Using high throughput computing to investigate the role of neural oscillations in visual working memory

Jacqueline M. Fulvio, PhD

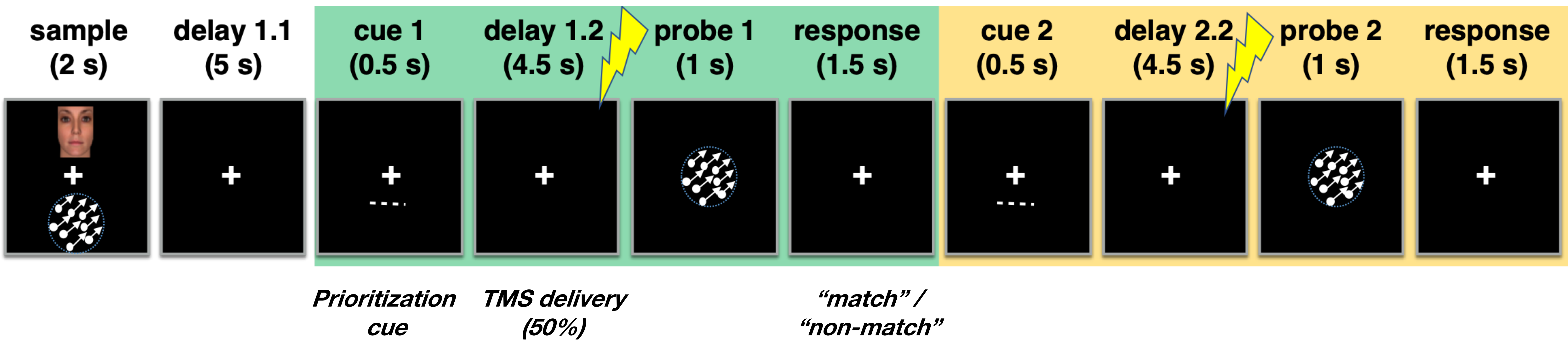
Postle Laboratory
Departments of Psychology & Psychiatry

neurogenbb.com

Visual working memory

- allows us to temporarily maintain and manipulate visual information in order to solve a task

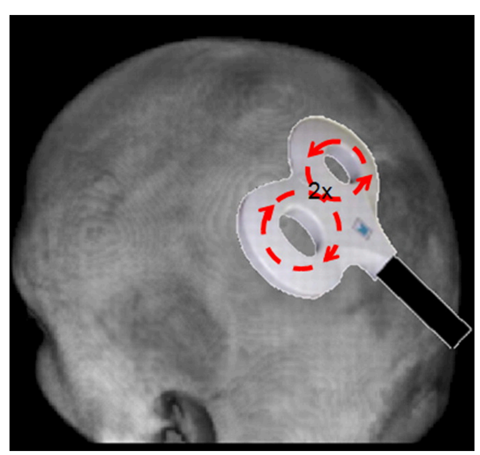
Double serial retrocuing (DSR) task

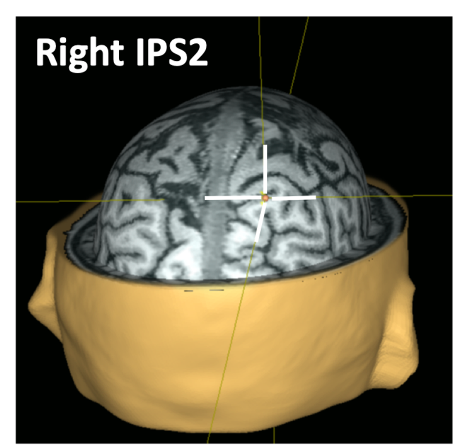


electroencephalogram (EEG)

brain vision solutions

transcranial magnetic stimulation (TMS)





