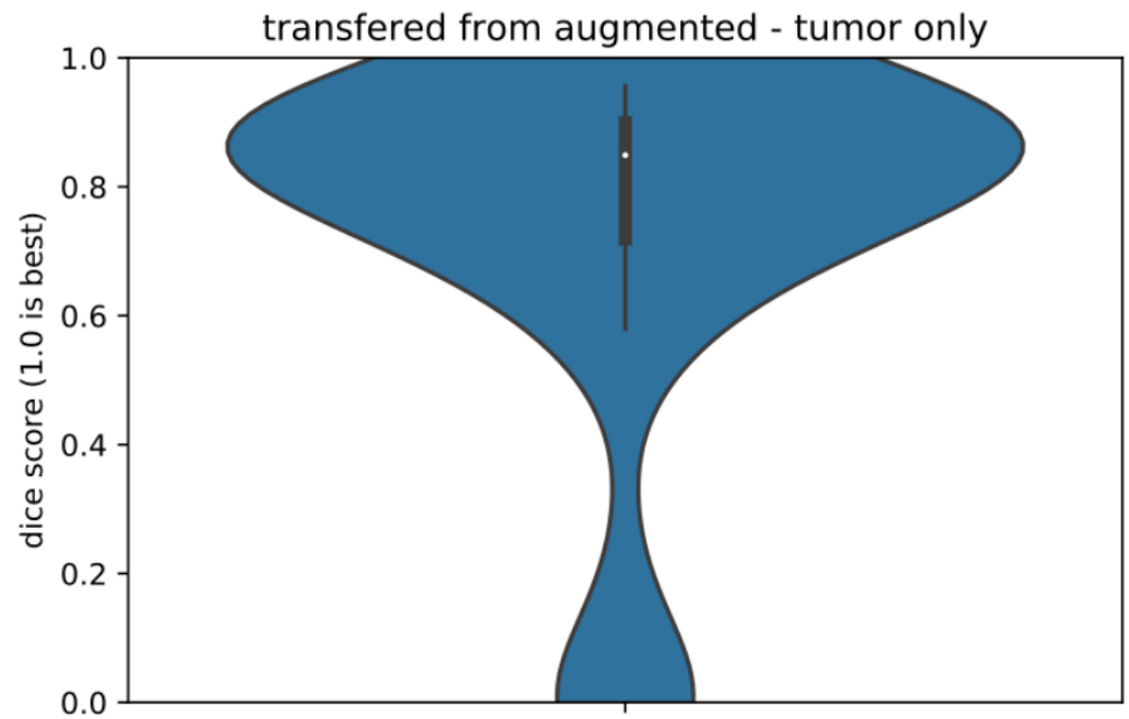




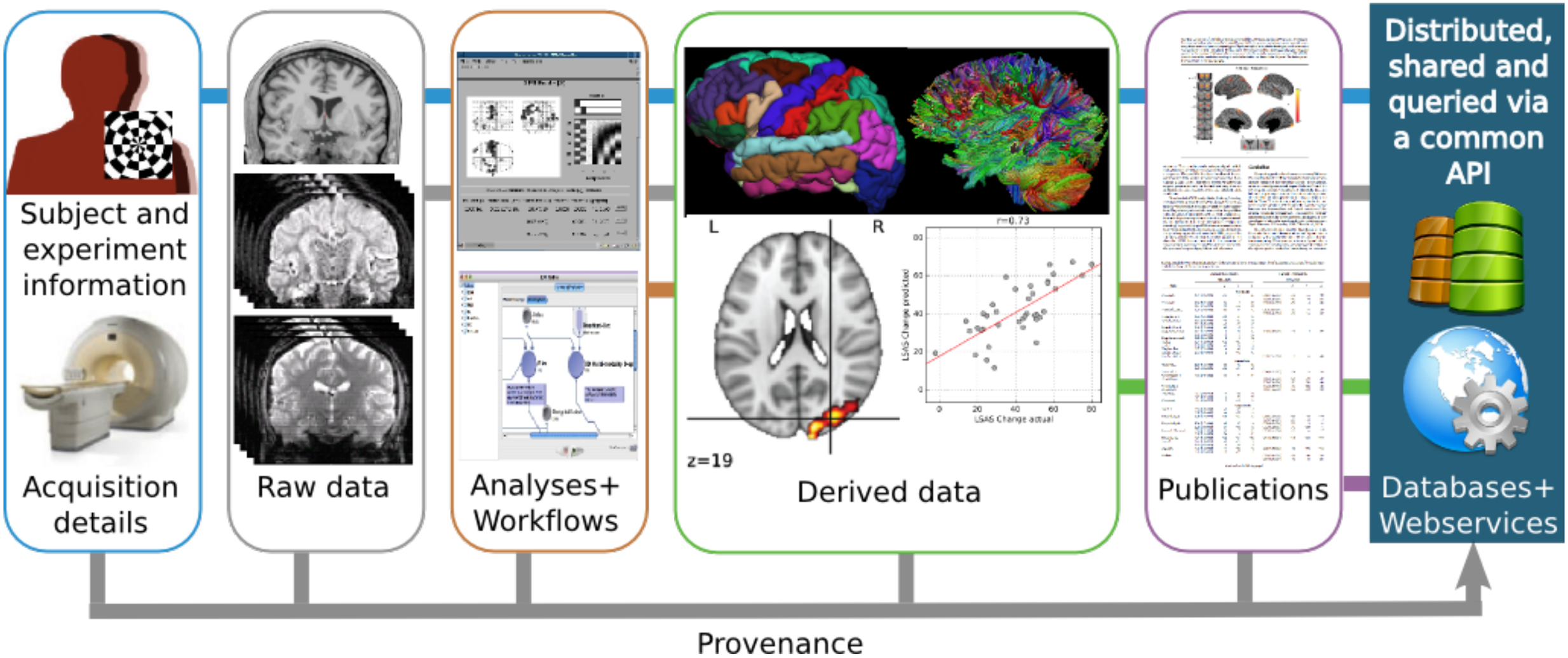
https://github.com/neuronets/nobrainer




no transfer benchmark (random weight init)
dice score (1.0 is best)
1.0
0.8
0.6
0.4
0.2
0.0


transfered from augmented - tumor only
dice score (1.0 is best)
1.0
0.8
0.6
0.4
0.2
0.0

The Dataflow in Brain Imaging



Demographics
Stimuli
Scanner
Sequence

Assays
- Clinical
- Neuropsych
- Behavioral

MRI
- T1
- T2
dMRI
fMRI
- Task
- Resting
EEG, MEG,
PET, MRS

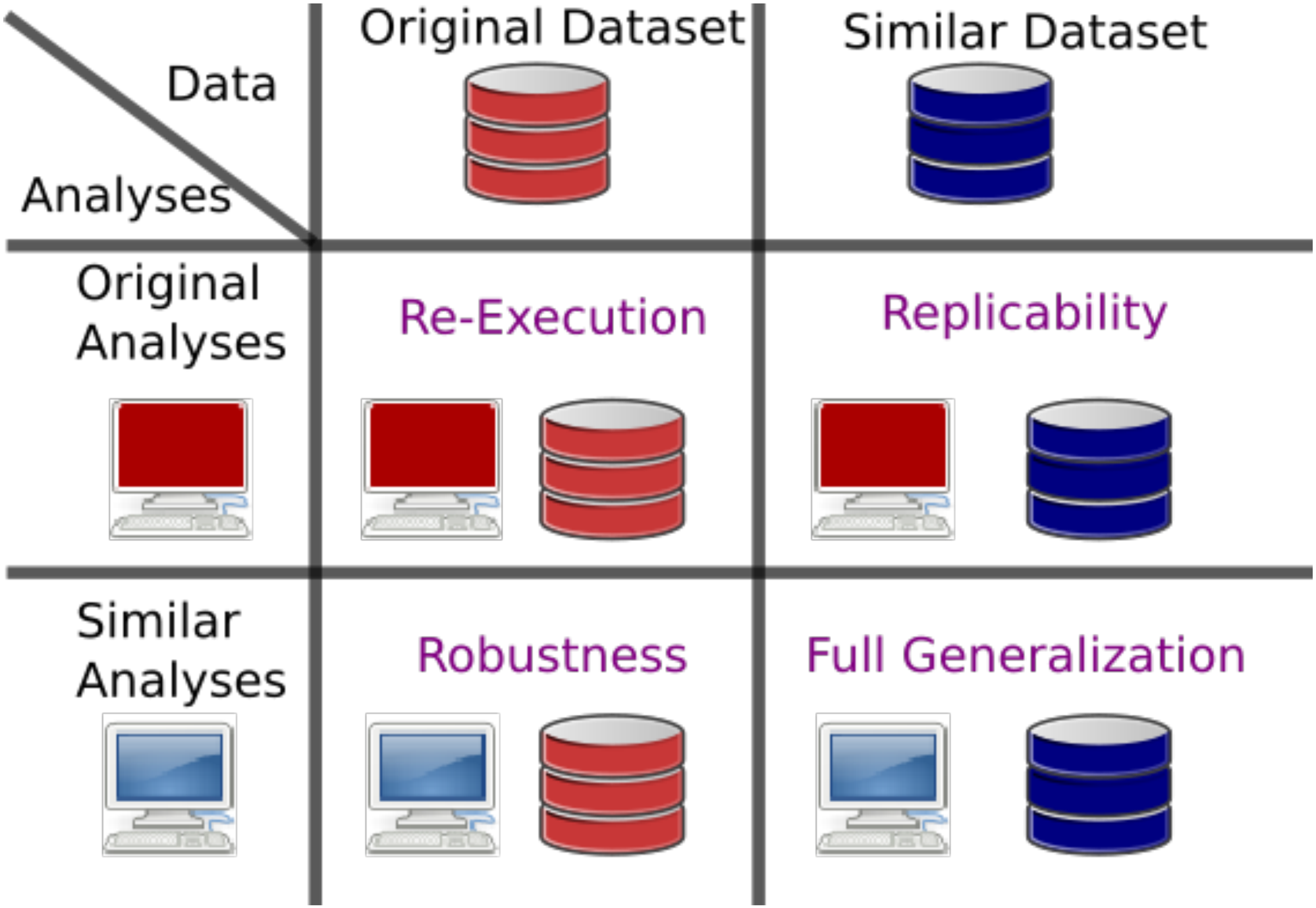
SPM
FSL
FreeSurfer
AFNI
...
Nipype
LONI
...
QA/QC

