

University of Memphis

University of Memphis Digital Commons

---

Electronic Theses and Dissertations

---

2020

## MULTIVARIATE MODELING OF COGNITIVE PERFORMANCE AND CATEGORICAL PERCEPTION FROM NEUROIMAGING DATA

Rakib Al-Fahad

Follow this and additional works at: <https://digitalcommons.memphis.edu/etd>

---

### Recommended Citation

Al-Fahad, Rakib, "MULTIVARIATE MODELING OF COGNITIVE PERFORMANCE AND CATEGORICAL PERCEPTION FROM NEUROIMAGING DATA" (2020). *Electronic Theses and Dissertations*. 2406.  
<https://digitalcommons.memphis.edu/etd/2406>

This Dissertation is brought to you for free and open access by University of Memphis Digital Commons. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of University of Memphis Digital Commons. For more information, please contact [khggerty@memphis.edu](mailto:khggerty@memphis.edu).

MULTIVARIATE MODELING OF COGNITIVE PERFORMANCE AND  
CATEGORICAL PERCEPTION FROM NEUROIMAGING DATA

by

Rakib Al-Fahad

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Engineering

The University of Memphis

May 2020

Copyright© Rakib Al-Fahad  
All rights reserved

*To Mom and Dad for your true, unconditional and  
never-ending love and support.*



## **Acknowledgments**

First, I am extremely grateful to my adviser, Dr. Mohammed Yeasin, for his continuous support and invaluable mentoring. I would like to express my deepest appreciation to my committee members Dr. Gavin M. Bidelman, Dr. Madhusudhanan Balasubramanian, and Dr. Wilburn E. Reddick, for their valuable suggestion and directions. The completion of my dissertation would not have been possible without the support and encouragement of my parents and family members.

Special thanks to funding organizations specifically Electrical and Computer Engineering Department, The University of Memphis; Departments of Diagnostic Imaging and Psychology, St. Jude Children's Research Hospital as well as in part by R01 CA090246 (PI: Reddick) from the National Cancer Institute; National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under award number NIH/NIDCD R01DC016267 (G.M.B.).

I thank my lab mates at CVPIA for being such good friends and helping me whenever I needed it specially Mahbubur Rahman, Iftekhar Anam, Shahinur Alam, Faruk Ahmed, Bahareh Elahian, Pouya Bashivan, Mohammed Adel Bany Muhammad, MD Sultan Mahmud, Kazi Ashraf Moinuddin.

## Abstract

State-of-the-art cognitive-neuroscience mainly uses hypothesis-driven statistical testing to characterize and model neural disorders and diseases. While such techniques have proven to be powerful in understanding diseases and disorders, they are inadequate in explaining causal relationships as well as individuality and variations. In this study, we proposed multivariate data-driven approaches for predictive modeling of cognitive events and disorders. We developed network descriptions of both structural and functional connectivities that are critical in multivariate modeling of cognitive performance (i.e., fluency, attention, and working memory) and categorical perceptions (i.e., emotion, speech perception). We also performed dynamic network analysis on brain connectivity measures to determine the role of different functional areas in relation to categorical perceptions and cognitive events.

Our empirical studies of structural connectivity were performed using Diffusion Tensor Imaging (DTI). The main objective was to discover the role of structural connectivity in selecting clinically interpretable features that are consistent over a large range of model parameters in classifying cognitive performances in relation to Acute Lymphoblastic Leukemia (ALL). The proposed approach substantially improved accuracy (13% - 26%) over existing models and also selected a relevant, small subset of features that were verified by domain experts. In summary, the proposed approach produced interpretable models with better generalization.

Functional connectivity is related to similar patterns of activation in different brain regions regardless of the apparent physical connectedness of the regions. The proposed data-driven approach to the source localized electroencephalogram (EEG) data includes an array of tools such as graph mining, feature selection, and multivariate analysis to determine the functional connectivity in categorical perceptions. We used the network description to

correctly classify listeners behavioral responses with an accuracy over 92% on 35 participants.

State-of-the-art network description of human brain assumes static connectivities. However, brain networks in relation to perception and cognition are complex and dynamic. Analysis of transient functional networks with spatiotemporal variations to understand cognitive functions remains challenging. One of the critical missing links is the lack of sophisticated methodologies in understanding dynamics neural activity patterns. We proposed a clustering-based complex dynamic network analysis on source localized EEG data to understand the commonality and differences in gender-specific emotion processing. Besides, we also adopted Bayesian nonparametric framework for segmentation neural activity with a finite number of microstates. This approach enabled us to find the default network and transient pattern of the underlying neural mechanism in relation to categorical perception.

In summary, multivariate and dynamic network analysis methods developed in this dissertation to analyze structural and functional connectivities will have a far-reaching impact on computational neuroscience to identify meaningful changes in spatiotemporal brain activities.

## Table of Contents

<b>List of Tables.....</b>	<b>x</b>
<b>List of Figures.....</b>	<b>xi</b>
<b>Chapter 1 - Introduction .....</b>	<b>1</b>
Specific aims.....	2
Main results .....	4
Broader impacts and novelty .....	6
<b>Chapter 2 - Research Context.....</b>	<b>8</b>
Structural connectivity in classifying cognitive capacity .....	8
Functional connectivity in classifying categorical speech perception .....	11
Dynamic network analysis in discovering temporal patterns in categoric perception.....	14
<b>Chapter 3 - Structural Connectivity Analysis.....</b>	<b>18</b>
Data collection .....	18
Data visualization.....	19
Graph mining .....	21
Feature selection.....	24
Stability selection with error control.....	25
How to select $EV$ , $\pi thr$ and $q\Lambda$ : .....	26
Stability selection with Randomized Lasso.....	28
Modeling.....	30
Empirical analysis .....	30
Discussion .....	35

<b>Conclusion.....</b>	<b>37</b>
<b>Appendix.....</b>	<b>38</b>
Characteristics path.....	38
Global efficiency.....	39
Average clustering coefficient.....	39
Transitivity.....	39
Small-worldness .....	40
Assortativity coefficient .....	40
Modularity index.....	41
<b>Chapter 4 -     <i>Functional Connectivity Analysis</i>.....</b>	<b>42</b>
<b>Dataset .....</b>	<b>42</b>
Participants .....	42
Speech stimulus continuum and behavioral task.....	44
Behavioral data analysis.....	44
EEG recording and preprocessing .....	46
EEG functional connectivity and graph analyses.....	48
<b>Identifying behaviorally relevant aspects of functional connectivity.....</b>	<b>49</b>
<b>Feature selection.....</b>	<b>51</b>
<b>Results .....</b>	<b>54</b>
<b>Discussion .....</b>	<b>57</b>
<b>Conclusion.....</b>	<b>63</b>
<b>Appendix.....</b>	<b>63</b>
<b>Chapter 5 -     <i>Dynamic Network Analysis</i> .....</b>	<b>66</b>
<b>Dataset .....</b>	<b>66</b>
DEAP dataset.....	66

CP dataset .....	68
<b>Adaptive segmentation based dynamic network analysis in understanding affect.....</b>	<b>68</b>
K-means clustering .....	69
Graph mining.....	70
Empirical analysis on DEAP dataset.....	70
<b>HDP-HMM based dynamic network analysis in understanding rapid speech categorization</b>	
<b>decisions .....</b>	<b>76</b>
Hierarchical Dirichlet Process Hidden Markov Models .....	76
Calculation of dwell time statistics.....	77
Empirical analysis on CP dataset .....	77
<b>Conclusion.....</b>	<b>85</b>
<b>Appendix.....</b>	<b>85</b>
Hierarchical Dirichlet Process Hidden Markov Models .....	85
Hierarchies of Dirichlet Processes. ....	85
Variational Inference:.....	86
Memoized and Stochastic Variational Inference: .....	87
<b><i>Relationship to published works .....</i></b>	<b><i>89</i></b>
<b><i>References .....</i></b>	<b><i>90</i></b>

## List of Tables

TABLE 1: TWO-SIDED P-VALUE OF WILCOXON RANK-SUM STATISTIC FOR DIFFERENT NETWORK FEATURES ACROSS GROUPS. ....	22
TABLE 2: OVERALL RESULT OF EMPIRICAL ANALYSIS, HERE ALL: WHOLE DATASET WITHOUT FEATURE EXTRACTION AND SELECTION, NET: NETWORK FEATURES, ND: NETWORK FEATURES AND DEMOGRAPHIC VARIABLES AND NDI: NETWORK FEATURES, DEMOGRAPHIC VARIABLES, AND CORTICAL THICKNESS, EC: STABILITY SELECTION WITH ERROR CONTROL, RL: STABILITY SELECTION WITH RANDOMIZED LASSO, AUC: AREA UNDER THE CURVE, ACC: ACCURACY, PS: PROCESSING SPEED COGNITIVE ABILITIES, DSB: DIGIT SPAN BACKWARDS, BRIEF-WORKING MEMORY: BEHAVIOR RATING INVENTORY EXECUTIVE FUNCTION CLASS. ....	34
TABLE 3: SIGNIFICANT (BOLD) GLOBAL NETWORK MEASURES (KRUSKAL-WALLIS H-TEST TESTS) (TRIAL-LEVEL). ....	55
TABLE 4: GROUP COMPARISON OF GRAPH MEASURES OF FUNCTIONAL CONNECTIVITY BETWEEN RT GROUPS.....	55
TABLE 5: EFFECT OF SELECTION THRESHOLD OF STABILITY SELECTION (THRESHOLD) ON MODEL PERFORMANCE. THE PAIRWISE CORRELATION BETWEEN TWO BRAIN REGIONS (FUNCTIONAL CONNECTIVITY EDGE) WERE CONSIDERED AS FEATURES. THE NUMBER OF UNIQUE NODES ARE THE BRAIN REGIONS ASSOCIATED WITH SELECTED FEATURES. ACC, ACCURACY; AUC, AREA UNDER CURVE. ....	57
TABLE 6: EIGHT MOST IMPORTANT EDGES THAT GOVERN SPEEDED SPEECH CLASSIFICATION. COLLECTIVELY, THESE EDGES ACHIEVE A MODEL ACCURACY OF 57% IN SEGREGATING LISTENERS' SPEEDED DECISIONS (RTs) IN THE PERCEPTUAL TASK. HERE, A SCORE OF 0.85 MEANS THAT OUT OF 1000 ITERATIONS, THE EDGE WAS SELECTED BY STABILITY SELECTION 850 TIMES.....	60
TABLE 7: TWO-SIDED P-VALUE OF WILCOXON RANK-SUM STATISTIC FOR DIFFERENT NETWORK MEASURES AMONG GROUPS. NO SIGNIFICANT MEASURE IS EVIDENT FROM CORTICAL SURFACE DATA, BUT MAXIMUM MODULARITY AND SMALL-WORLDS ARE SIGNIFICANTLY DIFFERENT AMONG GROUPS FOR SKULL SURFACE EEG DATA IN PROCESSING STIMULI WITH HIGH AROUSAL VALUE. ....	75
TABLE 8: TOP 10 NETWORK EDGES (WITH NODE A AND NODE B) AND THEIR WEIGHT (CORRELATION VALUES).....	75

## List of Figures

FIGURE 1: PIE PLOT SHOWS DIVERSITY IN THE DATASET. THE DATASET HAS DEMOGRAPHIC MEASURES OF DIFFERENT SEX, ETHNICITY, RACE, AND AGE GROUP. THE SAMPLE SIZE OF THE AVERAGE AND BELOW-AVERAGE GROUP OF DSB, PS, AND BRIEF-WORKING MEMORY MEASURES ARE RELATIVELY UNBALANCED. HERE, ETHNICITY GROUP1 REPRESENTS NON-HISPANIC, AND GROUP 2 REPRESENTS NOT OTHERWISE SPECIFIED SPANISH, HISPANIC, LATINO GROUP, RESPECTIVELY.....	21
FIGURE 2: THE T-SNE EMBEDDED HIGHER DIMENSIONAL FEATURES ARE REPRESENTED BY 2-DIMENSIONAL SCATTER AND KERNEL DENSITY ESTIMATION (KDE) PLOT. THE GREEN LINES WITH DOTS AND RED LINES WITH '+' SIGN REPRESENT AVERAGE AND BELOW-AVERAGE GROUP DATA, RESPECTIVELY.....	21
FIGURE 3: CORRELATION MATRIX AMONG FEATURES. MOST OF THE FEATURES ARE HIGHLY CORRELATED WITH EACH OTHER (CP: CHARACTERISTICS PATH, AC: AVERAGE CLUSTERING COEFFICIENT, SW: SMALL-WORLDNESS, AS: ASSORTATIVITY, GE: GLOBAL EFFICIENCY, TR: TRANSITIVITY, MODULARITY: MD).....	23
FIGURE 4: STABILITY PATH OF FEATURES MATRIX (FOR DSB) WITH A RANGE OF REGULARIZATION PARAMETER ( $\lambda = 0.01 \sim 100$ ) AS A FUNCTION OF $(A/AMAX)^{1/3}$ . THE POWER $1/3$ SCALES THE PATH AND ENABLES TO VISUALIZE THE PROGRESSION ALONG THE PATH.....	25
FIGURE 5: EXPECTED NUMBER OF FALSELY SELECTED VARIABLE $EV$ VS $q\lambda$ GRAPH AND STABILITY PATH FOR DSB CLASS. LEFT SIDE OF THE PLOT (A) SHOWS THE VARIATION OF $EV$ AND $\pi thr$ ON MODEL ACCURACY. RED SOLID LINES OF PLOT (B) SHOW THE RELEVANT FEATURES (63) FOR BEST $EV$ AND $\pi thr$ . THE BLACK DOTTED LINES REPRESENT STABILITY PATH FOR IRRELEVANT FEATURES (956) OVER A RANGE OF REGULARIZATION PARAMETERS. ....	27
FIGURE 6: EFFECT ON SECTION THRESHOLD OVER MODEL PERFORMANCE FOR BRIEF-WORKING MEMORY PREDICTION. THREE LINES OF X-LABEL REPRESENT THE RANGE OF EACH BIN OF FEATURES SCORE (RANGE: 0 TO 1), NUMBER AND PERCENT OF FEATURE FALL IN EACH BIN. ....	30
FIGURE 7: VIN DIAGRAM OF EC AND RL SELECTED FEATURES FOR (A): DSB, (B): BRIEF-WORKING MEMORY, (C): PS CLASS. CYAN, BROWN AND BLUE COLORED CIRCLE REPRESENT THE NUMBER OF STABLE FEATURES SELECTED BY EC, RL AND COMMON FEATURES AMONG METHODS. PREDICTION ACCURACY AND NUMBER OF SELECTED FEATURES ARE RELATIVELY BETTER FOR RL METHOD. HERE ACC REPRESENTS ACCURACY.....	32
FIGURE 8: SCHEMATIC DIAGRAM OF THE PROCESSING PIPELINE. THE FEATURE MATRIX IS RANDOMLY SHUFFLED AND SPLIT INTO 80% AND 20% AS TRAINING AND TESTING DATA. FEATURE SELECTION METHODS (EC AND RL) ARE APPLIED TO TRAINING DATA TO	



FIND THE STABLE FEATURES. THOSE SELECTED FEATURES WERE USED TO TUNE AND ESTIMATOR LEARNING USING SHUFFLE-SPLIT GRID SEARCH APPROACH, AND FINALLY, MODELS ARE EVALUATED ON TEST DATA. ....	33
FIGURE 9: CIRCULAR BRAIN CONNECTIVITY GRAPH FOR A): BRIEF-WORKING MEMORY, (B): DSB, AND (C): PS CLASS USING RL METHOD. LEFT AND RIGHT SIDE OF THE CIRCLE REPRESENTS LEFT AND RIGHT HEMISPHERE. THE INNER SQUARES, OUTER SQUARES, AND GREEN CONNECTED LINED INDICATE SELECTED ROIs, CORTICAL THICKNESS OF ROIs, AND CONNECTIVITY AMONG ROIs, RESPECTIVELY. SHAPE AND THE SIZE OF THE OUTER SQUARE VARY WITH RANK (IMPORTANCE) IN PREDICTING IMPAIRMENT. ....	35
FIGURE 10: (A, B) DEMOGRAPHIC GENDER AND AGE DISTRIBUTIONS. (C) ACOUSTIC SPECTROGRAMS OF THE SPEECH STIMULI: THE STIMULUS CONTINUUM WAS CREATED BY PARAMETRICALLY CHANGING VOWEL FIRST FORMANT FREQUENCY OVER FIVE EQUAL STEPS FROM 430 TO 730 Hz (►), RESULTING IN A PERCEPTUAL-PHONETIC CONTINUUM FROM /u/ TO /a/. (D) TOKEN WISE RESPONSE TIMES FOR AUDITORY CLASSIFICATION. LISTENERS ARE SLOWER TO LABEL SOUNDS NEAR THE CATEGORICAL BOUNDARY (I.E., TOKEN 3). FEMALES (F) HAVE SIGNIFICANTLY SLOWER RESPONSE TIMES THAN MALES (M). ....	43
FIGURE 11: CLUSTERING RT DATA USING GMM AND BIC CRITERIA. MODEL SELECTION CONCERNS BOTH THE COVARIANCE TYPE AND NUMBER OF COMPONENTS IN THE MODEL. BRUTE-FORCE BASED EMPIRICAL ANALYSIS SHOWS THAT N=4 COMPONENTS WITH UNIQUE COVARIANCE MATRIX IS OPTIMAL. THE ‘*’ MARKED POSITION OF (A) SHOWS THE OPTIMAL COMBINATION. (B): PROBABILITY OF TRIALS LOADING INTO EACH COMPONENT. (C): TOKEN WISE RT BROKEN DOWN BY COMPONENT. BASED ON BEHAVIORAL RTs, FOUR CLUSTERS ARE EVIDENT THAT DISTINGUISH SUBGROUPS OF LISTENERS BASED ON THEIR SPEECH IDENTIFICATION SPEEDS: FAST (CLUSTER 1): 120~476 ms, MEDIUM (CLUSTER 2): 478~722 ms, SLOW (CLUSTER 0): 724~1430 ms, AND OUTLIERS (CLUSTER 3): 1432~2500 ms. ....	50
FIGURE 12: THE T-SNE EMBEDDED HIGHER DIMENSIONAL FUNCTIONAL CONNECTIVITY DATA ARE REPRESENTED BY A 2-DIMENSIONAL SCATTER AND KERNEL DENSITY ESTIMATION (KDE) PLOT. THE GREEN LINES WITH ‘.’, BLUE LINES WITH ‘*’, AND RED LINES WITH ‘+’ SIGN REPRESENTS DATA POINTS FOR SLOW, MEDIUM, AND FAST RT PARTICIPANTS, RESPECTIVELY. ....	51
FIGURE 13: EFFECT OF SELECTION THRESHOLD ON MODEL PERFORMANCE PREDICTION. THE THREE X-LABELS REPRESENT (TOP) THE RANGE OF EACH BIN OF FEATURES SCORE (RANGE: 0~1), (MIDDLE) THE NUMBER OF FEATURES FALLING IN EACH BIN, AND (BOTTOM) THE CORRESPONDING PERCENTAGE. ....	52
FIGURE 14: SCHEMATIC DIAGRAM OF THE PROCESSING PIPELINE. THE 64-CH EEG DATA IS FIRST PREPROCESSED, AND THEN SOURCE LOCALIZATION IS ADAPTED TO CONVERT SKULL SURFACE DATA TO CORTICAL SURFACE TIME SERIES DATA (68 ROIs DEFINED BY THE DESIKAN-KILLANY ATLAS PARCELLATION). PAIRWISE CORRELATIONS WERE CALCULATED TO DERIVE THE CONNECTIVITY MATRIX FOR EACH TRIAL OF THE SPEECH CP TASK. BEHAVIORAL RESPONSE TIMES (RTs) WERE CLUSTERED WITH BAYESIAN NON-	

PARAMETRIC (GMM) CLUSTERING. THESE CLUSTERS WERE LABELED AS FAST, MEDIUM, AND SLOW RT. ANOVA ANALYSIS OF GRAPH MEASURES W ADOPTED TO TEST SIGNIFICANCE AMONG RT GROUPS. STABILITY SELECTION AND MACHINE LEARNING APPROACHES WERE THEN USED TO FIND SIGNIFICANT PROPERTIES OF THE BRAIN'S FUNCTIONAL CONNECTIVITY RELATED TO BEHAVIORAL SPEEDS (RTs) IN SPEECH CP. ....	53
FIGURE 15: BRAINNET VISUALIZATION (TOP TO BOTTOM: LATERAL, MEDIAL, AND DORSAL VIEW) OF THE BRAIN NETWORK (54 EDGES) IDENTIFIED VIA STABILITY SELECTION. COLOR MAP 1-6 INDICATES, 1: FRONTAL (22 ROI), 2: PARIETAL (10 ROI), 3: TEMPORAL (18 ROI), 4: OCCIPITAL (8 ROI), 5: CINGULATE (8 ROI), 6: INSULA (2 ROI) REGIONS. NODE SIZE VARIES WITH ITS DEGREE OF CONNECTIVITY. CONNECTIVITY AMONG THE SAME LOBE ARE COLORED WITH SIMILAR NODE COLOR. EDGE WIDTHS REPRESENT THE WEIGHT OF ABSOLUTE CORRELATION (CONNECTIVITY STRENGTH). ....	56
FIGURE 16: A SPARSE BRAIN NETWORK (8 EDGES) PREDICTS LISTENERS' SPEED (RTs) OF SPEECH CATEGORIZATION (57% MODEL ACCURACY). RED NUMBERS ARE THE RANKED IMPORTANCE OF THE EDGES DESCRIBING BEHAVIOR. OTHERWISE, AS IN FIGURE 15.....	59
FIGURE 17: ACCURACY CURVES OF STABILITY SELECTION (AS IN FIGURE 13). STABILITY SELECTION WAS APPLIED TO CORRELATION, CH, ICH, PLV PSI BASED-EDGE MATRIX, AS WELL AS COMBINATIONS OF CH AND ICH, COMBINATION OF CORRELATION, CH, ICH, AND PSI BASED-EDGE MATRIX. HERE 'MUL' AND 'FOU' REPRESENTS MULTITAPER AND FOURIER TRANSFORM METHODS. THE DOT POINT OF EACH ACCURACY CURVE INDICATES MAXIMUM ACCURACY OF THE OPTIMAL COMBINATION OF FEATURES. CORRELATION-BASED CONNECTIVITY OUTPERFORMS ALL OTHER MEASURES.....	64
FIGURE 18: BRAIN NETWORK UNDERLYING SLOW RT LISTENERS (LEFT), MEDIUM RT LISTENERS (MIDDLE), AND FAST RT LISTENERS. SHOWN HERE ARE THE MOST HIGHLY CORRELATED (ABSOLUTE CORRELATION $\geq 0.5$ ) NETWORK EDGES. OTHERWISE AS IN FIGS. 6-7. INS, INSULA; IST, ISTHMUS OF CINGULATE; TRANS, TRANSVERSE TEMPORAL GYRUS (AUDITORY CORTEX); POB, PARS ORBITALIS; PRC, PRECENTRAL GYRUS (MOTOR CORTEX); PHIP, PARAHIPPOCAMPAL GYRUS; PREC, PRECUNUS; L/R, LEFT/RIGHT HEMISPHERE. ...	65
FIGURE 19: DIVERSITY IN THE DEAP DATASET WITH DIFFERENT PHENOTYPES. ....	66
FIGURE 20: EEG ELECTRODE LOCATIONS AND CORTICAL VIEW OF DESIKAN-KILLIANY ROI TEMPLATE. ....	68
FIGURE 21: ESTIMATING OPTIMAL WINDOW SIZE FOR MICRO-STATES. THE X AND Y AXIS REPRESENTS NUMBER OF CLUSTERS AND DISTORTION SCORE, RESPECTIVELY. THE DISTORTION SCORE DOES NOT DIFFER MUCH FOR THE DIFFERENT WINDOW SIZE. HOWEVER, OPTIMAL "ELBOW" POINT IS NOT EVIDENT DUE TO THE NOISY NATURE OF CORTICAL SURFACE DATA.....	69

FIGURE 22: SCHEMATIC DIAGRAM OF THE PROCESSING PIPELINE FOR DEAP DATASET. THE 32 CHANNEL EEG DATA IS FIRST PREPROCESSED, AND THE SOURCE LOCALIZATION IS ADOPTED TO CONVERT SKULL SURFACE DATA TO CORTICAL SURFACE TIME SERIES DATA. AN OPTIMAL 4 SEC, 25% OVERLAPPING ROLLING WINDOW IS USED TO CHUNK EACH TRIAL. PAIR-WISE CORRELATION MATRIX, CLUSTERING-BASED APPROACH IS USED TO FIND MICRO-STATES. EXPLORATORY DATA ANALYSIS IS USED FOR FURTHER INTERPRETATION.....	70
FIGURE 23: THE T-SNE EMBEDDED HIGHER DIMENSIONAL FEATURES ARE REPRESENTED BY 2-DIMENSIONAL SCATTER AND KERNEL DENSITY ESTIMATION (KDE) PLOT. THE GREEN AND RED LINES WITH DOTS (FIGURE A) REPRESENT NORMAL MALE AND FEMALE CONNECTIVITY DATA POINTS. SIMILARLY, BLUE AND ORANGE DATA POINTS (FIGURE B) REPRESENT HIGH AND LOW AROUSAL. BOTH FIGURES SHOW A DISTINCT CLUSTERED PATTERN. ....	70
FIGURE 24: APPROACHES IN FINDING OPTIMAL NUMBER OF CLUSTERS (K). ....	73
FIGURE 25: MODEL SELECTION WITH GAUSSIAN MIXTURE MODELS USING BAYESIAN INFORMATION-THEORETIC CRITERIA (BIC). MODEL SELECTION CONCERNS BOTH THE COVARIANCE TYPE AND THE NUMBER OF COMPONENTS IN THE MODEL. IN THIS CASE, BIC IS ADOPTED TO DETERMINE THE RIGHT MODEL.....	73
FIGURE 26: KNN BASED MICRO-STATES SHOWS DISTINGUISHABLE PATTERN OF MICRO-STATES. ....	74
FIGURE 27: TRANSITION MATRIX VISUALIZATION FOR MALE (FIGURE A) AND FEMALE (FIGURE B) GROUP. THE HEAT MAPS SHOW A SIGNIFICANTLY DISTINCT PATTERN OF TRANSITION IN PROCESSING HIGH AROUSAL STIMULI.	74
FIGURE 28: SCHEMATIC DIAGRAM OF THE PROCESSING PIPELINE FOR CP DATASET. THE 64 CHANNEL EEG DATA IS FIRST PREPROCESSED, AND INDIVIDUAL TRIALWISE AVERAGED. STICKY HPD-HMM WITH MEMOIZED VARIATIONAL INFERENCE (MOVI) IS APPLIED FOR DATA SEGMENTATION. EXPLORATORY DATA ANALYSIS ON DWELL TIME PATTERN IS USED FOR FURTHER ANALYSIS AND INTERPRETATION.....	76
FIGURE 29: PROBABILITY OF DATA POINTS LOADING INTO EACH COMPONENT. ....	78
FIGURE 30. TOP: A SAMPLE TRIAL OF 64 CHANNEL EEG DATA (SUBJECT#1, SLOW RT), MIDDLE: TILE VISUALIZATION OF DWELL PATTERN, BOTTOM: TIME SERIES VISUALIZATION OF DWELL PATTERN.....	79
FIGURE 31: DWELL TIME STATISTICS. ....	80
FIGURE 32: FREQUENCY OF CHANGING STATES AMONG TRIALS. ENTROPY IN SLOWER RTs ARE HIGHER THAN FASTER RT (SLOW RT: 2.26, MEDIUM RT: 2.11, FAST RT: 2.03).....	80
FIGURE 33: STATE TOPOMAP. ....	82
FIGURE 34: VISUALIZATION CLARA ACTIVATION MAPS. ....	82

FIGURE 35: TRANSITION MATRIX. ....	83
FIGURE 36: GRAPH VISUALIZATION OF TRANSITION MATRIX META-ANALYSIS. EACH NODE REPRESENTS ONE STATE. THE SELF-LOOP OF NODES REPRESENTS AVERAGE TIME A STATE REMAINS STABLE IN CONDITION-SPECIFIC (RTs) TRIALS. EDGES REPRESENT STATE TRANSITION PROBABILITIES. PROBABILITY <0.03 ARE DISCARDED FOR BETTER VISUALIZATION WITH A SMALLER NUMBER OF NODES.....	84

## Chapter 1 - Introduction

A major aspect of the complexity of the nervous system relates to their sophisticated morphology, especially the interconnectivity of their neuronal processing elements. In more highly evolved nervous systems, brain connectivity can be described at several levels of scale. These levels include individual axon-dendritic connections at microscale. Millions of neurons make local networks that are interconnected with short range connectivity. Some neurons have long connections, projecting information from one side of the brain to the other. Those micro and macro scale connections make the brain a fully connected, always active, highly segregated, and densely integrated, complex neural network [1]–[3]. Brain connectivity patterns provide the basis for explaining neural information processing and also neurological diseases and disorders.

There are different levels of brain connectivity. For example, structural (anatomical) connectivity refers to axon-dendritic connections or linking sets of neurons or neuronal populations. Unlike structural connectivity, functional connectivity is related to similar patterns of activation in different brain regions regardless of the apparent physical connectedness of the regions. Fundamentally, functional connectivity is a statistical concept that captures deviations from statistical independence between distributed and often spatially remote neuronal units. Statistical dependence may be estimated by measuring correlation, coherence, or phase-locking. Effective connectivity provides a directional (causal) network description of cognitive events among different functional areas of the brain.

The network descriptions of brain connectivity can detect functional integration and segregation, quantify centrality of individual brain regions or pathways, and characterize patterns of local anatomical circuitry [4]. In this dissertation, our main goal is to develop network descriptions of both structural and functional connectivity and multivariate models for classifying cognitive capacity (i.e., fluency, attention, and working memory) and categorical perceptions (i.e., emotion, speech perception). In addition, we also perform

dynamic network analysis on brain connectivity measures to determine the role of different functional areas in relation to categorical perceptions and cognitive events.

### ***Specific aims***

#### ***Aim 1: Determine structural connectivity to identify deficits in cognitive performance in relation to therapy for Acute Lymphoblastic Leukemia (ALL)***

Neurotoxicity associated with cancer, radiation therapy, or chemotherapy plays a significant role in neurocognitive impairments among survivors due to a disruption of developing neural circuitry. Data from the broader research literature suggest that ALL survivors have reduced white matter volumes that correspond to decreased structural and functional connectivity within regions of the central executive and salience networks. This decreased connectivity may be associated with deficits in cognitive performance in the domains of fluency, attention, and working memory [5]. A plethora of studies used Diffusion Tensor Imaging (DTI) tractography data and graph-theoretic approaches to construct structural networks and compare the topological parameters of the networks between ALL patients and healthy controls. These properties are used to evaluate cognitive abilities based on statistical testing (e.g., p-value). The statistical testing methods provide concurrence of a fixed hypothesis with the available data points but fail to identify the best possible hypothesis among all alternatives. Significance testing does not describe how strongly two variables were related.