Nexstim NBS

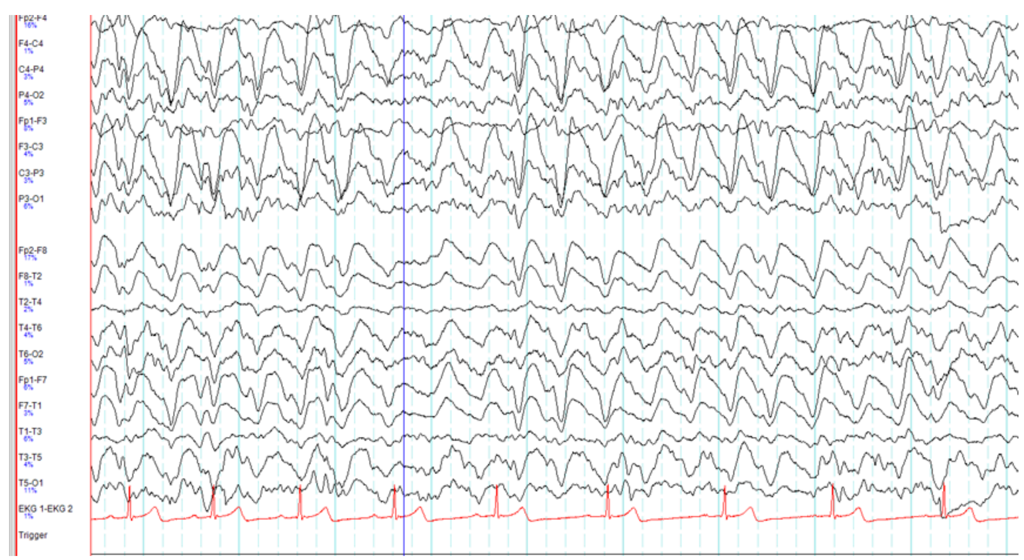
Rose et al., 2016; Fulvio & Postle, 2020

Visual working memory

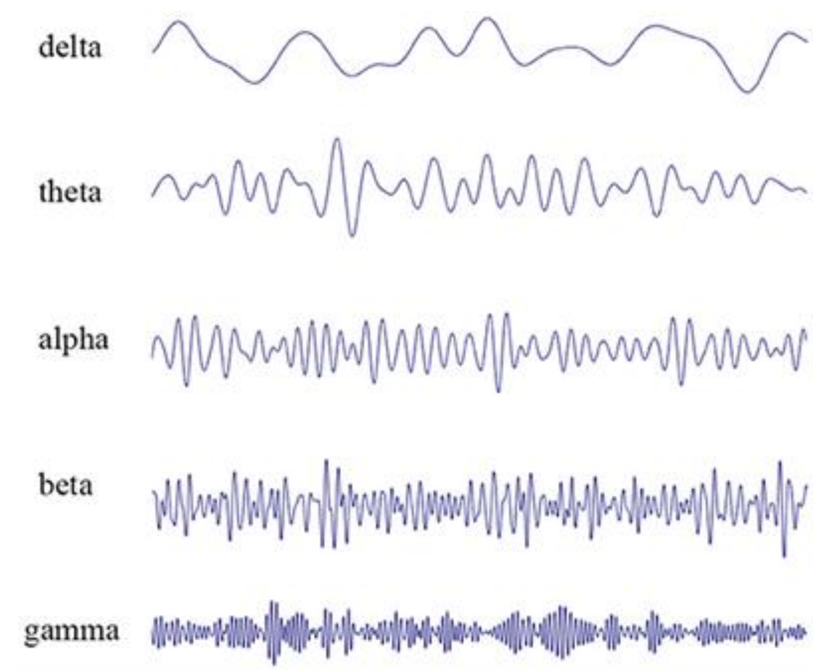
- the frequency of neural oscillations has been associated with distinct working memory processes

brain vision solutions

raw signal: 60 channels



neural oscillations



"SPACE-FSP" multi-way decomposition of high-dimensional EEG data

Spatially distributed **P**h**A**se **C**oupling **E**xtraction with a **F**requency-**S**pecific **P**hases model*

Motivation: task-related frequency oscillations are ubiquitous in EEG, including during the DSR task; however, standard analyses do not adjudicate between the oscillations arising from the frequency modulation of a single oscillating network, or multiple oscillating networks at different frequencies

Goal of the analysis: extract phase-coupled oscillatory networks at frequencies of interest (here, 4-40 Hz)

**van der Meij, et al. (2015; 2016)*

Why use HTCondor?

SPACE is a computationally-demanding analysis!

- The number of oscillatory networks extracted cannot be determined analytically and therefore must be ***estimated*** through decomposition.
    - This is done ***iteratively*** by starting at a set number of networks and increasing the number incrementally until a preset statistical criterion is no longer achieved.
    - A single decomposition can take days to weeks to months depending upon the number of epochs (i.e., chunks) of data fed to the algorithm, the number of networks ultimately extracted, and the hardware.

Our planned analyses required decomposition of at least 186 data sets!

