Unsupervised decoding of single-trial EEG reveals unique states of functional brain connectivity that drive rapid speech categorization decisions

Author(s): Rakib Al-Fahad[1], Mohammed Yeasin[1,2], Gavin M. Bidelman[2,3,4,†]

[1]Department of Electrical and Computer Engineering, University of Memphis, Memphis, 38152 TN, USA
[2]Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
[3]School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA
[4]University of Tennessee Health Sciences Center, Department of Anatomy and Neurobiology, Memphis, TN, USA

[†]**Address for editorial correspondence:**
Gavin M. Bidelman, PhD
School of Communication Sciences & Disorders
University of Memphis
4055 North Park Loop
Memphis, TN, 38152
TEL: (901) 678-5826
FAX: (901) 525-1282
EMAIL: gmbdlman@memphis.edu

ABSTRACT

Categorical perception (CP) is an inherent property of speech perception. The response time (RT) of listeners' perceptual speech identification are 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 perceivers tended to utilize functional brain networks excessively (or inappropriately) whereas fast perceivers (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 90% accuracy (AUC=0.9). Our results corroborate previous studies by confirming the engagement of similar temporal (STG), parietal, motor, and prefrontal regions in CP using an entirely data-driven approach.

78 **INTRODUCTION**
79
80   When identifying speech, listeners naturally group sounds into smaller sets of discrete
81 (phonetic) categories through the process of categorical perception (CP) (Harnad and Bureau,
82 1987; Liberman et al., 1967; Pisoni, 1973; Pisoni and Luce, 1987). Presumably, this type of
83 behavioral "downsampling" promotes speech comprehension by generating perceptual constancy
84 in the face of enormous physical variation in multiple acoustic dimensions, e.g., talker variability
85 in tempo, pitch, or timbre (Prather et al., 2009). CP is often characterized by sharp (stair-stepped)
86 identification and peaked (better) discrimination functions near the categorical boundary when
87 classifying an otherwise equidistant acoustic continuum.
88   Germane to the present study, response time (RT) data also reveal differences in the
89 speed of listeners' of categorical decisions (Bidelman et al., 2013; Pisoni and Tash, 1974). In
90 perceptual labeling tasks, for example, listeners categorize prototypical speech sounds (e.g.,
91 exemplars from their native language) much faster than their ambiguous or less familiar
92 counterparts (e.g., nonnative speech sounds) (Bidelman and Lee, 2015). RTs also slow near
93 perceptual speech boundaries, where listeners shift from hearing one linguistic class to another
94 (e.g., /u/ vs. /a/ vowel) and presumably require more time to access the "correct" speech template
95 (Bidelman et al., 2013; Liebenthal et al., 2010; Pisoni and Tash, 1974; Reetzke et al., 2018).
96 Relatedly, RTs vary with task manipulations and individual differences in speech perception in
97 different populations. Studies demonstrate listeners' speed in speech identification is highly
98 sensitive to stimulus familiarity (Bidelman and Walker, 2017; Liebenthal et al., 2010; Lively et
99 al., 1993), auditory plasticity of short- (Liberman et al., 1967) and long-term (Bidelman et al.,
100 2014b; Bidelman and Alain, 2015; Bidelman and Lee, 2015) experience, and neuropathologies
101 and language-learning disorders (e.g., (Bidelman et al., 2017, 2014a; Calcus Axelle et al., 2016;
102 Hakvoort Britt et al., 2016)). Given its fundamental role in the perceptual organization of speech,
103 understanding individual differences in CP and its underlying neurobiology is among the broad
104 interests to understand how sensory features are mapped to higher order perception (Bidelman et
105 al., 2013; Phillips, 2001; Pisoni and Luce, 1987).
106   The neuronal elements of the brain organize in complicated structural networks (Cajal,
107 1995). Increasingly, it is appreciated that anatomical substrates constrain the dynamic emergence
108 of coherent physiological activity that can span multiple spatially distinct brain regions (Bressler,
109 1995; Bullmore and Sporns, 2009; Fries, 2005). Such densely intra-connected, sparsely inter-
110 connected, dynamic connected networks are thought to provide the functional basis for
111 information processing, mental representations, and complex behaviors (Bassett and Bullmore,
112 2006; Honey et al., 2007; Newman, 2003; Tononi et al., 1994). In this regard, neuroimaging
113 studies have identified several functional brain regions that are important to CP including
114 primary auditory cortex, left inferior frontal areas (i.e., Broca's area), and middle temporal gyri
115 (e.g., Guenther Frank H. et al. 2004; Binder et al. 2004; Myers et al. 2009; Chang et al. 2010;
116 Liebenthal et al. 2010; Bidelman and Lee 2015; Alho et al. 2016; Toscano et al. 2018). Previous
117 studies also suggest that more neurons are preferentially activated by the prototypes of the
118 speech categories compared to those at category boundaries (Guenther and Gjaja, 1996).
119 Similarly, improved discriminability at category boundaries could reflect an increased number of
120 neurons encoding sensory cues at these perceptual transitions (Bauer and Der, 1996; Guenther et
121 al., 1999). Such neuronal overrepresentations warp the sensory space and may account for the
122 aforementioned RT effects in speech categorization. Still, while the neural correlates of CP have
123 been well studied in terms of the regional contributions to behavior, we are aware of no studies
124 that have investigated the mechanisms of speech CP from a full-brain (functional connectivity)