Contrasts
Correlation
Prediction

Journals
Preprint
archives
Websites
Blogs

XNAT
IDA
LORIS
COINS
Amazon S3
DropBox
Globus
Neurovault
Neurosynth

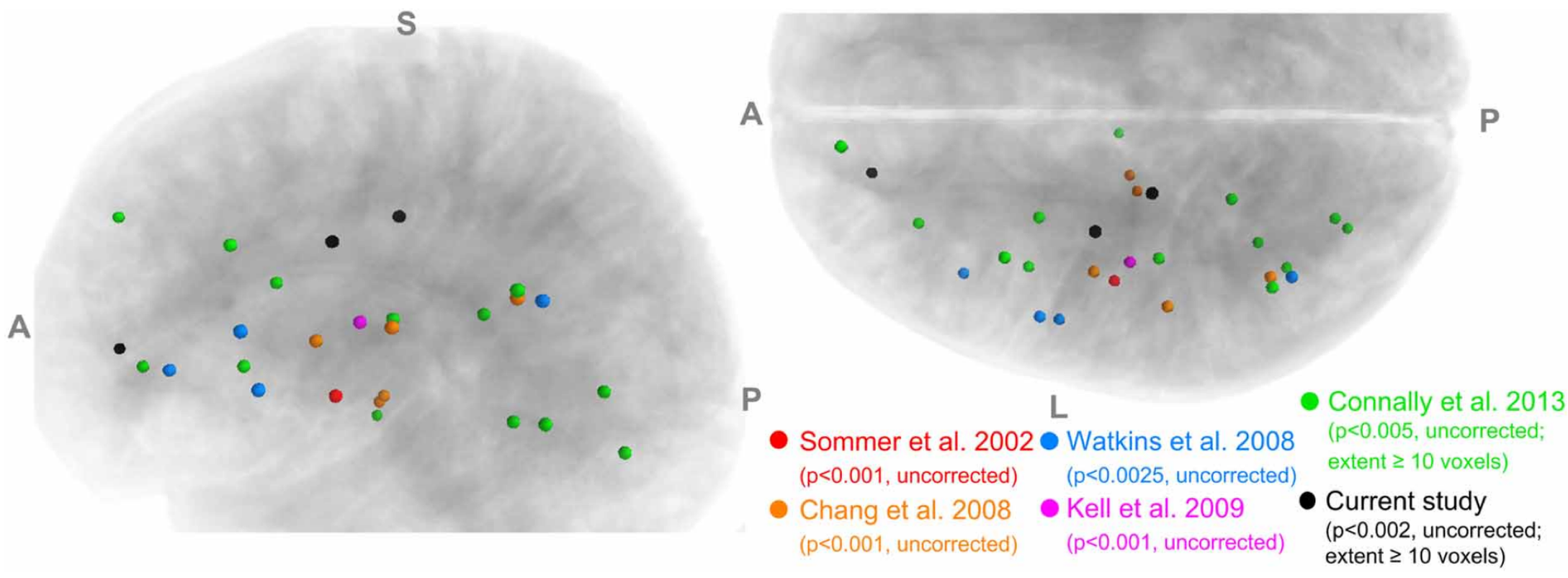
Poline et al. 2011



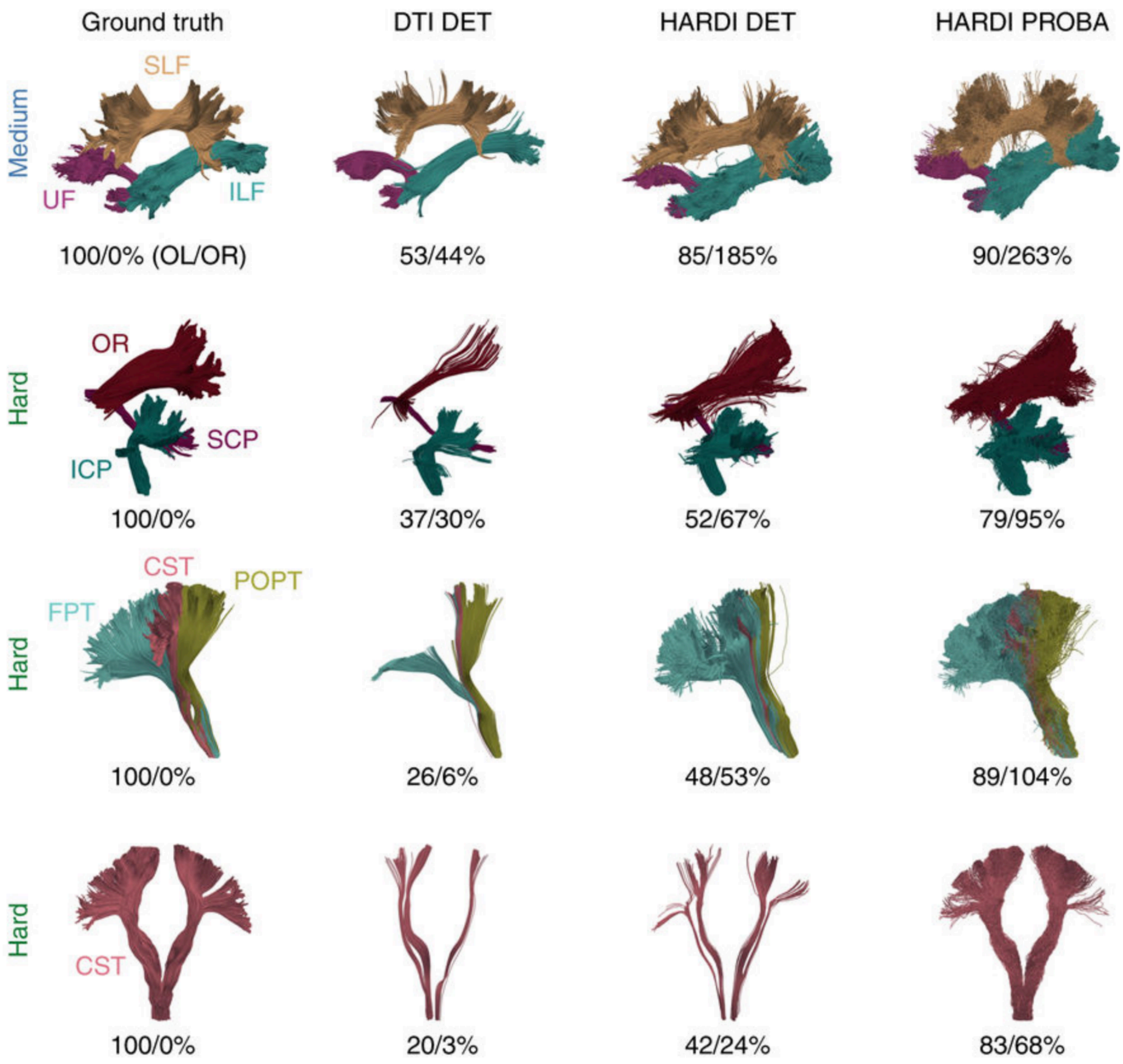
Everything matters. Kennedy et al. (2019)

Adapted from multiple sources, including (Whitaker 2016; Drummond 2009; Hong 2015; Goodman, Fanelli, and Ioannidis 2016; Peng 2011)

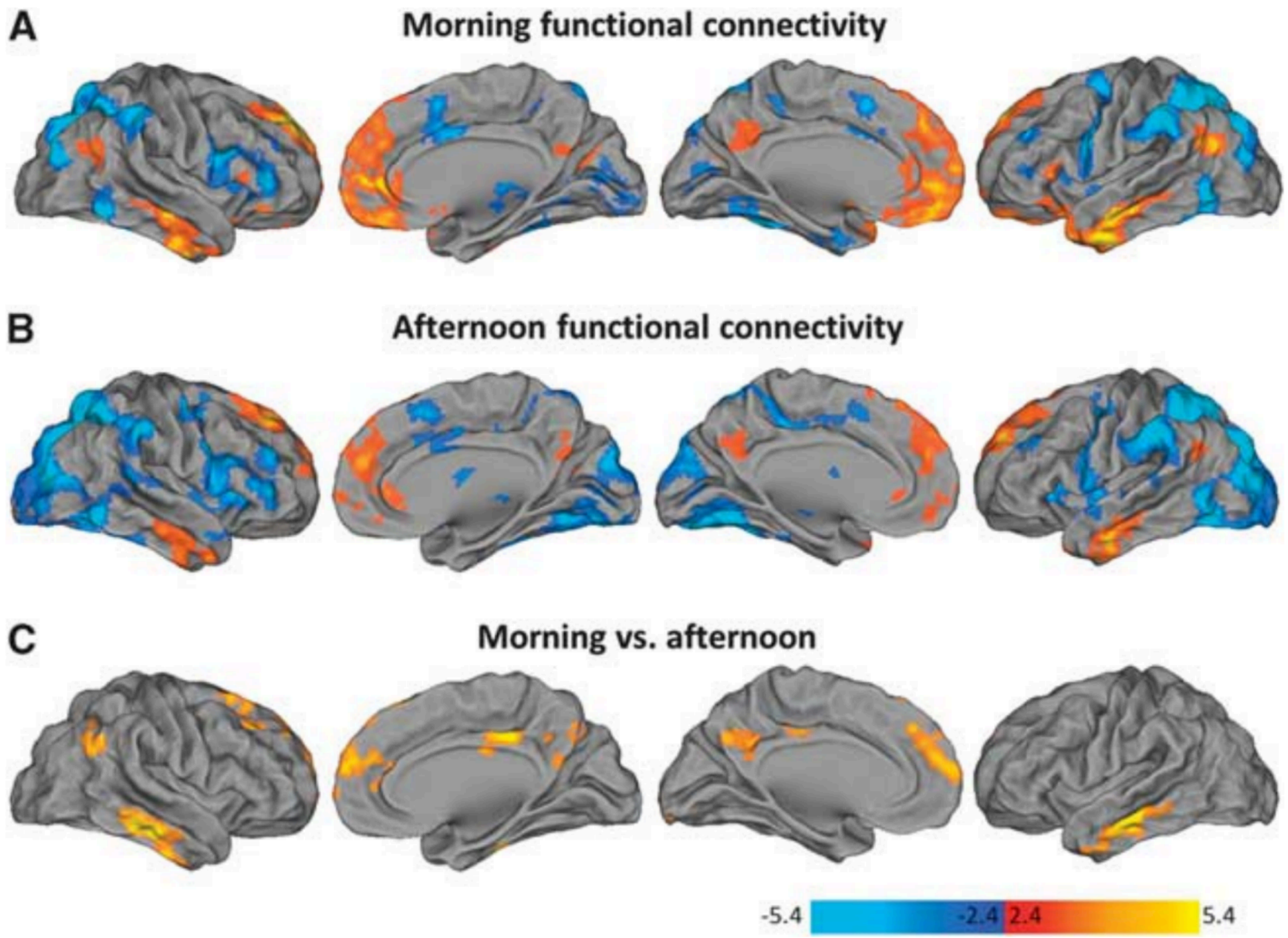
Variation in locations of reported white matter differences between people who stutter and people with fluent speech