On the contrary, multivariate analysis offers a unified framework to select *relevant and interpretable* features and also a *model* to explain the data. Widely used feature selection methods (filter-based and wrapper class) use weighted schemes to select features. However, the selection of features is inconsistent and changes with the type of model as well as a range of model parameters such as regularization, threshold selection, and hyperparameter tuning. Besides the conventional and graph mining approaches, we adopted stability selection and control [6] to select relevant and interpretable features in measuring cognitive deficits. This

feature selection process effectively reduces dimensions and reveals affected anatomical pathways that were verified by domain experts.

***Aim 2: Discover functional connectivity to explain individual response time (RT) variation in speech perception***

Categorical perception (CP) is an inherent property of speech perception. The response time (RT) of listeners' perceptual speech identification is highly sensitive to individual differences. While the neural correlates of CP have been well studied in terms of the regional contributions of the brain to behavior, functional connectivity patterns that signify individual differences in listeners' speed (RT) for speech categorization is less clear. To address these questions, we applied several computational approaches to EEG data, including graph mining, machine learning (i.e., support vector machine), and stability selection to investigate the unique brain states (functional neural connectivity) that predict the speed of listeners' behavioral decisions.

***Aim 3: Study dynamic network of brain activities to discover temporal patterns in categoric perception (i.e. emotion and speech perception)***

The human brain is a dynamic system, and its networks must dynamically self-(re)organize and coordinate to allow mental processes. Dynamic Functional Connectivity (dFC) is a recent expansion on conventional functional connectivity analysis, which assumes functional connectivity is static in time. dFC captures the spatiotemporal pattern of brain connectivity and has been suggested to be a more accurate representation of functional brain networks. Hence, dFC is widely used in resting-state fMRI analysis to find the default mode network as well as discover the dynamic network patterns of neurological diseases and disorders. Tracking the spatiotemporal fast transient networks remains challenging due to a limited understanding of neural activity dynamics as well as a lack of relevant sophisticated methodologies.

EEG microstate is a metrics-based approach that allows a dynamic view of coupling [7], [8]. This coupling refers to time-varying levels of correlated or mutually informed activity between brain regions. Conventional microstate based methods for investigating dFC mainly follow K-means clustering-based methods (e.g. [9]–[13]). Despite the usefulness of the K-means clustering-based method, it has several drawbacks. The method requires selection of an appropriate number of clusters (K). Choosing a proper number of clusters can be challenging, particularly if the data is dynamic, and prior knowledge is unknown. Besides, K-means clustering is a hard-clustering algorithm, cannot handle infinite number of clusters, sensitive to noise and outlier, and shows poor generalization over studies. In addition, this approach firmly relies on the window size, and the strategy for selecting a reasonable window size remains unsolved. To avoid these limitations, we adapted a sticky Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM) with memorized variational inference [14], which provides an elegant Bayesian Nonparametric framework for multivariate sequential data segmentation with finite numbers of microstates. Unlike conventional HDP-HMM, this approach is computationally feasible, scalable, reliable, and converges quickly.

### ***Main results***

Our structural connectivity analysis discovered many of the brain regions and connectivities that are known to be associated with executive functioning, including working memory, fluency, and attention. We also discovered the following results.

1. Patients with below-average working memory are less likely to have a complex, highly segregated, and densely integrated structural network as well as being unable to use brain connectivity effectively or adequately.
2. Fluency and working memory-related distinct structural brain networks are more intra-hemisphere centric, but intelligence is more inter-hemisphere centric.
3. It was observed that sex, race, and ethnicity were important demographic variables in modeling cognitive capacity.



Our functional connectivity analysis investigated the unique brain states that predict the speed of listeners' behavioral decisions. The results corroborate previous studies by supporting the engagement of similar temporal (STG), parietal, motor, and prefrontal regions in categorical perception using an entirely data-driven approach. Additionally, we also inferred the following.

1. Listeners' perceptual speed is directly related to dynamic variations in their brain connectomics.
2. Global network assortativity and efficiency distinguished fast, medium, and slow RT,
3. Functional networks underlying rapid decisions increased in negative assortativity (i.e., became disassortative) for slower RTs.
4. Slower categorical speech decisions caused excessive use of neural resources and more aberrant information flow within the CP circuitry.
5. Slower responders tended to utilize functional brain networks excessively (or inappropriately), whereas fast responders (with lower global efficiency) utilized the same neural pathways but with more restricted organization.

Human emotions change over time. The state-of-the-art data-driven emotion model ignored the inherent dynamic nature of emotions. We proposed K-means clustering and microstate based complex dynamic network analysis on cortical surface data to understand the connectives in modeling elicited emotion. Specifically, we focus on processing arousal in identifying the differences and similarities between males and females. Empirical analysis using the DEAP dataset revealed the following.

1. Males and females have mostly complimentary micro-states with some commonalities.
2. Males are more likely to stay in specific stable microstates, and females are more likely to stay in transient states.

3. Both groups utilize highly segregated and densely integrated network structure among brain regions in processing arousal.

Response time (RT) of listeners' perceptual speech identification is highly sensitive to individual differences. Dwell time patterns revealed from HDP-HMM analysis can easily explain individual differences in listeners' speed (RT) for speech categorization. This microstate-based analysis revealed the following.

1. Categorical speech perception with slower RT requires changing microstates more frequently than faster RT (higher uncertainty is involved in the decision-making process). Therefore, RT in categorical speech perception is inversely proportional to state transition frequency.
2. Stimulus coding, linguistic processing, response selection, and resting-state are the main components of the categorical speech perception related decision network (DN). Specifically, stimulus coding, response selection, and resting-state are strongly interconnected in this DN.
3. Besides transition frequency, listeners' perceptual speed depends on the duration they stay in DN. Perception with slower RT involved staying in DN longer than faster RT.

### ***Broader impacts and novelty***

The studies described here reflect an interdisciplinary blend of engineering and neuroscience. From structural to functional connectivity analysis, we developed a systematic data-driven computational framework for predictive modeling of cognitive events. In particular, we develop network descriptions of cognitive events such as executive function (working memory, fluency, and attention), speech perception, and emotion. Our proposed approach substantially improved model accuracy over existing models and improved our understanding of how neural activity processes cognitive events. Taken together, our novel approach to neuroimaging data demonstrates the derivation of small, yet highly meaningful patterns of brain connectivity that dictate speech behaviors using solely EEG. More broadly,

the functional connectivity and machine learning techniques used here could be deployed in future studies to identify the most meaningful changes in spatiotemporal brain activity that are modulated by development, normal learning, or those which decline in neuropathological states.

## Chapter 2 - Research Context

Neural activities are not only determined by a single neuron or a single brain region independently. The human brain contains about 100 billion neurons connected by about 100 trillion synapses, which are structurally and functionally organized over multiple scales [15], [16]. Recent studies [14]–[16] indicate important individual differences in structural and functional connectivity patterns even among persons with no diagnosable neurological or psychiatric disorders, and there is increasing evidence that this variability is associated with alterations in cognitive and behavioral variables that constrain real-world function. The variation or alteration in human structural and functional brain connectivity play a role in adult and pediatric neurological and psychiatric disorders that collectively incur a huge economic cost to the United States [19].

Revealing brain networks and different functional areas in relation to categorical perceptions and cognitive events has been a challenging scientific problem. Modern neuroimaging techniques such as diffusion MRI, functional MRI, PET, EEG or MEG produce increasingly large datasets of anatomical or functional connection patterns (e.g., Connectome Programs [20], [21], BRAIN Initiative[22]). Over the last decade, a new multidisciplinary approach is immersing into the study of complex systems. In this dissertation, our main goal is to develop network descriptions of both structural and functional connectivity and multivariate models for classifying cognitive capacity (i.e., fluency, attention, and working memory) and categorical perceptions (i.e., emotion, speech perception).

### ***Structural connectivity in classifying cognitive capacity***

Commonly affected neuro-biological substrates impacted by therapy for Acute Lymphoblastic Leukemia (ALL) include atrophy of grey matter (GM) and/or demyelination of the white matter (WM), suppression of neural progenitor proliferation, microvascular damage, dysregulation of proinflammatory cytokine cascades, oxidative stress, and general vulnerabilities [23]–[26]. A study of ALL survivors and controls revealed reduced gray

matter volumes in cortical regions associated with central executive and salience networks, as well as bilateral reductions in the periventricular and subcortical white matter volumes [27]. The most relevant study conducted, analyzed 31 ALL survivors and 39 matched healthy controls with a graph metric analysis of the diffusion-based structural connectome to demonstrate that ALL survivors had significantly lower small-worldness and network cluster coefficient [28]. Reductions in WM volume in the frontal lobes and significant bilateral reduction in prefrontal cortices have been shown to correspond with lower performances on tests of attention and short-term memory [29]. Diffusion Tensor Imaging (DTI) studies have shown that fractional anisotropy (FA) in right frontal, fronto-parietal, and temporal areas are associated with processing speed [30], [31] and working memory [32]. The differential in FA between patients and controls was proportional to both IQ and processing speed. Another study of ALL survivors 15 years off therapy and controls, demonstrated higher FA on the left but not the right and worse performance in processing speed and academics [33]. Taken together, data from the broader research literature and our studies suggest that ALL survivors have reduced WM volumes that correspond to decreased structural and functional connectivity within regions of the central executive and salience networks; this decreased connectivity may be associated with deficits in cognitive performance in the domains of processing speed, attention, and working memory.

In this dissertation, one of our goal was to develop a multivariate data-driven model of “cognitive abilities” of ALL patients from MRI-based volumetric measures, morphometry statistics (e.g., surface area and cortical thickness) from diffusion tensor imaging (DTI), and behavioral as well as demographic variables. We used

1. Wechsler Intelligence Scale: Digit Span Backwards (DSB),
2. Woodcock-Johnson Tests of Cognitive Abilities: Processing Speed (PS), and
3. The Behavior Rating Inventory of Executive Function (BRIEF-Working Memory)

to measure cognitive abilities (executive functions and processing speed) and categorized them into two categories: below-average and average. Performance below one standard deviation of the mean was considered to be below-average. The DSB and PS scores are performance measures completed by the patients, whereas the BRIEF-Working Memory scores are based on parent reports and were inversely coded relative to the DSB and PS scores such that higher scores indicated more difficulties. All scores were normalized for patient age. For a patient to participate in the neurocognitive testing, the patient must be English speaking or English language dominant.

A plethora of studies (e.g., [28], [34], [35] ) used DTI tractography data and graph-theoretic approach to construct structural networks and compare the topological parameters of the network between ALL patients and healthy controls. Commonly used network properties (features) are clustering coefficient, small-worldness index, characteristic path length, modularity, and nodal clustering. These properties are used to evaluate cognitive abilities based upon the p-value or correlation. For example, Zou et al. [34] reported significantly lower small-worldness and network clustering coefficient, in addition to greater cognitive impairments in the ALL subjects.

To understand the non-linear embedding of the data, we performed t-SNE visualization (t-distributed Stochastic Neighbor Embedding [36] ) with the LDA projection of high-dimensional DTI connectome data (i.e., cortical thickness, average length of all fibers that interconnect Region of Interests (ROIs), and demographic measurements). It is easy to note that data exhibit complex and linearly separable distributions. However, Popular machine learning algorithm shows poor performance. Presence of noise, high correlation among variables, reparation, small number of positive samples, and unbalanced distribution in ALL connectome data prevent further improvements. It was also observed that even the structural connectivity based network features are less informative, discriminative, and unable to describe the variability and structures inherent in connectome data.

To overcome such limitations, we adopted a multivariate, wrapper-based feature selection method called stability selection [6]. It not only works efficiently in the high-dimensional data but also provides finite sample control for some error rates of false discoveries in structure estimation. Besides the error control approach, we also applied Randomized Lasso for feature selection. The stability selection not only significantly improved the model performances. it effectively reduces the feature dimension and selected features that were verified by the domain expert. It was observed that few demographic variables, morphometry statistics, and structural connectivity among ROIs are consistent and relevant features that are invariant and stable over range of model parameters. Besides, stability section ranks the importance of features, hence helps us to interpret the relation between structural brain connectivity and cognitive abilities.

### ***Functional connectivity in classifying categorical speech perception***

When identifying speech, listeners naturally group sounds into smaller sets of discrete (phonetic) categories through the process of categorical perception (CP) [37]–[40]. Presumably, this type of behavioral “downsampling” promotes speech comprehension by generating perceptual constancy in the face of enormous physical variation in multiple acoustic dimensions, e.g., talker variability in tempo, pitch, or timbre [41]. CP is often characterized by sharp (stair-stepped) identification and peaked (better) discrimination functions near the categorical boundary when classifying an otherwise equidistant acoustic continuum.

Germane to the present study, response time (RT) data also reveal differences in the speed of listeners’ categorical decisions [42], [43]. In perceptual labeling tasks, for example, listeners categorize prototypical speech sounds (e.g., exemplars from their native language) much faster than their ambiguous or less familiar counterparts (e.g., nonnative speech sounds) [44]. RTs also slow near perceptual speech boundaries, where listeners shift from hearing one linguistic class to another (e.g., /u/ vs. /a/ vowel) and presumably require more

time to access the “correct” speech template [42], [43], [45], [46]. Relatedly, RTs vary with task manipulations and individual differences in speech perception in different populations. Studies demonstrate listeners’ speed in speech identification is highly sensitive to stimulus familiarity [45], [47], [48], auditory plasticity of short- [37] and long-term [44], [49], [50] experience, and neuropathologies and language-learning disorders (e.g., [51]–[54]). Given its fundamental role in the perceptual organization of speech, understanding individual differences in CP and its underlying neurobiology is among the broad interests to understand how sensory features are mapped to higher-order perception [40], [43], [55].

The neuronal elements of the brain organize in complicated structural networks [56]. Increasingly, it is appreciated that anatomical substrates constrain the dynamic emergence of coherent physiological activity that can span multiple spatially distinct brain regions [57]–[59]. Such densely intra-connected, sparsely inter-connected, dynamic connected networks are thought to provide the functional basis for information processing, mental representations, and complex behaviors [1]–[3], [60]. In this regard, neuroimaging studies have identified several functional brain regions that are important to CP including primary auditory cortex, left inferior frontal areas (i.e., Broca’s area), and middle temporal gyri (e.g., [44], [45], [61]–[67]). Previous studies also suggest that more neurons are preferentially activated by the prototypes of the speech categories compared to those at category boundaries [68]. Similarly, improved discriminability at category boundaries could reflect an increased number of neurons encoding sensory cues at these perceptual transitions [69], [70]. Such neuronal overrepresentations warp the sensory space and may account for the aforementioned RT effects in speech categorization. Still, while the neural correlates of CP have been well studied in terms of the regional contributions to behavior, we are aware of no studies that have investigated the mechanisms of speech CP from a full-brain (functional connectivity) perspective. Here, we focus on the speed (RT) of the listener’s perceptual speech



identification as RTs are highly sensitive to individual differences in CP [48]–[51] and reflect an objective, continuous measure of perceptual categorization skill.

Functional connectivity matrices derived from neuroimaging data are highly sparse and reflect high dimensional data. Hence, finding RT-related network edges is challenging. To solve that problem, we used stability selection with randomized lasso. Besides, we propose a systematic approach to determine and rank RT-related functional connectivity among brain regions that are consistent across model parameters. In doing so, we identify, objectively, the most important properties (i.e., features) of the functional EEG connectome that describe perceptual categorization.

Our recent EEG studies have characterized the neural underpinnings and plasticity in speech categorization using hypothesis-based approaches (e.g., Bidelman and Walker, 2019; N. Price et al., 2019 (in press)). Here, we take an entirely different, comprehensive *data-driven* approach to test whether individual differences in speed for speech categorization can be decoded from network-level descriptions of brain activity. Based on prior work, we expected machine learning to minimally decode brain regions previously identified in rapid categorical decisions (e.g., inferior frontal gyrus; Binder et al., 2004), thereby corroborating hypothesis-driven accounts of CP using an entirely data-driven, machine learning approach.

Our goal was to focus on graph theoretical approaches to analyze the complex networks that could provide a powerful new way of quantifying individual differences in speech perception. Another goal was to discover which aspects of those functional connectivity networks best explained the variation and diversity in listeners’ perceptual responses during speech sound categorization. We recorded high-density electroencephalograms (EEGs) while listeners rapidly classified speech in a rapid vowel identification task [43], [48]. We then applied graph analyses to source-localized EEG responses to derive the underlying functional brain networks related to speech categorization. Using Bayesian non-parametric modeling, we then show that speed for categorical decisions

unfold in three RT clusters that distinguish subgroups of listeners based on their behavioral performance (i.e., slow, medium, and fast responders). Applying state-of-the-art machine learning and stability selection analyses to neural data, we further show that local and global network properties of brain connectomics can decode group differences in behavioral CP performance with 92% accuracy (AUC=0.9). Our findings demonstrate that slow RT decisions related to categorical speech perception involve improper (or excessive) utilization of functional brain networks underlying speech. In contrast, fast and medium responders show less utilization.

### ***Dynamic network analysis in discovering temporal patterns in categoric perception***

Functional connectivity (FC) is assumed to have a stationary nature for a long period of time. But, assuming stationarity over entire measurements may be too simplistic to reveal the full extent of neuronal activity. However, human brain is a dynamic system, and its networks must dynamically self-(re) organize and coordinate to allow mental processes [72].

Therefore, neuronal dFC has recently been hypothesized to show fast fluctuations of FCs.

Emotions change over time [73]. Disturbances in emotion dynamics are key features of various mental health disorders, including post-traumatic stress disorder, depression, personality disorder [74]. Though, a state-of-the-art data-driven emotion model (e.g.[75]–[77]) ignored the inherent dynamic nature of emotions. Recently, EEG and neuroimaging-based tools were widely used to identify the neural regions or networks that underlie these dynamics. EEG microstate is a metrics-based approach that allows a dynamic view of a coupling [7], [8]. This coupling refers to possibly time-varying levels of correlated or mutually informed activity between brain regions. Microstates are a global topography but occur on such small time scales and change rapidly. Van et al. [78] hypothesized that these “atoms of thoughts” are fractal-like in the temporal dimension and characterized by scale-invariant dynamics hence have a wide range of applications. (E.g. detect schizophrenia [7], panic disorders [79], and sleep disorders [80]). Arousal is important in regulating

consciousness, information processing, and attention. It is an essential element in many influential theories of emotion, such as the “James- Lange theory of emotion” [81] or the “Circumplex model of affect” [82]. Emotional stimuli are thought to gain rapid and privileged access to processing resources in the brain. In this dissertation, we introduce a data-driven approach to discover how and where brain connectivity changes dynamically. Here, cortical surface data (using EEG source localization) is used to overcome the limitation of skull surface EEG data and answer neuroscience questions. We adopted the microstates-based approach, popular clustering algorithm, and complex-network analysis to discover a distinct dynamic brain connectivity pattern.

However, conventional methods for investigating dFC in EEG recordings mainly follows two common strategies: (i). Temporal sliding windows approach (e.g [10], [11], [72], [83]–[85]) and (ii). Adaptive segmentation using clustering approach (e.g. [9]–[13]). The main drawback of the first approach is connectivity analysis firmly relies on the window size, and the strategy for selecting the reasonable window size remains unsolved. The latter is based on clustering the time-varying FC graphs, FC matrices, or topological features across time and subjects with an assumption that the network is in one state at a specific time, and the states vary across times and subjects. State of the art practices use overlapping windows and K-means clustering [86]. Despite the usefulness of the K-means based method, it has several drawbacks. First, the method requires appropriate number of clusters ( $K$ ). Choosing a proper number of clusters can be difficult, particularly in case the data is dynamic, and prior knowledge is unknown. Second, K-means clustering starts with random initialization of the centroid, which leads to unrepeatable and inconsistent results. Third, K-means clustering is sensitive to noise and outliers, hence showing poor generalization over studies. Fourth, K-means is a hard-clustering algorithm. It cannot handle infinite number of clusters. Centroids may be influenced by noisy data when used with small number of trials. Lastly, some K-

means algorithms first employ sliding windows to construct dFC networks. Consequently, they also have the limitation of selecting the reasonable window size.

Some recent advanced approaches are taken to avoid these limitations. For example, combination of synchronization likelihood analysis (measure for detecting non-linear, dynamical coupling between pairs of recording series) and evolutionary clustering of the FC edge time series have been used to analyze resting-state EEG functional connectivity networks FCNs [87]. Tensor decomposition-based approaches [88], [89] and principal component analysis (PCA)-based approaches [90] have also been used. Nguyen et al. [91] introduced a stochastic multivariate Gaussian hidden Markov model (MGHMM) to unveil the multi-subject Event-Related Potential (ERP) into distinct EEG microstates (ERP components). Microstate functional Phase-locked Value (PLV) based connectivity are analyzed to reveal cognitive tasks.

The Hierarchical Dirichlet process HMM (HDP-HMM) [92]–[94] provides an elegant Bayesian nonparametric framework for sequential data segmentation with different numbers of states. State-of-the-art inference algorithms for HMMs and HDP-HMMs have enormous drawbacks. They cannot efficiently learn from large datasets, often trapped at local optima, and cannot effectively explore segmentations with varying finite numbers of states [14]. However, stochastic optimization methods [95], [96] cannot change the number of states during execution, therefore they are limited in large datasets and often converge to poor local optima. To overcome these limitations, Monte Carlo proposals [97]–[99] use the entire dataset, but require all sequences to fit in memory. Hence this process is not scalable and computationally inefficient. To overcome this limitation, generalizing memoized variational inference has been used for dirichlet process mixture models [100] and HDP topic models [101]. Michael et al. proposed an inference algorithm for the sticky HDP-HMM that scales to big datasets by processing a few sequences at a time [14]. Hence this algorithm is scalable,

reliable, and converges fast. In this dissertation, we adopted sticky HDP-HMM with memoized variational inference to discover temporal patterns in speech perception.

### **Chapter 3 - Structural Connectivity Analysis**

In the United States, Acute Lymphoblastic Leukemia (ALL), the most common child and adolescent malignancy, accounts for roughly 25% of childhood cancers diagnosed annually with a 5-year survival rate as high as 94% [102]. This improved survival rate comes with an increased risk for delayed neurocognitive effects in attention, working memory, and processing speed [33]. Predictive modeling and characterization of neurocognitive effects are critical to inform the family and also to identify patients for interventions targeting. Current state-of-the-art methods mainly use hypothesis-driven statistical testing methods to characterize and model such cognitive events. While these techniques have proven to be useful in understanding cognitive abilities, they are inadequate in explaining causal relationships, as well as individuality and variations. We developed multivariate data-driven models to predict the late neurocognitive effects of ALL patients using behavioral phenotypes, Diffusion Tensor Magnetic Resonance Imaging (DTI) based tractography data, morphometry statistics, tractography measures, behavioral, and demographic variables. Alongside conventional machine learning and graph mining, we adopted “Stability Selection” to select the most relevant features and choose models that are consistent over a range of parameters. The proposed approach demonstrated substantially improved accuracy (13% - 26%) over existing models and also yielded relevant features that were verified by domain experts.

#### ***Data collection***

Survivors of childhood ALL treated on a chemotherapy-only protocol (Total XVI([NCT00549848]), were prospectively evaluated (N=200; age on protocol  $7.2 \pm 4.4$  years; 61% male; 96 low-risk, 104 standard/high-risk). Subjects underwent MRI within six months after beginning treatment and neurocognitive testing two years later at the end of all protocol directed therapy. Working memory and decision speed were assessed using:

1. Wechsler Intelligence Scale: Digit Span Backwards (DSB),

2. Woodcock-Johnson Tests of Cognitive Abilities: Processing Speed (PS),
3. The Behavior Rating Inventory of Executive Function (BRIEF-Working Memory).

All MR examinations were performed on a Siemens 3T scanner. A T1-weighted imaging set was acquired with a 3D MPRAGE sequence, which provides excellent tissue contrasts among white matter, gray matter, and CSF as well as high spatial resolution (1x1x1 mm). Diffusion tensor imaging was acquired with 1.8\*1.8\*3.0 mm resolution with 12 directions, a, b=700, and 4 averages to increase signal-to-noise. For each scan, the anatomic imaging set was processed using FreeSurfer [103] obtain the 82 cortical and sub-cortical regions. Cortical thickness measures were evaluated for each of these regions. DTI processing was performed using the FSL FMRIB Toolbox[104]. To establish a reproducible network graph for each exam, probabilistic fiber tracking was then performed using FSL with 500 permutations from each voxel of the anatomic structures. The connection pathway between two nodes, which was the volume in image space that the connection fibers passed through, was extracted for each valid connection using a previously developed adaptation of the probabilistic fiber tracing technique [104]. The mean fractional anisotropy (FA) values of the connection pathway served as the quantitative measure for each edge. All processing was performed in the patient's native space. Overall, the following three types of features were used for model development:

1. Thickness of cortical regions (e.g., thickness of Left Cuneus),
2. DTI measures (undirected, weighted ROI connectivity),
3. Demographic and clinical variables (e.g., Sex, race, ethnicity).

### ***Data visualization***

The dataset we used was 61% male, 94% non-Hispanic, with the racial categories white, black, Asian, and multiple being 81%, 13%, 1%, and 5%, respectively. The feature matrix has 186, 126, and 182 samples for PS, DSB, and BRIEF-Working Memory measures. The number of samples in the below-average group is relatively low in our dataset (24 to 36%).

Hence, it is relatively unbalanced. Pie plots in Figure 1 show the data distribution for the different demographic variables and cognitive measures. It should also be noted that while the prevalence of below-average performance is relatively low for this application, it is substantially greater than normative expectation (16%).

Our dataset has 1019 variables overall. Before applying any machine learning algorithm, it is expected to check the assumptions required for model fitting and hypothesis testing. The t-distributed stochastic neighbor embedding or t-SNE [36] is a widely used unsupervised learning algorithm used to visualize high-dimensional data. It converts similarities between higher dimensional data points to joint probabilities. Thus, provides a faithful representation of those data points in a lower-dimensional human interpretable 2D or 3D plane. Such a projection brings insight on whether the data is separable, the data lies in multiple different clusters or inspecting the nature of those clusters. We applied LDA on our two-class dataset and considered 50 dimensions for t-SNE visualization. The LDA based t-SNE approach shows two distinct clusters for normal and below-normal groups in cognitive measurements. Figure 2 shows the t-SNE embedded scatter and kernel density estimation (KDE) plot of our data distribution. KDE plot is a non-parametric way to represent the probability density function. Besides, the scatter plot, the KDE plot is used here to visualize the trend of data distribution. The green dot and red '+' sign represent data points for normal and below-normal groups respectively. It is evident that the distribution of PS and DSB is complex and linearly non-separable. This necessitates the use of stability selection and control to choose features that are relevant and stable.



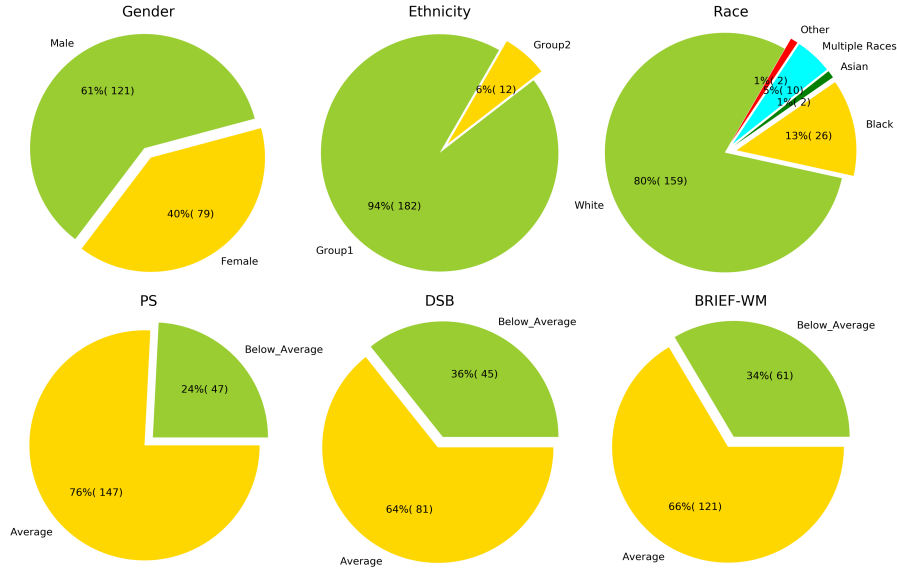


Figure 1: Pie plot shows diversity in the dataset. The dataset has demographic measures of different sex, ethnicity, race, and age group. The sample size of the average and below-average group of DSB, PS, and BRIEF-Working Memory measures are relatively unbalanced. Here, ethnicity Group1 represents Non-Hispanic, and Group 2 represents not otherwise specified Spanish, Hispanic, Latino group, respectively.

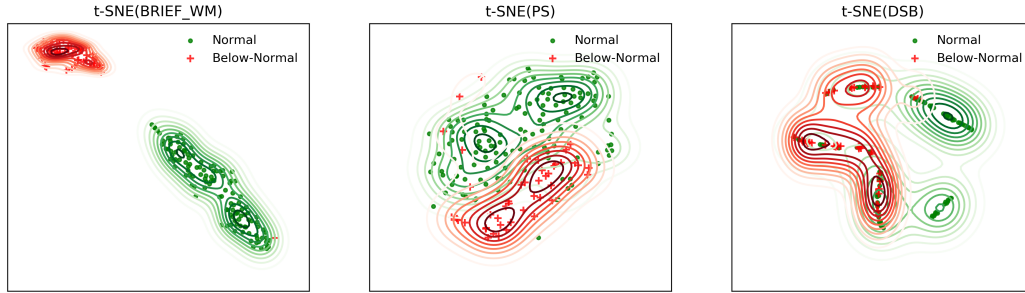


Figure 2: The t-SNE embedded higher dimensional features are represented by 2-dimensional scatter and kernel density estimation (KDE) plot. The green lines with dots and red lines with '+' sign represent average and below-average group data, respectively.

## Graph mining

Cognitive function is supported by distributed neural networks with highly segregated and integrated “small-world” organizations or clusters [1]–[3], [60]. More specifically, those organizations of neurons are densely intra-connected and sparsely interconnected. We applied graph theory to construct and analyze the brain connectome from DTI data. The 82\*82 undirected and weighted adjacency connectivity matrix from DTI FA data is used to calculate

seven basic global network features using BCT tools [4]. See Appendix section for mathematical definitions and interpretation of these network features. Those features are:

1. Characteristics path,
2. Global efficiency,
3. Average clustering coefficient,
4. Transitivity,
5. Small-worldness,
6. Assortativity coefficient, and
7. Modularity.

We computed the Wilcoxon rank-sum statistic on network features to evaluate significance differences across groups for DSB, BRIEF-Working Memory, and PS. Summary results are listed in Table 1. It was observed that, there is no significant difference in measurement across groups (except DSB: Transitivity ( $p < 0.048$ ), Global efficiency ( $p < 0.030$ ), Characteristics path ( $p < 0.046$ )). Those global measurements are based on normalized or averaged versions of clustering and community structure. Therefore, we found a strong correlation between the measurements except for modularity. Correlation measures among features are shown on Figure 3.