3

125   perspective. Here, we focus on the speed (RT) of listeners perceptual speech identification as
126   RTs are highly sensitivity to individual differences in CP (Bidelman et al., 2014b, 2014a;
127   Bidelman and Alain, 2015; Bidelman and Walker, 2017) and reflect an objective, continuous
128   measure of perceptual categorization skill.
129   Functional connectivity matrices derived from neuroimaging data are highly sparse and
130   reflect high dimensional data. Hence, finding RT-related network edges is challenging. State-of-
131   the-art studies usually use naive approaches to discover and analyze each edge individually and
132   then compensate for possible errors arising from multiple comparisons (e.g., family-wise error or
133   false discovery rate). These studies mostly yield an unstable set of network edges that are highly
134   sensitive to changes in the hyperparameters within and between datasets (e.g., neural responses
135   from different populations). In this regard, variable selection attempts to identify the most salient
136   subset of variables from a larger set of features mixed with irrelevant variables. This problem is
137   especially challenging when the number of available data samples is smaller compared to the
138   number of possible predictors. Using generic subsampling and high-dimensional selection
139   algorithms, stability selection can yield a stable set of features that distinguish subgroups of the
140   data (e.g., here, listeners with slow vs. fast perceptual decisions). It has widely been used in
141   diverse fields of science including gene selection and neuroimaging. One of the downsides of
142   multivariate approaches is that outcomes often depend on model parameters (e.g., regularization
143   factor). Compared to conventional multivariate approaches, stability selection produces more
144   reliable estimations because of its internal randomization implemented as bootstrap-based
145   subsampling (Meinshausen and Bühlmann, 2010; Shah and Samworth, 2013). Here, we propose
146   a systematic approach to determine and rank RT-related functional connectivity among brain
147   regions that are consistent across model parameters. In doing so, we identify, objectively, the
148   most important properties (i.e., features) of the functional EEG connectome that describe
149   perceptual processing with regard to categorization.
150   The primary aim of this study was to test whether individual differences in speeded
151   speech categorization could be explained in terms of network-level descriptions of brain activity.
152   Our first goal was to focus on graph theoretical approaches to analyze the complex networks that
153   could provide a powerful new way of quantifying individual differences in speech perception. A
154   second goal was to discover which aspects of those functional connectivity networks best
155   explained the variation and diversity in listeners' perceptual responses during speech sound
156   categorization. We recoded high-density electroencephalograms (EEGs) while listeners rapidly
157   classified speech in a speeded vowel identification task (Bidelman et al., 2013; Bidelman and
158   Walker, 2017). We then applied graph analyses to source-localized EEG responses to derive the
159   underlying functional brain networks related to speech categorization. Using Bayesian non-
160   parametric modeling, we then show that speeded categorical decisions unfold in three RT
161   clusters that distinguish subgroups of listeners based on their behavioral performance (i.e., slow,
162   medium, and fast perceivers). Applying state-of-the-art machine learning and stability selection
163   analyses to neural data we further show that local and global network properties of brain
164   connectomics can decode group differences in behavioral CP performance with 92% accuracy
165   (AUC=0.9). Our findings demonstrate that slow RT decisions related to categorical speech
166   perception involve improper (or excessive) utilization of functional brain networks underlying
167   speech whereas fast and medium perceivers show less utilization.
168
169
170