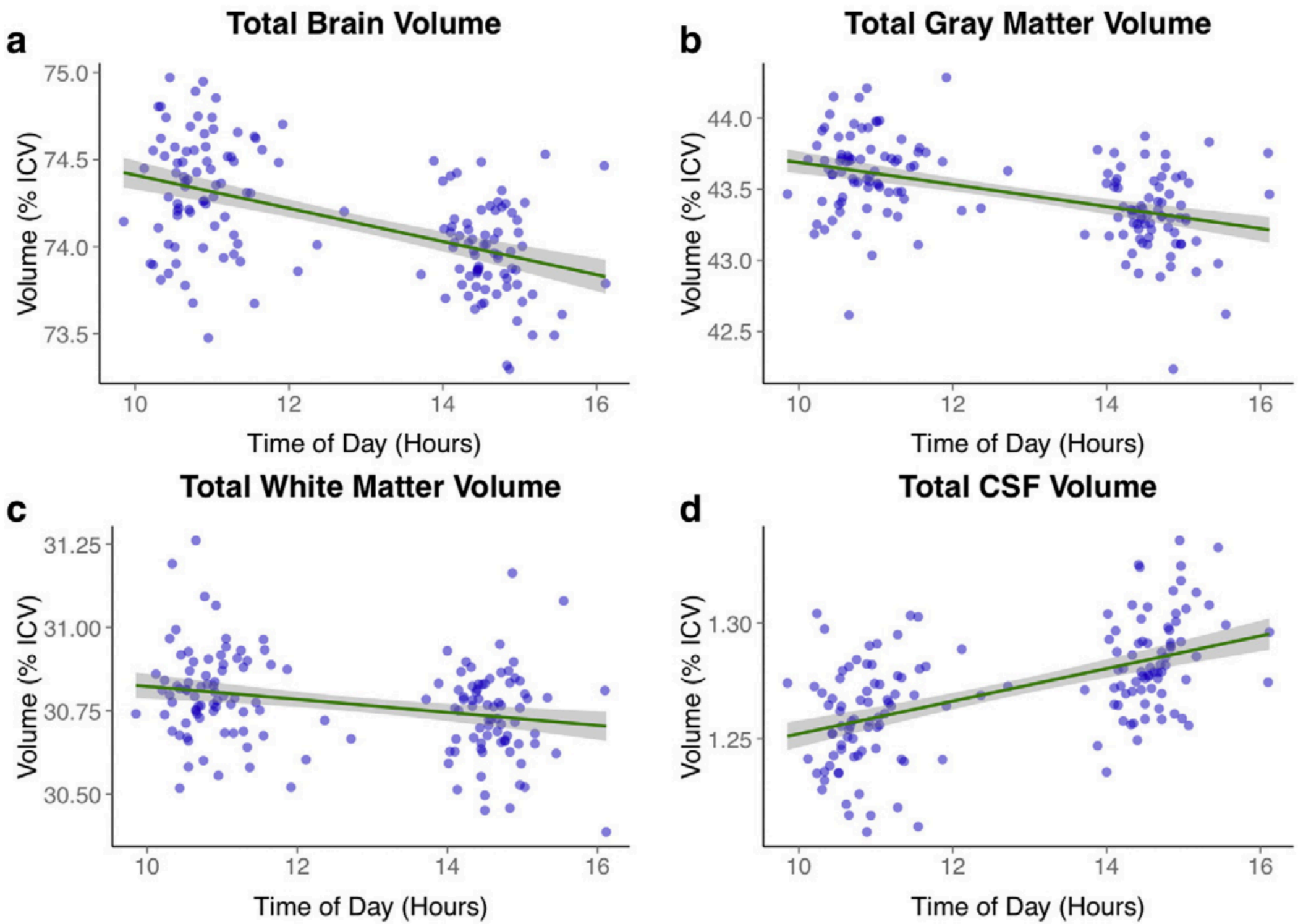
Cai et al. (2014)



Maier-Hein et al. (2017)



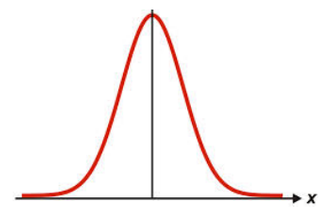
Hodkinson et al. (2014)



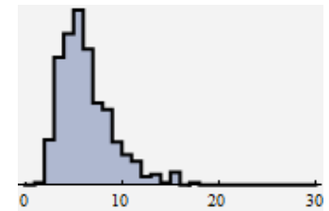
Trefler et al. (2016)

Model of the error

- Prob. distribution:

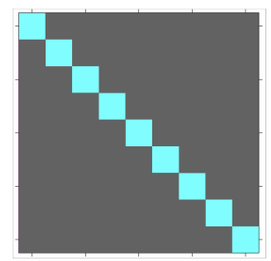


Gaussian

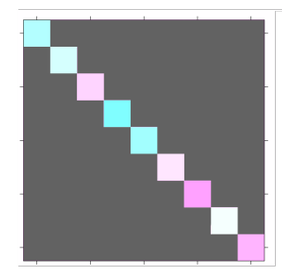


Non-Parametric

- Variance:



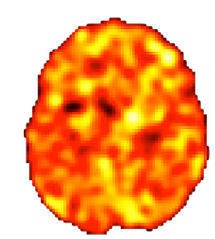
homogeneous



heterogeneous

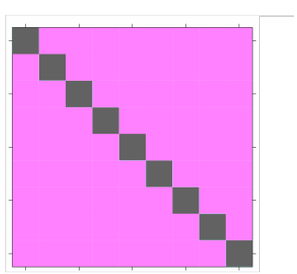
global

local

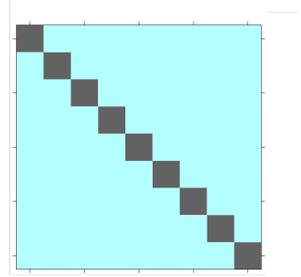


regularized

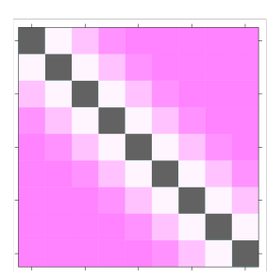
- Dependence:



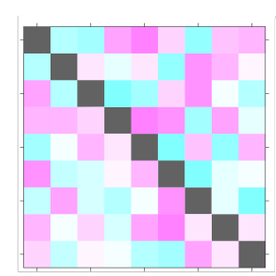
Independent noise



Compound Symmetry



Serially correlated

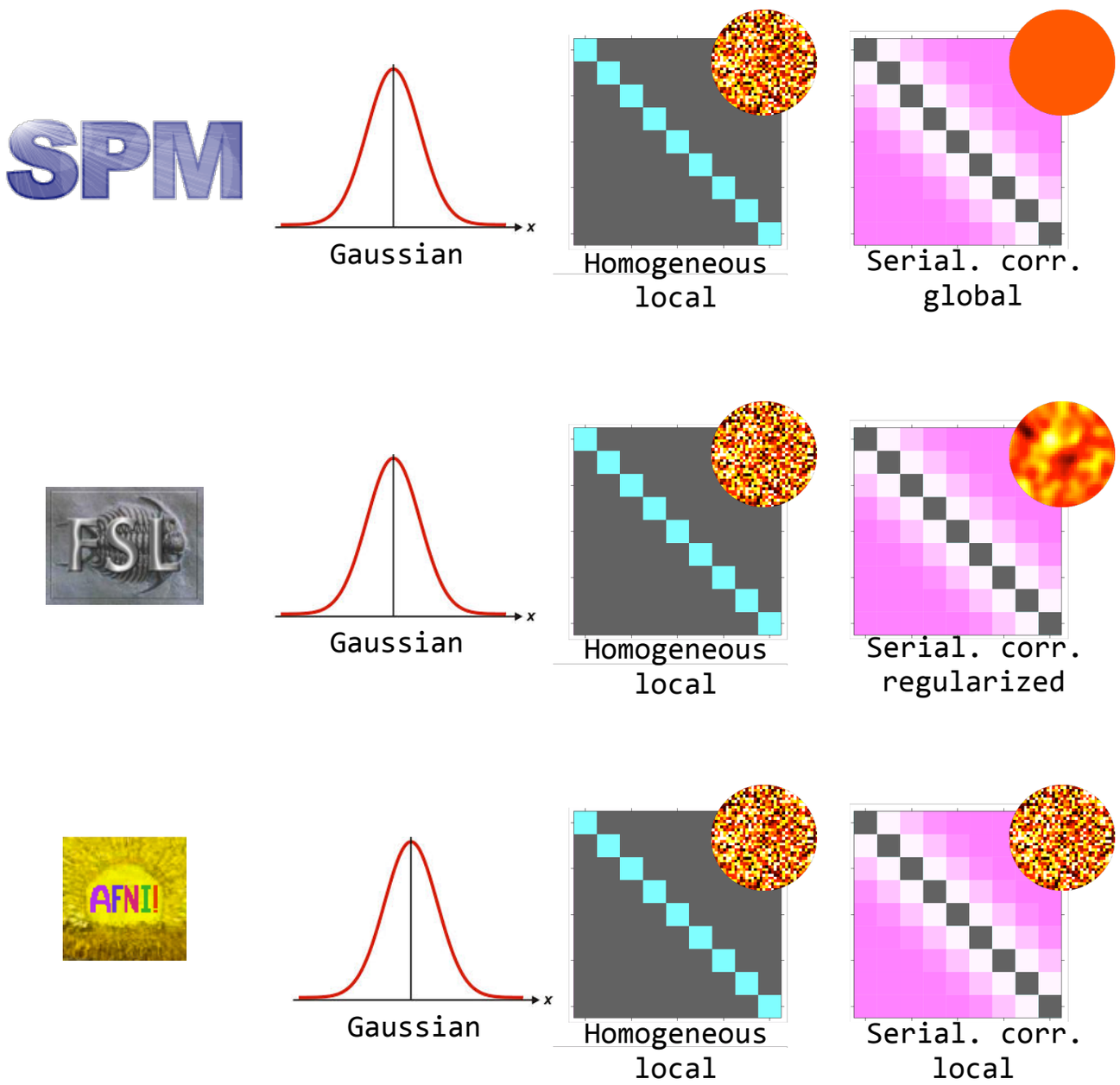


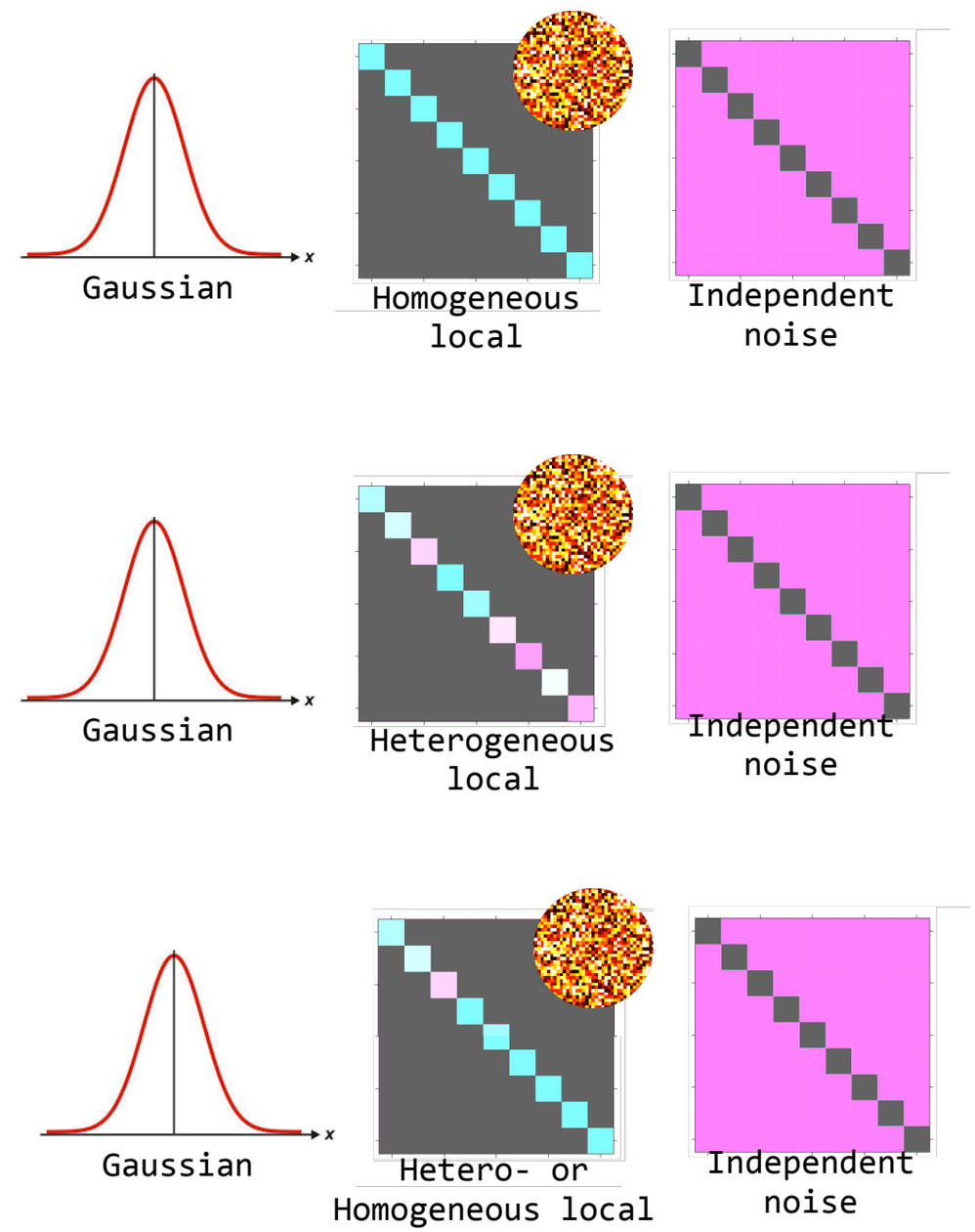
Arbitrarily correlated

slides courtesy of Camille Maumet

1st level

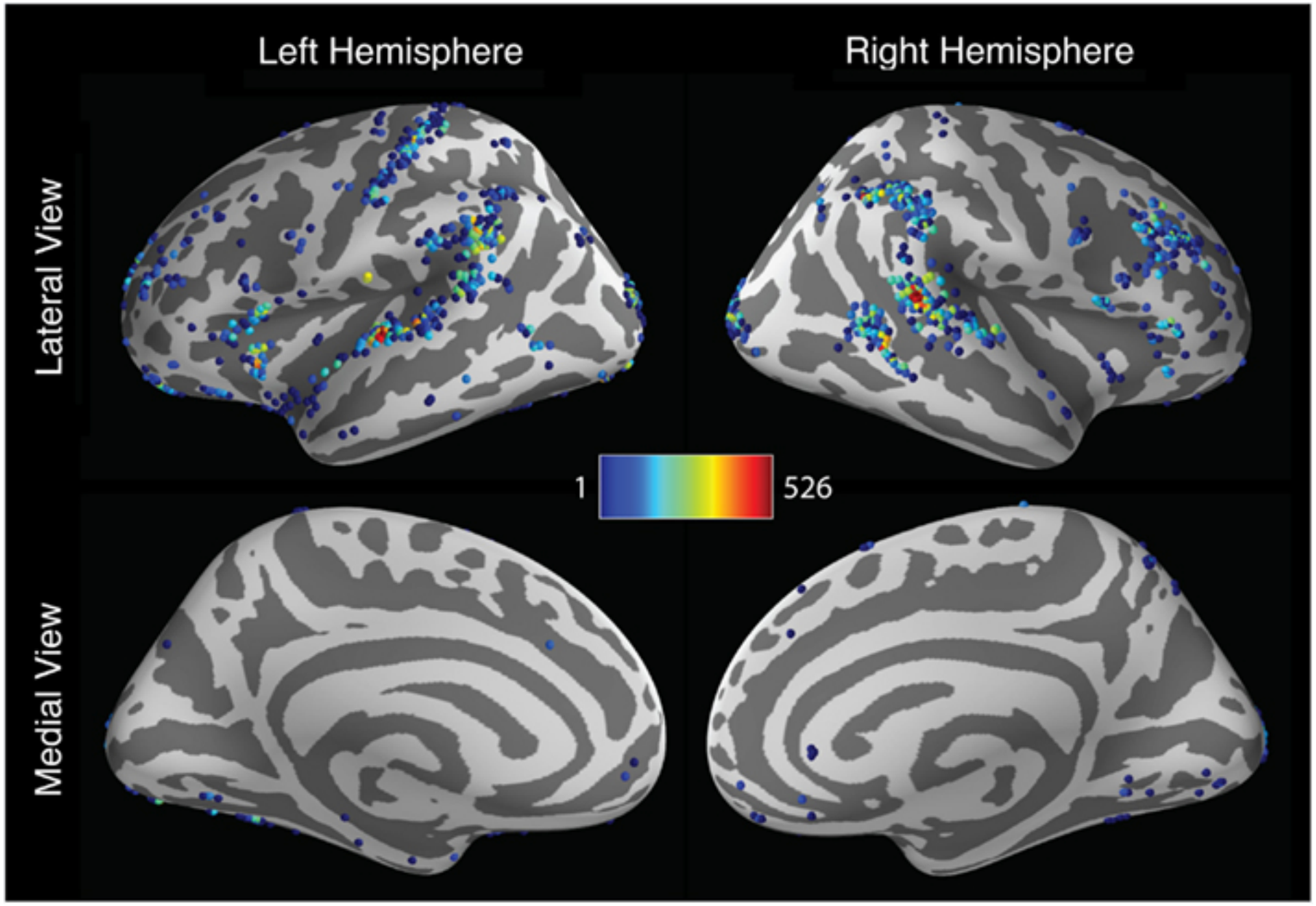
2nd level





slides courtesy of Camille Maumet

Variation in functional activity when the same task data was analyzed by different workflows



Carp (2013)