Table 1: Two-sided p-value of Wilcoxon rank-sum statistic for different network features across groups.

Feature	DSB	BRIEF Working Memory	PS
Characteristics path	0.75	0.16	0.97
Average clustering coefficient	0.81	0.17	0.83
Small-worldness	0.5	0.51	0.07
Assortativity	0.32	0.2	0.11
Global efficiency	0.78	0.13	0.97
Transitivity	0.75	0.2	0.88
Modularity	0.63	0.36	0.62

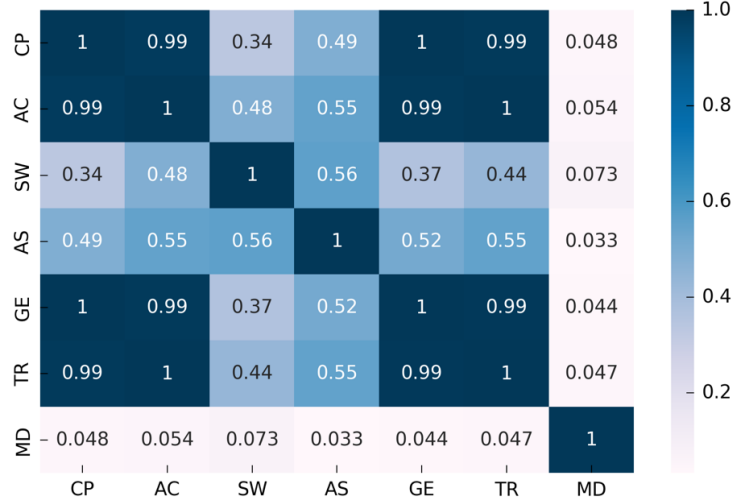


Figure 3: Correlation matrix among features. Most of the features are highly correlated with each other (CP: Characteristics path, AC: Average clustering coefficient, SW: Small-worldness, AS: Assortativity, GE: Global efficiency, TR: Transitivity, Modularity: MD)

In the next step, we concatenated network features, applied different classifiers, and observed their performance (details of parameter tuning and model fitting are explained in the appendix). We also evaluated the performance of the combination of network features and demographic variables (named as ND). The combination of network features and demographic variables with cortical thickness (named as NDI) were also similarly assessed. The summary results of overall empirical analysis are listed in Table 2. It was observed that the best classification accuracy among ND and NDI features are 74%, 69%, and 62% for PS, DSB, and BRIEF-Working Memory, respectively. However, AUC scores (0.5~0.62) of this model indicate that performances are not better than random guess. The mathematical definition of network features, p-values, high correlation among features, and poor model performances indicate overfitting, the presence of noise, and repetition. Overall, the model's performances differ from previous studies [28] because of :

1. Less number of trials for average group rather than below-average,
2. Connectivity matrix is highly sparse,

3. Network measurements are in average form. Average value over space matrix with lots of outliers making the features less discriminative,
4. Network features represent global properties rather than local, and
5. Few numbers of highly correlated features are used for modeling.

Hence there is scope for improvement using multimodal features, feature fusion, or decision fusion. Therefore, we applied conventional machine learning on weighted connectivity matrix with a stability selection. The details of this approach are discussed below.

### ***Feature selection***

Feature selection is used to reduce the dimensionality, improve the estimator's accuracy, and enhance generalizations by reducing overfitting in high-dimensional datasets [76], [105], [106]. Widely used filter-based methods identify consistent set of variables outside of the predictive model based on filtering criteria. Highly-correlated or redundant features may be selected, and significant interactions and relationships between variables may not be able to be quantified in those methods. However, one of the downsides of the multivariate approaches is that outcomes often depend on model parameters (e.g., regularization factor, hyperparameters) and needs massive computational resources, time, and there is a risk of over-fitting. Stability selection is a combination of subsampling and high-dimensional feature selection algorithms. Despite its simplicity, it is consistent for variable selection. The main advantages of this algorithm are:

1. It works efficiently in the high-dimensional data with less number of samples,
2. Stability selection provides finite sample control for some error rates of false discoveries and hence a transparent principle to choose a proper amount of regularization for structure estimation,
3. The method is extremely general and has an extensive range of applicability.

### Stability selection with error control

The concept of ‘stability path’ comes from the regularization path of regression analysis. A regularization path is defined as the coefficient value ( $\beta_k^\lambda$ ) of each features of a feature matrix over a range of regularization parameters ( $\lambda \in \Lambda$ ) [6]. Let  $I$  be a random subsample of  $n \times p$  feature matrix ( $n$ = number of samples and  $p$  is the dimension of features) of size  $n/2$  is drawn without replacement. The random sample size of  $n/2$  resembles most closely the bootstrap [107], [108], which is not worse than random guessing [6]. For every set  $K \subseteq \{1, \dots, p\}$ , the probability of being in the selected set  $S^\lambda(I)$  is

$$\Pi_k = P^*(S^\lambda(I)). \quad 1$$

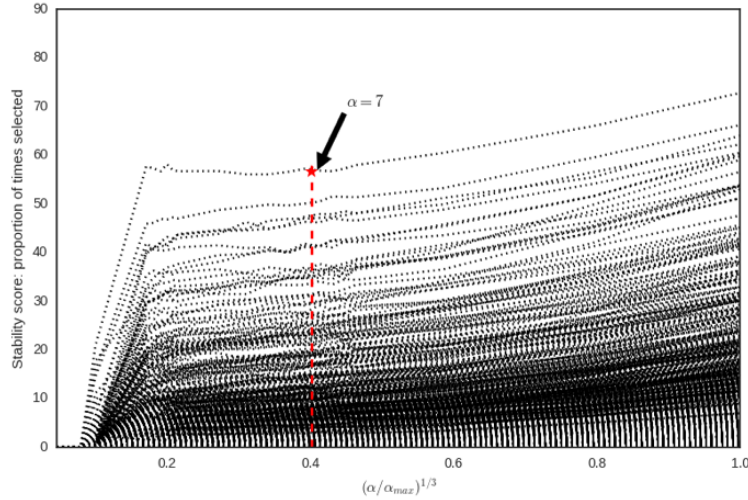


Figure 4: Stability path of features matrix (for DSB) with a range of regularization parameter ( $\alpha = 0.01 \sim 100$ ) as a function of  $(\alpha/\alpha_{max})^{1/3}$ . The power  $1/3$  scales the path and enables to visualize the progression along the path.

With stability selection, data are perturbed many times, and features are selected that occur in a large fraction of the resulting selection sets. Those variables are called the set of stable variables. Figure 4 shows the stability path of the feature matrix for the DSB class. L1 penalized logistic regression was used to discard the feature with zero coefficient and to consider the non-zero coefficient over a range of regularization parameters. This process is iterated over 10,000 times. Each of the black dotted lines represents a stability path of one

feature out of 1,019 features, and each dot represents the percentage of times it was selected out of all iterations. A red broken vertical line was drawn for the regularization parameter  $\alpha=7$ .

For a range of regularization parameter  $\Lambda$  and a cutoff threshold  $\pi_{thr}$  with  $0 < \pi_{thr} < 1$ , the set of stable variables is defined as

$$S^{stable} = \{\max_{\lambda \in \Lambda} \Pi_k^\lambda \geq \pi_{thr}\}. \quad 2$$

In Figure 4, the star (\*) point indicates the best stable feature (proportion of times selected  $\approx 57\%$ ). For stability selection, we keep variables with a high selection probability. Stability selection with error control ([24], Theorem-1) provides an upper bound to the expected number of falsely selected variables  $E(V)$ . The boundary equation can be defined as:

$$E(V) \leq \frac{1}{2\pi_{thr}-1} \frac{q_\Lambda^2}{p}. \quad 3$$

Where  $E(V)$  is the expected number of falsely selected variables,  $q_\Lambda$  is the average number of selected variables of a range of regularization parameter  $\Lambda$ . The threshold value  $\pi_{thr}$  in the equation 3 is a tuning parameter. In this study, we named this stability selection approach as Stability Selection with Error Control (EC).

*How to select  $E(V)$ ,  $\pi_{thr}$  and  $q_\Lambda$ :*

The influence of  $\pi_{thr}$  parameter in equation 3 is negligible [6], [109]. For a value, ranges from 0.6 to 0.9 results tend to be very similar. The value of  $E(V)$  is a design specification and can be controlled at the desired level. For a specific value of  $E(V)$ ,  $\pi_{thr}$  and regularization parameter  $\Lambda$ , the amount of stable features  $p$  can be calculated from the equation 3. The stable features are those which enter the regularization path first. For  $E(V) = 10$ ,  $\pi_{thr} = 0.70$  the average selected feature is = 63. The solid red lines of Figure 5(b) are the 63 features that come first in the regularization path; therefore, they can be considered as the most relevant features. The rest of the black dotted lines are irrelevant features. We can see some black dotted lines are mixed with red solid lines, but those can be treated as false positives for error tolerance  $E(V)$ .

To find the effect of the average number of selected variables (and size of feature matrix) in modeling, we tuned different combination of  $E(V)$  and  $\pi_{thr}$  and observed the model performance (The modeling and performance evaluation process is briefly described in modeling section). The points indicated by \* in Figure 5(a) are accuracy and AUC for different values of  $q$ . It is apparent that for a fixed value of  $E(V)$  the impact of  $\pi_{thr}$  is similar. On the other hand, variations of  $E(V)$  does not change model performance significantly. Though  $E(V)$  gives stability selection more freedom or error tolerance; after a certain level, the stability selection starts selecting more noise features, hence degrading the

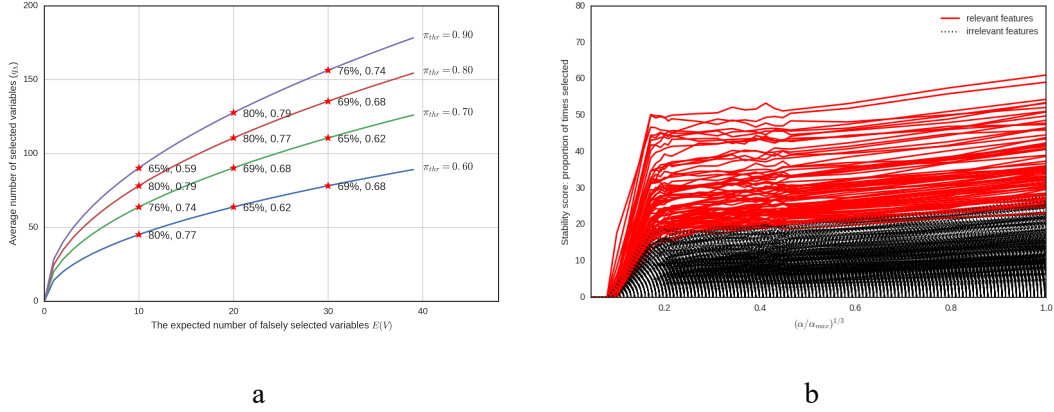


Figure 5: Expected number of falsely selected variable  $E(V)$  VS  $q_\Lambda$  graph and stability path for DSB class. Left side of the plot (a) shows the variation of  $E(V)$  and  $\pi_{thr}$  on model accuracy. Red solid lines of plot (b) show the relevant features (63) for best  $E(V)$  and  $\pi_{thr}$ . The black dotted lines represent stability path for irrelevant features (956) over a range of regularization parameters.

model performance. Though this algorithm allows freedom for error control, the bound has some drawbacks. First, it applies to the population version of the subsampling process. For a data set with small sample size, it is unrealistic to use it in practice. Second, the bound is derived under a very strong exchangeability assumption on the selection of noise variables and a weak assumption upon the quality of the original selection procedure. Shah et al. [109] claimed that this process is worse than random guessing.

Randomized Lasso (RL) [6] is a straightforward two-step approach. Instead of applying a specific algorithm to the whole data set to determine the selected set of variables based on the weight of coefficient, RL applied randomized lasso several times to random subsamples of the data of size  $n/2$  ( $n$  = number of samples) and chose those variables that are selected consistently across subsamples. By performing this double randomization several times, the method assigns high scores to features that are repeatedly selected across randomizations. In short, features selected more often are considered good features even though the “irrepresentable condition” [110] is violated. This approach is similar to the concept of bagging [111] and sub-bagging [108] algorithm.

We know, Lasso has sparse solutions. For higher-dimensional data, many estimated coefficients of variables become zero. Removing the variables can be used to reduce the dimensionality of the data. There are some limitations of Lasso-based feature selection are:

1. Lasso has a tendency to select an individual variable out of a group of highly correlated features.
2. When the correlation between features is low, the performance of Lasso is restrictive.

Lasso penalizes the absolute value of coefficients  $|\beta|_k$  of every component with a penalty term proportional to the regularization parameter  $\lambda \in \mathbb{R}$ . On the other hand, Randomized Lasso penalizes using randomly chosen values in a range  $[\lambda, \lambda/\alpha]$  where,  $\alpha \in (0,1)$  is the weakness parameter. The concept of weakness parameter is closely related to weak greedy algorithms [112]. Let  $W_k$  be i.i.d. random variable in a range from  $(\alpha, 1)$  for  $k = 1, \dots, p$ . The estimator of Randomized Lasso can be written as [6]:

$$\check{\beta}^{\lambda, W} = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \|Y - X\beta\|_2^2 + \lambda \sum_{k=1}^p \frac{|\beta_k|}{W_k} \quad 4$$



Here,  $Y$  and  $X$  are the class label and feature matrix, respectively. Implementation of equation: 4 is a straightforward two-stage process:

1. Re-scaling of the feature variables (with scale factor  $W_k$  for the  $k$ -th variable),
2. LARS algorithm is applied on re-scaled variables [113].

In this approach, the reweighting is simply chosen at random. It is not sensible to expect improvement from randomization with one random perturbation. However, applying Randomized Lasso with many iterations (e.g., 1000 times) and looking for variables that are chosen frequently is a useful tool to identify stable features [6].

RL assigns feature scores between 0 and 1 based on the frequency of selection over 10,000 iterations. We need to specify the score above which features should be selected to find out the best representative stable features. Threshold selection is a design parameter. We varied the selection threshold (i.e., the number of selected features) and observed the effect on model performance.

Figure 6 shows the effect of different selection threshold on modeling. The histogram illustrates the distribution of the score. The first line of x label shows the bin ranges of scores (0 to 1), second and third line shows the amount and percent of features that have a nearly same score for a specific bin. It was observed that 53% of features have the score of 0 to 0.1. That means, out of 10,000 iterations, they were selected between 0 to 10% of the time. For a specific selection threshold, e.g., 0.46, this algorithm selected 29 features. We built a model using those 29 features, which then gave us 89% accuracy (best model performance) with  $AUC=0.87$  for BRIEF-Working Memory class. The bell-shaped solid black and red dotted lines shows the accuracy and AUC curves for different selection thresholds. It was observed that the selection threshold was higher than the optimal value (0.46), which allowed the model to consider more noise variables and thus degrading the model performance significantly.

## Modeling

Modeling from selected features has several steps including

1. Use mean imputation to remove missing or NaN values,
2. Apply z-score normalization (center to the mean and component-wise scale to unit variance) to normalize the data,

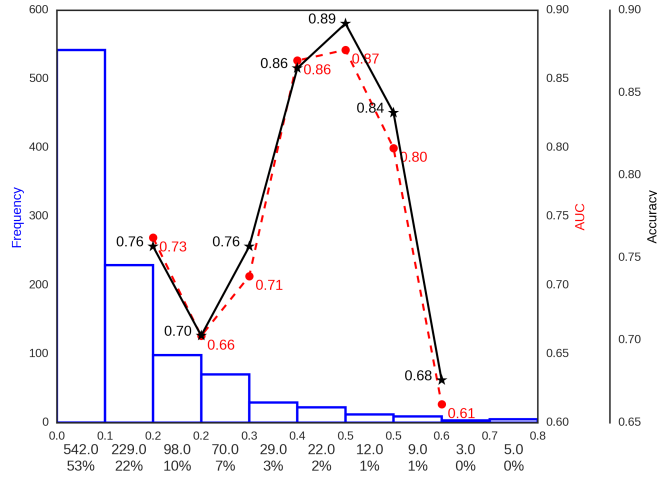


Figure 6: Effect on section threshold over model performance for BRIEF-Working Memory prediction. Three lines of x-label represent the range of each bin of features score (range: 0 to 1), number and percent of feature fall in each bin.

3. Test train splitting (80% for training and validation, 20% for testing),
4. SVM, Random Forest (RF) and Bagging (BAG) classifiers were used as estimators,
5. Hyperparameter tuning and model fitting using best estimator, and
6. Performance evaluation.

More about modeling is explained in the appendix.

## Empirical analysis

In this section, we will discuss the processing pipeline and results from the different experiments. The first step was data preprocessing. Missing values (NaN) of the feature matrix were replaced using the mean imputation along the column, and z-score normalization was used for data standardization. Preprocessed and standardized features matrix are then randomly shuffled and split into 80% training and 20% test examples. This testing data was kept unseen and used only for the final model evaluation.

The condition number of a matrix  $X$  is defined as the norm of  $X$  times the norm of the inverse of  $X$  [114] such that Condition Number =  $\|X\|\|X^{-1}\|$ . The condition number was computed using singular value decomposition and  $l_2$  normalization. If the condition number is less than infinity, the matrix is invertible. There is no hardbound; the higher the condition number (ill-condition matrix), the greater the error in the calculation. The condition number of our feature matrix is moderately higher for PS and BRIEF-Working Memory than the DSB class (PS: 61.17, DSB: 22.63, BRIEF-Working Memory: 56.11). However, stability selection can work perfectly on the ill-conditioned feature matrix, so we applied it with EC and RL on this training data. For EC,  $l_1$  penalized logistic regression with 10,000 iterations with 22 regularization parameters (ranges from  $10^{-2} \sim 10^2$ ) was used to get the stability path for each class label.

In this study, we did not specify the tolerance of error. Hence, we let the empirical analysis find optimality. Different combinations of  $E(V)$  and  $\pi_{thr}$  of equation 3 were evaluated to get the best accuracy and AUC. Our search grid approach indicated, the optimal  $E(V) = 10$  and  $\pi_{thr} = 0.8, 0.7, 0.8$  produced an accuracy of 87%, 81% and 86% for PS, DSB and BRIEF-Working Memory, respectively. Overall, 78, 63, and 78 features can be considered stable. Besides EC, RL approach was also evaluated on training data using the same range of regularizations parameter (22 continuous values) over 10,000 iterations. It was observed that RL selected a small subset of features (except PS class) compared to EC. Overall 32, 136, and 29 features were the optimal number of features for PS, DSB, and BRIEF-Working Memory class, respectively.

Though accuracy and AUC vary for the two-selection method, there is a significant commonality between selection. The Venn diagram of Figure 7 shows the set of selected and

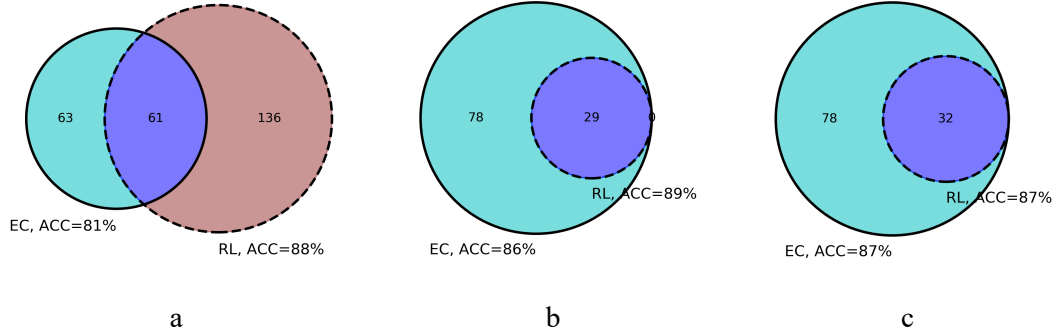


Figure 7: Venn diagram of EC and RL selected features for (a): DSB, (b): BRIEF-Working Memory, (c): PS class. Cyan, Brown and blue colored circle represent the number of stable features selected by EC, RL and common features among methods. Prediction accuracy and number of selected features are relatively better for RL method. Here ACC represents accuracy.

common features among methods. The cyan, brown, and blue circles represent ES, RL, and common features among two methods. As we allowed some errors in selection ( $E(V) = 10$ ), the EC method selected more features (except for PS) than RL method. This method selected nearly 8% of features from the feature matrix as stable features. However, RL method selected nearly 4%. Those selected variables were then used to train estimators. Estimator learning has three steps:

1. Reshape the feature matrix with stability selection (reduce the dimension),
2. Random shuffle and split the selected feature matrix into 80-20% training and validation set and
3. Iterative grid search approach was used to find the model with the best accuracy.

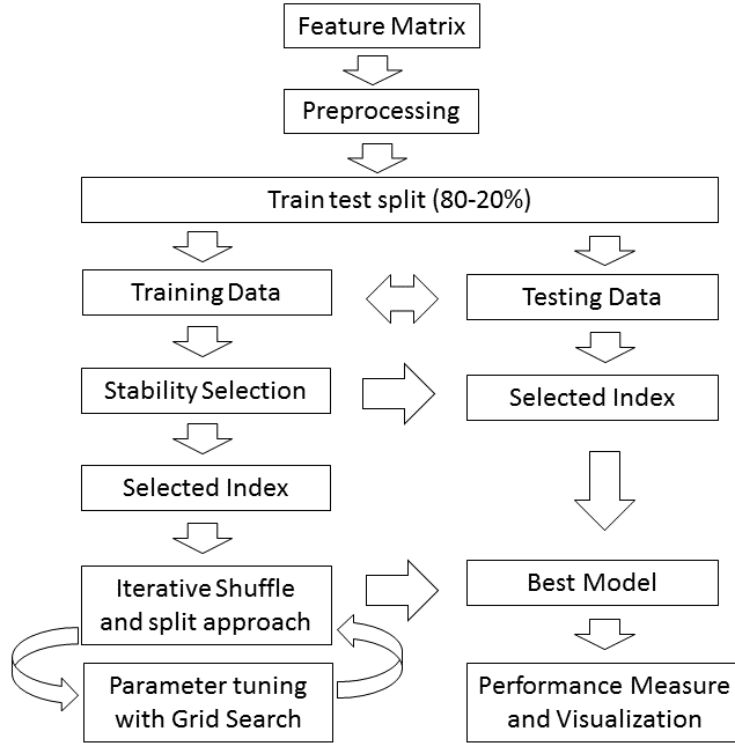


Figure 8: Schematic diagram of the processing pipeline. The feature matrix is randomly shuffled and split into 80% and 20% as training and testing data. Feature selection methods (EC and RL) are applied to training data to find the stable features. Those selected features were used to tune and estimator learning using shuffle-split grid search approach, and finally, models are evaluated on test data.

The same steps were applied for ensemble methods (RF and BAG). The best tuned model was then evaluated on test data. Data processing and modeling pipelines are shown in Figure 8. Table 2 shows the performance of different methods. It was observed that the performance of RL method is not only better than EC but also selected less number of stable and robust features. On the other hand, SVM shows better performance than the ensemble methods.

The best F1 score for the average category is greater than or equal to 0.90 for all three estimators, which means SVM with RL has fewer false negatives. On the other hand, the best F1 score for the below-average category was 0.71, 0.86, and 0.87 for PS, DSB, and BRIEF-Working Memory, respectively. Though this score for DSB and BRIEF-Working Memory class was at a satisfactory level (less than false positive), as well as the score for the PS class

(0.71) with an accuracy 87% has scope for further improvement. The main reason for this poor performance is a fewer number of negative examples. The PS class has only 24% negative examples. The estimators got very few (only 35) training examples after 80-20% split. Therefore, we need more negative instances for further improvement.

Besides directionality reduction and model improvement, Stability selection can be used for interpreting important features and their rank. Class label correlated shadow features get a high score even though necessary conditions (regularization parameter or estimator) change. This illustrates their strong and stable relationship with the response (class label).

Figure 9 shows the circular visualization of anatomical connectivity between different ROIs that are closely related to class labels. Those connectivities are selected by RL method. Left and right sides represent left and right hemisphere accordingly. The width of connection varies with the rank of importance. Similarly, the outer green square represents the related cortical thickness. The width of those squares varies with their rank as well. It was observed that among the 311 possible connections among ROIs, 29.90% 7.3%, 7.71% anatomical connectivity is important in modeling PS, DSB, and BRIEF-Working Memory, respectively.

Table 2: Overall result of empirical analysis, here All: whole dataset without feature extraction and selection, Net: Network features, ND: Network features and demographic variables and NDI: Network features, demographic variables, and cortical thickness, EC: Stability selection with Error Control, RL: Stability selection with Randomized Lasso, AUC: Area Under the Curve, ACC: Accuracy, PS: Processing Speed Cognitive Abilities, DSB: Digit Span Backwards, BRIEF-Working Memory: Behavior Rating Inventory Executive Function class.

Class	Selection Method	Number of Variables	SVM				RF				BAG			
			ACC		AUC		F1		ACC		AUC		F1	
			Average	Below-Average	Average	Below-Average	Average	Below-Average	Average	Below-Average	Average	Below-Average	Average	Below-Average
PS	All	1019	74%	0.50	0.85	0.00	71%	0.48	0.83	0.00	66%	0.45	0.79	0.00
	Net	7	74%	0.50	0.85	0.00	74%	0.50	0.85	0.00	74%	0.60	0.83	0.37
	ND	12	74%	0.50	0.85	0.00	74%	0.50	0.85	0.00	71%	0.71	0.83	0.15
	NDI	93	74%	0.50	0.85	0.00	59%	0.49	0.74	0.12	71%	0.48	0.83	0.00
	EC	78	87%	0.78	0.92	0.71	66%	0.45	0.79	0.00	71%	0.50	0.83	0.00
	RL	32	87%	0.78	0.92	0.71	74%	0.50	0.85	0.00	74%	0.50	0.85	0.00
DSB	All	1019	62%	0.50	0.76	0.00	65%	0.57	0.77	0.31	58%	0.49	0.72	0.15
	Net	7	62%	0.50	0.76	0.00	62%	0.54	0.74	0.29	69%	0.62	0.79	0.43
	ND	12	62%	0.50	0.76	0.00	65%	0.59	0.76	0.40	65%	0.55	0.78	0.18
	NDI	93	62%	0.50	0.76	0.00	65%	0.57	0.77	0.31	65%	0.65	0.76	0.40
	EC	63	81%	0.79	0.85	0.74	62%	0.52	0.75	0.17	62%	0.54	0.74	0.29
	RL	136	88%	0.89	0.90	0.86	65%	0.59	0.76	0.40	62%	0.59	0.69	0.50
BRIEF_WM	All	1019	62%	0.50	0.77	0.00	62%	0.53	0.75	0.22	57%	0.48	0.70	0.20
	Net	7	57%	0.46	0.72	0.00	57%	0.46	0.72	0.00	51%	0.43	0.67	0.10
	ND	12	62%	0.50	0.77	0.00	35%	0.31	0.48	0.14	35%	0.31	0.48	0.14
	NDI	93	62%	0.50	0.77	0.00	59%	0.49	0.74	0.12	57%	0.47	0.71	0.11
	EC	78	86%	0.82	0.90	0.78	68%	0.59	0.79	0.33	70%	0.64	0.79	0.48
	RL	29	89%	0.87	0.92	0.85	68%	0.59	0.79	0.33	65%	0.55	0.77	0.24

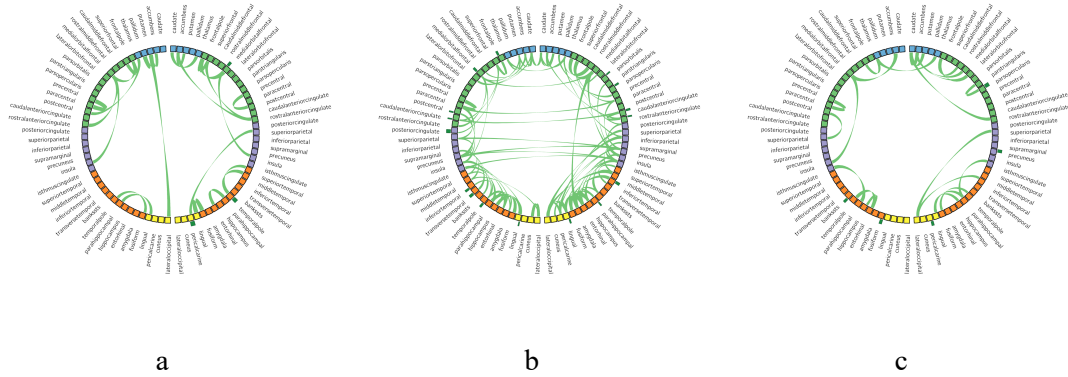


Figure 9: Circular brain connectivity graph for a): BRIEF-Working Memory, (b): DSB, and (c): PS class using RL method. Left and right side of the circle represents left and right hemisphere. The inner squares, outer squares, and green connected lines indicate selected ROIs, cortical thickness of ROIs, and connectivity among ROIs, respectively. Shape and the size of the outer square vary with rank (importance) in predicting impairment.

## Discussion

The hippocampus serves a critical function in long-term memory (LTM), navigation, cognition, and working memory maintenance. An increasing amount of evidence shows that the hippocampus is involved during the processing of spatial and spatiotemporal discontinuity, and relational memory [115], [116]. Specifically, CA1 neurons in the hippocampus are critical for autobiographical memory, autonoetic consciousness, and mental time travel [117]. The medial orbitofrontal cortex is necessary for the coordination of working memory, manipulation, maintenance, and monitoring processes [118]. Stability selection ranked the volume of left and right CA1, left CA2 of the hippocampus, right Medial Orbitofrontal (RMO), and right hippocampus as a very important feature for working memory classification. Significant differences ( $p < 0.0001$ ) in these areas were detected in the below-average BRIEF-Working Memory group.

On the other hand, memory span is the longest list of items that a person can repeat back in correct order immediately after the presentation. It is a standard measure of short-term memory. Once a digit sequence is presented, the participant is asked to recall the sequence in

reverse order in DSB related task (to assess working memory). The Posterior Cingulate Cortex (PCC) has a central role in supporting internally directed cognition [119]. The PCC shows increased activity when individuals retrieve autobiographical memories, plan for the future, and regulate the focus of attention[120], [121]. We found the volume of left PCC is a highly ranked feature in DSB classification and was significantly smaller ( $p<0.02$ ) for the below-average group. Hence, working memory is strongly related to the volume of the left PCC.

However, CA4 neurons of the hippocampus in the perikaryon area and dendritic branching of both CA4 and CA1 neurons are less in autistic children [122]. We observed that right CA4 and dentate gyrus of the hippocampus (CA4-DG) is an essential, highly ranked, significantly distinguishable feature to predict PS and BRIEF-Working Memory. The decreased volume in CA4-DG volume is observed in the below-average group (PS, DSB, and BRIEF-Working Memory).

Global efficiency is used to find how cost-efficient and fault-tolerant a particular network construction is. We found Global efficiency of DSB related network is significantly ( $p<0.03$ ) lower in the below-average group. This indicates that patients with below-average working memory are unable to use brain connectivity effectively or adequately. Besides Global efficiency, below-average performing groups exhibited significantly ( $p<0.04$ ) reduced transitivity. Lower transitivity indicates loose connectivity and less potential for integration among nodes to create a clique or complete graph. Hence, the group with the below-average working memory is less likely to have a complex, highly segregated, and densely integrated structural network.