METHODS

Participants

Thirty-five adults (12 male, 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± 36.4 laterality index (Oldfield, 1971)), and had obtained a collegiate level of education (17.2 ± 2.9 years). None had any history of neuropsychiatric illness. On average, participants had 5.1± 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 1 (A, B) shows the distribution of demographic measures (gender and age) of participants.
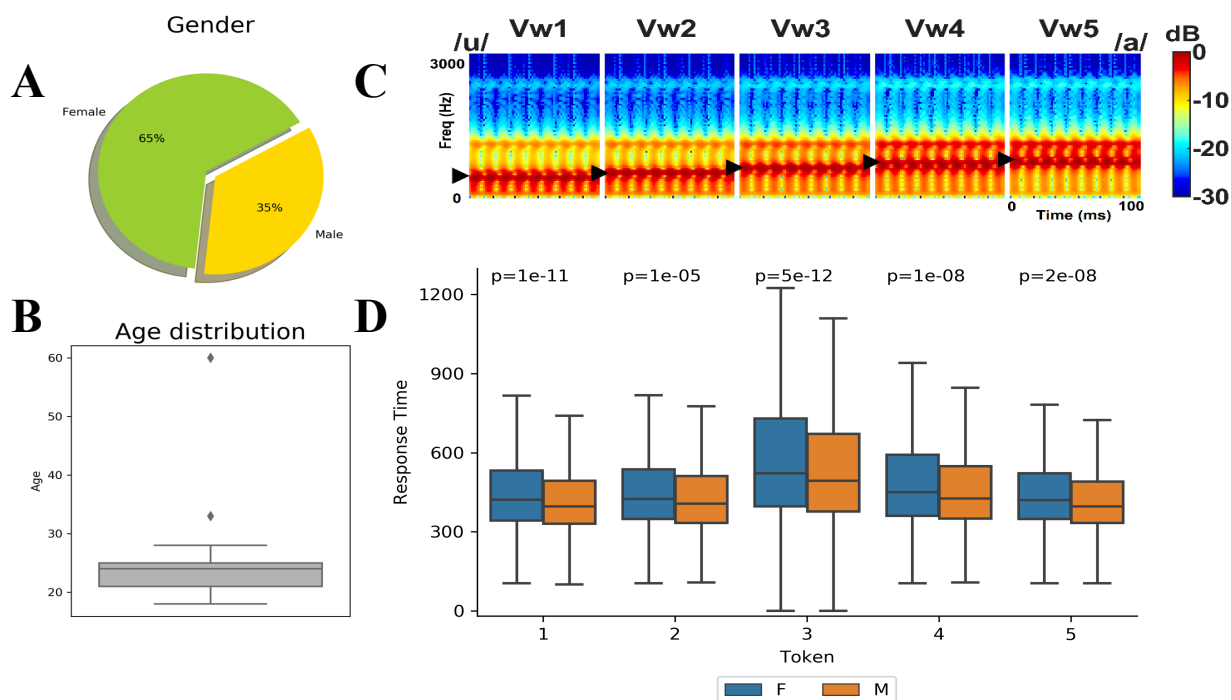

Figure 1: (A, B) Demographic gender and age distributions. (C) Acoustic spectrograms of the speech stimuli: The stimulus continuum was created by parametrically changing vowel first formant frequency over five equal steps from 430 to 730 Hz (▶), resulting in a perceptual-phonetic continuum from /u/ to /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 (Oldfield, 1971) (Figure 1C). 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

5

198 spanned a perceptual phonetic continuum from /u/ to /a/ (Bidelman et al., 2013). Speech stimuli
199 were delivered binaurally at 83 dB SPL through shielded insert earphones (ER-2; Etymotic
200 Research) coupled to a TDT RP2 processor (Tucker Davis Technologies).
201   During EEG recording, listeners heard 150-200 trials of each individual speech token. On
202 each trial, they were asked to label the sound with a binary response ("u" or "a") as quickly and
203 accurately as possible (speeded classification task). Reaction times (RTs) were logged,
204 calculated as the timing difference between stimulus onset and listeners' behavioral response.
205 Following their keypress, the inter-stimulus interval (ISI) was jittered randomly between 800 and
206 1000 ms (20 ms steps, uniform distribution) and the next trial was commenced.