		Reference Run (Docker)		Mac OSX (10.12.4)			Difference				
Hemisphere	Region	Mean Volume (mm3)	STD	Mean Volume (mm3)	STD	Correlation	Mean Volume (mm3)	STD	Range	Mean Absolute Volume	Percent of Reference
Left	Accumbens	471.2	178.8	468.4	181.0	0.980	2.8	35.7	[-72.0, 75.6]	25.8	5.5
	Amygdala	840.2	257.9	849.2	262.9	0.963	-9.1	70.8	[-303.0, 84.0]	32.1	3.8
	Caudate	3842.4	650.4	3840.5	630.7	0.980	2.0	130.9	[-320.6, 518.4]	56.1	1.5
	Hippocampus	3276.2	794.9	3264.5	783.6	0.997	11.7	62.0	[-147.0, 185.0]	44.2	1.3
	Pallidum	1683.5	316.0	1668.8	313.6	0.995	14.7	30.7	[-16.8, 99.0]	20.2	1.2
	Putamen	4881.1	965.4	4890.2	952.7	0.998	-9.2	59.5	[-145.0, 144.0]	45.4	0.9
	Thalamus	8088.6	1240.8	8108.8	1239.0	0.998	-20.2	71.4	[-135.0, 194.0]	54.5	0.7
Right	Accumbens	389.1	147.6	402.2	148.5	0.966	-13.1	38.5	[-145.5, 32.0]	24.1	6.2
	Amygdala	882.3	297.8	897.6	290.7	0.918	-15.3	119.3	[-464.0, 221.0]	68.2	7.7
	Caudate	3781.0	762.1	3780.3	767.7	0.999	0.7	27.1	[-60.1, 48.0]	20.6	0.5
	Hippocampus	3433.4	796.9	3453.3	793.2	0.997	-19.9	60.6	[-190.0, 78.0]	42.4	1.2
	Pallidum	1694.5	293.8	1695.2	289.5	0.997	-0.7	21.8	[-64.0, 60.9]	12.9	0.8
	Putamen	4965.3	1017.3	4942.8	1008.0	0.993	22.5	118.0	[-201.6, 397.0]	78.4	1.6
	Thalamus. Proper	7723.6	1120.6	7724.3	1129.4	0.999	-0.7	52.9	[-145.0, 118.0]	33.6	0.4
Total	CSF	189856.5	26936.5	190209.3	26668.1	0.999	-352.8	1025.2	[-4274.2, 20.0]	359.7	0.2
	Gray Matter	684781.8	86048.8	684111.7	86173.4	1.000	670.1	2009.7	[-11.3, 7113.4]	676.3	0.1
	White Matter	513866.1	52348.3	513802.8	52341.8	1.000	63.3	467.3	[-541.2, 2192.4]	125.3	0.0
	Brain	1388504.4	132850.8	1388123.8	132962.5	1.000	380.6	1511.6	[-534.0, 6956.6]	425.1	0.0

Ghosh et al, 2017

Create and use open software

http://neuro.debian.net

neurodebian

scikits learn

PyMVPA

jupyter

PsychoPy

Psychology software in Python

R

RStudio

AFNI

ANTs

Advanced Normalization Tools

FreeSurfer

FSL

SPM

and many more

To install this dataset in your current directory use

```
datalad install ///
```

To install with all sub-datasets add `-r`. To get all the data add `-g`. To get data in parallel add `-JN`, where `N` would be the number of parallel downloads.

For more information about DataLad and installation instructions visit datalad.org

datasets.datalad.org

Search:

Name	Last Modified	Size	Description
./	2019-03-18 14:23:26	303.1 GB/59.0 TB	The Ultimate DataLad collection of datasets
abide/ @0.0.20161216-9-gf660781	2017-12-08 00:31:23	1.2 MB/197.7 GB	
abide2/	2017-12-08 00:27:09	-/110.6 GB	
adhd200/	2018-06-11 10:41:17	-/516.4 GB	
allen-brain-observatory/	2018-09-04 13:37:20	6.1 MB/39.3 TB	
cifar/	2019-03-18 14:23:22	1.1 GB	
corr/	2017-05-24 15:14:23	2.2 kB/937.1 GB	Consortium for Reliability and Reproducibility (CoRR)
crcns/	2019-03-27 15:39:58	18.8 kB/12.0 TB	Primarily electrophysiology data from CRCNS initiative
dbic/	2017-05-09 21:57:19	39.5 GB	Dartmouth Brain Imaging Center (DBIC) datasets
devel/	2017-05-25 11:27:02	466.1 MB	Various datasets used for development
dicoms/	2018-12-20 13:39:41	1.5 GB/6.2 GB	Various collections of DICOM files which could be used for software and methods testing

http://datasets.datalad.org

Brain Imaging Data Structure (or BIDS)

1200 Subject Referenc

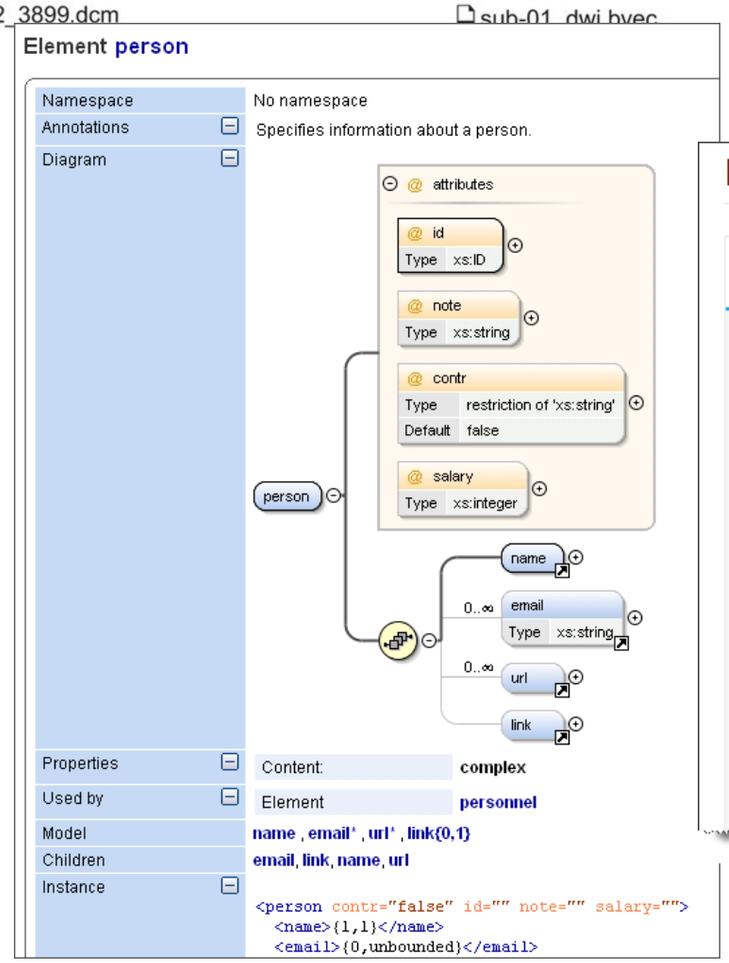
PDF Download (3 MB)

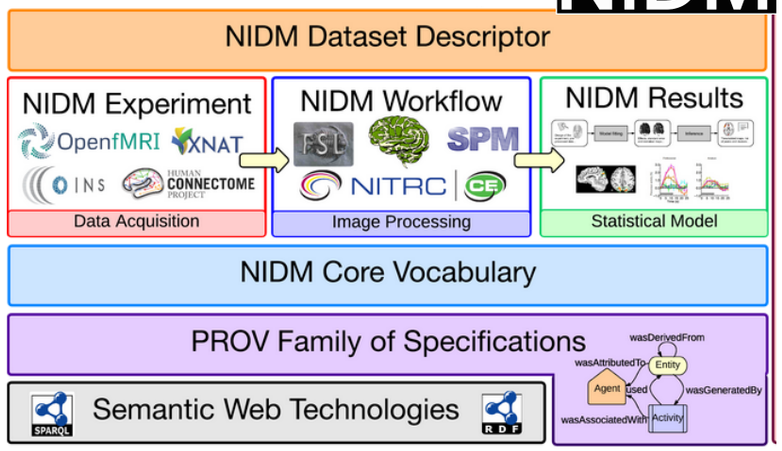
Appendices:

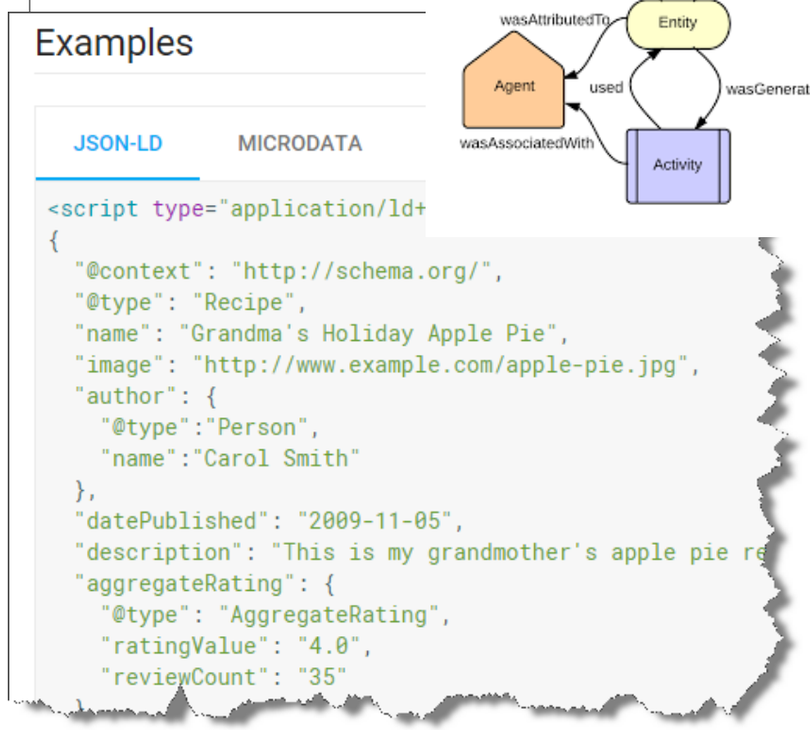
- Appendix 1: HCP scan pro (PDF)
- Appendix 2: Matlab code for voxel-wise correction of dMRI gradients (PDF)
- Appendix 3: File names and directory structure for 1200 subjects data (PDF)
- Appendix 4: HCP protocol standard operating procedures (SOPs) (PDF)
- Appendix 5: MR Data acquisition information for an exemplar subject(PDF)
- Appendix 6: Task fMRI and tMEG E-Prime key variables (PDF)
- Appendix 7: tMEG: Extracting trial subjects from data produced by tmegproc pipeline (PDF)