Selected network edges presented in Figure 9 showed remarkable connectivity patterns. Highly distinguishable inter-hemispheric connectivity is evident among patients with below-average working memory (DSB network). However, cognitive abilities and working memory-related distinct structural brain networks are more intra-hemisphere centric.



Networks presented in Figure 9(a) and Figure 9(b) contain many of the brain regions known to be associated with executive functioning, including working memory, fluency, and attention. Involvement of the superior and middle frontal regions, the ventrolateral frontal regions of parstriangularis and parsopercularis, anterior cingulate, insula, and superior parietal are critical components of the central executive and salience networks. While these networks form the most reproducible basis for these functions in both fMRI and cognitive neuroscience, the involvement of the temporal lobe regions is also consistent with short-term storage of information for manipulation in the frontal lobes during the working memory tasks.

On the other hand, networks presented in Figure 9 contain many brain regions, which would be associated with processing speed. Involvement of the superior and middle frontal regions, the orbital frontal regions, anterior cingulate, insula, and superior parietal regions are consistent with regions, which would potentially be engaged during the processing speed tasks. Some of the regions, such as the insula and anterior cingulate, would be engaged in active switching between tasks such as surveillance and response. While these networks form the most reproducible basis for these functions in both fMRI and cognitive neuroscience, the involvement of the temporal lobe regions is also consistent with short-term storage of information for manipulation during the evaluation process. Besides cortical thickness and structural connectivity, it was observed that sex, race, and ethnicity were also important demographic variables in modeling those cognitive functions.

### ***Conclusion***

The aim of this study was to develop a data-driven multivariate approach to accurately classify cognitive abilities in ALL patients at the end of therapy. The state-of-the-art cognitive neuroscience mainly uses hypothesis-driven statistical testing to characterize and model neural disorders and diseases, and while these methods provide concurrence of a fixed hypothesis with the available data points, they fail to evaluate all possible hypotheses. In this

study, we developed models with stability selection using MRI-based volumetric measures, morphometry statistics, and behavioral as well as demographic variables. Stability selection not only reduced feature dimension and improved model accuracy, but it selected consistent and relevant connectome features that were invariant and stable over a range of model parameters. This approach also discovered brain regions and structural connectivity, which were strongly associated with processing speed and executive functions, including working memory, fluency, and attention. The findings of this study suggest that the performance and generalization capability of stability selection-based models are superior compared to the classical machine learning and graph mining approach. Since this study was limited to DTI based structural connectivity, it is inadequate in explaining causal relationships among brain regions as well as individuality and variations. Furthermore, a number of possible fMRI based future studies using the same experimental set up with a larger population are necessary for further improvement.

### ***Appendix***

Mathematical definitions and interpretation of network features are given below:

#### ***Characteristics path***

A fundamental property of brain networks is functional integration, which indicates how integrated a network is and, thus, how easily information flows [4] among nodes. A widely used approach to estimate properties of functional integration between nodes is based on the concept of characteristic path length. The characteristic path length is defined as the average shortest path length in the network [123]. Hence, small characteristic path values imply dense connectivity and stronger potential for integration among nodes. Let,  $L_i$  is the average distance between node  $i$  and all other nodes of a network, Average Characteristic path is defined as:

$$L = \frac{1}{n} \sum_{i \in N} L_i = \frac{1}{n} \sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{ij}}{n-1} \quad 5$$

Where,  $d_{ij}$  is the shortest distance between node  $i, j$  (shortest path can be calculated using any popular shortest path algorithm),  $N$  is the set of all nodes, and  $n$  is the total number of nodes.

#### *Global efficiency*

Global efficiency (E) is used to find how cost-efficient a particular network construction and how fault tolerant the network is. Hence, high global efficiency, implying the excellent use of resources. In brain connectivity analysis, structural and effective networks are similarly organized and share high global efficiency. On the other hand, functional networks have weaker connections and consequently share lower global efficiency [3]. Global efficiency is the average of inverse shortest path length hence inversely related to the average characteristic path length. E is defined as:

$$E = \frac{1}{n} \sum_{i \in N} E_i = \frac{1}{n} \sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{ij}^{-1}}{n-1} . \quad 6$$

#### *Average clustering coefficient*

The average clustering coefficient for the network reflects, how close its neighbors are to being a clique or complete graph. The average clustering coefficient of a node is defined as the fraction of triangles around a node [123] and defined as:

$$C = \frac{1}{n} \sum_{i \in N} C_i . \quad 7$$

Here,  $C_i$  is the clustering coefficient of  $i^{\text{th}}$  node. Let  $k_i$  is the number of neighborhood nodes, and  $t_i$  is the number of triangles created around  $i^{\text{th}}$  node. If a node has  $k$  neighbors, there are  $k(k-1)/2$  edges could exist among the nodes within the neighborhood. Hence,  $C$  can be defined as:

$$C = \frac{1}{n} \sum_{i \in N} \frac{2t_i}{k_i(k_i-1)} . \quad 8$$

#### *Transitivity*

Transitivity is a classical variant of average clustering coefficient and defied as:

$$T = \frac{\sum_{i \in N} 2t_i}{\sum_{i \in N} k_i(k_i - 1)} . \quad 9$$

The value of average clustering coefficient can be influenced by nodes with a low degree. But transitivity is normalized collectively and, consequently hence, does not have such problem [60].

#### *Small-worldness*

Small-world network (S) is formally defined as networks that are significantly densely clustered and have larger characteristic path length than random networks [123].

Mathematically S can be expressed as:

$$S = \frac{C/C_{random}}{L/L_{random}} . \quad 10$$

Where  $C$  and  $C_{rand}$  are the clustering coefficients, and  $L$  and  $L_{rand}$  are the characteristic path lengths of the test network and an equivalent random network with the same degree on average respectively. For a small world network  $S > 1$ ,  $C \gg C_{random}$  and  $L \gg L_{random}$ . Such network tends to contain more densely connected cliques/near-cliques/sub-networks than random network. Those sub-networks are interconnected by one or more edge.

#### *Assortativity coefficient*

Despite the importance of local and community structure, it is essential to study global diversity in networks. Hence the tendency to connect nodes with similar numbers of edges. This tendency, called assortativity, described crucial dynamic and structural properties of real-world networks, such as epidemic spreading or error tolerance [124]. A positive assortativity coefficient indicates that nodes tend to link to other nodes with the same or similar degree; on the other hand, negative values indicate relationships between nodes of different degree. Biological networks typically show negative assortativity coefficient as high degree nodes tend to attach to low degree nodes [125]. Mathematically, the assortativity coefficient is the Pearson correlation coefficient of degree between pairs of linked nodes [126]. Consider an undirected graph of  $N$  vertices and  $M$  edges with degree distribution  $p_j$ .

That is  $p_j$  is the probability that a randomly chosen node on the graph will have degree  $k$  and  $q_k$  is the distribution of the remaining degree. This  $q_k$  captures the number of edges leaving the node, other than the one that connects the pair. The assortativity coefficient ( $r$ ) is defined as:

$$r = \frac{\sum_{jk} jk(e_{jk} - p_j q_k)}{\sigma_q^2} . \quad 11$$

Where,  $\sigma_q^2$  is the variance of distribution  $p_k$  and  $e_{jk}$  refers to the joint probability distribution of the remaining degrees of the two nodes.

#### *Modularity index*

Modularity refers to the ability of subdivision the network into non-overlapping groups of nodes (known as modules or community) in a way that maximizes the number of within-group edges. Networks with high modularity have dense connections between the nodes within the modules but sparse connections between nodes in different modules. Hence, modularity quantifies the community strength of a test network by comparing the fraction of edges within the community with respect to random network [127]. It is widely used to discover anatomical modules correspond to groups of specialized functional area which is previously determined by physiological recordings. Usually, anatomical, effective, and functional modules in brain connectivity show extensive overlap [4]. The modularity index of a given network is the fraction of the edges that fall within the given groups minus the expected fraction if edges were distributed at random. Finding optimal modular structure is an optimization problem. Any optimization approach generally sacrifices some degree of accuracy for computational speed. Widely used algorithm to find optimal modular structure are proposed by Newman et al. [128], and Blondel et al. [129].

## Chapter 4 - Functional Connectivity Analysis

Categorical perception (CP) is an inherent property of speech perception. The response time (RT) of listeners' perceptual speech identification is highly sensitive to individual differences. While the neural correlates of CP have been well studied in terms of the regional contributions of the brain to behavior, functional connectivity patterns that signify individual differences in listeners' speed (RT) for speech categorization is less clear. To address these questions, we applied several computational approaches to the EEG, including graph mining, machine learning (i.e., support vector machine), and stability selection to investigate the unique brain states (functional neural connectivity) that predict the speed of listeners' behavioral decisions. We infer that (i) the listeners' perceptual speed is directly related to dynamic variations in their brain connectomics, (ii) global network assortativity and efficiency distinguished fast, medium, and slow RT, (iii) the functional network underlying speeded decisions increases in negative assortativity (i.e., became disassortative) for slower RTs, (iv) slower categorical speech decisions cause excessive use of neural resources and more aberrant information flow within the CP circuitry, (v) slower responders tended to utilize functional brain networks excessively (or inappropriately) whereas fast responders (with lower global efficiency) utilized the same neural pathways but with more restricted organization. Our results showed that neural classifiers (SVM) coupled with stability selection correctly classify behavioral RTs from functional connectivity alone with over 92% accuracy (AUC=0.9). Our results corroborate previous studies by supporting the engagement of similar temporal (STG), parietal, motor, and prefrontal regions in CP using an entirely data-driven approach.

### ***Dataset***

#### *Participants*

Thirty-five adults (12 males, 23 females) were recruited from the University of Memphis student body and Greater Memphis Area to participate in the experiment. All but

one participant was between the age of 18 and 35 years ( $M = 24.5$ ,  $SD = 6.9$  years). All exhibited normal hearing sensitivity confirmed via audiometric screening (i.e.,  $< 20$  dB HL, octave frequencies 250 - 8000 Hz), were strongly right-handed ( $77.1 \pm 36.4$  laterality index [130]), and had obtained a collegiate level of education ( $17.2 \pm 2.9$  years). None had any history of neuropsychiatric illness. On average, participants had a median of 1.0 year ( $SD=7.5$  years) of formal music training. All were paid for their time and gave informed consent in compliance with a protocol approved by the Institutional Review Board at the University of Memphis. Figure 10 (A, B) shows the distribution of demographic measures (gender and age) of participants.

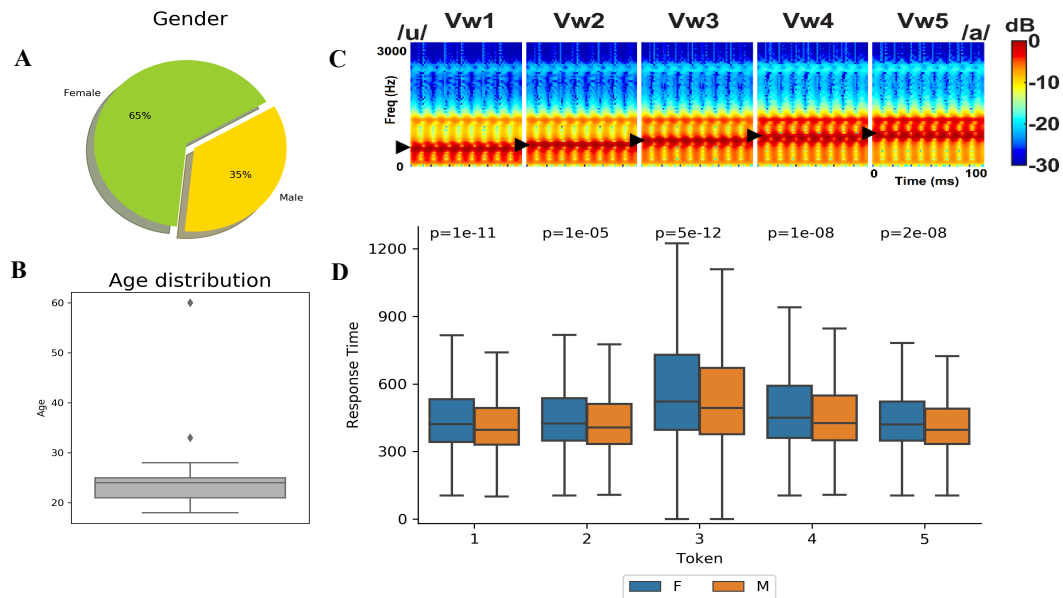


Figure 10: (A, B) Demographic gender and age distributions. (C) Acoustic spectrograms of the speech stimuli: The stimulus continuum was created by parametrically changing vowel first formant frequency over five equal steps from 430 to 730 Hz (►), resulting in a perceptual-phonetic continuum from /u/ to /a/. (D) Token wise response times for auditory classification. Listeners are slower to label sounds near the categorical boundary (i.e., Token 3). Females (F) have significantly slower response times than males (M).

### *Speech stimulus continuum and behavioral task*

We used a synthetic five-step vowel continuum previously used to investigate the neural correlates of CP [130] (Figure 10C). Each token of the continuum was separated by equidistant steps acoustically based on first formant frequency (F1) yet was perceived categorically from /u/ to /a/. Tokens were 100 ms, including 10 ms of rise/fall time to reduce spectral splatter in the stimuli. Each contained an identical voice fundamental (F0), second (F2), and third formant (F3) frequencies (F0: 150, F2: 1090, and F3: 2350 Hz). The F1 was parameterized over five equal steps between 430 and 730 Hz such that the resultant stimulus set spanned a perceptual phonetic continuum from /u/ to /a/ [43]. Speech stimuli were delivered binaurally at 83 dB SPL through shielded insert earphones (ER-2; Etymotic Research) coupled to a TDT RP2 processor (Tucker Davis Technologies).

During EEG recording, listeners heard 150-200 trials of each individual speech token. On each trial, they were asked to label the sound with a binary response (“u” or “a”) as quickly and accurately as possible (speeded classification task). Reaction times (RTs) were logged, calculated as the timing difference between stimulus onset and listeners’ behavioral response. Following their keypress, the inter-stimulus interval (ISI) was jittered randomly between 800 and 1000 ms (20 ms steps, uniform distribution), and the next trial was commenced.

Our speech categorization task requires listeners to make a binary judgment on what they hear. As such, it is a subjective task that does not have true accuracy, *per se*. Consequently, we chose to decode RTs since they are a continuous, more objective measure that provides much richer decoding of listeners’ behavioral decision.

### *Behavioral data analysis*

We adopted classical Gaussian mixture model (GMM) with expectation-maximization (EM) to identify an optimal number of clusters (i.e., subgroups of listeners) from the



distribution of their RT speeds (see Figure 10D). GMMs are probabilistic models that assume the data are generated from a mixture of a finite number of Gaussian distributions (components) with unknown parameters. Mixture models generalize k-means clustering to incorporate information about the covariance structure of the data as well as the centers of the latent Gaussians. Unlike Bayesian procedures, such inferences are prior-free. However, finding an optimal number of components is challenging. The Bayesian Information Criterion (BIC) can be used to select the number of components in a GMM if data is generated from an independent and identically distributed mixture of Gaussian distributions. In this study, we used brute-force and BIC based approaches as an alternative solution to the Variational Bayesian Gaussian mixture model. In this exhaustive parameter search, the hyperparameters were (1) Number of components (clusters), (ranges from 1 to 14); (2) Type of covariance parameters ('full': each component has its own general covariance matrix; 'tied': all components share the same general covariance matrix; 'diag': each component has its own diagonal covariance matrix; or 'spherical': each component has its own single variance). This identified an optimal combination of four components with the unique covariance matrix.

Figure 11A shows the BIC scores while tuning parameters. The '\*' indicates the optimal combination of components. The probability of each component (see Figure 11B) shows that most trials fall into components 1- 3 ranging from 17% - 47% of the total trials in the speech identification task. Component 4 has the fewest number of trials (1.6%). Based on the interpretation of RTs, we categorized these components as Fast RT (Cluster 2, 120 - 476 ms), Medium RT (Cluster 3, 478 - 722 ms), Slow RT (Cluster 1, 724 -1430 ms), and Outliers (Cluster 4, 1432 - 2500 ms). The outliers (Cluster 4) were discarded for further analysis, given the low trial counts loading into this cluster. The boxplot in Figure 11C shows token wise response times. Each speech token can be broken down into a combination of the three RT clusters, meaning that speech categorization speeds could be objectively clustered into

fast, medium, slow (and outliers) responses via the GMM. These cluster divisions were then used in subsequent EEG analyses to determine if functional brain connectomics differentiated these subgroups of CP performers.

### *EEG recording and preprocessing*

**Recording and preprocessing:** EEG recording procedures were identical to our previous neuroimaging studies on CP (e.g., Bidelman et al., 2013; Bidelman and Alain, 2015; Bidelman and Walker, 2017). Briefly, neuroelectric activity was recorded from 64 sintered Ag/AgCl electrodes at standard 10-10 locations around the scalp [131]. Continuous data were digitized using a sampling rate of 500 Hz (SynAmps RT amplifiers; Compumedics Neuroscan) and an online passband of DC-200 Hz. Electrodes placed on the outer canthi of the eyes and the superior and inferior orbit monitored ocular movements. Contact impedances were maintained  $< 10 \text{ k}\Omega$  during data collection. During acquisition, electrodes were referenced to an additional sensor placed  $\sim 1 \text{ cm}$  posterior to the Cz channel.

Subsequent pre-processing was performed in BESA® Research (v7) (BESA, GmbH). Ocular artifacts (saccades and blinks) were first corrected in the continuous EEG using a principal component analysis (PCA) [132]. Cleaned EEGs were then filtered (bandpass: 1-100 Hz; notch filter: 60 Hz), epoched (-200-800 ms)<sup>1</sup> into single trials, baseline corrected to

---

<sup>1</sup> To measure functional connectivity, the epoch window was set wide enough (-200 to 800 ms) to include all sensory (auditory), post-perceptual (linguistic), and response (motor) ERP components relevant to our speech identification task. Therefore, the late endpoint of the analysis window included task-relevant responses, which is likely why we see parietal, motor, and even prefrontal regions that define the CP network (see Figure 15). We did not limit our search analysis window because we explicitly wanted to maintain the richness of the data and decode any and all task-relevant nodes of the brain without *a priori* biases to isolate auditory, language, or motor components, *per se*.

the pre-stimulus interval and re-referenced to the common average of the scalp. This resulted in between 750 and 1000 single trials of EEG data per subject (i.e., 150-200 trials per speech token).

**Source analysis:** Following our previous neuroimaging studies on speech processing [133], [134], we performed a distributed source analysis to more directly assess the neural generators underlying behavioral decisions related to CP. Source reconstruction was implemented in the MATLAB package Brainstorm [135]. We used a realistic, boundary element model (BEM) volume conductor [136], [137] standardized to the MNI template brain [138]<sup>2</sup>. The BEM head model was created using the OpenMEEG [141] as implemented in Brainstorm on the MNI template brain [135]. A BEM is less prone to spatial errors than other head models (e.g., concentric spherical conductor) [137]. The sLORETA allowed us to estimate the distributed neuronal current density underlying the measured sensor data. The resulting activation maps (akin to fMRI) represent the transcranial current source density underlying the scalp-recorded potentials as seen from the cortical surface. We used the default settings in Brainstorm’s implementation of sLORETA [135]. The sLORETA provides a smoothness constraint that ensures the estimated current changes little between neighboring neural populations [132], [142]. This method is better than other inverse solutions because of its smaller average localization error. While higher channel counts improve source

---

<sup>2</sup> Spatial accuracy of inverse source modeling from EEG can be improved by incorporating MRIs and electrode digitization at the single subject level. Our source reconstruction pipeline was applied uniformly across listeners so our use of template brain anatomies is thus a source of noise in our data. While a template brain was expected to reduce the absolute precision of localization by  $\sim 5$  mm [139] this error was uniform across individuals and critically, much smaller than the distance between the broad ROIs of the DK atlas which we aimed to localize. Indeed, source localization from *macroscopic* brain structures is not necessarily improved by individual anatomical constraints [140].

localization, for a 64-ch electrode array as used here, best-case estimates of localization error for sLORETA are as low  $\sim 1$  mm [143].

From each single-trial sLORETA map, we extracted the time-course of source activity within 68 regions of interest (ROI) defined by the Desikan-Killany Atlas parcellation [144] as implemented in Brainstorm. Single-trial source waveforms (derived per subject and speech token) were then submitted to functional connectivity analyses. We have recently used a similar approach to successfully decode single-trial EEG and predict individual differences in other cognitive domains (e.g., working memory capacity [145]), motivating its use here.

#### *EEG functional connectivity and graph analyses*

**Bootstrapping:** Functional connectivity measures are more accurate when calculated using source localized compared to scalp-recorded (sensor-level) EEG [146]. Still, to ensure the robustness of our connectivity measures, we used bootstrapping to reduce the uncertainty of our connectivity estimates [147]. This method involved repeatedly taking small samples with replacement, calculating the statistics, and averaging over the calculated statistics. We applied a mean based bootstrap approach on 35106 trials. For each RT class, 100 random trials from each individual participant were chosen as a bootstrap sample (with replacement). We calculated the mean source amplitude in each of the 68 ROIs for each bootstrap sample. This process was then iterated 30 times to derive the final estimate of the mean source signal in each ROI. Overall, 3150 trials were generated (1050 trials of each RT class) in this process for further analysis.

**Functional connectivity:** A graph network is defined by a collection of nodes (vertices) and links (edges) between pairs of nodes. Nodes in large-scale brain networks usually represent brain regions (ROIs), while links represent anatomical, functional, or effective connections [148]. Anatomical connections typically correspond to white matter tracts

between pairs of brain regions. However, functional connections correspond to the strength of temporal correlations between pairs of anatomically connected/unconnected regions.

Depending on the measure, functional connectivity may reflect linear or nonlinear interactions, as well as interactions at different time scales [149]. Popular approaches to quantify functional connectivity are Correlation, Coherence (CH), imaginary part of coherency (iCH), Phase Locked Value (PLV), Phase Slope Index (PSI) [150]–[152]. A comprehensive comparison of these methods showed that correlation-based connectivity outperformed the others in classifying behavioral RTs (see appendix for details).

We measured pair-wise Pearson product-moment correlation coefficients among the 68 brain regions (ROIs). This resulted in connectivity matrix describing the weighted strength (undirected network) between all pairwise nodes ( ${}^{68}C_2 = 2278$  edges) for each trial. Diagonal and upper diagonal elements of the connectivity matrices were discarded to avoid spurious self and repeated connectivity. Matrices were then concatenated to a vector to describe the connectivity across all brain nodes and trials (e.g.,  $3150 \times 2278$ ) for each participant.

Seven global network connectivity features were estimated from each network graph using the BCT toolbox [4]: (i) Characteristics path, (ii) Global efficiency, (iii) Average clustering coefficient, (iv) Transitivity, (v) Small-worldness, (vi) Assortativity coefficient, and (vii) Maximized modularity (see Appendix for mathematical definitions and interpretation of these network features).

### ***Identifying behaviorally relevant aspects of functional connectivity***

To visualize the data, we used the t-distributed stochastic neighbor embedding (t-SNE) [36] similar to the previous chapter. The hyperparameters of t-SNE were tuned with a grid search approach. Figure 12 shows the t-SNE embedded scatter and kernel density estimation (KDE) plot of our data distribution. The t-SNE visualization confirms three nearly distinct clusters of functional connectivity for the different RT groups in speech categorization. Unrelated or noisy edges may exist in the higher dimensional functional

connectivity matrices. This undesired contamination necessitates the use of feature selection methods to choose functional connectivity metrics that are relevant and can be modeled robustly over a range of model parameters.

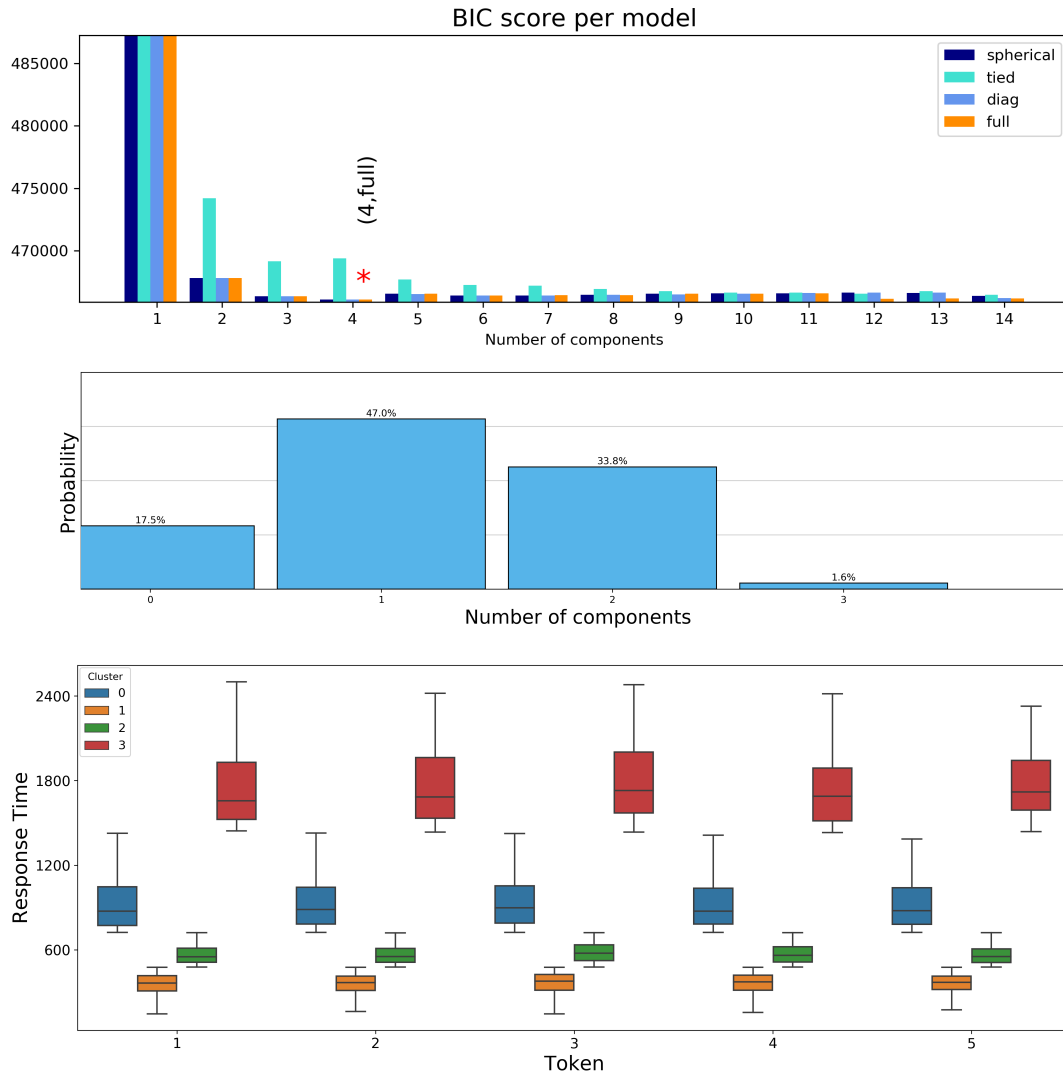


Figure 11: Clustering RT data using GMM and BIC criteria. Model selection concerns both the covariance type and number of components in the model. Brute-force based empirical analysis shows that  $n=4$  components with unique covariance matrix is optimal. The '\*' marked position of (A) shows the optimal combination. (B): Probability of trials loading into each component. (C): Token wise RT broken down by component. Based on behavioral RTs, four clusters are evident that distinguish subgroups of listeners based on their speech identification speeds: Fast (Cluster 1): 120~476 ms, Medium (Cluster 2): 478~722 ms, Slow (Cluster 0): 724~1430 ms, and Outliers (Cluster 3): 1432~2500 ms.

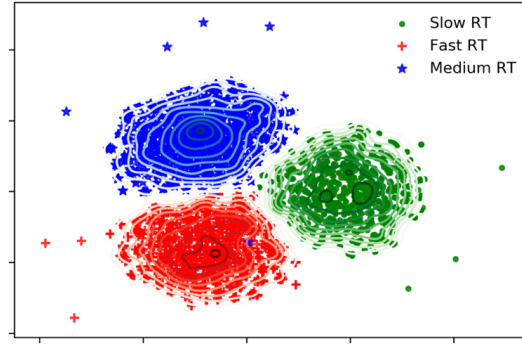


Figure 12: The t-SNE embedded higher dimensional functional connectivity data are represented by a 2-dimensional scatter and kernel density estimation (KDE) plot. The green lines with '.', blue lines with '\*', and red lines with '+' sign represents data points for slow, medium, and fast RT participants, respectively.

### ***Feature selection***

For feature selection, we used Randomized Logistic Regression for stability selection with randomized lasso. We considered sample fraction = 0.75, number of resampling = 1000 with tolerance = 0.001. This algorithm assigns feature scores between 0 and 1 based on frequency of selection over 1000 iterations. We needed to specify a threshold score to identify the best representative set of stable features. Hence, a threshold selection is an essential design parameter. We varied different selection thresholds (i.e., the number of selected features) and observed the effect on model performance. Modeling involved four steps:

1. Randomly shuffle and split the dataset into training and test set (80% and 20%),
2. Consider Support Vector Machine with "RBF" kernel as a base estimator,
3. Tune hyperparameter (i.e., C and Gamma) on training data using grid search approach and 10-fold cross-validation, and
4. Select best models evaluated on unseen test data. Accuracy (ACC) and Area Under Curve (AUC) were considered for performance measures,

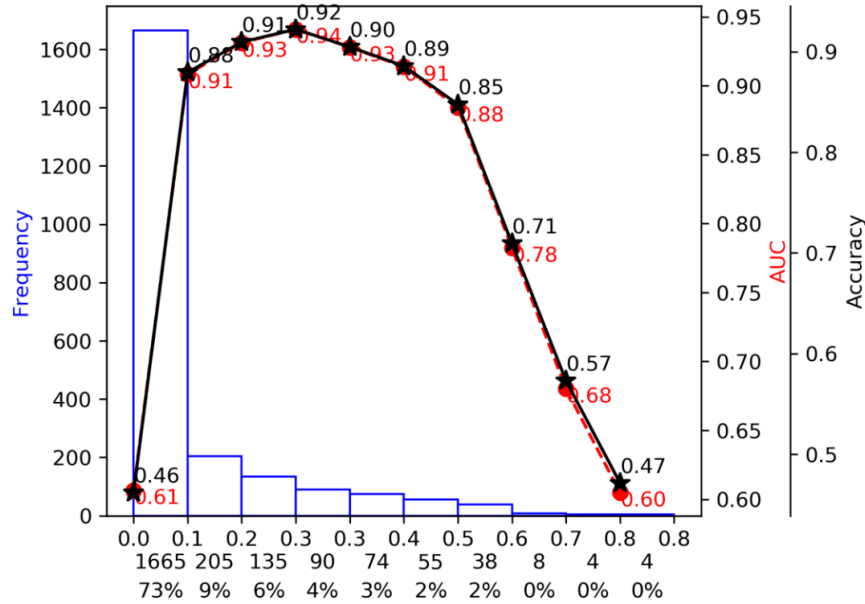


Figure 13: Effect of selection threshold on model performance prediction. The three x-labels represent (top) the range of each bin of features score (range: 0~1), (middle) the number of features falling in each bin, and (bottom) the corresponding percentage.