Behavioral data analysis

208   We adopted classical Gaussian mixture modelling (GMM) with expectation-
209 maximization (EM) to identify an optimal number of clusters (i.e., subgroups of listeners) from
210 the distribution of their RT speeds (see Figure 1D). GMMs are probabilistic models that assume
211 the data are generated from a mixture of a finite number of Gaussian distributions (components)
212 with unknown parameters. Mixture models generalize k-means clustering to incorporate
213 information about the covariance structure of the data as well as the centers of the latent
214 Gaussians. Unlike Bayesian procedures, such inferences are prior-free. However, finding an
215 optimal number of components is challenging. The Bayesian Information Criterion (BIC) can be
216 used to select the number of components in a GMM, if data is generated from an i.i.d. mixture of
217 Gaussian distributions. In this study, we used brute-force and BIC based approaches as an
218 alternative solution to the Variational Bayesian Gaussian mixture model. In this exhaustive
219 parameter search, the hyper parameters were (1) Number of components (clusters), (ranges from
220 1 to 14); (2) Type of covariance parameters ('full': each component has its own general
221 covariance matrix; 'tied': all components share the same general covariance matrix; 'diag': each
222 component has its own diagonal covariance matrix; or 'spherical': each component has its own
223 single variance).
224   This identified an optimal combination four components with unique covariance matrix.
225 Figure 2A shows the BIC scores while tuning parameters. The '*' indicates the optimal
226 combination of components. The probability of each component (see Figure 2B) shows that most
227 trials fall into components 1- 3 ranging from 17% - 47% of the total trials in the speech
228 identification task. Component 4 has the fewest number of trials (1.6%). Based on the
229 interpretation of RTs, we categorized these components as Fast RT (Cluster 2, 120 - 476 ms),
230 Medium RT (Cluster 3, 478 - 722 ms), Slow RT (Cluster 1, 724 -1430 ms), and Outliers (Cluster
231 4, 1432 - 2500 ms). The outliers (Cluster 4) were discarded for further analysis given the low
232 trial counts loading into this cluster. The boxplot in Figure 2C shows token wise response times.
233 Each speech token can be broken down into a combination of the three RT clusters, meaning that
234 speech categorization speeds could be objectively clustered into fast, medium, slow (and
235 outliers) responses via the GMM. These cluster divisions were then used in subsequent EEG
236 analyses to determine if functional brain connectomics differentiated these subgroups of CP
237 performers.
238
EEG recording and preprocessing
240 *Recording and preprocessing.* EEG recording procedures were identical to our previous
241 neuroimaging studies on CP (e.g., (Bidelman et al., 2013; Bidelman and Alain, 2015; Bidelman
242 and Walker, 2017)). Briefly, neuroelectric activity was recorded from 64 sintered Ag/AgCl
243 electrodes at standard 10-10 locations around the scalp (Oostenveld and Praamstra, 2001).

6

244 Continuous data were digitized using a sampling rate of 500 Hz (SynAmps RT amplifiers;
245 Compumedics Neuroscan) and an online passband of DC-200 Hz. Electrodes placed on the outer
246 canthi of the eyes and the superior and inferior orbit monitored ocular movements. Contact
247 impedances were maintained < 10 kΩ during data collection. During acquisition, electrodes were
248 referenced to an additional sensor placed ~ 1 cm posterior to the Cz channel.
249   Subsequent pre-processing was performed in BESA® Research (v7) (BESA, GmbH).
250 Ocular artifacts (saccades and blinks) were first corrected in the continuous EEG using a
251 principal component analysis (PCA) (Picton et al., 2000). Cleaned EEGs were then filtered
252 (bandpass: 1-100 Hz; notch filter: 60 Hz), epoched (-200-800 ms) into single trials, baseline
253 corrected to the pre-stimulus interval, and re-referenced to the common average of the scalp.
254 This resulted in between 750 and 1000 single trials of EEG data per subject (i.e., 150-200 trials
255 per speech token).