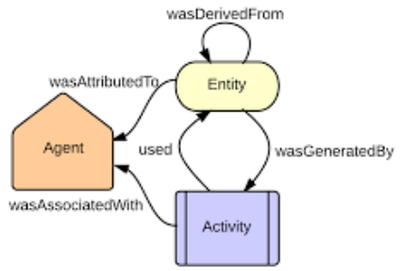
```
dicomdir/
  1208200617178_22/
    1208200617178_22_8973.dcm
    1208200617178_22_8943.dcm
    1208200617178_22_2973.dcm
    1208200617178_22_8923.dcm
    1208200617178_22_4473.dcm
    1208200617178_22_8783.dcm
    1208200617178_22_7328.dcm
    1208200617178_22_9264.dcm
    1208200617178_22_9967.dcm
    1208200617178_22_3894.dcm
    1208200617178_22_3899.dcm
  1208200617178_23/
  1208200617178_24/
  1208200617178_25/
```

```
my_dataset/
  participants.tsv
  sub-01/
    anat/
      sub-01_T1w.nii.gz
    func/
      sub-01_task-rest_bold.nii.gz
      sub-01_task-rest_bold.json
    dwi/
      sub-01_dwi.nii.gz
      sub-01_dwi.json
      sub-01_dwi.bval
      sub-01_dwi.bvec
```







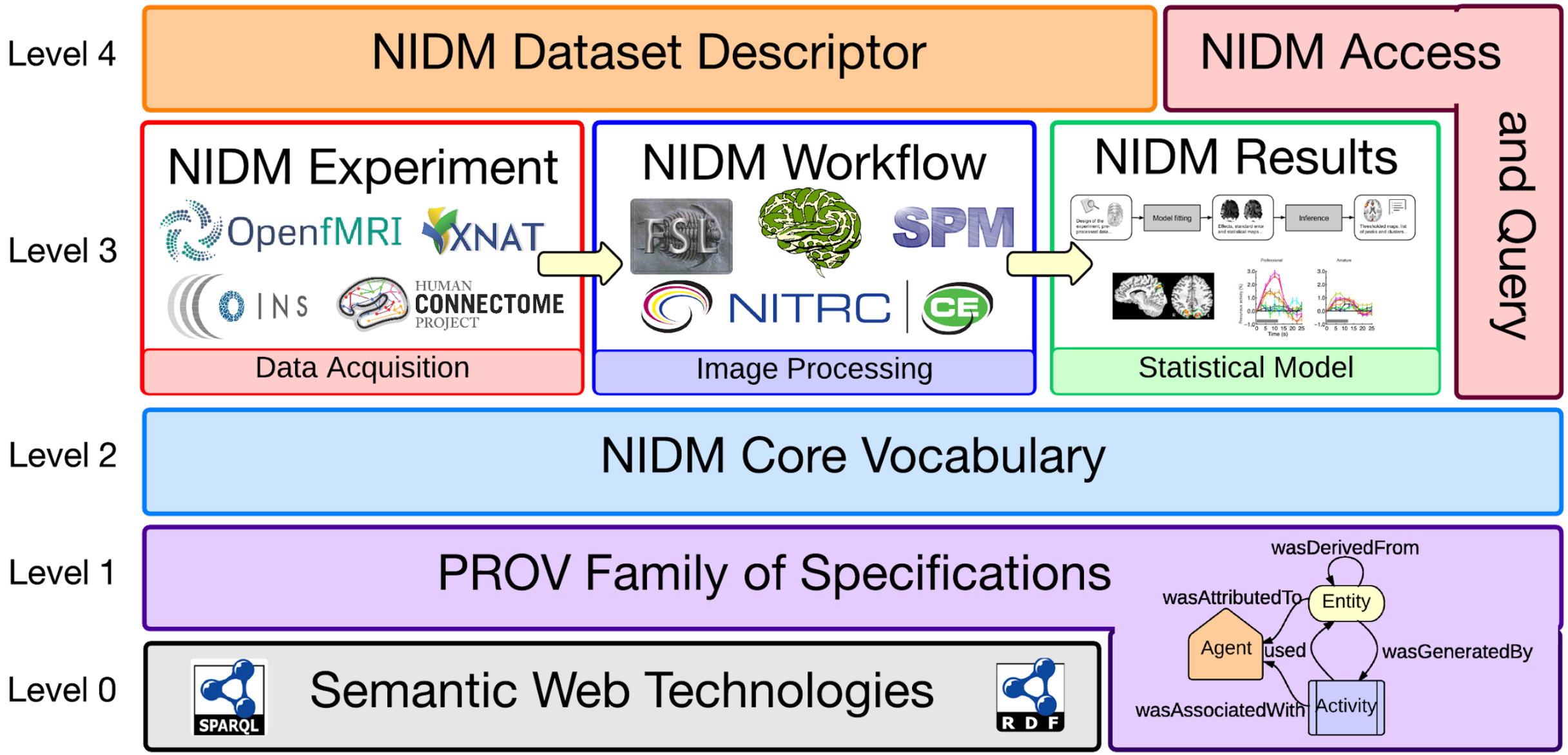


Unstructured

Structured

Structured
(with semantics)

Neuroimaging Data Model



http://nidm.nidash.org

Vague but exciting ...

CERN DD/OC

Information Management: A Proposal

Tim Berners-Lee, CERN/DD

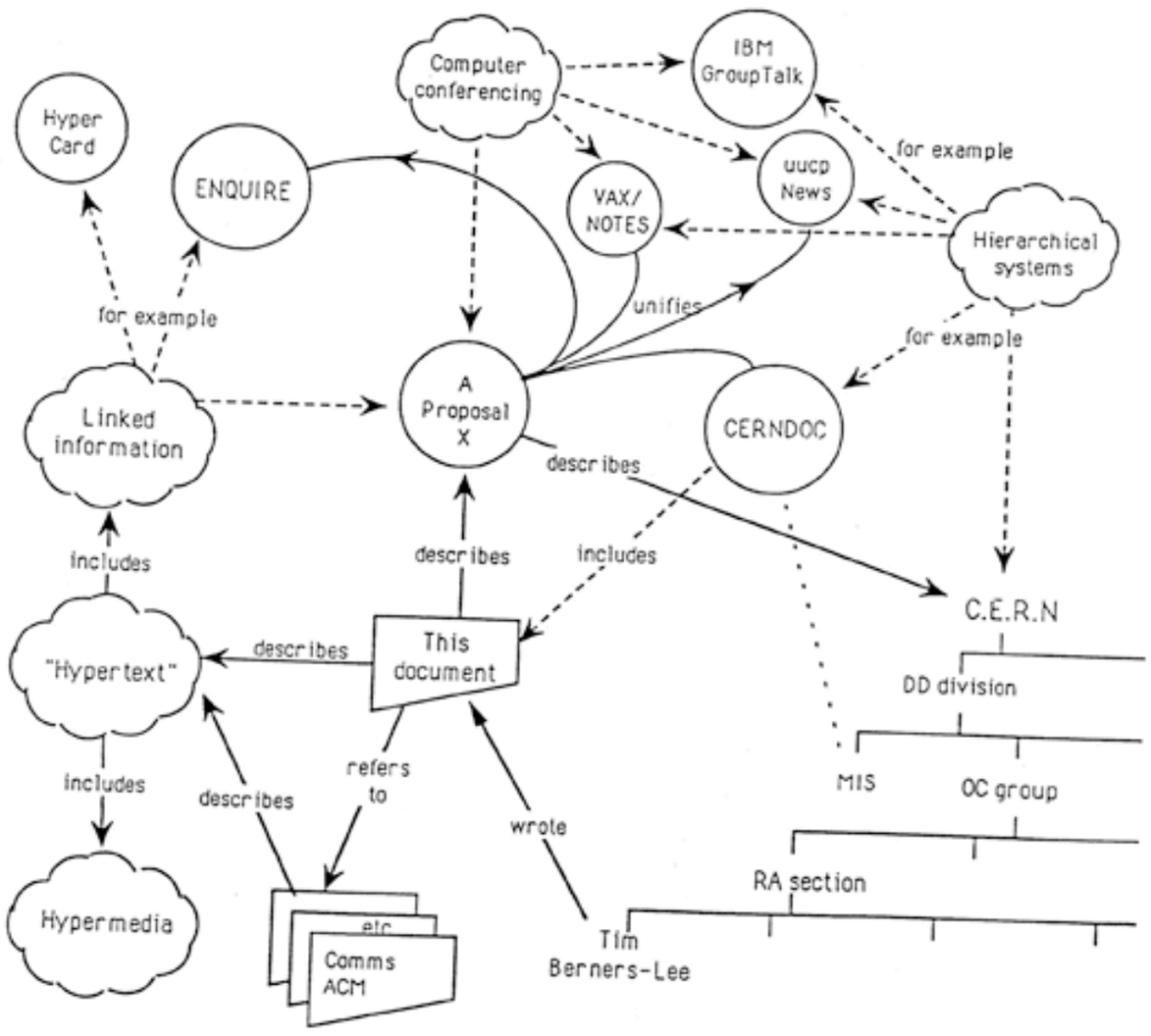
March 1989

Information Management: A Proposal

Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext sytstem.

Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control



http://info.cern.ch/Proposal.html

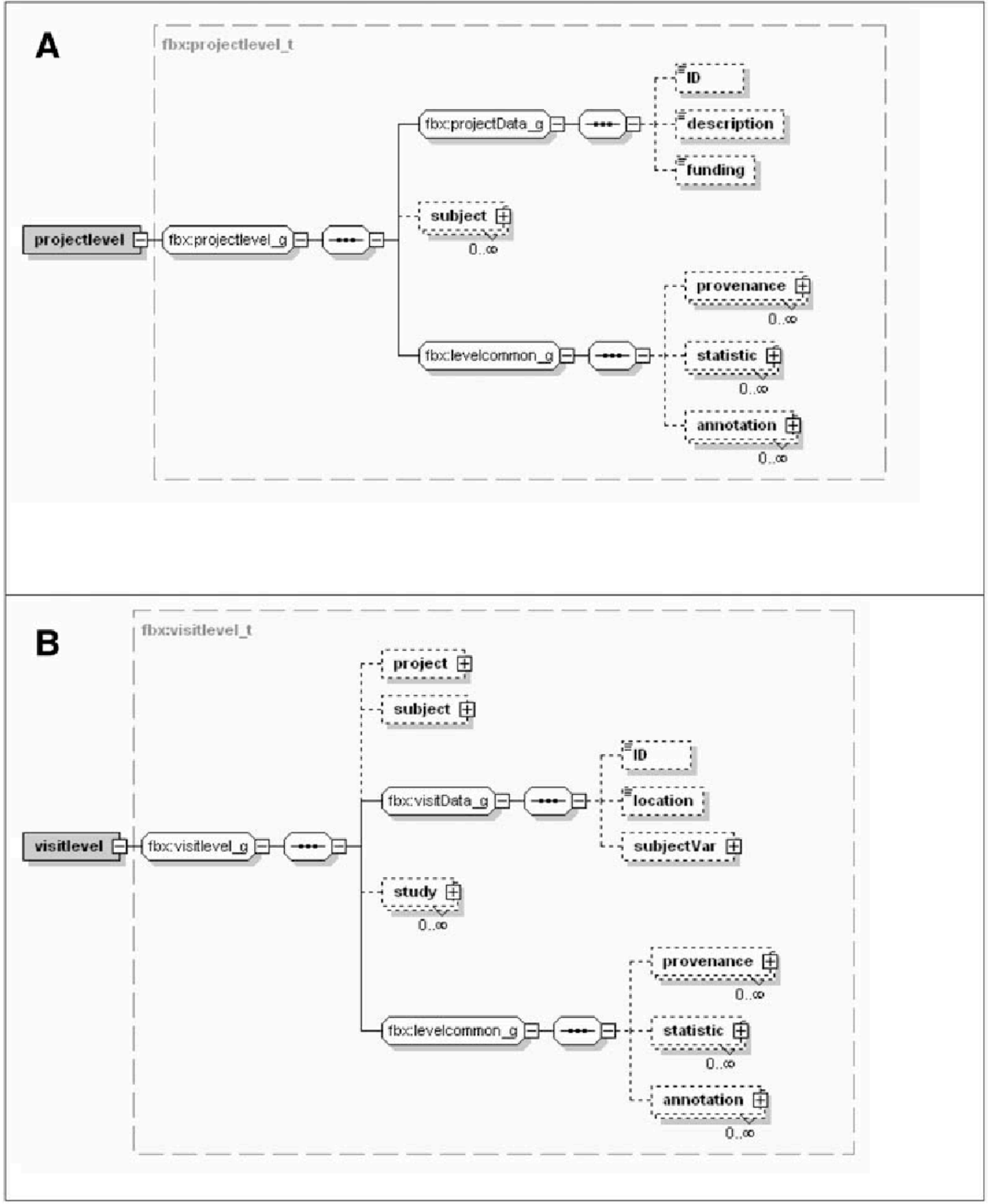


Fig. 1. *(Continued)*

Keator et al, 2006

Activities

English

- Screening
- Demographics
- General anxiety
- Behavioral
- PROMIS Measure
- Voice Tasks

ReproNim

MindLogger

LORIS

Home

For each task, you should hit the record button before speaking and then stop once you are done speaking. You may hit play to hear what was recorded.

Say 'ahhh' for 10 seconds

record

Don't Know Skip

Say 'pataka' as fast as you can for a minute

record

Don't Know Skip

Look at the map in the image below, and give directions in 1 min to another person over the phone as to how you would go from Start to End.

Stanford Center For
Reproducible Neuroscience

The Growing Ecosystem

General Purpose

Pydra: File I/O Workflow engine

TestKraken: Testing framework

Neurodocker: Analysis environments

Nobrainer: Brain DNN library

Specialized

Heudiconv, MRIQC, fMRIPrep, Niflows

Modularity and Robustness!

Neuroscout

Browse Help Sign in Sign up

Neuroscout

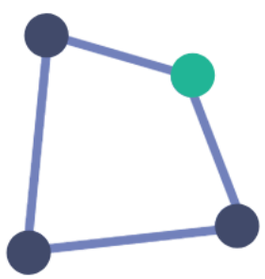
A platform for fast and flexible re-analysis of (naturalistic) fMRI studies

Browse public analyses

Re-use public data

Select from openly available naturalistic fMRI datasets, from sources such as OpenNeuro and DataLad.

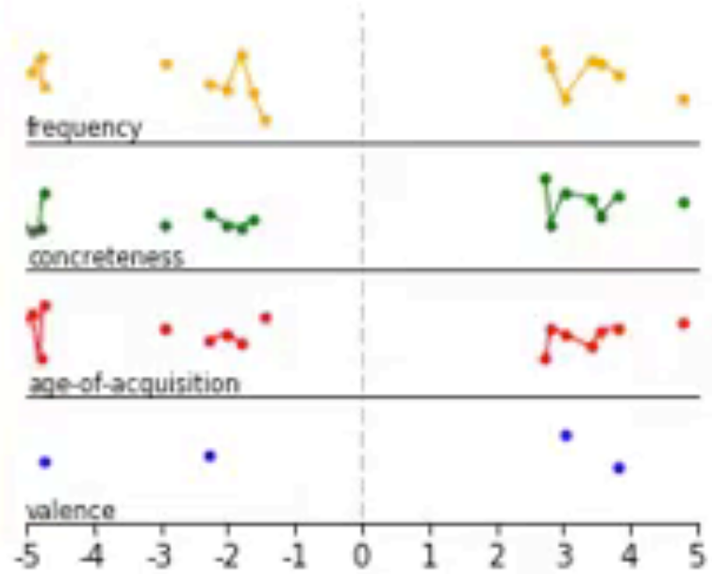
Design your analysis



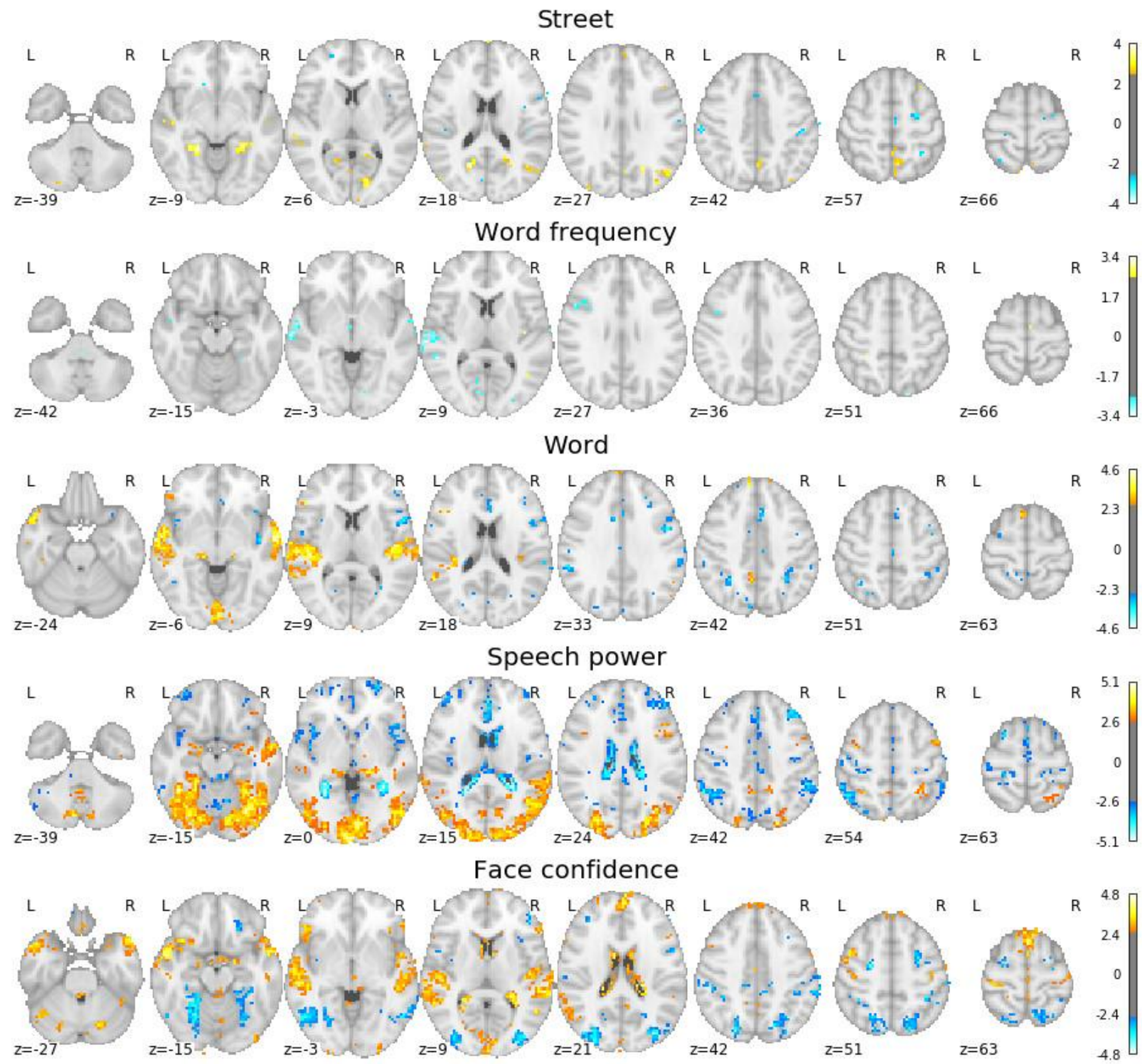
Browse hundreds of annotations automatically extracted from stimuli using **state-of-the-art machine learning** algorithms, such as Google Cloud Vision, IBM Watson, and more.