Figure 13 shows the effect of different selection thresholds on modeling. The histogram illustrates the distribution of the feature score. The first line of the x-axis shows the bin ranges of scores (0 to 1). The second and third lines show the amount and percent of features that had nearly the same score for a specific bin. We found that 73% of the features had scores of 0-0.1, meaning the majority of connectivity measures were not selected even once (i.e., the coefficient was zero) among 1000 model iterations. That is, 73% of functional connectivity metrics explored in our search space were not related to speed of speech categorization (i.e., behavioral RTs).

For a specific selection threshold of 0.26, the algorithm selected 227 edge features that collectively achieved 92% accuracy (best model performance) with an AUC=0.9. The bell-shaped solid black and red dotted lines of Figure 13 show the accuracy and AUC curves for different selection thresholds. Note that selection thresholds higher than the optimal value (0.26) allowed the model to consider more noise variables, degrading model performance



significantly. On the other hand, selection thresholds higher than the optimal value discard behaviorally relevant features and reduce model performance. Table 5 details the effect of the selection threshold on model performance. Here, the number of unique edges represents correlation-based connectivity between two brain nodes (features), and the number of unique nodes represents brain regions associated with those selected edges. Overall, we leveraged different machine learning techniques to address different steps in the data pipeline (i.e., data preprocessing, visualization, feature selection, modeling). The LDA based t-SNE was only used for data visualization. Randomized lasso (Stability selection with L1-penalized logistic regression) was used for feature selection. Moreover, a SVM was used for classification and evaluating the performance of stability selection. Our process, leveraging different techniques for each respective stage data analysis, follows widely used conventions in the EEG related machine learning field [5], [153], [154]. A schematic diagram of the method pipeline is shown in Figure 14.

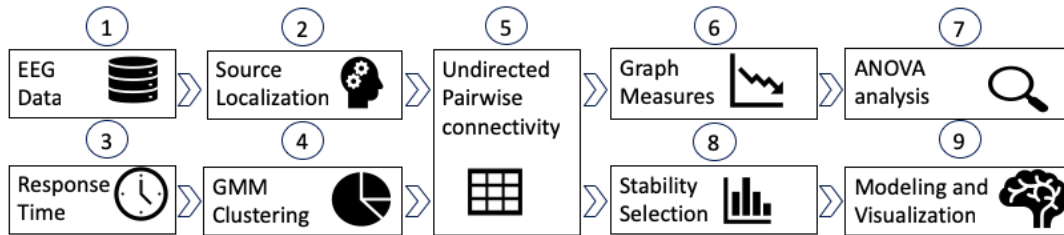


Figure 14: Schematic diagram of the processing pipeline. The 64-ch EEG data is first preprocessed, and then source localization is adapted to convert skull surface data to cortical surface time series data (68 ROIs defined by the Desikan-Killany Atlas parcellation). Pairwise correlations were calculated to derive the connectivity matrix for each trial of the speech CP task. Behavioral response times (RTs) were clustered with Bayesian non-parametric (GMM) clustering. These clusters were labeled as Fast, Medium, and Slow RT. ANOVA analysis of Graph measures was adopted to test significance among RT groups. Stability selection and machine learning approaches were then used to find significant properties of the brain's functional connectivity related to behavioral speeds (RTs) in speech CP.

## ***Results***

Figure 10D shows behavioral results in the speech categorization task. Generally speaking, listeners were slower to label sounds near the categorical boundary (token 3), consistent with the higher ambiguity of the mid-continuum stimuli [42], [43], [45], [46]. On average, females also showed slower RTs than males across the continuum (Welch's t-test;  $p < 0.0001$ ). Bayesian nonparametric clustering revealed four distinct clusters in the speed (RTs) of listeners' CP (Fast: 120~476 ms, Medium: 478~722 ms, Slow: 724~1430 ms, and Outliers: 1432~2500 ms) (Figure 11C). These clusters were even present at the individual token level.

Having established that listeners could be distinguished based on their speed in speech categorization, our next goal was to determine whether network properties of the brain accounted for these behavioral differences. We applied graph theory techniques to construct and analyze the functional brain connectome underlying CP. We considered both individual trials- as well as group-based analyses. For group-based analysis, data were averaged across subjects within each RT cluster. Group means were computed by concatenating group-wise trials and calculating their mean. We then calculated seven global network connectivity features using the BCT toolbox [4] similar to the previous chapter.

We used non-parametric ANOVAs (Kruskal-Wallis H-test) to determine if individual trial-based global graph measures varied across RTs (Table 3). This non-parametric test was used given the unequal sample size per group [155]. These analyses revealed that Assortativity and Global Efficiency were modulated depending on behavior speed. Table 4 shows a comparison of the graph measures across three RT groups. Global efficiency measures were relatively small, and assortativity had a negative tendency. All other network features were not discriminatory among the RT groups. Therefore, modeling with those features (using SVM with 'RBF' kernel) showed expectedly poor accuracy (38%).

Table 3: Significant (bold) global network measures (Kruskal-Wallis H-test tests) (trial-level).

Measures	p-value
Characteristics Path	0.1359
Average Clustering Coefficient	0.8286
Small Worldness	0.0815
<b>Assortativity</b>	<b>0.0052</b>
<b>Global Efficiency</b>	<b>0.0290</b>
Transitivity	0.8424
Maximized Modularity	0.6617

Table 4: Group comparison of graph measures of functional connectivity between RT groups.

Measures	Fast RT	Medium RT	Slow RT
Characteristics Path	0.1473	0.1507	0.1504
Average Clustering Coefficient	0.1327	0.1358	0.1352
Small Worldness	1.1516	1.1522	1.1497
<b>Assortativity</b>	<b>-0.0086</b>	<b>-0.0128</b>	<b>-0.0118</b>
<b>Global Efficiency</b>	<b>0.1909</b>	<b>0.1934</b>	<b>0.1944</b>
Transitivity	0.1329	0.1362	0.1354
Maximized Modularity	0.1872	0.1845	0.1875

Besides analyzing global network properties, we next aimed to identify the most significant properties of functional brain connectivity that were related to behavioral RTs. Functional connectivity for each trial is a high dimensional sparse matrix. Some studies have suggested that properties of functional brain networks are most consistent with the actual brain anatomy when network density is 8% to 16% [156]–[158]. To determine the most behaviorally relevant arrangement of sparse connectivity, we used stability selection with Randomized Lasso to detect and rank the most important, consistent, and relevant functional connectivity measures that were invariant (stable) over a range of model parameters. Stability selection discarded 88% (total 273) of network edges that were not related to behavioral RTs, but still achieved 92% classification accuracy with AUC=0.9. From Table 5, It was observed

that 7% error tolerance from the optimal value (accuracy from 92% to 85%) allowed 80% fewer edges and 22% fewer associated nodes. Hence, the selection threshold 0.51 with reasonable performance (Accuracy=85%, AUC=0.9) were chosen for network visualization as performance declined precipitously above this threshold (as shown in Figure 13).

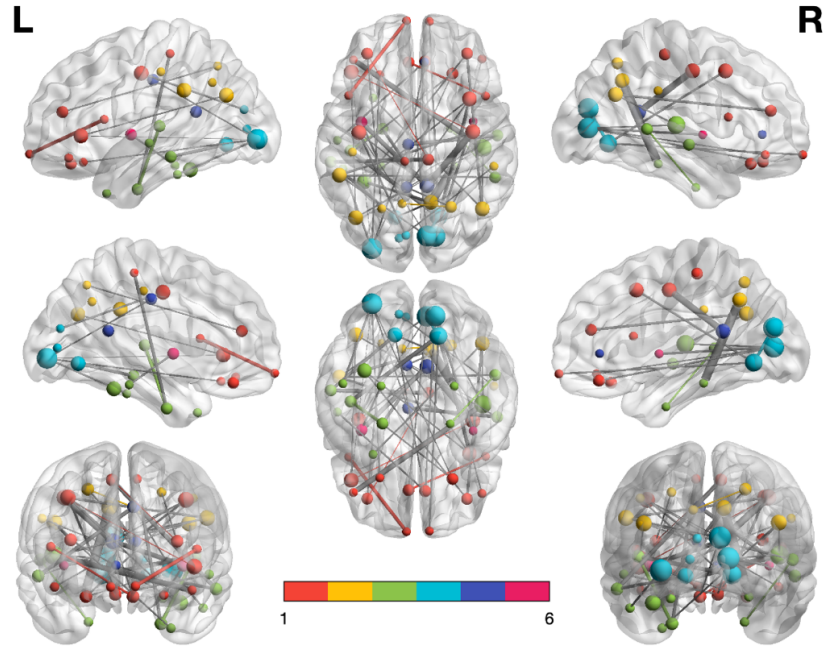


Figure 15: BrainNet visualization (top to bottom: lateral, medial, and dorsal view) of the brain network (54 edges) identified via stability selection. Color map 1-6 indicates, 1: Frontal (22 ROI), 2: Parietal (10 ROI), 3: Temporal (18 ROI), 4: Occipital (8 ROI), 5: Cingulate (8 ROI), 6: Insula (2 ROI) regions. Node size varies with its degree of connectivity. Connectivity among the same lobe are colored with similar node color. Edge widths represent the weight of absolute correlation (connectivity strength).

Figure 15 shows a visualization of the 54 nodes among 53 ROIs identified via stability selection using BrainNet [159]. The resulting network revealed a highly dense connectome reflective of listeners' behavioral RTs in speech categorization. Connectivity was particularly strong between the occipital, parietal, and bilateral frontal lobes. As an additional means of data reduction, we further thresholded ( $\geq 0.68$ ) the stability-selected connectome. This resulted in eight highly ranked connectivity edges among 13 nodes across the brain (shown in Figure 16). Even with this sparse network of only eight edges, model classification was still 57%, meaning this small set of features accurately predicted RTs. We then ranked the

contribution of these stable nodes in Table 6. We found that three edges (rank: 3, 4, and 6) were in left hemisphere, two edges were in the right hemisphere (rank: 2, and 5), and three edges were inter-hemispheric (rank: 1, 7 and 8). Notably, these edges included connections between motor (paracentral), visual (lateral occipital/ lingual), linguistic (left IFG, pars triangularis), auditory (superior temporal gyrus), and parietal areas both within and between hemispheres.

Table 5: Effect of selection threshold of stability selection (Threshold) on model performance. The pairwise correlation between two brain regions (functional connectivity edge) were considered as features. The number of unique nodes are the brain regions associated with selected features. ACC, accuracy; AUC, area under curve.

Threshold	ACC	AUC	Number of Unique Edges	Number of Unique Nodes
0	46%	0.6	2278	68
0.08	88%	0.9	613	68
0.17	91%	0.9	408	68
0.26	92%	0.9	273	68
0.34	90%	0.9	183	68
0.42	89%	0.9	109	64
0.51	85%	0.9	54	53
0.59	71%	0.8	16	24
0.68	57%	0.7	8	13
0.76	47%	0.6	4	8

## ***Discussion***

The present study evaluated whether individual differences in a core operation of speech and language function (i.e., categorization) could be explained in terms of network-level descriptions of brain activity. By applying machine learning classification techniques to functional connectivity data derived from EEG, our data show that the speed of listeners' ability to categorize and properly label speech sounds is directly related to dynamic variations in their brain connectomics.

It has been suggested that important cognitive functions are supported by distributed neural networks with highly segregated and integrated “small-world” organizations or clusters [1]–[3], [60]. However, in relation to distinguishing listeners’ perceptual speed for categorized speech, we did not find differences in network properties of Characteristics Path, Average Clustering Coefficient, Small Worldness, Transitivity, and Maximized Modularity clearly indicates (Table 3). Instead, global network assortativity and efficiency distinguished fast, medium, and slow RT individuals. In network science, assortativity refers to the tendency of “like to connect with like.” That is, at the macroscopic level, high degree nodes attach to other high degree nodes and similarly, low to low [152]. In our study, functional brain networks were defined via task-based co-activations. Hence, they were expected to exhibit some assortativity as co-activation means that regions of the network were engaged by the same task. Previous studies have shown that the property of assortative tendency changes with task demands [160]. The resting state brain functional network is largely assortative. Higher order association areas exhibit non-assortative organization tendency and form periphery-core topologies. However, assortative structures break down during tasks and is supplanted by periphery, core, and disassortative communities.

In addition, we found that the functional CP network underlying speed of decisions increased in negative assortativity (i.e., became disassortative) for slower RTs. This indicates that brain nodes were more likely to connect with nodes having different degree during slower RTs, implying that important hubs of the CP network communicated with insignificant hubs during states of slower decisions. Based on the interpretation of these graph metrics (see Appendix), we infer that slower, more taxing categorical speech decisions cause excessive use of neural resources and more aberrant information flow within the CP circuitry. Supporting this interpretation, we found that network utilization (Global efficiency) also differentiated RT groups. Higher Global efficiency indicates that the routing of information among nodes with different degrees was significantly higher for slow RT trials.

In short, we find that slower responders tended to utilize functional brain networks excessively (or inappropriately) whereas fast responders (with lower global efficiency) utilized the same neural pathways but with more restricted organization. Presumably, these dynamic changes in brain connectivity account for the variations in RTs we find during speech categorization at the behavioral level (Figure 10D).

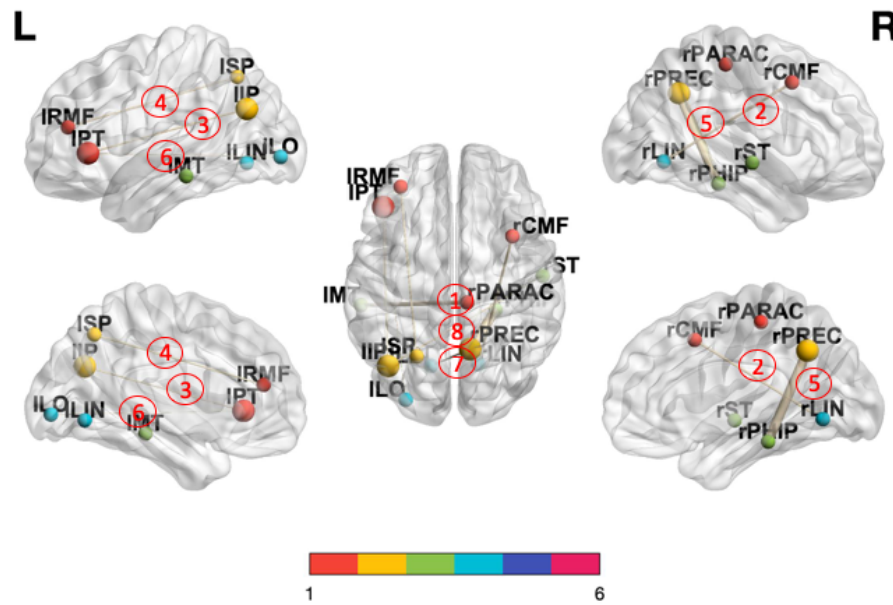


Figure 16: A sparse brain network (8 edges) predicts listeners' speed (RTs) of speech categorization (57% model accuracy). Red numbers are the ranked importance of the edges describing behavior. Otherwise, as in Figure 15.

Our data show that global graph measures fail to fully explain the behavioral relevance of important connectivity edges. We observed that the functional connectivity matrix underlying speech CP is highly sparse and dynamic. Indeed, only ~12% of all possible edges in the Desikan-Killany Atlas was needed to explain variation in behavioral RTs. In this vein, we adopted stability selection to find edges that were most consistent in distinguishing different network states related to perception. By performing this two-stage randomization iteratively (e.g., 1000 bootstraps), stability selection with randomized lasso assigned high

scores to features that were repeatedly selected across randomizations, yielding the most meaningful connections within the CP connectome that describes behavior.

Table 6: Eight most important edges that govern speeded speech classification. Collectively, these edges achieve a model accuracy of 57% in segregating listeners' speeded decisions (RTs) in the perceptual task. Here, a score of 0.85 means that out of 1000 iterations, the edge was selected by stability selection 850 times.

Edge	Score	Rank
Paracentral R-Middletemporal L	0.85	1
Lingual R-Caudalmiddlefrontal R	0.845	2
Parstriangularis L-Inferiorparietal L	0.785	3
Superiorparietal L-Rostralmiddlefrontal L	0.785	4
Precuneus R-Parahippocampal R	0.725	5
Parstriangularis L-Lateraloccipital L	0.705	6
Precuneus R-Lingual L	0.705	7
Superiortemporal R-Inferiorparietal L	0.695	8

Collectively, our results showed that neural classifiers (SVM) coupled with stability selection could correctly classify behavioral RTs related to CP from functional connectivity alone with over 90% accuracy (AUC=0.9). The resulting edges composing the RT-related networks were distributed in both hemispheres, and both intra- and inter-hemispheric edges were evident. More interestingly, we found that only eight edges among 13 ROIs were needed to distinguish RTs well above chance. ROIs composing this sparse but behaviorally-relevant network included (1) Caudalmiddlefrontal R, (2) Inferiorparietal L, (3) Lateraloccipital L, (4) Lingual L, (5) Lingual R, (6) Middletemporal L, (7) Paracentral R, (8) Parahippocampal R, (9) Parstriangularis L, (10) Precuneus R, (11) Rostralmiddlefrontal L, (12) Superiorparietal L, (13) and Superiortemporal R. Previous neuroimaging studies have demonstrated a distributed fronto-temporo-parietal neural network supporting auditory categorization (e.g., [44], [45], [61], [63], [64], [66], [161]–[165]). Our data corroborate these previous studies by confirming engagement of similar temporal (STG), parietal, motor, and



prefrontal regions in CP using an entirely data-driven approach (machine learning with stability selection).

Notably, we found functional connectivity between right paracentral and left middletemporal gyrus (MTG) was the most important connection describing the speed of behavioral CP (Table 6). MTG has been associated with accessing word meaning while reading [166] and has been described as an early lexical interface that is heavily involved in sound-to-meaning inference [167], [168]. Some studies indicate that lesions of the posterior region of the middle temporal gyrus, in the left cerebral hemisphere, may result in certain forms of alexia and agraphia [169], indicating its role in the language production network [170]. The strong link between MTG and paracentral gyrus implies a direct pathway between the neural substrates that map sounds to meaning and sensorimotor regions that execute the motor command and therefore govern response speeds (indexed by RTs). The leftward laterality of the MTG node is consistent with the left lateralized nature of language processing in the brain. Still, why left MTG so strongly interfaces with right motor areas in our data is unclear, especially given the right-handedness of our participants and expected left (contralateral) motor involvement. Differences in brain connectivity have been observed between sexes [171], and females may have a more diffuse, bilateral neural system for language processing than males [172]. Speculatively, the strong communication between left linguistic (MTG) and right motor brain areas we find may reflect the higher preponderance of females in our sample.

Relatedly, stability selection identified the second-ranked edge between lingual and caudal-middlefrontal gyrus. While the functional role of lingual (occipital) gyrus in speech processing is not apparent *prima facie*, this region is involved in visual word processing, especially letters [173]. It has also implicated in stimulus naming [174], [175], an operation at the core of our speech categorization (i.e., sound labeling) task. We also found a third ranked edge predictive of behavioral CP between parstriangularis and inferior parietal cortex.

Previous functional neuroimaging and connectivity studies have shown strong engagement of frontal-parietal networks during CP [45], [164], [165]. Our results corroborate these findings by similarly implicating a prominent interface between linguistic (IFG) and parietal (IPL) brain regions in modulating the speed of listeners' categorical decisions. Indeed, decision loads IFG during effortful speech listening [63], [176], [177] and the IFG-IPL pathway is upregulated when speech material is perceptually confusable [164]. Therefore, the network organization of brain connectivity observed for slow RTs and importance of IFG-IPL in describing behavior may reflect a similar state of perceptual confusion during rapid categorical speech labeling.

One limitation of our study was that our sample contained more females than males (2:1 ratio). This is relevant since RTs were differed among genders (Figure 10D). Thus, a natural question that emerges from our data is the degree to which our machine learning techniques segregated data based on gender rather than different RTs (i.e., fast vs. slow responders), per se. Still, this is probably not the case. Conventional filter-based group analysis can bias classification and feature selection results, whereas with our Lasso-based bootstrapped analysis, this becomes less likely [178]. Moreover, stability selection with randomized lasso is a similar but more robust approach that produces consistent variable selection with minimal bias. Hence, the impact of our unbalanced sample size on feature selection is probably negligible.

Taken together, our novel approach to neuroimaging data demonstrates the derivation of small, yet highly meaningful patterns of brain connectivity that dictate speech behaviors using solely EEG. More broadly, the functional connectivity and machine learning techniques used here could be deployed in future studies to identify the most meaningful changes in spatiotemporal brain activity that are modulated by development, normal learning, or those which decline in neuropathological states.

## ***Conclusion***

We developed an efficient computational framework to investigate whether individual differences in speed of speech categorization can be decoded from network-level descriptions of brain activity. We adopted appropriate best practices in machine learning and data analysis to visualize very noisy high dimensional data using a combination of supervised and unsupervised techniques to understand the embedding and linear separability of the data. We further used stability selection to determine the set of features over a range of model parameters. This is critical for interpretation and validation and identifying unique states of functional brain connectivity. Our EEG data-driven approach reveals that the speed of listeners' ability to categorize and properly label speech sounds is directly related to dynamic variations in their brain connectomics. These findings contribute in several ways to our understanding of how the brain works in categorical perception and provide a basis for further research. In future iterations of the work, we plan to improve our approach by including directional and dynamic connectivity analysis to better delineate the temporal emergence of the phenomena observed here.

## ***Appendix***

### **Connectivity matrix**

We calculated Correlation, Coherence (CH), Imaginary coherence (iCH), phase locked value (PLV) to construct undirected graph matrices. Phase Slope Index (PSI) worked better than conventional directed graph matrix measure (e.g., Granger causality) [179]. Therefore, PSI is considered as directed graph matrix for directed graph analysis. Here, the spectral densities were estimated using a “multitaper” method with digital prolate spheroidal sequence (DPSS) windows and a discrete Fourier transform with Hanning windows.

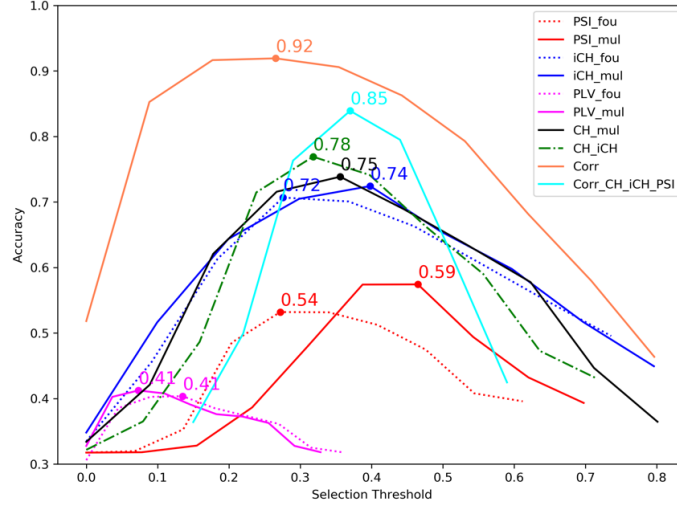


Figure 17: Accuracy curves of stability selection (as in Figure 13). Stability selection was applied to Correlation, CH, iCH, PLV PSI based-edge matrix, as well as combinations of CH and iCH, combination of correlation, CH, iCH, and PSI based-edge matrix. Here ‘mul’ and ‘fou’ represents multitaper and Fourier transform methods. The dot point of each accuracy curve indicates maximum accuracy of the optimal combination of features. Correlation-based connectivity outperforms all other measures.

We considered average connectivity scores for each frequency band. Let,  $S_{xy}$  is cross-spectral densities and  $S_{xx}$ ,  $S_{yy}$  is the auto spectral density of x and y respectively. Coherence is calculated using this equation:

$$CH = \frac{E(S_{xy})}{\sqrt{E(S_{xx}) * E(S_{yy})}} \quad 12$$

The equation of Imaginary coherence [151] is given by:

$$iCH = \frac{I_m(E(S_{xy}))}{\sqrt{E(S_{xx}) * E(S_{yy})}} \quad 13$$

The equation of Phase-Locking Value [150] is given by:

$$PLV = |E[S_{xy}/|S_{xy}|]| \quad 14$$

Each line in Figure 17 represents the effect of the selection threshold over classification accuracy. A higher threshold value selects a fewer number of features. For a specific selection threshold, we found that correlation-based connectivity out-performed CH, iCH, PLV, and PSI in segregating speech RTs. Why correlation works better in classifying

behavioral RTs is an empirical question that needs to be further evaluated in future signal processing studies. Based on results from our empirical comparisons, we adopted correlation-based connectivity throughout the remainder of the study.

## Figure

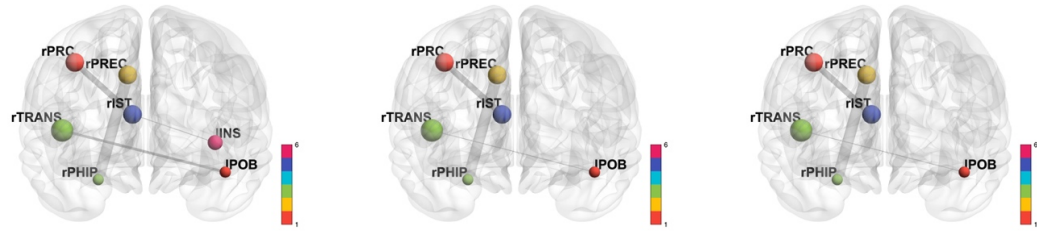


Figure 18: Brain network underlying Slow RT listeners (left), Medium RT listeners (middle), and Fast RT listeners. Shown here are the most highly correlated (absolute correlation  $\geq 0.5$ ) network edges. Otherwise as in Figs. 6-7. INS, insula; IST, isthmus of cingulate; TRANS, transverse temporal gyrus (auditory cortex); POB, pars orbitalis; PRC, precentral gyrus (motor cortex); PHIP, parahippocampal gyrus; PREC, precunus; l/r, left/right hemisphere.

## Chapter 5 - Dynamic Network Analysis

In this chapter, we perform dynamic network analysis to discover temporal patterns in categoric perception. Specifically, DEAP [180] and CP dataset (used in the previous chapter) is used here to analyze affect and speech perception.

### **Dataset**

#### *DEAP dataset*

In DEAP (A Database for Emotion Analysis using Physiological Signals) dataset [3], thirty-two healthy participants participated in experiment where male-female ratio was 50% and aged between 19 and 37 (mean age=26.9). Forty selected music video clips were used as the visual stimuli to elicit participant's emotions. Participants EEG and peripheral physiological signals were recorded using Biosemi Active System. They rated video stimuli on a discrete 9-point likert scale (Valence, Arousal, Dominance, and Like rating) using Self-Assessment Manikins during the experiment. The data was recorded in two separate geographical locations. Twenty-two participants were recorded in Twente, Netherlands, and rest of the participants in Geneva, Switzerland. We considered only the 22 Dutch participant's data to keep consistency.

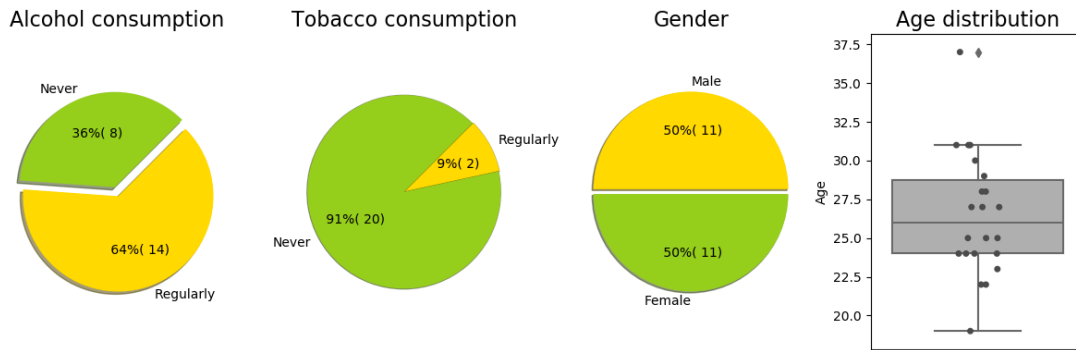


Figure 19: Diversity in the DEAP dataset with different phenotypes.

Overall, 880 (22\*40) trials were taken for further analysis. The 32 Channel EEG data was first preprocessed, and the source estimation was adopted to convert skull surface time-series data to cortical (68 ROI, Desikan-Killiany template) surface time series data. An optimal 4

Sec, 25% overlapping rolling window was used to chunk each trial. The pairwise correlation matrix, clustering-based approach was applied to find micro-states. Exploratory data analysis was used for further interpretation. The schematic diagram of the processing pipeline is shown in Figure 22. The 32 Channel EEG data were preprocessed with several steps by applying:

1. Notch filter (60, 120, and 180 Hz): Notch filters were used for removing well-identified contaminations from systems oscillating,
2. Band-pass: (0.5-60 Hz),
3. Detect and remove eye blinks using EOG channels and SSP (Signal-Space Projection)
4. Import EEG stimulus events, and
5. DC offset correction (-3 to -0.002 sec.).

Reconstructing the activity of the brain from EEG recordings involves several sophisticated steps. The Brainstorm [135] simplifies the procedures. We need to consider two distinct modeling problems to estimated cerebral currents from the EEG recordings:

1. Modeling of the electromagnetic properties of the head and the sensor array (head model or forward model), and
2. The estimation of the brain sources which produced the data, according to a specific head model. That second step is known as source modeling or solving an inverse problem.

Overall, we followed four steps:

- a. Sensors/MRI registration,
- b. Compute covariance (noise or data),
- c. Compute source (using sLORETA),
- d. Recording ROI (Desikan-Killiany template with 68 regions) based mean time series.

The cortical template and its corresponding EEG electrode location are shown in Figure 20.

### *CP dataset*

A description of this dataset was given in the previous chapter. In this chapter, instead of applying bootstrap based approach, we averaged individual RT wise data. Therefore, from 35 participants, we got  $(35 \times 3)$  105 trails for further analysis.

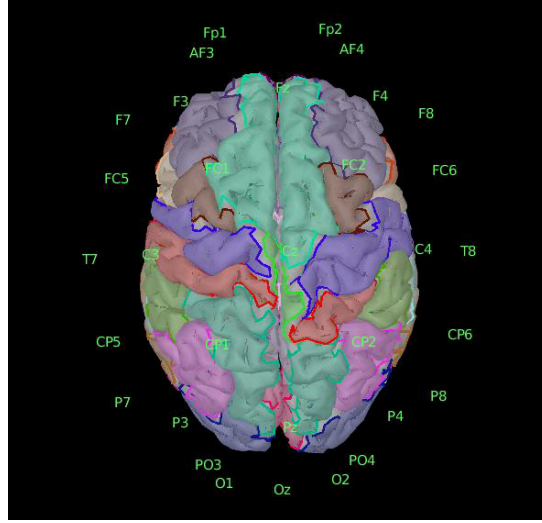


Figure 20: EEG electrode locations and cortical view of Desikan-Killiany ROI template.