256 *Source analysis*. Following our previous neuroimaging studies on speech processing(Bidelman
257 and Dexter, 2015; Bidelman and Howell, 2016), we performed a distributed source analysis to
258 more directly assess the neural generators underlying behavioral decisions related to CP. Source
259 reconstruction was implemented in the MATLAB package Brainstorm (Tadel et al., 2011). We
260 used a realistic, boundary element model (BEM) volume conductor (Fuchs et al., 2002, 1998)
261 standardized to the MNI brain (Mazziotta et al., 1995). The BEM head model was created using
262 the OpenMEEG (Gramfort et al., 2010) as implemented in Brainstorm (Tadel et al., 2011). A
263 BEM is less prone to spatial errors than other head models (e.g., concentric spherical conductor)
264 (Fuchs et al., 2002). sLORETA allowed us to estimate the distributed neuronal current density
265 underlying the measured sensor data. The resulting activation maps (akin to fMRI) represent the
266 transcranial current source density underlying the scalp-recorded potentials as seen from the
267 cortical surface. We used the default settings in Brainstorm's implementation of sLORETA
268 (Tadel et al., 2011). From each single-trial sLORETA map, we extracted the time-course of
269 source activity within 68 regions of interest (ROI) defined by the Desikan-Killany Atlas
270 parcellation (Desikan et al., 2006) as implemented in Brainstorm. Single-trial source waveforms
271 (derived per subject and speech token) were then submitted to functional connectivity analyses.
272 We have recently used a similar approach to successfully decode single-trial EEG and predict
273 individual differences in other cognitive domains (e.g., working memory capacity (Bashivan et
274 al., 2017)), motivating its use here.

275 **EEG functional connectivity and graph analyses**
276 *Bootstrapping*. Functional connectivity measures are more accurate when calculated using
277 source localized compared to scalp-recorded (sensor-level) EEG (Brunner et al., 2016). Still, to
278 ensure robustness of our connectivity measures, we used bootstrapping to reduce the uncertainty
279 of our connectivity estimates (James et al., 2013). This method involved repeatedly taking small
280 samples with replacement, calculating the statistics, and averaging over the calculated statistics.
281 We applied a mean based bootstrap approach on 35106 trials. For each RT class, 100 random
282 trials from each individual participant were chosen as a bootstrap sample (with replacement). We
283 calculated the mean source amplitude in each of the 68 ROIs for each bootstrap sample. This
284 process was then iterated 30 times to derive the final estimate of the mean source signal in each
285 ROI. Overall, 3150 trials were generated (1050 trials of each RT class) in this process for further
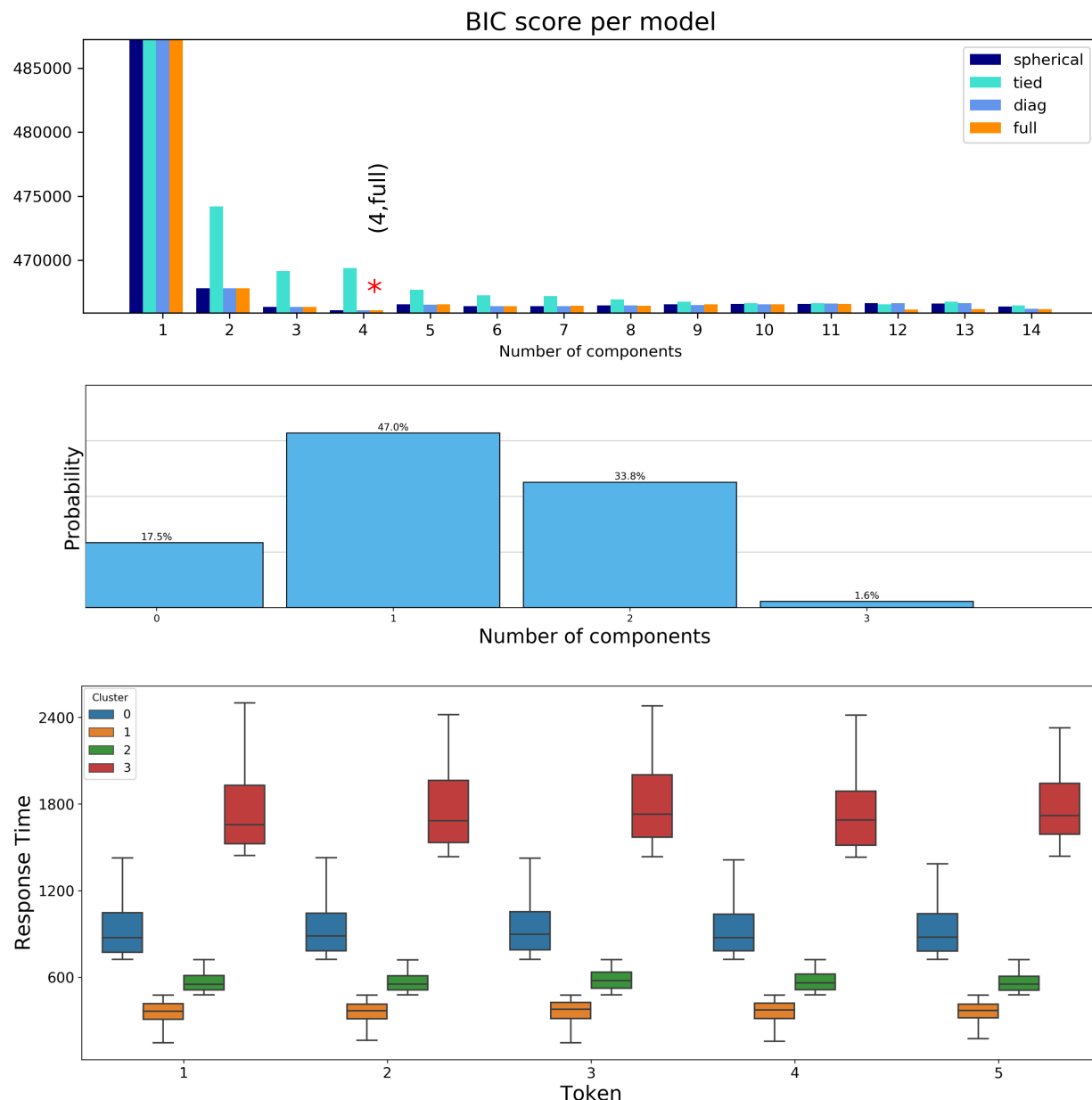286 analysis.