- HTC offered the opportunity to run the decompositions in parallel on better hardware than our lab server was built upon…*
    - …with MATLAB parallel pool compatibility

How we got started with HTC…

- Lab (P.I. Dr. Bradley Postle) history of using CHTC resources

- Initial consultation with staff to discuss planning and basic details about the system and helpful resources on the website

- Most useful/helpful resource has been the online help guide (https://chtc.cs.wisc.edu/uw-research-computing/guides.html) with the sample code and detailed explanations

- Also very helpful: CHTC office hours
  - visits were necessary for assistance with code compiling and parameter setting to take advantage of Matlab-based functionality

Submitting our first jobs…

- Compiled a Matlab standalone executable including relevant code/toolboxes

- Submitted jobs with the Matlab executable and the particular data file (e.g., from a particular subject+condition+pipeline) to be decomposed

- Optimization: Took advantage of Matlab's distributed computing as part of the Parallel Computing Toolbox

  - Requested the maximum number of "workers" in a parallel pool supported by the cluster (12) with equal number of CPUs requested

- Used "LongJob" flag

Input data file size	Requested memory per job	Requested disk space per job	Total number of jobs completed	Typical job time
39 MB - 59 MB (Rose et al., 2016; two analysis "pipelines")	20-40 GB	5 - 10 GB	42	2 days - 2 weeks (compared to ~1 month on our lab server)

Updating the analyses….

- The initial results were promising, but the two data analysis pipelines we tried were insufficient to address some of our key questions
  - dataset did not contain “built-in” control conditions for more rigorous statistical analysis
- Re-ran the analyses using a newer data set that I had collected (Fulvio & Postle, 2020) that overcame some of the limitations

Input data file size	Requested memory per job	Requested disk space per job	Total number of jobs completed	Typical job time
39 MB - 59 MB (Rose et al., 2016; two analysis “pipelines”)	20 - 40 GB	5 - 10 GB	42	2 days - 2 weeks (compared to ~1 month on our lab server)
25 MB - 60 MB (Fulvio & Postle, 2020; same two analysis pipelines)	20 - 40 GB	5 - 10 GB	72	2 days - 2 weeks

Updating the analyses…again…

- Some tweaks to the data structure were still necessary
  - needed to extend data epochs from 500 ms to 1s
  - needed to include an additional fixation period (i.e., baseline) epoch

Input data file size	Requested memory per job	Requested disk space per job	Total number of jobs completed	Typical job time
39 MB - 59 MB (Rose et al., 2016; two analysis "pipelines")	20 - 40 GB	5 - 10 GB	42	2 days - 2 weeks (compared to ~1 month on our lab server)
25 MB - 60 MB (Fulvio & Postle, 2020; same two analysis pipelines)	20 - 40 GB	5 - 10 GB	72	2 days - 2 weeks
200 MB - 500 MB (Fulvio & Postle, 2020; new combined pipeline)	50 - 75 GB	10 - 15 GB	72	few weeks - few months*

Summary of computing experience to date

From our final analysis, we obtained 1,690 components!

Positive points

- Helpful “hands-on” support from staff along with the ability to run independent jobs to test various aspects of the pipeline

- Years of computing on lab machines (projected) were condensed to months (for each iteration of the analysis)

- Since have been able to generalize experience to carry out a different, more “HTC-friendly” analysis as a scientific control to back up claims about results from initial analyses

Pain points

- Still took longer than necessary because:

  - high throughput framework is not ideal for optimization problems like this: with the iterative nature, the problem/code cannot be further optimized (and jobs cannot be started where they left off if interrupted).

  - our jobs seemed to often be sent to slower machines or busy machines that would kick the job part-way through, which added many hours to the total time we used HTC.

Bigger picture

- From the broad perspective of our research-group, OSPool resources have significantly expanded our computational capabilities
    - applying computationally-demanding, but better-suited, analysis methods to address questions not well-answered by more commonly used approaches.
- From a more focused research-perspective, OSPool resources have provided the ability to adjudicate between different possible sources of common EEG findings
    - specifically, the source of changes in EEG in response to memory cues and TMS pulses appears to be modulations of existing oscillations
- From a more personal perspective, OSPool resources have improved my skills and resume
    - early results were presented at the Cognitive Neuroscience Society 2020 (virtual) conference
    - latest results will be presented at the Society for Neuroscience 2022 conference

Thank you!

jacqueline.fulvio@wisc.edu

https://postlab.psych.wisc.edu/people/west/jacqueline-fulvio1/

neurogenbb.com