### ***Adaptive segmentation based dynamic network analysis in understanding affect***

In this dissertation, the DEAP dataset and adaptive segmentation based dynamic network is used to understand dynamic properties of affect. Here, time-varying change among brain node time courses was captured by calculating cross-correlations between brain networks over time using a tapered windowing. Here, the cortical based time series data was z-score normalized, and a sliding window with 25% overlap was used. The pairwise correlation for each time window was then calculated. Correlation-based connectivity is a non-directional measure; hence  ${}^{68}C_2 = 2278$  data points were calculated from each time window. Then k-means clustering was applied to those data to find the micro-states. We perform an empirical analysis to find an optimal window size using the elbow method. We varied window size (2, 4, 6, 8, 10 seconds) as well as the number of the cluster (2 to 60) and measured the distortion score (the sum of squared distances from each point to its assigned center).



Figure 21 shows that the effect of window size is very negligible for 4 and 10 seconds. In this study, we consider 4-second as the optimal window size for further analysis.

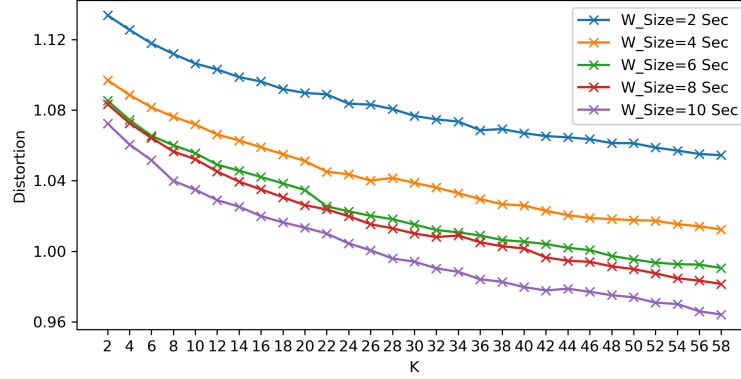


Figure 21: Estimating optimal window size for micro-states. The X and Y axis represents number of clusters and distortion score, respectively. The distortion score does not differ much for the different window size. However, optimal “elbow” point is not evident due to the noisy nature of cortical surface data.

### *K-means clustering*

Mainly K-means clustering with the “elbow” method was used to identify the optimal number of clusters. The centroids of each cluster (micro-states) can be thought of as average patterns that subjects tend to return to during the experiment. The Gaussian mixture model (GMM) is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. GMM is generalizing K-means clustering to incorporate information about the covariance structure of the data as well as the centers of the latent Gaussians. Besides the K-means approach, we also used GMM with the expectation-maximization (EM) algorithm for clustering. Hidden Markov Model with Gaussian emissions was adopted to find the transition probability among the micro-states.

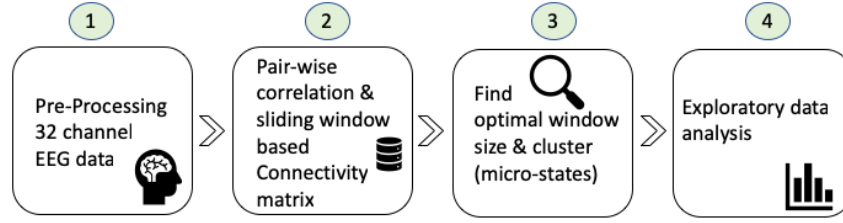


Figure 22: Schematic diagram of the processing pipeline for DEAP dataset. The 32 Channel EEG data is first preprocessed, and the source localization is adopted to convert skull surface data to cortical surface time series data. An optimal 4 Sec, 25% overlapping rolling window is used to chunk each trial. Pair-wise correlation matrix, clustering-based approach is used to find micro-states. Exploratory data analysis is used for further interpretation.

### Graph mining

The undirected and weighted correlation-based adjacency connectivity matrix is then used to calculate seven basic global network features as before. The features are: (i) Characteristic path, (ii) Global efficiency, (iii) Average clustering coefficient, (iv) Transitivity, (v) Small-worldness, (vi) Assortativity coefficient, and (vii) Modularity

### Empirical analysis on DEAP dataset

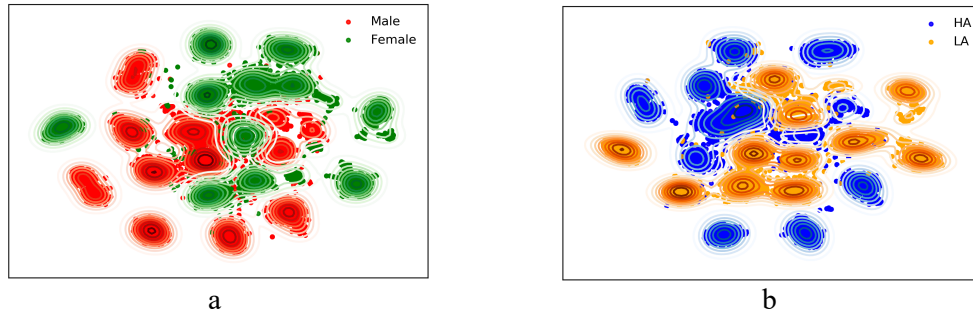


Figure 23: The t-SNE embedded higher dimensional features are represented by 2-dimensional scatter and kernel density estimation (KDE) plot. The green and red lines with dots (figure a) represent normal male and female connectivity data points. Similarly, blue and orange data points (figure b) represent high and low arousal. Both figures show a distinct clustered pattern.

Figure 23 shows the t-SNE embedded scatter and kernel density estimation (KDE) plot of our 4-sec connectivity data distribution. KDE plot is a non-parametric way to represent the probability density function. Besides scatter plot, KDE plot is used here to visualize the trend of data distribution for each different class. In Figure 23a, the green and red dots represent data points for male and female connectivity. Similarly, Figure 23b shows the connectivity pattern for high and low arousal. The t-SNE with PCA and LDA approach shows two distinct state clusters among groups.

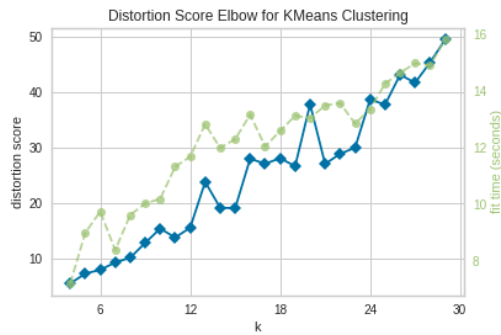
Our next goal was to identify connectivity states. We applied K-means clustering. From Figure 21, it was observed that there was no specific 'elbow' point present in any time window data. Figure 24 shows more details about how we select the optimal number of clusters (23) by fitting the model with a range of values for K. The elbow method runs K-means clustering on the dataset for a range of values for K (say from 2-30) and then for each value of K computes an average score for all clusters. Here, the distortion score is the sum of squared distances from each point to its assigned center. The dashed green line of Figure 24a displays the amount of time needed to train the clustering model, and blue lines indicate the distortion score per K. It was observed that with the increase of 'K', the distortion score tended to increase, hence we were not able to find specific 'elbow' point. Based on the visualization from Figure 24, we considered  $K=23$ .

Figure 24b, Figure 24c, and Figure 24d shows: (i) Inter-cluster distance maps (ii) Silhouette analysis, and (iii) Cluster probability visualization, respectively. Inter-cluster distance maps display an embedding of the cluster centers in 2-dimensions with the distance to other centers preserved. E.g., the closer to centers are in the visualization, the closer they are in the original feature space. The clusters are sized according to a scoring metric. By default, they are sized by the membership, e.g., the number of instances that belong to each center. This gives a sense of the relative importance of clusters. However, because two clusters overlap in the 2D space, it does not imply that they overlap in the original feature space. We used Multi-Dimensional Scaling (MDS) to reduce the feature dimension for visualization purposes.

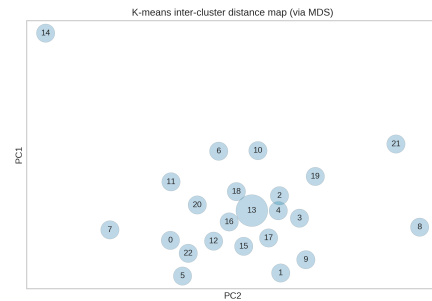
On the other hand, silhouette analysis can be used to study the separation distance between the resulting clusters. The silhouette plot displays the measure of how close each point in one cluster is to points in the neighboring clusters and thus provides a way to assess parameters like the number of clusters visually. This measure has a range of  $[-1; +1]$ . Also, from the thickness of the silhouette plot the cluster size can be visualized. The negative

silhouette coefficients indicate clusters are not well balanced. The cluster probability of Figure 24d show the probability of each cluster. For a sanity check and as a comparative study of clustering, we adopted GMM based method. The information-theoretic criteria (BIC) criterion was used to select the number of components in a classical Gaussian Mixture Model (unlike Bayesian procedures, such inferences are prior-free). In theory, GMM recovers the actual number of components only in the asymptotic regime (i.e., if much data is available and assuming that the data was generated from a mixture of Gaussian distributions). Here, we estimated model parameters with the EM algorithm for different covariance types (spherical, tied, diag, full). It was observed that for the best number of components = 27 and different covariance types, the measures are 'tied' (the winner is indicated with red '\*' in Figure 25).

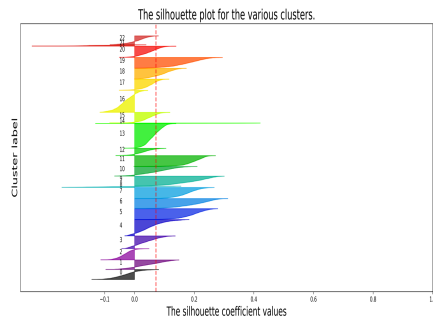
The result from GMM approach indicated K-means approach is near-optimal. Cluster labels were predicted from K-means clustering for each time window. Dwell time (how long an individual spends in a given state [9] ) was then calculated. Figure 26 shows the dwell time pattern for male and female groups. Each of the bars represents the probability that a participant was more likely to stay in a specific state. The state labels are given on the X-axis. A significant distinguishable pattern of microstates occupancy was evident in both male and female groups. That indicates dynamic brain connectivities are different among groups. For each trial, those dwell time series (e.g., for each trial, we get 19 cluster labels) can be used as feature vector. We applied an SVM with an 'RBF' kernel. Hyperparameters of the classifier were tuned on a random shuffle of 80% training data with a grid search-based approach using 10-fold cross-validation. The test accuracy is 84% with ROC = 0.83. Therefore those patterns are significantly different among groups.



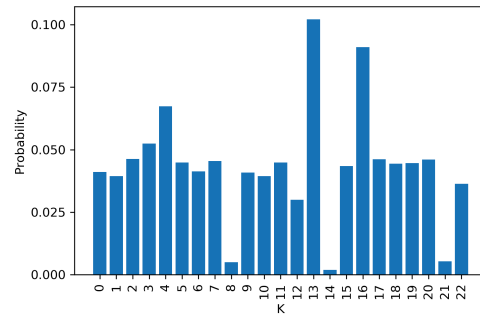
a. Elbow Method



b. Inter-cluster Distance Maps



c. Silhouette Visualization



d. Cluster Probability

Figure 24: Approaches in finding optimal number of clusters (k).

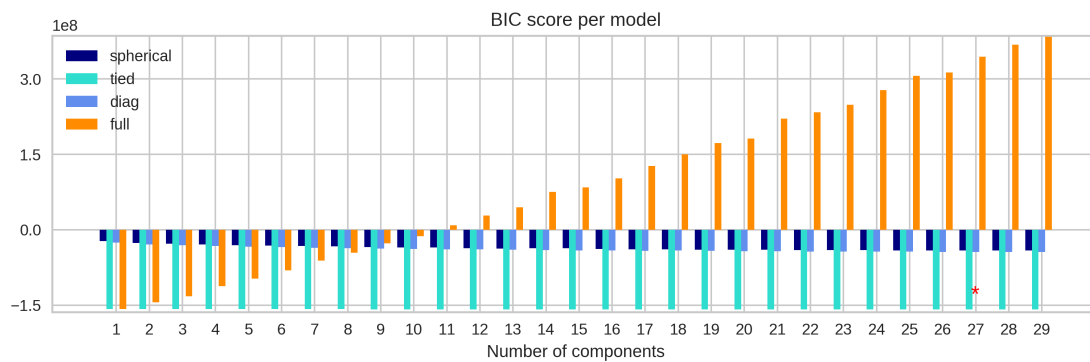


Figure 25: Model selection with Gaussian Mixture Models using Bayesian information-theoretic criteria (BIC). Model selection concerns both the covariance type and the number of components in the model. In this case, BIC is adopted to determine the right model.

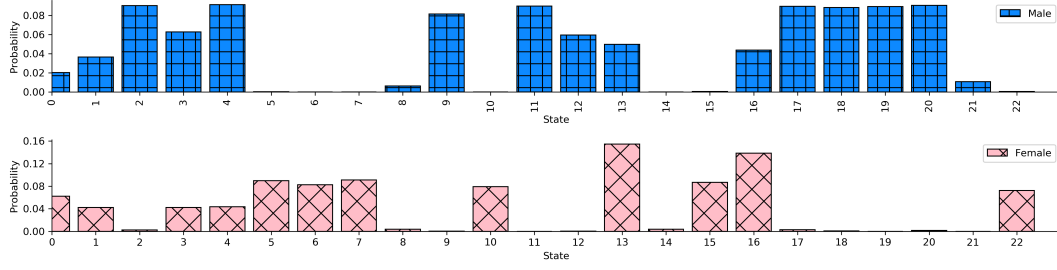


Figure 26: KNN based micro-states shows distinguishable pattern of micro-states.

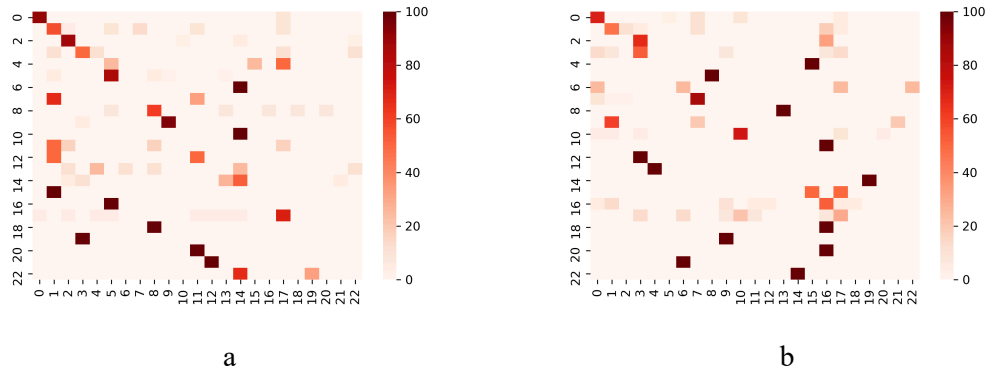


Figure 27: Transition matrix visualization for male (figure a) and female (figure b) group. The heat maps show a significantly distinct pattern of transition in processing high arousal stimuli.

Next, we computed transition probabilities for each state with high arousal stimuli, which tells us whether a subject with a certain spatial pattern at a certain time is more or less likely to transition to another spatial pattern at a future time. Results indicated that males have the less probability of transition between states and more likely to stay in specific stable state (specifically state 0, 2, 5 had a higher probability (greater than 80%)) than switch among states. However, for females, it is entirely the opposite. The transition probability is highly transient. This pattern is significantly distinguishable with some commonality. If we apply popular graph mining measures to observe the characteristics of those microstates, we did not find any significant results for cortical surface data. Though, we observed, maximum modularity and small-worldness are significantly different between groups for skull surface data (EEG data). Hence, males and females use significantly distinguishable, highly

segregated, and densely integrated network structures among brain regions in affect processing.

Table 7: Two-sided p-value of Wilcoxon rank-sum statistic for different network measures among groups. No significant measure is evident from cortical surface data, but maximum modularity and small-worldness are significantly different among groups for skull surface EEG data in processing stimuli with high arousal value.

Measures	Cortical Surface	Skull Surface
Characteristics path	0.522	0.352
Average clustering coefficient	0.841	0.317
Small worldness	0.446	0.036
Assortativity	0.432	0.702
Global efficiency	0.407	0.277
Transitivity	0.832	0.328
Maximized modularity	0.406	0.03

Table 7 shows the p-values among groups (male/female) for different network measures for the skull and cortical surface micro-states. We computed the mean of all centroids and Wilcoxon rank-sum statistics to find out significant connectivity (correlation) among brain regions among groups (male/female). The top ten highly significant connections are given in Table 8. Here, ‘Node A’ and ‘Node B’ indicates two nodes (ROI) of an edge, and “Correlation” is the connectivity weight of the edge.

Table 8: Top 10 network edges (with node A and node B) and their weight (correlation values).

Node A	Node B	Correlation
Rostral Anterior Cingulate Right	Medial Orbitofrontal Right	0.88
Rostral Anterior Cingulate Left	Medial Orbitofrontal Right	0.82
Pericalcarine Left	Lingual Left	0.81
Isthmus Cingulate Left	Precuneus Right	0.77
Pars Triangularis Left	Lateral Orbitofrontal Left	0.76
Inferior Temporal Right	Entorhinal Right	0.76
Superior Frontal Left	Superior Frontal Right	0.74
Cuneus Left	Cuneus Right	0.74
Paracentral Left	Posterior Cingulate Right	0.73
Pericalcarine Left	Lateral Occipital Left	0.72
Precuneus Left	Isthmus Cingulate Left	0.71

We developed an efficient micro-state based computational framework to study the dynamic brain network in processing emotion. We found a significantly distinct network pattern among male and female groups in processing arousal. However, the number of states and their strength of coupling may change with population size as well as the diversity of data due to the limitation of hard clustering. In the next step, we applied Hierarchical Dirichlet Process Hidden Markov Models (HDP-HMM) on CP data set as an alternative solution to address this issue and further improvement models.

### ***HDP-HMM based dynamic network analysis in understanding rapid speech categorization decisions***

In this dissertation, Sticky HDP-HMM with memoized variational inference based dynamic network analysis is used to understand rapid speech categorization decision from CP dataset. The schematic diagram of data processing pipeline is given in Figure 28.

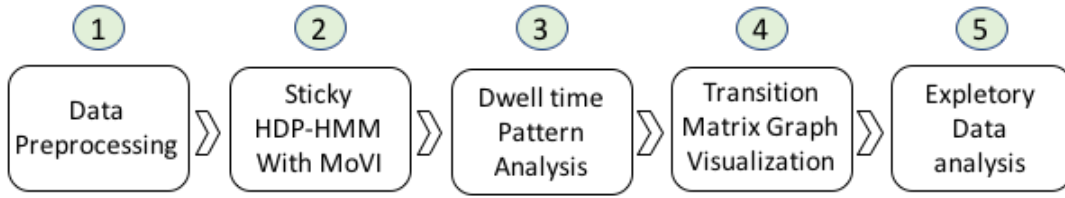


Figure 28: Schematic diagram of the processing pipeline for CP dataset. The 64 Channel EEG data is first preprocessed, and individual trialwise averaged. Sticky HPD-HMM with Memoized Variational inference (MoVI) is applied for data segmentation. Exploratory data analysis on dwell time pattern is used for further analysis and interpretation.

### ***Hierarchical Dirichlet Process Hidden Markov Models***

We used bnpy [181] for spatiotemporal EEG data segmentation. The bnpy inference engine is an opensource Python library for unsupervised learning from big data. This framework supports Bayesian nonparametric clustering that captures multidimensional, sequential, spatial, and hierarchical structures. To run inference on a dataset, bnpy requires an allocation model, a data-generation method, and the inference algorithm. The allocation model describes the generative process that allocates cluster assignments to individual data



points. Here we used HDP-HMM (Markov sequence models with an infinite number of states).

Observation models define a likelihood for producing data from a cluster-specific density. We used Diagonal-Covariance Gaussian as observation models. However, the inference algorithm optimizes a variational-bound objective function. To achieve scalability, we focus on modern optimization-based approaches that can process batched data, particularly memoized variational inference. Mathematical definition and interpretation part of Sticky HDP-HMM with Memoized Variational Inference (MoVI) are described in the appendix.

#### *Calculation of dwell time statistics.*

Dwell time counts the time the brain spends in a microstate before it transitions to another brain state. It reflects the characteristics of neural activity [182], [183]. Widely used dwell time statistics are (i) Duration: average duration that a given microstate remains stable. It reflects the stability of potential neural assembly. (ii) Occurrence: the frequency of occurrence for each microstate independent of its individual duration which reflects the activation trend of a potential neural source. (iii) Time coverage: the fraction of total recording time for which a given microstate is dominant. It reflects the occurrence percentage of potential neural source and amplitude reflects the intensity of a potential nerve source. (iv) Global variance: the global variance explained by each microstate (v) Transition probabilities: the transition probabilities of a given microstate to any other microstate. [7], [182].

#### *Empirical analysis on CP dataset*

The HDP-HMM consider infinite number of states. The Birth, Merge, and delete proposals are widely used to remove ineffective states [14]. To get interpretable less number of states, we varied the number of cluster from 5 to 30 with a step of 5 and observed the cluster probability. Figure 29 shows the number of clusters (K) vs. cluster probability graph. Each bar show number of data point loads in that specific cluster of a specific model (out of 5

models). The first cluster (0) has most of the data points. It was observed that up to 10 clusters can accommodate an average of 93% of data. Cluster numbers more than 10 had few numbers of data points. The K=10 is enough to explain 93% variability on average. The rest of the cluster is unnecessary or useless. Hence, we select a finite model with K=10 for further analysis.

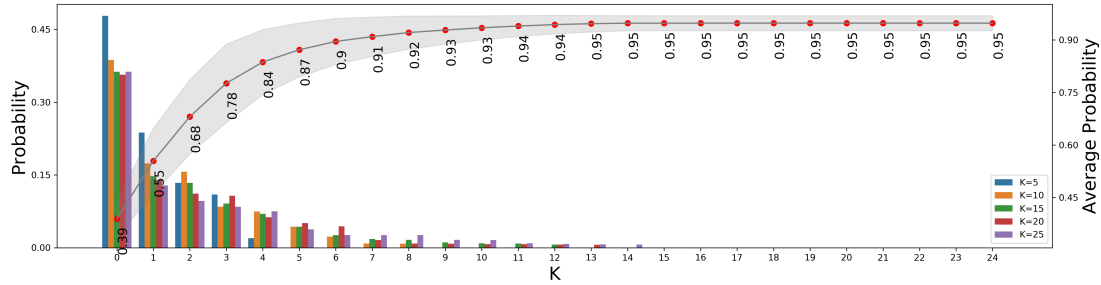


Figure 29: Probability of data points loading into each component.

HDP-HMM segments each EEG time series data using these 10 states. EEG time series data of a sample trial (Slow-RT of Subject #1) and dwell time (tile and time-series) patterns are shown in Figure 30. Each color of the tile represents the amount of time a state remains stable. Those colors are also marked with state number text. Dwell time series showed the pattern of state transition.

We calculated trial wise individual dwell time statistics described in section 0. Figure 31 shows the summary statistics of state-wise dwell time pattern analysis. It was observed that the frequency of occurrence of state 0, 2, 6, 7, 8, 9 and duration to stay in state 4, 5, 6, 7, 8 are significantly different among RTs. Figure 33 shows the topomap of the states. Though it is possible to see the pattern of activation from topomap, we need more understandable visualization of brain activation. We applied a Classical Low Resolution Electromagnetic Tomography Analysis Recursively Applied (CLARA) [BESA Research (v7); BESA, GmbH] [184]–[186] to provide a qualitative description of the underlying brain sources that generate each state-specific scalp topography. CLARA renders more focal source reconstructions by iteratively reducing the source space during repeated estimations of the inverse solution. On

each step, a spatially smoothed LORETA solution was recomputed, and voxels below a 1% max amplitude threshold were removed. This process provided a spatial weighting term for each voxel of the LORETA image on the subsequent step. Two iterations were used with a voxel size of 7 mm in Talairach space and regularization (parameter accounting for noise) set at 0.01% singular value decomposition. CLARA activation maps were overlaid onto the BESA adult MRI template for visualization with respect to the brain anatomy. Figure 34 shows the CLARA visualization of the 4 main states. Here, STG = superior temporal gyrus (home of auditory cortex; stimulus coding), IFG = inferior frontal gyrus (home of important language regions like Broca's area), and SMG = supramarginal gyrus (implicated in lexical/semantic decisions including making phonological word choices).

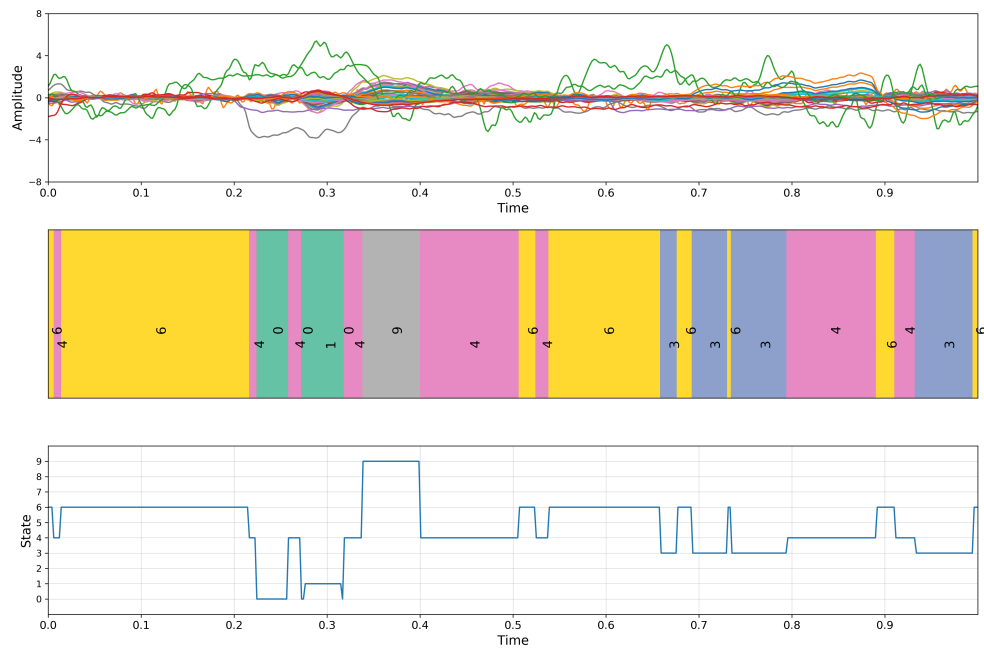


Figure 30. Top: A sample trial of 64 channel EEG Data (Subject#1, slow RT), middle: tile visualization of dwell pattern, bottom: time series visualization of dwell pattern.

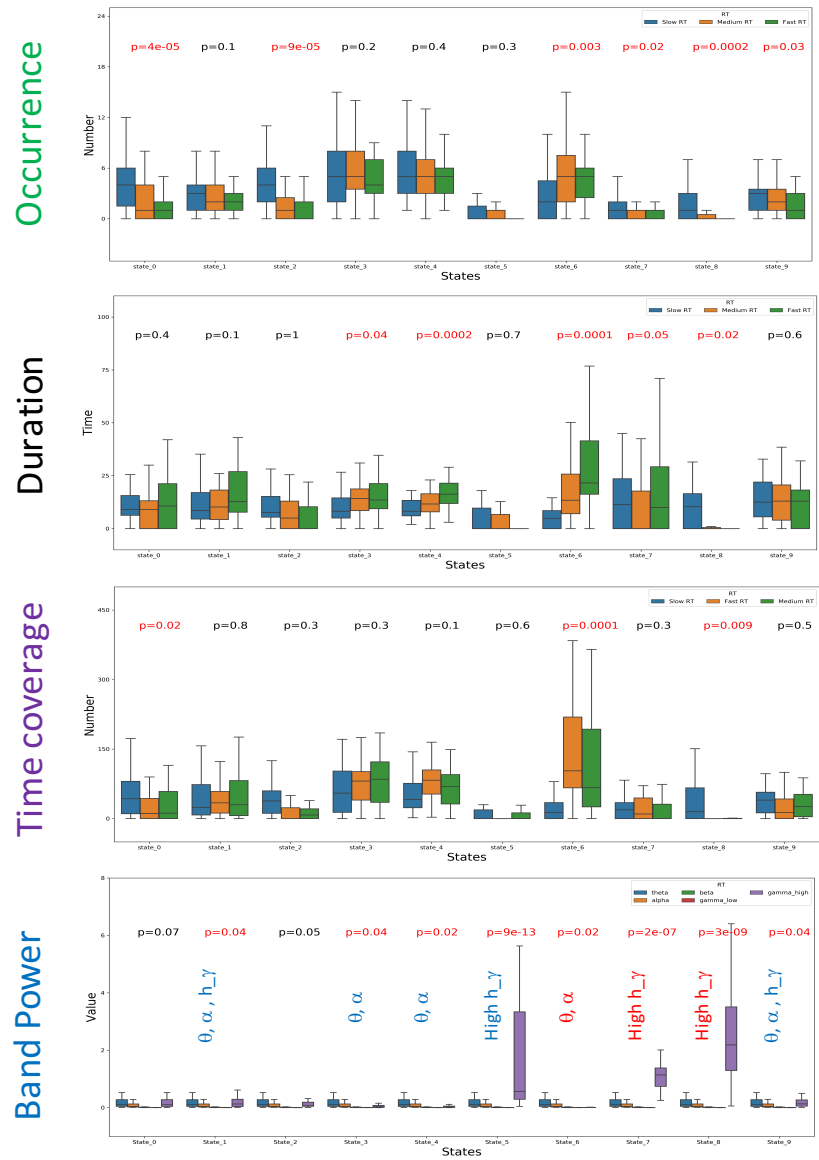


Figure 31: Dwell time statistics.

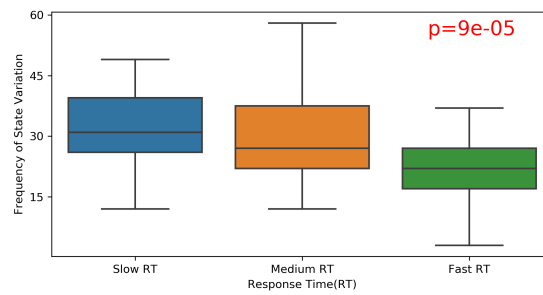


Figure 32: Frequency of changing states among trails. Entropy in slower RTs are higher than faster RT (Slow RT: 2.26, Medium RT: 2.11, Fast RT: 2.03).