Figure 2: 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.

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 (Friston et al., 1994). 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

8

301   interactions at different time scales (Zhou et al., 2009). To quantify functional connectivity, we
302   measured pair-wise Pearson product-moment correlation coefficients among the 68 brain regions
303   (ROIs). This resulted in connectivity matrix describing the weighted strength (undirected
304   network) between all pairwise nodes ($^{68}C_2 = 2278$ edges) for each trial. Diagonal and upper
305   diagonal elements of the connectivity matrices were discarded to avoid spurious self and
306   repeated connectivity. Matrices were then concatenated to a vector to describe the connectivity
307   across all brain nodes and trials (e.g., 3150*2278) for each participant.
308   Seven global network connectivity features were estimated from each network graph
309   using the BCT toolbox (Rubinov and Sporns, 2010): (i) Characteristics path, (ii) Global
310   efficiency, (iii) Average clustering coefficient, (iv) Transitivity, (vi) Small-worldness, (vi)
311   Assortativity coefficient, and (vii) Maximized modularity (see Appendix for mathematical
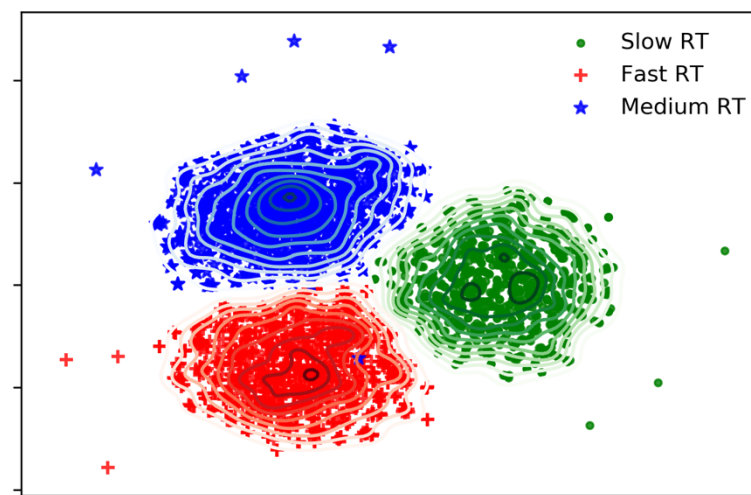312   definitions and interpretation of these network features).