Execute and share

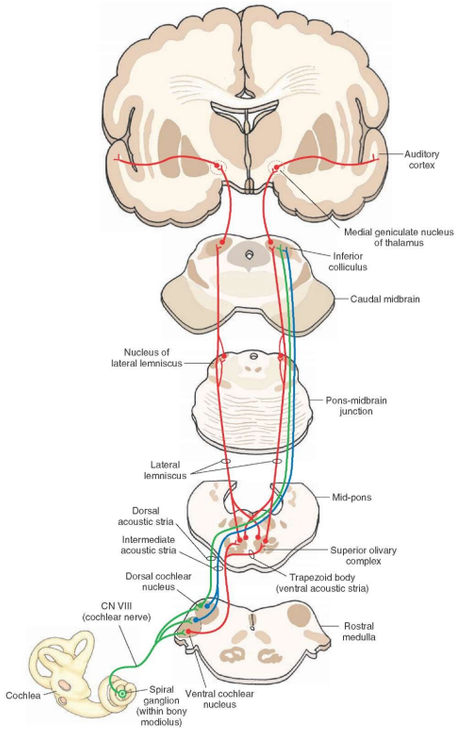
Portable BIDS pipelines enable execution with no configuration. Results are automatically uploaded to NeuroVault for easy sharing.

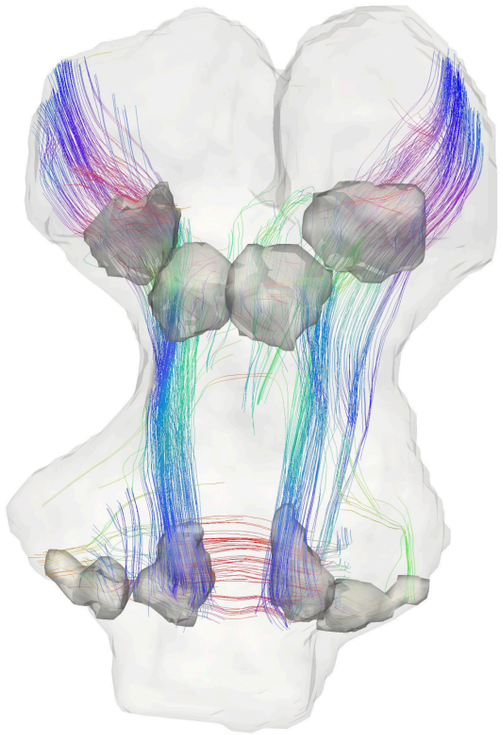


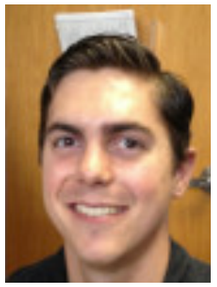
Neuroscout: Tal Yarkoni, Alejandro de la Vega, Michael Hanke

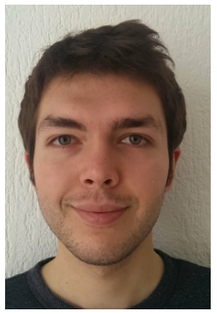


slides: Alejandro de la Vega

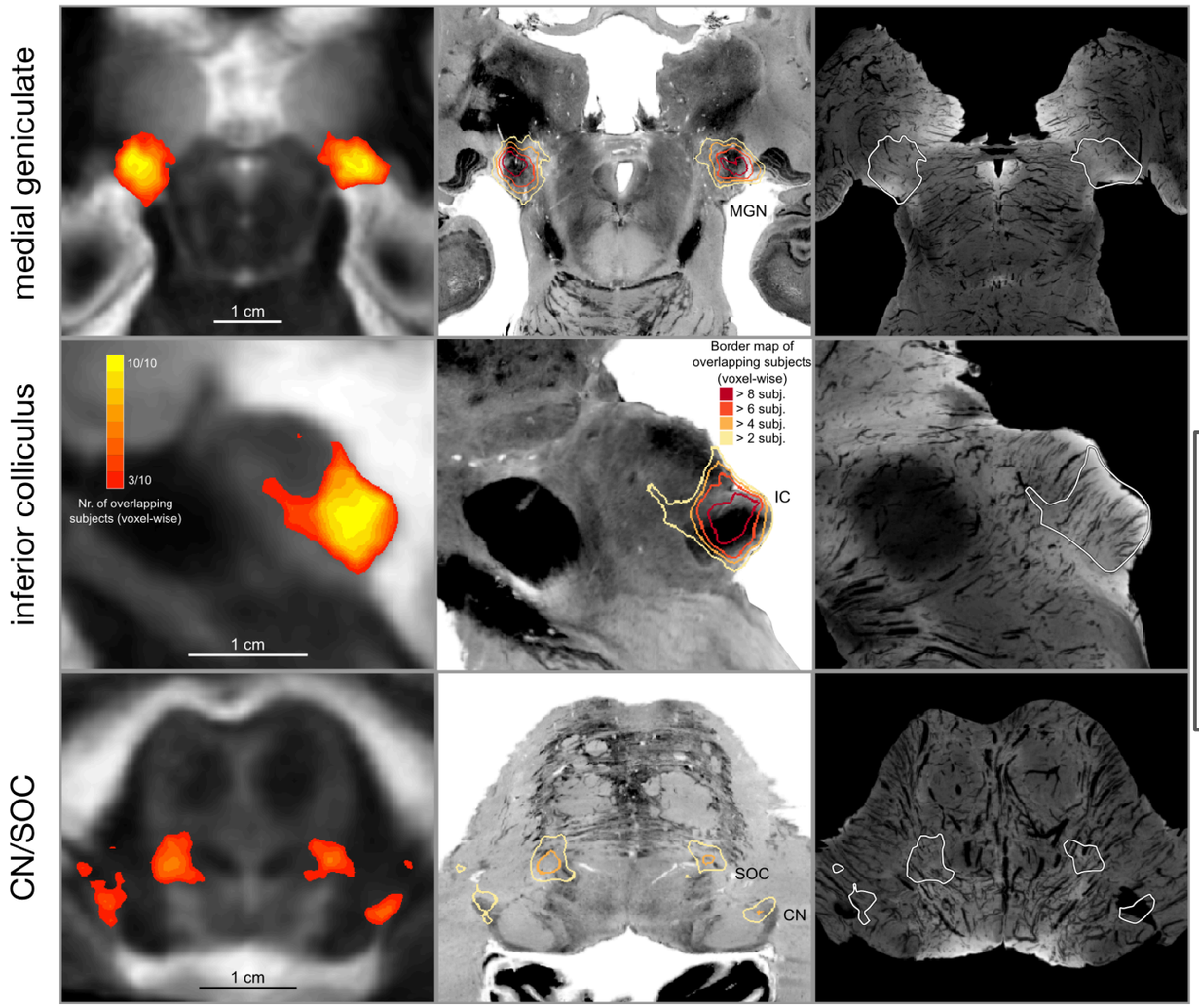


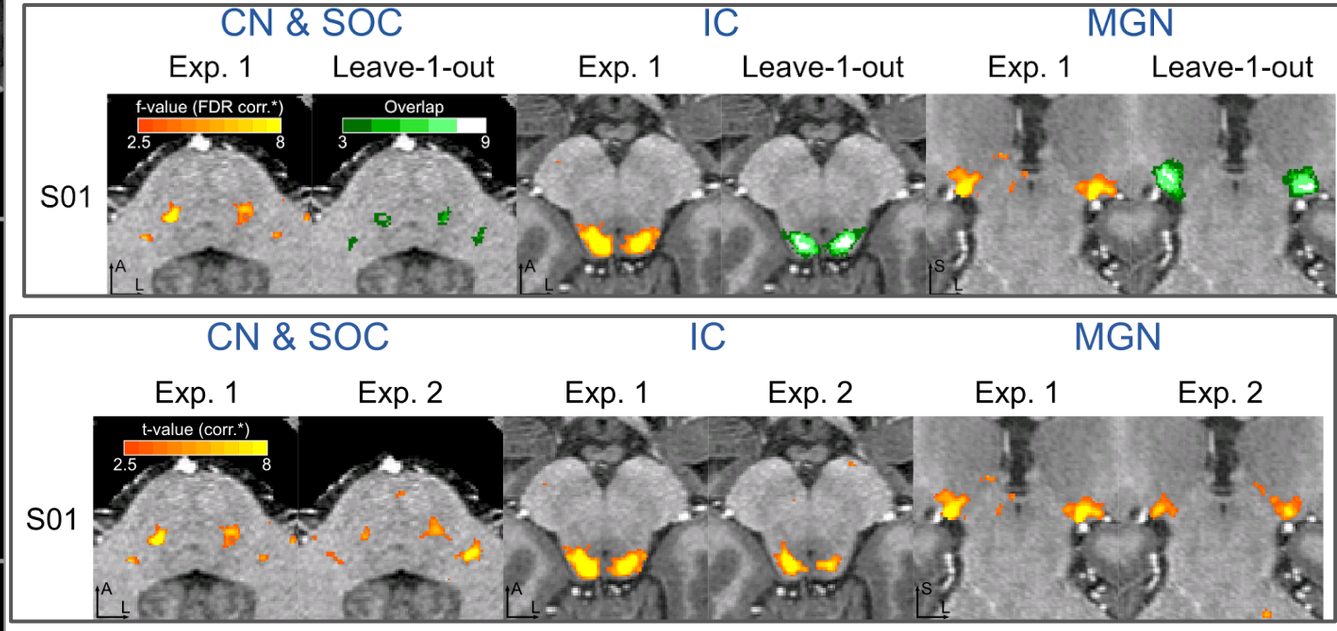






In vivo MRI BigBrain Post mortem MRI





Sitek, Gulban et al. (In press)

Publicly available data and resources	
In vivo data	openneuro.org/datasets/ds001942
Atlases & derivatives	osf.io/c4m82/
Analysis code	github.com/sitek/subcortical-auditory-atlas
preprint (bioarxiv)	tinyurl.com/sitek-gulban-preprint
poster	doi.org/10.6084/m9.figshare.8226053

wikipedia

National Institutes of Health

NIDASH

ReproNim

McGOVERN INSTITUTE FOR BRAIN RESEARCH AT MIT

MIT Lincoln Laboratory

McGovern Institute NeuroTechnology Fund - MIT LL

Questions?

Data courtesy of: Caroline Mangin, MGH