We can easily identify states 6, 3, 4, and 9 are related to resting-state, stimulus coding, response selection, and linguistic processing. However, State 6 and State 1 showed a similar pattern. Hence, we can also consider state-1 as a state for stimulus decoding. To find the properties of those state we calculated the average band power of those states. State-wise average band power visualization are shown in Figure 31. Here  $\theta$ ,  $\alpha$ ,  $\beta$ , Low  $\gamma$ , and High  $\gamma$  are in ranges from 4 to 7 Hz, from 8 to 15 Hz, from 16 to 31 Hz, from 31 to 60 Hz and from 60 to 200 Hz, respectively [187]. It was observed that states 5, 7, and 8 have significantly high  $\gamma$  band activity. However,  $\theta$  and  $\alpha$  band activity was dominant in stimulus encoding, response selection, linguistic processing, and resting state. Slightly high  $\gamma$  activity was also evident in those states.

In a task-related experiment, significantly higher theta band activity was shown during the encoding, which could be remembered in the later recall task [188]. Frontal theta is a compelling candidate mechanism by which emergent processes such as ‘cognitive control’(novelty, conflict, punishment, and error) may be biophysically realized [189]. Cognitive control is the broader construct of information prioritization for goal-driven decision-making [190]. On the other hand,  $\alpha$  is a relevant aspect of visual scene processing, integral to spatial attention. The  $\alpha$  waves are used to predict mistakes too. A recent study showed that both  $\theta$  and  $\alpha$  band was active in resting state. But, high  $\gamma$  represents conscious attention, which easily validates our claims.

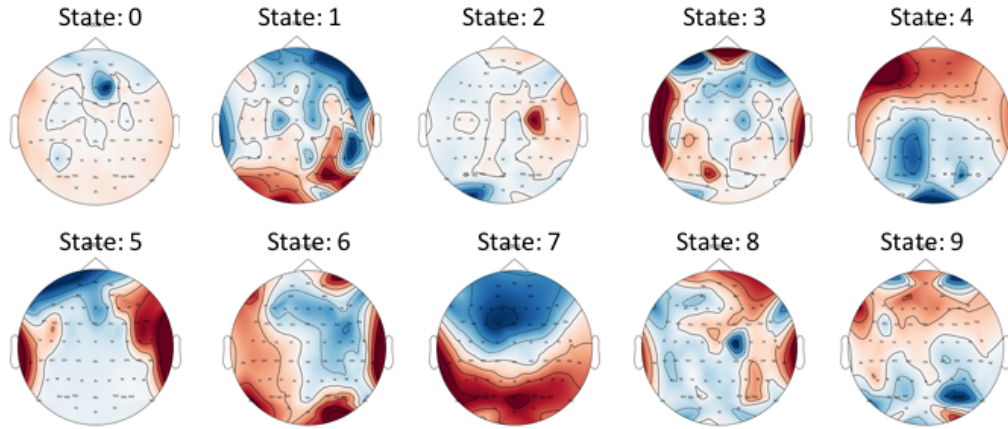


Figure 33: State topomap.

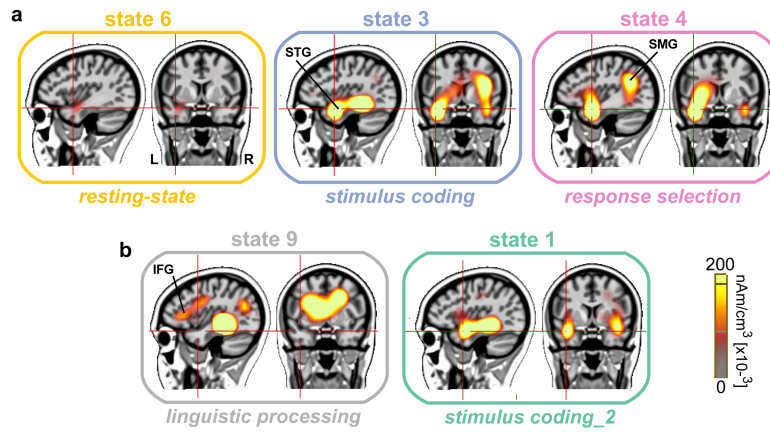


Figure 34: Visualization CLARA activation maps.

Besides, dwell time pattern also shows that:

1. Trials with faster RTs spend significantly less time in stimulus encoding (state 3), response selection (state 4), and resting-state (state 6),
2. RT wise, stimulus encoding, and response selection occurrence is similar,
3. Combining 1,2 we found: it is not as important how frequently people stay in stimulus encoding and response selection states, but rather how long they stay in those states,
4. Trials with faster RT spend significantly more time on resting state,
5. In RT wise group analysis (considering all trials), we found that transition frequency was significantly higher in slower RTs (shown in Figure 32). Therefore, entropy in slower

RTs is higher than faster RT (Slow RT: 2.26, Medium RT: 2.11, Fast RT: 2.03). Hence, decisions making process in slower RT is more uncertain than usual.

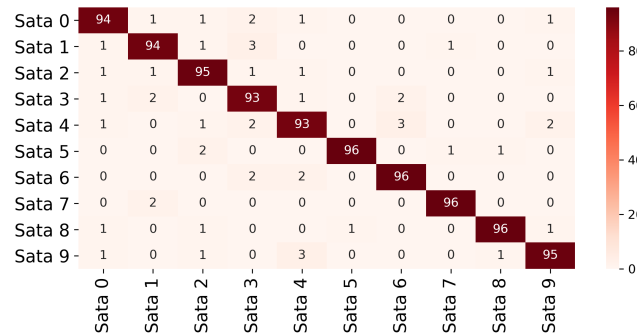


Figure 35: Transition matrix.

It is possible to get a transition matrix from Sticky HDP-HMM analysis. From the transition matrix (shown in Figure 35), it can be observed that states are more likely to stay consistent rather than transition. Besides the heat map visualization of the transition matrix (shown in Figure 35), we also graphed RT differences (shown in Figure 36). Here each node represents one state. The self-loop of nodes represents the average time a state remains stable in condition-specific (RTs) trials. Edges represent state transition probabilities. Probability  $< 0.03$  were discarded for better visualization with a smaller number of nodes. This graph visualization gives us a more clear view of how dynamically brain changes states and explain the reasons behind variations in RT.

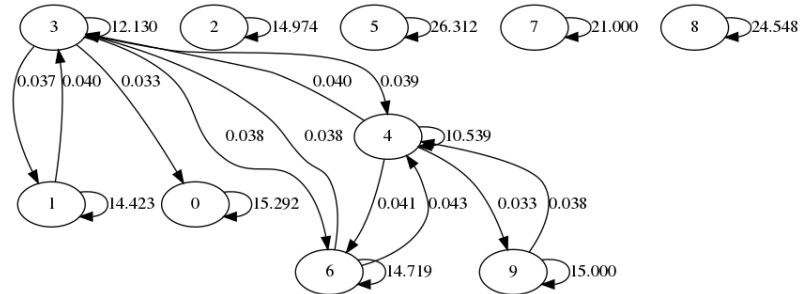
There is a common delta (triangular) graph pattern that consists of resting-state (state 6), stimulus encoding (state 3), and resting-state (state 4) present in all RT related graphs. To make a speech perception related decision, these three functions are very important. Beside that delta connection, the response selection nodes are also connected with stimulus encoding\_2 (state 1) node, which is slightly different than state 4. It was also observed that the stimulus coding node is strongly connected with linguistic processing states. Overall, we observed that

1. stimulus coding, linguistic processing, resting state, and response selection state comprise a CP related decision-making network (DN),

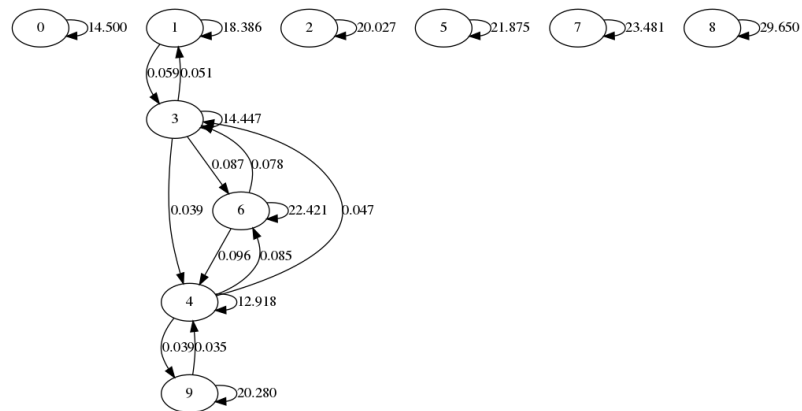
2. From Figure 36, it was observed that slower RTs are more likely to stay in this DN.

Hence the speed of rapid speech categorization decisions is inversely proportional to the time listeners stay in the DN.

Slow RT



Medium RT



Fast RT

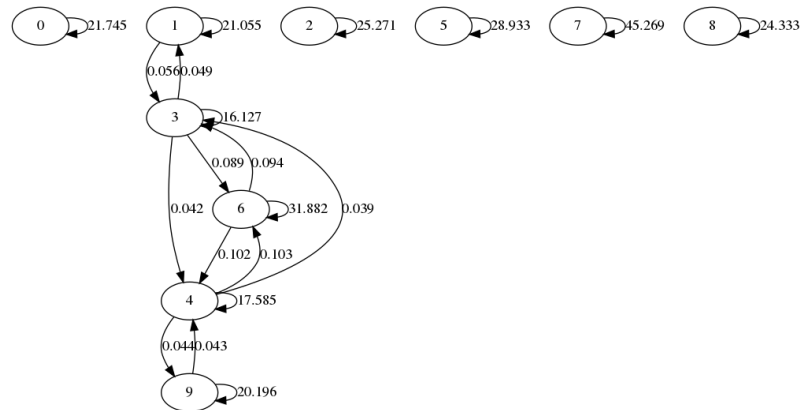


Figure 36: Graph visualization of transition matrix meta-analysis. Each node represents one state. The Self-loop of nodes represents average time a state remains stable in condition-specific (RTs) trials. Edges represent state transition probabilities. Probability <0.03 are discarded for better visualization with a smaller number of nodes.



## Conclusion

Brain networks in relation to perception and cognition, are complex and dynamic in nature. Here, we developed an efficient micro-state based computational framework to study the dynamic brain network in processing emotion and speech perception. We found a significantly distinct network pattern among male and female groups in processing arousal. However, the number of states and their strength of coupling may change with population size as well as the diversity of data due to the limitation of hard clustering. To overcome this limitation, we adopted a Bayesian nonparametric framework for segmentation neural activity with a finite number of microstates. This approach enabled us to find default network and transient pattern of the underlying neural mechanism in relation to categorical perception.

## Appendix

### *Hierarchical Dirichlet Process Hidden Markov Models*

Let EEG has  $n$  number of trials and data is represented as  $x_n = [x_{n1}, x_{n2}, x_{n3} \dots x_{nT}]$ . Observation  $x_{nt}$  is a vector representing at time  $t$  and  $x_{nt} \in \mathbb{R}^D$ . For 64 channel EEG data  $D=64$ . The HDP-HMM explains this data by assigning each observation  $x_{nt}$  to a single hidden state  $Z_{nt}$ . The chosen state comes from a countably infinite set of cluster  $K \in \{1, 2, \dots\}$ , generated via Markovian dynamics with initial state distributions  $\pi_0$  and transition distributions  $\{\pi_k\}_{k=1}^\infty$ .

$$\mathcal{P}(Z_{n1} = k) = \pi_{0k} \quad 15$$

$$\mathcal{P}(Z_{nt} = \ell \mid Z_{n,t-1} = k) = \pi_{k\ell} \quad 16$$

### *Hierarchies of Dirichlet Processes.*

The number of states is unbounded under the HDP-HMM prior and posterior. The hierarchical Dirichlet process (HDP) shares states over time via a latent root probability vector  $\beta$  over the infinite set of states. The stick-breaking representation of the prior on  $\beta$  first draws independent variables  $\mu_k = \text{Beta}(1, \gamma)$  for each state  $k$  and then set  $\beta_k =$

$\mu_k \prod_{l=1}^{k-1} (1 - \mu_l)$ . The  $\mu_k$  can be interpreted as conditional probability of choosing  $k$  th state among states. The HDP-HMM generates transition distributions  $\mu_k$  for each state  $k$  from a Dirichlet with mean equal to  $\beta$  and variance governed by concentration parameter  $\alpha$ .

$$[\pi_{k1}, \pi_{k2}, \pi_{k3} \dots \pi_{kk}] = \text{Dir}(\alpha\beta_1, \alpha\beta_2, \alpha\beta_3 \dots \alpha\beta_k) \quad 17$$

The  $\pi_{k0}$  is the starting probability vector with  $\pi_{k0} \sim \text{Dir}(\alpha_o\beta_1)$ . Where  $\alpha_o \gg \alpha$ .

*Variational Inference:*

The inferential goal of HDP-HMM is to get posterior knowledge of top-level conditional probabilities  $\mu_k$ , HMM parameters: cluster probability  $\pi$ , cluster shape  $\phi$  and cluster assignments  $z$  after observing data  $x$ . Parameter  $\mu, \pi, \phi$  are considered as global parameter parameters because they generalize to new data sequences. The cluster assignments  $z_n$  is a local parameter specific to data sequence  $x_n$ .

Variational methods frame posterior inference as an optimization problem [191]. Here, we seek a distribution  $q(\mu, \pi, \phi)$  over the unobserved variables that is close to the true posterior i.e.  $q(\mu, \pi, \phi, z) \sim p(\mu, \pi, \phi, z | x)$ . We can re-present  $q(\mu, \pi, \phi, z)$  as simpler factorized family  $q(\mu, \pi, \phi, z) \cong q(\mu) q(\pi) q(\phi) q(z)$ . Inference algorithms update these parameters to minimize the Kullback-Leibler (KL) divergence. Equation for KL divergence is given by:

$$KL(q(\mu, \pi, \phi, z) || p(\mu, \pi, \phi, z | x)) = \mathcal{L}(\cdot) = E_q \left[ \log \frac{q(\mu, \pi, \phi, z)}{p(\mu, \pi, \phi, z | x)} \right] \quad 18$$

To get best  $q^*$  distribution, we need to optimize this objective function  $\mathcal{L}(\cdot)$ . This equation can be simplified with four components.

$$\mathcal{L}(\cdot) = \mathcal{L}_{\text{data}} + \mathcal{L}_{\text{entropy}} + \mathcal{L}_{\text{hdp-local}} + \mathcal{L}_{\text{hdp-global}} \quad 19$$

Here

$$\mathcal{L}_{\text{data}}(x, \hat{r}, \hat{t}) \cong E_q \left[ \log p(x|z, \phi) + \log \frac{p(\phi)}{q(\phi)} \right] \quad 20$$

$$\mathcal{L}_{\text{entropy}}(\hat{s}) \cong E_q [\log q(z)] \quad 21$$

$$\mathcal{L}_{\text{hdp-local}}(\hat{s}, \hat{\theta}, \hat{\rho}, \hat{\omega}) \cong E_q \left[ \log p(z|\pi) + \log \frac{p(\pi)}{q(\pi)} \right] \quad 22$$

$$\mathcal{L}_{\text{hdp-global}} \cong E_q \left[ \log \frac{p(u)}{q(u)} \right] \quad 23$$

*Memoized and Stochastic Variational Inference:*

Common variational inference algorithms maximize  $\mathcal{L}(\cdot)$  using coordinate ascent optimization. Here optimal value of each parameter is keep fixed while optimizing other parameters. For the sticky HDP-HMM variational objective, each sequence is randomly assigned to one of  $B$  batches initially. The algorithm repeatedly and random visits batches one at a time. Each full pass through the complete set of  $B$  batches a is considered as a lap. At each visit to batch  $b$ , sticky HDP-HMM perform a local step for all sequences  $n$  in batch  $b$  and then a global step. The batch optimization of  $\mathcal{L}(\cdot)$  is possible by exploiting the additivity of statistics  $M, S$ . For each statistic, this algorithm track batch-specific quantity  $M^b$ , and a summary of whole-dataset  $M = \sum_{b=1}^B M^b$ . After a local step at batch  $b$ , yields  $\widehat{s}^b, \widehat{r}^b$  and update  $M^b(\widehat{s}^b)$  and  $S^b(\widehat{r}^b)$ , increment each whole-dataset statistic by adding the new batch summary and subtracting the summary stored in memory from the previous visit and store (or memoize) the new statistics for future iterations. It is possible to evaluate  $\mathcal{L}(\cdot)$  at any point during memoized execution except  $\mathcal{L}_{\text{entropy}}(\cdot)$  term. To compute it, a  $(K + 1) \times K$  dimensional matrix  $H^b$  is tracked at each batch  $b$ . Where:

$$H_{ol}^b = - \sum_n \widehat{r_{n1l}} \log \widehat{r_{n1l}} \quad 24$$

$$H_{kl}^b = - \sum_n \sum_{t=1}^{T_n-1} \widehat{s_{ntkl}} \log \frac{\widehat{s_{ntkl}}}{\widehat{r_{ntk}}} \quad 25$$

For whole dataset entropy matrix:  $H = \sum_{b=1}^B H^b$

$$\mathcal{L}_{\text{entropy}} = \sum_{k=1}^K \sum_{l=0}^K H_{kl} \quad 26$$

## Relationship to published works

### Chapter 3 -

1. Al-Fahad, Rakib, et al. "Early Imaging-Based Predictive Modeling of Cognitive Performance Following Therapy for Childhood ALL." *IEEE Access* 7 (2019): 146662-146674.

### Chapter 4 -

1. Al-Fahad, Rakib, Mohammed Yeasin, and Gavin M. Bidelman. "Decoding of single-trial EEG reveals unique states of functional brain connectivity that drive rapid speech categorization decisions." *Journal of Neural Engineering* (2019).

### Chapter 5 -

1. Al-Fahad, Rakib, and Mohammed Yeasin. "Micro-states based dynamic brain connectivity in understanding the commonality and differences in gender-specific emotion processing." *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019.
2. Al-Fahad, Rakib, and Mohammed Yeasin. "Robust modeling of continuous 4-d affective space from eeg recording." *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2016.
3. Al-Fahad, Rakib, et al. "Selection of stable features for modeling 4-D affective space from EEG recording." *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017.

## References

- [1] G. Tononi, O. Sporns, and G. M. Edelman, “A measure for brain complexity: relating functional segregation and integration in the nervous system,” *Proceedings of the National Academy of Sciences*, vol. 91, no. 11, pp. 5033–5037, 1994.
- [2] D. S. Bassett and E. D. Bullmore, “Small-world brain networks,” *The neuroscientist*, vol. 12, no. 6, pp. 512–523, 2006.
- [3] C. J. Honey, R. Kötter, M. Breakspear, and O. Sporns, “Network structure of cerebral cortex shapes functional connectivity on multiple time scales,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 24, pp. 10240–10245, 2007.
- [4] M. Rubinov and O. Sporns, “Complex network measures of brain connectivity: uses and interpretations,” *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.
- [5] R. Al-Fahad, M. Yeasin, J. O. Glass, H. M. Conklin, L. M. Jacola, and W. E. Reddick, “Early Imaging-Based Predictive Modeling of Cognitive Performance Following Therapy for Childhood ALL,” *IEEE Access*, vol. 7, pp. 146662–146674, 2019, doi: 10.1109/ACCESS.2019.2946240.
- [6] N. Meinshausen and P. Bühlmann, “Stability selection,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 4, pp. 417–473, 2010.
- [7] T. Koenig, D. Lehmann, M. C. G. Merlo, K. Kochi, D. Hell, and M. Koukkou, “A deviant EEG brain microstate in acute, neuroleptic-naïve schizophrenics at rest,” *European Archives of Psychiatry and Clinical Neurosciences*, vol. 249, no. 4, pp. 205–211, Aug. 1999, doi: 10.1007/s004060050088.
- [8] V. D. Calhoun, R. Miller, G. Pearlson, and T. Adalı, “The Chronnectome: Time-Varying Connectivity Networks as the Next Frontier in fMRI Data Discovery,” *Neuron*, vol. 84, no. 2, pp. 262–274, Oct. 2014, doi: 10.1016/j.neuron.2014.10.015.
- [9] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, “Tracking Whole-Brain Connectivity Dynamics in the Resting State,” *Cereb Cortex*, vol. 24, no. 3, pp. 663–676, Mar. 2014, doi: 10.1093/cercor/bhs352.
- [10] R. M. Hutchison *et al.*, “Dynamic functional connectivity: promise, issues, and interpretations,” *Neuroimage*, vol. 80, pp. 360–378, 2013.
- [11] S. Shakil, C.-H. Lee, and S. D. Keilholz, “Evaluation of sliding window correlation performance for characterizing dynamic functional connectivity and brain states,” *NeuroImage*, vol. 133, pp. 111–128, Jun. 2016, doi: 10.1016/j.neuroimage.2016.02.074.
- [12] E. Damaraju *et al.*, “Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia,” *NeuroImage: Clinical*, vol. 5, pp. 298–308, Jan. 2014, doi: 10.1016/j.nicl.2014.07.003.
- [13] A. Mheich, M. Hassan, M. Khalil, C. Berrou, and F. Wendling, “A new algorithm for spatiotemporal analysis of brain functional connectivity,” *Journal of Neuroscience Methods*, vol. 242, pp. 77–81, Mar. 2015, doi: 10.1016/j.jneumeth.2015.01.002.
- [14] M. C. Hughes, W. T. Stephenson, and E. Sudderth, “Scalable Adaptation of State Complexity for Nonparametric Hidden Markov Models,” in *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, Inc., 2015, pp. 1198–1206.
- [15] F. A. Azevedo *et al.*, “Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain,” *Journal of Comparative Neurology*, vol. 513, no. 5, pp. 532–541, 2009.
- [16] S. Herculano-Houzel, “The human brain in numbers: a linearly scaled-up primate brain,” *Frontiers in human neuroscience*, vol. 3, p. 31, 2009.
- [17] M. Song *et al.*, “Brain spontaneous functional connectivity and intelligence,” *Neuroimage*, vol. 41, no. 3, pp. 1168–1176, 2008.

- [18] M. P. van den Heuvel, C. J. Stam, R. S. Kahn, and H. E. H. Pol, "Efficiency of functional brain networks and intellectual performance," *Journal of Neuroscience*, vol. 29, no. 23, pp. 7619–7624, 2009.
- [19] "Assessing the Economic Costs of Serious Mental Illness | American Journal of Psychiatry." <https://ajp.psychiatryonline.org/doi/full/10.1176/appi.ajp.2008.08030366> (accessed Apr. 12, 2018).
- [20] "The Human Connectome Project." <https://www.neuroscienceblueprint.nih.gov/connectome/> (accessed Jan. 23, 2018).
- [21] D. C. Van Essen *et al.*, "The WU-Minn human connectome project: an overview," *Neuroimage*, vol. 80, pp. 62–79, 2013.
- [22] L. A. Jorgenson *et al.*, "The BRAIN Initiative: developing technology to catalyze neuroscience discovery," *Philos Trans R Soc Lond B Biol Sci*, vol. 370, no. 1668, May 2015, doi: 10.1098/rstb.2014.0164.
- [23] K. R. Krull *et al.*, "Genetic mediators of neurocognitive outcomes in survivors of childhood acute lymphoblastic leukemia," *Journal of Clinical Oncology*, vol. 31, no. 17, pp. 2182–2188, 2013.
- [24] R. Seigers and J. E. Fardell, "Neurobiological basis of chemotherapy-induced cognitive impairment: a review of rodent research," *Neuroscience & Biobehavioral Reviews*, vol. 35, no. 3, pp. 729–741, 2011.
- [25] M. Monje and J. Dietrich, "Cognitive side effects of cancer therapy demonstrate a functional role for adult neurogenesis," *Behavioural brain research*, vol. 227, no. 2, pp. 376–379, 2012.
- [26] F. Altiparmak, M. Gen, L. Lin, and T. Paksoy, "A genetic algorithm approach for multi-objective optimization of supply chain networks," *Computers & Industrial Engineering*, vol. 51, no. 1, pp. 196–215, 2006.
- [27] L. Porto *et al.*, "Voxel-based morphometry and diffusion-tensor MR imaging of the brain in long-term survivors of childhood leukemia," *European radiology*, vol. 18, no. 11, p. 2691, 2008.
- [28] S. R. Kesler, M. Gugel, E. Huston-Warren, and C. Watson, "Atypical structural connectome organization and cognitive impairment in young survivors of acute lymphoblastic leukemia," *Brain connectivity*, vol. 6, no. 4, pp. 273–282, 2016.
- [29] M. E. Carey, M. W. Haut, S. L. Reminger, J. J. Hutter, R. Theilmann, and K. L. Kaemingk, "Reduced frontal white matter volume in long-term childhood leukemia survivors: a voxel-based morphometry study," *American Journal of Neuroradiology*, vol. 29, no. 4, pp. 792–797, 2008.
- [30] D. J. Mabbott, M. D. Noseworthy, E. Bouffet, C. Rockel, and S. Laughlin, "Diffusion tensor imaging of white matter after cranial radiation in children for medulloblastoma: correlation with IQ," *Neuro-oncology*, vol. 8, no. 3, pp. 244–252, 2006.
- [31] D. J. Mabbott, M. Noseworthy, E. Bouffet, S. Laughlin, and C. Rockel, "White matter growth as a mechanism of cognitive development in children," *Neuroimage*, vol. 33, no. 3, pp. 936–946, 2006.
- [32] D. J. Mabbott, J. Rovet, M. D. Noseworthy, M. L. Smith, and C. Rockel, "The relations between white matter and declarative memory in older children and adolescents," *Brain research*, vol. 1294, pp. 80–90, 2009.
- [33] M. N. Edelmann *et al.*, "Diffusion tensor imaging and neurocognition in survivors of childhood acute lymphoblastic leukaemia," *Brain*, vol. 137, no. 11, pp. 2973–2983, 2014.
- [34] L. Zou *et al.*, "Abnormal topological organization in white matter structural networks in survivors of acute lymphoblastic leukaemia with chemotherapy treatment," *Oncotarget*, vol. 8, no. 36, pp. 60568–60575, Jul. 2017, doi: 10.18632/oncotarget.19104.

- [35] A. Amidi *et al.*, “Changes in Brain Structural Networks and Cognitive Functions in Testicular Cancer Patients Receiving Cisplatin-Based Chemotherapy,” *J Natl Cancer Inst*, vol. 109, no. 12, Dec. 2017, doi: 10.1093/jnci/djx085.
- [36] L. van der Maaten and G. Hinton, “Visualizing data using t-SNE,” *Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2579–2605, 2008.
- [37] A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy, “Perception of the speech code,” *Psychol Rev*, vol. 74, no. 6, pp. 431–461, Nov. 1967.
- [38] D. B. Pisoni, “Auditory and phonetic memory codes in the discrimination of consonants and vowels,” *Perception & Psychophysics*, vol. 13, no. 2, pp. 253–260, Jun. 1973, doi: 10.3758/BF03214136.
- [39] S. Harnad and A. S. L. Bureau, *Categorical perception : the groundwork of cognition*, 2nd ed. Cambridge ; New York : Cambridge University Press, 1987.
- [40] D. B. Pisoni and P. A. Luce, “Acoustic-phonetic representations in word recognition,” *Cognition*, vol. 25, no. 1, pp. 21–52, Mar. 1987, doi: 10.1016/0010-0277(87)90003-5.
- [41] J. F. Prather, S. Nowicki, R. C. Anderson, S. Peters, and R. Mooney, “Neural correlates of categorical perception in learned vocal communication,” *Nature Neuroscience*, vol. 12, no. 2, pp. 221–228, Feb. 2009, doi: 10.1038/nn.2246.
- [42] D. B. Pisoni and J. Tash, “Reaction times to comparisons within and across phonetic categories,” *Perception & Psychophysics*, vol. 15, no. 2, pp. 285–290, Mar. 1974, doi: 10.3758/BF03213946.
- [43] G. M. Bidelman, S. Moreno, and C. Alain, “Tracing the emergence of categorical speech perception in the human auditory system,” *NeuroImage*, vol. 79, pp. 201–212, Oct. 2013, doi: 10.1016/j.neuroimage.2013.04.093.
- [44] G. M. Bidelman and C.-C. Lee, “Effects of language experience and stimulus context on the neural organization and categorical perception of speech,” *NeuroImage*, vol. 120, pp. 191–200, Oct. 2015, doi: 10.1016/j.neuroimage.2015.06.087.
- [45] E. Liebenthal, R. Desai, M. M. Ellingson, B. Ramachandran, A. Desai, and J. R. Binder, “Specialization along the Left Superior Temporal Sulcus for Auditory Categorization,” *Cereb Cortex*, vol. 20, no. 12, pp. 2958–2970, Dec. 2010, doi: 10.1093/cercor/bhq045.
- [46] R. Reetzke, Z. Xie, F. Llanos, and B. Chandrasekaran, “Tracing the Trajectory of Sensory Plasticity across Different Stages of Speech Learning in Adulthood,” *Current Biology*, vol. 28, no. 9, pp. 1419–1427.e4, May 2018, doi: 10.1016/j.cub.2018.03.026.
- [47] S. E. Lively, J. S. Logan, and D. B. Pisoni, “Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories,” *The Journal of the Acoustical Society of America*, vol. 94, no. 3, pp. 1242–1255, Sep. 1993, doi: 10.1121/1.408177.
- [48] G. M. Bidelman and B. S. Walker, “Attentional modulation and domain-specificity underlying the neural organization of auditory categorical perception,” *European Journal of Neuroscience*, vol. 45, no. 5, pp. 690–699, 2017, doi: 10.1111/ejn.13526.
- [49] G. M. Bidelman, M. W. Weiss, S. Moreno, and C. Alain, “Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians,” *European Journal of Neuroscience*, vol. 40, no. 4, pp. 2662–2673, 2014, doi: 10.1111/ejn.12627.
- [50] G. M. Bidelman and C. Alain, “Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception,” *Journal of Neuroscience*, vol. 35, no. 3, pp. 1240–1249, 2015.
- [51] G. M. Bidelman, J. W. Villafuerte, S. Moreno, and C. Alain, “Age-related changes in the subcortical–cortical encoding and categorical perception of speech,” *Neurobiology of*



- Aging*, vol. 35, no. 11, pp. 2526–2540, Nov. 2014, doi: 10.1016/j.neurobiolaging.2014.05.006.
- [52] G. M. Bidelman, J. E. Lowther, S. H. Tak, and C. Alain, “Mild Cognitive Impairment Is Characterized by Deficient Brainstem and Cortical Representations of Speech,” *J. Neurosci.*, vol. 37, no. 13, pp. 3610–3620, Mar. 2017, doi: 10.1523/JNEUROSCI.3700-16.2017.
  - [53] Calcus Axelle, Lorenzi Christian, Collet Gregory, Colin Cécile, and Kolinsky Régine, “Is There a Relationship Between Speech Identification in Noise and Categorical Perception in Children With Dyslexia?,” *Journal of Speech, Language, and Hearing Research*, vol. 59, no. 4, pp. 835–852, Aug. 2016, doi: 10.1044/2016\_JSLHR-H-15-0076.
  - [54] Hakvoort Britt *et al.*, “The Role of Categorical Speech Perception and Phonological Processing in Familial Risk Children With and Without Dyslexia,” *Journal of Speech, Language, and Hearing Research*, vol. 59, no. 6, pp. 1448–1460, Dec. 2016, doi: 10.1044/2016\_JSLHR-L-15-0306.
  - [55] C. Phillips, “Levels of representation in the electrophysiology of speech perception,” *Cognitive Science*, vol. 25, no. 5, pp. 711–731, 2001, doi: 10.1207/s15516709cog2505\_5.
  - [56] S. R. Cajal, “Histology of the nervous system of man and vertebrates,” *History of Neuroscience (Oxford Univ Press, New York)*, no. 6, 1995.
  - [57] S. L. Bressler, “Large-scale cortical networks and cognition,” *Brain Research Reviews*, vol. 20, no. 3, pp. 288–304, 1995.
  - [58] P. Fries, “A mechanism for cognitive dynamics: neuronal communication through neuronal coherence,” *Trends in cognitive sciences*, vol. 9, no. 10, pp. 474–480, 2005.
  - [59] E. Bullmore and O. Sporns, “Complex brain networks: graph theoretical analysis of structural and functional systems,” *Nature Reviews Neuroscience*, vol. 10, no. 3, p. 186, 2009.
  - [60] M. E. Newman, “The structure and function of complex networks,” *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.
  - [61] J. Alho *et al.*, “Early-latency categorical speech sound representations in the left inferior frontal gyrus,” *Neuroimage*, vol. 129, pp. 214–223, 2016.
  - [62] G. M. Bidelman and B. Walker, “Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network,” *NeuroImage*, vol. 201, p. 116022, 2019.
  - [63] J. R. Binder, E. Liebenthal, E. T. Possing, D. A. Medler, and B. D. Ward, “Neural correlates of sensory and decision processes in auditory object identification,” *Nature Neuroscience*, vol. 7, no. 3, p. 295, Mar. 2004, doi: 10.1038/nn1198.
  - [64] E. F. Chang, J. W. Rieger, K. Johnson, M. S. Berger, N. M. Barbaro, and R. T. Knight, “Categorical speech representation in human superior temporal gyrus,” *Nature Neuroscience*, vol. 13, no. 11, pp. 1428–1432, Nov. 2010, doi: 10.1038/nn.2641.
  - [65] Guenther Frank H., Nieto-Castanon Alfonso, Ghosh Satrajit S., and Tourville Jason A., “Representation of Sound Categories in Auditory Cortical Maps,” *Journal of Speech, Language, and Hearing Research*, vol. 47, no. 1, pp. 46–57, Feb. 2004, doi: 10.1044/1092-4388(2004/005).
  - [66] E. B. Myers, S. E. Blumstein, E. Walsh, and J. Eliassen, “Inferior Frontal Regions Underlie the Perception of Phonetic Category Invariance,” *Psychol Sci*, vol. 20, no. 7, pp. 895–903, Jul. 2009, doi: 10.1111/j.1467-9280.2009.02380.x.
  - [67] J. C. Toscano, N. D. Anderson, M. Fabiani, G. Gratton, and S. M. Garnsey, “The time-course of cortical responses to speech revealed by fast optical imaging,” *Brain and Language*, vol. 184, pp. 32–42, Sep. 2018, doi: 10.1016/j.bandl.2018.06.006.
  - [68] F. H. Guenther and M. N. Gjaja, “The perceptual magnet effect as an emergent property of neural map formation,” *The Journal of the Acoustical Society of America*, vol. 100, no. 2, pp. 1111–1121, 1996.