313   *Figure 3: The t-SNE embedded higher dimensional functional connectivity data are represented by a 2-dimensional scatter and*
314   *kernel density estimation (KDE) plot. The green lines with '.', blue lines with '*', and red lines with '+'sign represents data*
315   *points for slow, medium, and fast RT participants, respectively.*
316

Machine learning: identifying behaviorally-relevant aspects of functional connectivity
318   The t-distributed stochastic neighbor embedding (t-SNE) (Maaten and Hinton, 2008) is a
319   widely used unsupervised learning algorithm to visualize high-dimensional data. t-SNE converts
320   similarities between higher dimensional data points to joint probabilities, providing a faithful
321   representation of those data points in a lower-dimensional human interpretable 2D or 3D plane.
322   Such a projection brings insight on whether the data is separable, the data lies in multiple
323   different clusters, or inspecting the nature of those clusters. We adopted LDA on our three-class
324   connectivity dataset (i.e., fast, medium, slow perceivers identified from the behavioral data) and
325   considered 50 dimensions for t-SNE. The hyper parameters of t-SNE were tuned with a grid
326   search approach. Figure 3 shows the t-SNE embedded scatter and kernel density estimation
327   (KDE) plot of our data distribution. KDE plot is a non-parametric way to represent the
328   probability density function and is used here to visualize the trend of the data distribution for
329   each different class (data points for fast, medium, and slow RTs). The t-SNE visualization shows

9

330 three nearly distinct clusters of functional connectivity for the different RT groups in speech
331 categorization. Unrelated or noisy edges may exist in the higher dimensional functional
332 connectivity matrices. This necessitates the use of feature selection methods to choose functional
333 connectivity metrics that are relevant and can be modeled robustly over a range of model
334 parameters.
335
336 **Feature selection (stabilty selection)**
337   Feature selection attempts to identify the most salient subset of variables from a larger set
338 of features mixed with irrelevant variables. This problem is especially challenging when the
339 number of available data samples are smaller compared to the number of possible features.
340 Conventional filter methods identify a consistent set of variables outside of the predictive model
341 based on some filtering criteria, e.g., the variables are individually evaluated to check the
342 probable relationship between classes. The sets of variables in this technique are selected based
343 on a threshold of importance. Commonly filter-based methods include correlation, F-test, chi-
344 square test, ANOVA analysis. The highly-correlated or redundant features may be selected, and
345 significant interactions and relationships between variables may not be able to be quantified.
346 However, one of the downsides of the multivariate approaches (e.g. , PCA, LDA, Lasso,
347 Elasticnet SVM ranking, Wrapper based methods, GA Wrapper, Forward Backward based
348 methods) is that outcomes often depend on model parameters (e.g., regularization factor).
349 Compared to conventional filter and multivariate approaches, stability selection produces more
350 reliable estimations and yields a stable set of features because of its internal randomization
351 implemented as bootstrap based subsampling. It was reported that even if the necessary
352 conditions needed for consistency of the original Lasso (L1 norm penalized linear models)
353 method are violated, stability selection will be consistent in variable selection (Meinshausen and
354 Bühlmann, 2010). The main advantages of this algorithm are (1) it works efficiently with the
355 high-dimensional data, (2) stability selection provides finite sample control with error rates of
356 false discoveries and is a transparent method to choose a amount of regularization for structure
357 estimation; and (3) it is extremely general and has a very wide range of applicability.
358   An attractive feature of Lasso (L1 regularization on least squares) is its computational
359 feasibility for the high-dimensional data with many more variables than samples since the
360 optimization problem of lasso estimator is convex. Furthermore, the Lasso can select variables
361 by shrinking certain estimated coefficients exactly to 0. Hence Lasso was used for stability
362 selection. Applying Randomized Lasso many times and looking for variables that are chosen is a
363 very powerful procedure tool to select consistent or stable features (Al-Fahad et al., 2017;
364 Meinshausen and Bühlmann, 2006; Shah and Samworth, 2013; Tibshirani, 1996). Despite its
365 simplicity, it is consistent for variable selection even though the 'neighborhood stability'
366 condition is violated. More about stability section, interpretation and mathematical definition are
367 explained in the appendix.
368   We used Randomized Logistic Regression for stability selection with randomized lasso. It
369 works by subsampling the training data and fitting an L1-penalized Logistic Regression model
370 where the penalty of a random subset of coefficients has been scaled. We considered sample
371 fraction = 0.75, number of resampling =1000 with tolerance=0.001. This algorithm assigns
372 feature scores between 0 and 1 based on frequency of selection over 1000 iterations. We need to
373 specify the score to find out the best representative set of stable features. Hence, threshold
374 selection is a design parameter. We varied different selection thresholds (i.e., the number of
375 selected features) and observed the effect on model performance. Modeling involved four steps:
376