- [69] H.-U. Bauer and R. Der, “Controlling the magnification factor of self-organizing feature maps,” *Neural computation*, vol. 8, no. 4, pp. 757–771, 1996.
- [70] F. H. Guenther, F. T. Husain, M. A. Cohen, and B. G. Shinn-Cunningham, “Effects of categorization and discrimination training on auditory perceptual space,” *The Journal of the Acoustical Society of America*, vol. 106, no. 5, pp. 2900–2912, 1999.
- [71] C. N. Price, C. Alain, and G. M. Bidelman, “Auditory-frontal channeling in  $\alpha$  and  $\beta$  bands is altered by age-related hearing loss and relates to speech perception in noise,” *Neuroscience*, vol. 423, pp. 18–28, 2019.
- [72] M. G. Preti, T. A. Bolton, and D. Van De Ville, “The dynamic functional connectome: State-of-the-art and perspectives,” *NeuroImage*, vol. 160, pp. 41–54, Oct. 2017, doi: 10.1016/j.neuroimage.2016.12.061.
- [73] M. Résibois *et al.*, “The neural basis of emotions varies over time: different regions go with onset-and offset-bound processes underlying emotion intensity,” *Social cognitive and affective neuroscience*, vol. 12, no. 8, pp. 1261–1271, 2017.
- [74] P. Pichot, “[DSM-III: the 3d edition of the Diagnostic and Statistical Manual of Mental Disorders from the American Psychiatric Association],” *Rev. Neurol. (Paris)*, vol. 142, no. 5, pp. 489–499, 1986.
- [75] S. Koelstra *et al.*, “Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos,” in *Brain informatics*, Springer, 2010, pp. 89–100.
- [76] R. Al-Fahad, M. Yeasin, A. I. Anam, and B. Elahian, “Selection of stable features for modeling 4-D affective space from EEG recording,” in *Neural Networks (IJCNN), 2017 International Joint Conference on*, 2017, pp. 1202–1209, [Online]. Available: <http://ieeexplore.ieee.org/abstract/document/7965989/>.
- [77] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, “Emotion recognition based on EEG using LSTM recurrent neural network,” *Emotion*, vol. 8, no. 10, pp. 355–358, 2017.
- [78] D. Van de Ville, J. Britz, and C. M. Michel, “EEG microstate sequences in healthy humans at rest reveal scale-free dynamics,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 42, pp. 18179–18184, 2010.
- [79] M. Kikuchi *et al.*, “EEG microstate analysis in drug-naive patients with panic disorder,” *PLoS One*, vol. 6, no. 7, 2011.
- [80] J. L. Cantero, M. Atienza, R. M. Salas, and C. M. Gómez, “Brain spatial microstates of human spontaneous alpha activity in relaxed wakefulness, drowsiness period, and REM sleep,” *Brain topography*, vol. 11, no. 4, pp. 257–263, 1999.
- [81] W. B. Cannon, “The James-Lange theory of emotions: A critical examination and an alternative theory,” *The American journal of psychology*, vol. 39, no. 1/4, pp. 106–124, 1927.
- [82] J. A. Russell, “A circumplex model of affect,” *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.
- [83] N. Karamzadeh, A. Medvedev, A. Azari, A. Gandjbakhche, and L. Najafizadeh, “Capturing dynamic patterns of task-based functional connectivity with EEG,” *NeuroImage*, vol. 66, pp. 311–317, Feb. 2013, doi: 10.1016/j.neuroimage.2012.10.032.
- [84] R. Hindriks *et al.*, “Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI?,” *NeuroImage*, vol. 127, pp. 242–256, Feb. 2016, doi: 10.1016/j.neuroimage.2015.11.055.
- [85] G. C. O’Neill *et al.*, “Measurement of dynamic task related functional networks using MEG,” *NeuroImage*, vol. 146, pp. 667–678, Feb. 2017, doi: 10.1016/j.neuroimage.2016.08.061.
- [86] C. M. Michel and T. Koenig, “EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review,” *NeuroImage*, vol. 180, pp. 577–593, Oct. 2018, doi: 10.1016/j.neuroimage.2017.11.062.

- [87] R. F. Betzel, M. A. Erickson, M. Abell, B. F. O'Donnell, W. P. Hetrick, and O. Sporns, "Synchronization dynamics and evidence for a repertoire of network states in resting EEG," *Front. Comput. Neurosci.*, vol. 6, 2012, doi: 10.3389/fncom.2012.00074.
- [88] A. G. Mahyari, D. M. Zoltowski, E. M. Bernat, and S. Aviyente, "A Tensor Decomposition-Based Approach for Detecting Dynamic Network States From EEG," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 225–237, Jan. 2017, doi: 10.1109/TBME.2016.2553960.
- [89] A. Ozdemir, E. M. Bernat, and S. Aviyente, "Recursive Tensor Subspace Tracking for Dynamic Brain Network Analysis," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 4, pp. 669–682, Dec. 2017, doi: 10.1109/TSIPN.2017.2668146.
- [90] N. Leonardi *et al.*, "Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest," *NeuroImage*, vol. 83, pp. 937–950, Dec. 2013, doi: 10.1016/j.neuroimage.2013.07.019.
- [91] N. T. Duc and B. Lee, "Microstate functional connectivity in EEG cognitive tasks revealed by a multivariate Gaussian hidden Markov model with phase locking value," *J. Neural Eng.*, vol. 16, no. 2, p. 026033, Feb. 2019, doi: 10.1088/1741-2552/ab0169.
- [92] M. J. Beal, Z. Ghahramani, and C. E. Rasmussen, "The Infinite Hidden Markov Model," in *Advances in Neural Information Processing Systems 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. MIT Press, 2002, pp. 577–584.
- [93] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Sharing Clusters among Related Groups: Hierarchical Dirichlet Processes," in *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou, Eds. MIT Press, 2005, pp. 1385–1392.
- [94] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "A STICKY HDP-HMM WITH APPLICATION TO SPEAKER DIARIZATION," *The Annals of Applied Statistics*, vol. 5, no. 2A, pp. 1020–1056, 2011.
- [95] N. Foti, J. Xu, D. Laird, and E. Fox, "Stochastic variational inference for hidden Markov models," in *Advances in neural information processing systems*, 2014, pp. 3599–3607.
- [96] M. J. Johnson and A. S. Willsky, "Stochastic Variational Inference for Bayesian Time Series Models," p. 9.
- [97] C. Wang and D. M. Blei, "A split-merge MCMC algorithm for the hierarchical Dirichlet process," *arXiv preprint arXiv:1201.1657*, 2012.
- [98] J. Chang and J. W. Fisher III, "Parallel Sampling of HDPs using Sub-Cluster Splits," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 235–243.
- [99] E. B. Fox, M. C. Hughes, E. B. Sudderth, and M. I. Jordan, "Joint modeling of multiple time series via the beta process with application to motion capture segmentation," *Ann. Appl. Stat.*, vol. 8, no. 3, pp. 1281–1313, Sep. 2014, doi: 10.1214/14-AOAS742.
- [100] M. C. Hughes and E. Sudderth, "Memoized Online Variational Inference for Dirichlet Process Mixture Models," in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2013, pp. 1133–1141.
- [101] M. Hughes, D. I. Kim, and E. Sudderth, "Reliable and scalable variational inference for the hierarchical dirichlet process," in *Artificial Intelligence and Statistics*, 2015, pp. 370–378.
- [102] C.-H. Pui *et al.*, "Treating childhood acute lymphoblastic leukemia without cranial irradiation," *New England Journal of Medicine*, vol. 360, no. 26, pp. 2730–2741, 2009.

- [103] B. Fischl *et al.*, “Automatically parcellating the human cerebral cortex,” *Cereb. Cortex*, vol. 14, no. 1, pp. 11–22, Jan. 2004.
- [104] T. E. J. Behrens, H. J. Berg, S. Jbabdi, M. F. S. Rushworth, and M. W. Woolrich, “Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?,” *Neuroimage*, vol. 34, no. 1, pp. 144–155, Jan. 2007, doi: 10.1016/j.neuroimage.2006.09.018.
- [105] M. L. Bermingham *et al.*, “Application of high-dimensional feature selection: evaluation for genomic prediction in man,” *Scientific reports*, vol. 5, 2015, [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4437376/>.
- [106] R. Al-Fahad and M. Yeasin, “Robust Modeling of Continuous 4-D Affective Space from EEG Recording,” in *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*, 2016, pp. 1040–1045, [Online]. Available: <http://ieeexplore.ieee.org/abstract/document/7838292/>.
- [107] D. Freedman, “A remark on the difference between sampling with and without replacement,” *Journal of the American Statistical Association*, vol. 72, no. 359, pp. 681–681, 1977.
- [108] P. Büchlmann and B. Yu, “Analyzing bagging,” *Annals of Statistics*, pp. 927–961, 2002.
- [109] R. D. Shah and R. J. Samworth, “Variable selection with error control: Another look at Stability Selection,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 75, no. 1, pp. 55–80, Jan. 2013, doi: 10.1111/j.1467-9868.2011.01034.x.
- [110] P. Zhao and B. Yu, “On model selection consistency of Lasso,” *Journal of Machine learning research*, vol. 7, no. Nov, pp. 2541–2563, 2006.
- [111] L. Breiman, “Using adaptive bagging to debias regressions,” Technical Report 547, Statistics Dept. UCB, 1999. Accessed: Feb. 21, 2017. [Online]. Available: <https://statistics.berkeley.edu/sites/default/files/tech-reports/547.pdf>.
- [112] V. N. Temlyakov, “Weak greedy algorithms [\*] This research was supported by National Science Foundation Grant DMS 9970326 and by ONR Grant N00014-96-1-1003,” *Advances in Computational Mathematics*, vol. 12, no. 2, pp. 213–227, 2000.
- [113] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, and others, “Least angle regression,” *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [114] N. J. Rose, “Linear Algebra and Its Applications (Gilbert Strang),” *SIAM Rev.*, vol. 24, no. 4, pp. 499–501, Oct. 1982, doi: 10.1137/1024124.
- [115] B. P. Staresina and L. Davachi, “Mind the gap: binding experiences across space and time in the human hippocampus,” *Neuron*, vol. 63, no. 2, pp. 267–276, 2009.
- [116] H. Eichenbaum, “Hippocampus: cognitive processes and neural representations that underlie declarative memory,” *Neuron*, vol. 44, no. 1, pp. 109–120, 2004.
- [117] T. Bartsch, J. Döhring, A. Rohr, O. Jansen, and G. Deuschl, “CA1 neurons in the human hippocampus are critical for autobiographical memory, mental time travel, and autonoetic consciousness,” *PNAS*, vol. 108, no. 42, pp. 17562–17567, Oct. 2011, doi: 10.1073/pnas.1110266108.
- [118] A. K. Barbey, M. Koenigs, and J. Grafman, “Orbitofrontal Contributions to Human Working Memory,” *Cereb Cortex*, vol. 21, no. 4, pp. 789–795, Apr. 2011, doi: 10.1093/cercor/bhq153.
- [119] M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman, “A default mode of brain function,” *Proceedings of the National Academy of Sciences*, vol. 98, no. 2, pp. 676–682, 2001.
- [120] D. A. Gusnard and M. E. Raichle, “Searching for a baseline: functional imaging and the resting human brain,” *Nature reviews neuroscience*, vol. 2, no. 10, p. 685, 2001.

- [121] B. Hahn, T. J. Ross, and E. A. Stein, "Cingulate activation increases dynamically with response speed under stimulus unpredictability," *Cerebral cortex*, vol. 17, no. 7, pp. 1664–1671, 2006.
- [122] G. V. Raymond, M. L. Bauman, and T. L. Kemper, "Hippocampus in autism: a Golgi analysis," *Acta Neuropathol*, vol. 91, no. 1, pp. 117–119, Dec. 1995, doi: 10.1007/s004010050401.
- [123] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [124] J. G. Foster, D. V. Foster, P. Grassberger, and M. Paczuski, "Edge direction and the structure of networks," *Proceedings of the National Academy of Sciences*, vol. 107, no. 24, pp. 10815–10820, 2010.
- [125] M. Piraveenan, M. Prokopenko, and A. Zomaya, "Assortative mixing in directed biological networks," *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, vol. 9, no. 1, pp. 66–78, 2012.
- [126] M. E. Newman, "Assortative mixing in networks," *Physical review letters*, vol. 89, no. 20, p. 208701, 2002.
- [127] M. Chen, K. Kuzmin, and B. K. Szymanski, "Community detection via maximization of modularity and its variants," *IEEE Transactions on Computational Social Systems*, vol. 1, no. 1, pp. 46–65, 2014.
- [128] M. E. Newman, "Fast algorithm for detecting community structure in networks," *Physical review E*, vol. 69, no. 6, p. 066133, 2004.
- [129] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.
- [130] R. C. Oldfield, "The assessment and analysis of handedness: The Edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, Mar. 1971, doi: 10.1016/0028-3932(71)90067-4.
- [131] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clinical Neurophysiology*, vol. 112, no. 4, pp. 713–719, Apr. 2001, doi: 10.1016/S1388-2457(00)00527-7.
- [132] T. W. Picton, P. van Roon, M. L. Armilio, P. Berg, N. Ille, and M. Scherg, "The correction of ocular artifacts: a topographic perspective," *Clinical Neurophysiology*, vol. 111, no. 1, pp. 53–65, Jan. 2000, doi: 10.1016/S1388-2457(99)00227-8.
- [133] G. M. Bidelman and L. Dexter, "Bilinguals at the 'cocktail party': Dissociable neural activity in auditory–linguistic brain regions reveals neurobiological basis for nonnative listeners' speech-in-noise recognition deficits," *Brain and language*, vol. 143, pp. 32–41, 2015.
- [134] G. M. Bidelman and M. Howell, "Functional changes in inter-and intra-hemispheric cortical processing underlying degraded speech perception," *Neuroimage*, vol. 124, pp. 581–590, 2016.
- [135] F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy, "Brainstorm: a user-friendly application for MEG/EEG analysis," *Computational intelligence and neuroscience*, vol. 2011, p. 8, 2011.
- [136] M. Fuchs, R. Drenckhahn, H. Wischmann, and M. Wagner, "An improved boundary element method for realistic volume-conductor modeling," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 8, pp. 980–997, Aug. 1998, doi: 10.1109/10.704867.
- [137] M. Fuchs, J. Kastner, M. Wagner, S. Hawes, and J. S. Ebersole, "A standardized boundary element method volume conductor model," *Clinical Neurophysiology*, vol. 113, no. 5, pp. 702–712, May 2002, doi: 10.1016/S1388-2457(02)00030-5.

- [138] J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster, "A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM)," *Neuroimage*, vol. 2, no. 2, pp. 89–101, Jun. 1995.
- [139] Z. A. Acar and S. Makeig, "Effects of forward model errors on EEG source localization," *Brain topography*, vol. 26, no. 3, pp. 378–396, 2013.
- [140] S. Y. Shirazi and H. J. Huang, "More Reliable EEG Electrode Digitizing Methods Can Reduce Source Estimation Uncertainty, But Current Methods Already Accurately Identify Brodmann Areas," *bioRxiv*, p. 557074, 2019.
- [141] A. Gramfort, T. Papadopoulou, E. Olivi, and M. Clerc, "OpenMEEG: opensource software for quasistatic bioelectromagnetics," *Biomedical engineering online*, vol. 9, no. 1, p. 45, 2010.
- [142] C. M. Michel, M. M. Murray, G. Lantz, S. Gonzalez, L. Spinelli, and R. Grave de Peralta, "EEG source imaging," *Clin Neurophysiol*, vol. 115, no. 10, pp. 2195–2222, Oct. 2004, doi: 10.1016/j.clinph.2004.06.001.
- [143] J. Song *et al.*, "EEG source localization: Sensor density and head surface coverage," *Journal of Neuroscience Methods*, vol. 256, pp. 9–21, Dec. 2015, doi: 10.1016/j.jneumeth.2015.08.015.
- [144] R. S. Desikan *et al.*, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.
- [145] P. Bashivan, M. Yeasin, and G. M. Bidelman, "Temporal progression in functional connectivity determines individual differences in working memory capacity," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 2943–2949.
- [146] C. Brunner, M. Billinger, M. Seeber, T. R. Mullen, and S. Makeig, "Volume Conduction Influences Scalp-Based Connectivity Estimates," *Front. Comput. Neurosci.*, vol. 10, 2016, doi: 10.3389/fncom.2016.00121.
- [147] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*, vol. 112. Springer, 2013.
- [148] K. J. Friston, P. Jezzard, and R. Turner, "Analysis of functional MRI time-series," *Human Brain Mapping*, vol. 1, no. 2, pp. 153–171, 1994, doi: 10.1002/hbm.460010207.
- [149] D. Zhou, W. K. Thompson, and G. Siegle, "MATLAB toolbox for functional connectivity," *NeuroImage*, vol. 47, no. 4, pp. 1590–1607, Oct. 2009, doi: 10.1016/j.neuroimage.2009.05.089.
- [150] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Human brain mapping*, vol. 8, no. 4, pp. 194–208, 1999.
- [151] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, "Identifying true brain interaction from EEG data using the imaginary part of coherency," *Clinical neurophysiology*, vol. 115, no. 10, pp. 2292–2307, 2004.
- [152] C. J. Stam, P. Tewarie, E. Van Dellen, E. C. W. van Straaten, A. Hillebrand, and P. Van Mieghem, "The trees and the forest: Characterization of complex brain networks with minimum spanning trees," *International Journal of Psychophysiology*, vol. 92, no. 3, pp. 129–138, Jun. 2014, doi: 10.1016/j.ijpsycho.2014.04.001.
- [153] H.-J. Hwang, S. Kim, S. Choi, and C.-H. Im, "EEG-Based Brain-Computer Interfaces: A Thorough Literature Survey," *International Journal of Human-Computer Interaction*, vol. 29, no. 12, pp. 814–826, Dec. 2013, doi: 10.1080/10447318.2013.780869.
- [154] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," *J. Neural Eng.*, vol. 15, no. 3, p. 031005, Apr. 2018, doi: 10.1088/1741-2552/aab2f2.
- [155] R. Lowry, "Concepts and applications of inferential statistics," 2014.

- [156] R. Salvador, J. Suckling, M. R. Coleman, J. D. Pickard, D. Menon, and E. D. Bullmore, "Neurophysiological architecture of functional magnetic resonance images of human brain," *Cerebral cortex*, vol. 15, no. 9, pp. 1332–1342, 2005.
- [157] L. Wang, Y. Li, P. Metzack, Y. He, and T. S. Woodward, "Age-related changes in topological patterns of large-scale brain functional networks during memory encoding and recognition," *Neuroimage*, vol. 50, no. 3, pp. 862–872, 2010.
- [158] W. Li, M. Wang, Y. Li, Y. Huang, and X. Chen, "A novel brain network construction method for exploring age-related functional reorganization," *Computational Intelligence and Neuroscience*, vol. 2016, p. 5, 2016.
- [159] M. Xia, J. Wang, and Y. He, "BrainNet Viewer: a network visualization tool for human brain connectomics," *PloS one*, vol. 8, no. 7, p. e68910, 2013.
- [160] R. F. Betzel, M. A. Bertolero, and D. S. Bassett, "Non-assortative community structure in resting and task-evoked functional brain networks," *bioRxiv*, p. 355016, 2018.
- [161] N. Golestani, T. Paus, and R. J. Zatorre, "Anatomical Correlates of Learning Novel Speech Sounds," *Neuron*, vol. 35, no. 5, pp. 997–1010, Aug. 2002, doi: 10.1016/S0896-6273(02)00862-0.
- [162] N. Golestani and R. J. Zatorre, "Learning new sounds of speech: reallocation of neural substrates," *NeuroImage*, vol. 21, no. 2, pp. 494–506, Feb. 2004, doi: 10.1016/j.neuroimage.2003.09.071.
- [163] Y.-S. Lee, P. Turkeltaub, R. Granger, and R. D. S. Raizada, "Categorical Speech Processing in Broca's Area: An fMRI Study Using Multivariate Pattern-Based Analysis," *J. Neurosci.*, vol. 32, no. 11, pp. 3942–3948, Mar. 2012, doi: 10.1523/JNEUROSCI.3814-11.2012.
- [164] G. Feng, Z. Gan, S. Wang, P. C. M. Wong, and B. Chandrasekaran, "Task-General and Acoustic-Invariant Neural Representation of Speech Categories in the Human Brain," *Cereb Cortex*, vol. 28, no. 9, pp. 3241–3254, Sep. 2018, doi: 10.1093/cercor/bhx195.
- [165] S. Luthra, S. Guediche, S. E. Blumstein, and E. B. Myers, "Neural substrates of subphonemic variation and lexical competition in spoken word recognition," *Language, Cognition and Neuroscience*, vol. 34, no. 2, pp. 151–169, Feb. 2019, doi: 10.1080/23273798.2018.1531140.
- [166] D. J. Acheson and P. Hagoort, "Stimulating the brain's language network: syntactic ambiguity resolution after TMS to the inferior frontal gyrus and middle temporal gyrus," *Journal of cognitive neuroscience*, vol. 25, no. 10, pp. 1664–1677, 2013.
- [167] G. Hickok and D. Poeppel, "Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language," *Cognition*, vol. 92, no. 1, pp. 67–99, May 2004, doi: 10.1016/j.cognition.2003.10.011.
- [168] G. Hickok and D. Poeppel, "The cortical organization of speech processing," *Nature Reviews Neuroscience*, vol. 8, no. 5, pp. 393–402, May 2007, doi: 10.1038/nnr2113.
- [169] Y. Sakurai, I. Mimura, and T. Mannen, "Agraphia for kanji resulting from a left posterior middle temporal gyrus lesion," *Behavioural neurology*, vol. 19, no. 3, pp. 93–106, 2008.
- [170] S. C. Blank, S. K. Scott, K. Murphy, E. Warburton, and R. J. Wise, "Speech production: Wernicke, Broca and beyond," *Brain*, vol. 125, no. 8, pp. 1829–1838, 2002.
- [171] M. Ingallhalikar *et al.*, "Sex differences in the structural connectome of the human brain," *PNAS*, vol. 111, no. 2, pp. 823–828, Jan. 2014, doi: 10.1073/pnas.1316909110.
- [172] B. A. Shaywitz *et al.*, "Sex differences in the functional organization of the brain for language," *Nature*, vol. 373, no. 6515, p. 607, Feb. 1995, doi: 10.1038/373607a0.
- [173] A. Mechelli, G. W. Humphreys, K. Mayall, A. Olson, and C. J. Price, "Differential effects of word length and visual contrast in the fusiform and lingual gyri during,"

- Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 267, no. 1455, pp. 1909–1913, 2000.
- [174] D. Howard *et al.*, “THE CORTICAL LOCALIZATION OF THE LEXICONSPOSITRON EMISSION TOMOGRAPHY EVIDENCE,” *Brain*, vol. 115, no. 6, pp. 1769–1782, Dec. 1992, doi: 10.1093/brain/115.6.1769.
  - [175] S. Y. Bookheimer, T. A. Zeffiro, T. Blaxton, W. Gaillard, and W. Theodore, “Regional cerebral blood flow during object naming and word reading,” *Human Brain Mapping*, vol. 3, no. 2, pp. 93–106, 1995, doi: 10.1002/hbm.460030206.
  - [176] Y. Du, B. R. Buchsbaum, C. L. Grady, and C. Alain, “Noise differentially impacts phoneme representations in the auditory and speech motor systems,” *PNAS*, vol. 111, no. 19, pp. 7126–7131, May 2014, doi: 10.1073/pnas.1318738111.
  - [177] S. Bouton *et al.*, “Focal versus distributed temporal cortex activity for speech sound category assignment,” *PNAS*, vol. 115, no. 6, pp. E1299–E1308, Feb. 2018, doi: 10.1073/pnas.1714279115.
  - [178] F. Bach, “Bolasso: model consistent Lasso estimation through the bootstrap,” *arXiv:0804.1302 [cs, math, stat]*, Apr. 2008, Accessed: May 12, 2019. [Online]. Available: <http://arxiv.org/abs/0804.1302>.
  - [179] S. Haufe, V. V. Nikulin, K.-R. Müller, and G. Nolte, “A critical assessment of connectivity measures for EEG data: a simulation study,” *Neuroimage*, vol. 64, pp. 120–133, 2013, doi: 10.1016/j.neuroimage.2012.09.036.
  - [180] S. Koelstra *et al.*, “Deap: A database for emotion analysis; using physiological signals,” *Affective Computing, IEEE Transactions on*, vol. 3, no. 1, pp. 18–31, 2012.
  - [181] M. C. Hughes and E. B. Sudderth, “Bnpy: Reliable and scalable variational inference for bayesian nonparametric models,” in *Proceedings of the NIPS Probabilistic Programming Workshop, Montreal, QC, Canada*, 2014, pp. 8–13.
  - [182] A. Khanna, A. Pascual-Leone, C. M. Michel, and F. Farzan, “Microstates in resting-state EEG: Current status and future directions,” *Neuroscience & Biobehavioral Reviews*, vol. 49, pp. 105–113, Feb. 2015, doi: 10.1016/j.neubiorev.2014.12.010.
  - [183] K. J. Miller, G. Schalk, E. E. Fetz, M. den Nijs, J. G. Ojemann, and R. P. N. Rao, “Cortical activity during motor execution, motor imagery, and imagery-based online feedback,” *PNAS*, vol. 107, no. 9, pp. 4430–4435, Mar. 2010, doi: 10.1073/pnas.0913697107.
  - [184] T. Iordanov, H. Bornfleth, K. Hoechstetter, B. Lanfer, and M. Scherg, “Performance of cortical LORETA and cortical CLARA applied to MEG data,” in *Biomag*, 2016.
  - [185] R. D. Pascual-Marqui, M. Esslen, K. Kochi, and D. Lehmann, “Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review,” *Methods and findings in experimental and clinical pharmacology*, vol. 24, no. Suppl C, pp. 91–95, 2002.
  - [186] J. E. Richards, C. Sanchez, M. Phillips-Meek, and W. Xie, “A database of age-appropriate average MRI templates,” *Neuroimage*, vol. 124, pp. 1254–1259, 2016.
  - [187] P. J. Uhlhaas, G. Pipa, S. Neuenschwander, M. Wibral, and W. Singer, “A new look at gamma? High-(> 60 Hz)  $\gamma$ -band activity in cortical networks: function, mechanisms and impairment,” *Progress in biophysics and molecular biology*, vol. 105, no. 1–2, pp. 14–28, 2011.
  - [188] W. Klimesch, M. Doppelmayr, H. Russegger, and T. Pachinger, “Theta band power in the human scalp EEG and the encoding of new information,” *NeuroReport*, vol. 7, no. 7, 1996, [Online]. Available: [https://journals.lww.com/neuroreport/Fulltext/1996/05170/Theta\\_band\\_power\\_in\\_the\\_human\\_scalp\\_EEG\\_and\\_the.2.aspx](https://journals.lww.com/neuroreport/Fulltext/1996/05170/Theta_band_power_in_the_human_scalp_EEG_and_the.2.aspx).



- [189] J. F. Cavanagh and M. J. Frank, “Frontal theta as a mechanism for cognitive control,” *Trends in Cognitive Sciences*, vol. 18, no. 8, pp. 414–421, Aug. 2014, doi: 10.1016/j.tics.2014.04.012.
- [190] M. I. Posner and C. R. R. Snyder, *Attention and cognitive control. Information processing and cognition: The Loyola symposium*. Hillsdale NJ: Erlbaum, 1975.
- [191] M. J. Wainwright and M. I. Jordan, “Graphical Models, Exponential Families, and Variational Inference,” *MAL*, vol. 1, no. 1–2, pp. 1–305, Nov. 2008, doi: 10.1561/22000000001.