10

377 1. Randomly shuffle and split the dataset in to training and test set (80% and 20%).
378 2. Consider Support Vector Machine with "RBF" kernel as a base estimator.
379 3. Tune hyper parameter (i.e. C and Gamma) on training data using grid search approach
380   and10-fold cross validation.
381 4. Selected best models are evaluated on unseen test data. Accuracy (ACC) and Area Under
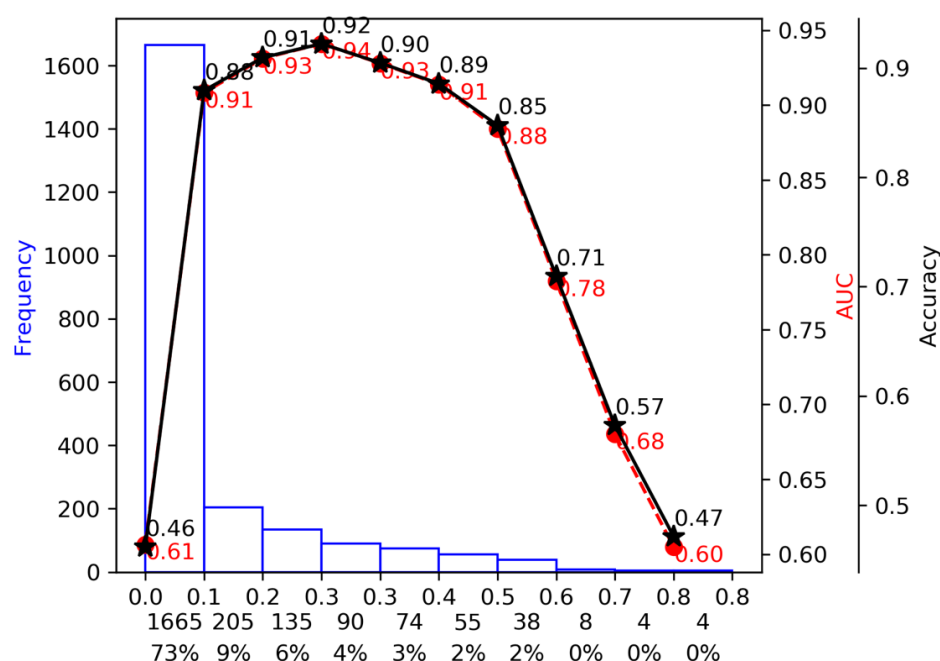382   Curve (AUC) were considered for performance measure.
383

384 *Figure 4: Effect of selection threshold on model performance prediction. The three x-labels represent (top) the range of each bin*
385 *of features score (range: 0~1), (middle) the number of features falling in each bin, and (bottom) the corresponding percentage.*
386

387   Figure 4 shows the effect of different selection thresholds on modeling. The histogram
388 illustrates the distribution of the feature score. The first line of the x axis shows the bin ranges of
389 scores (0 to 1. The second and third lines show the amount and percent of features that had
390 nearly the same score for a specific bin. We found 73% of the features had scores of 0-0.1,
391 meaning the majority of connectivity measures were not selected even once (i.e., coefficient was
392 zero) among 1000 model iterations. That is, 73% of functional connectivity metrics explored in
393 our search space were not related to speeded speech categorization (i.e., behavioral RTs).
394   For a specific selection threshold of 0.26, the algorithm selected 227 edge features that
395 collectively achieved 92% accuracy (best model performance) with AUC=0.9. The bell shaped
396 solid black and red dotted lines of Figure 4 shows the Accuracy and AUC curves for different
397 selection thresholds. Note that selection thresholds higher than the optimal value (0.26) allowed
398 the model to consider more noise variables, degrading model performance significantly. On the
399 other hand, selection thresholds higher than the optimal value discard behaviorally relevant
400 features and reduce model performance. Table 3 details the effect of selection threshold on
401 model performance. Here, the number of unique edges represents correlation-based connectivity
402 between two brain nodes (features) and the number of unique nodes represents brain regions
403 associated with those selected edges. A schematic diagram of the method pipeline is shown in
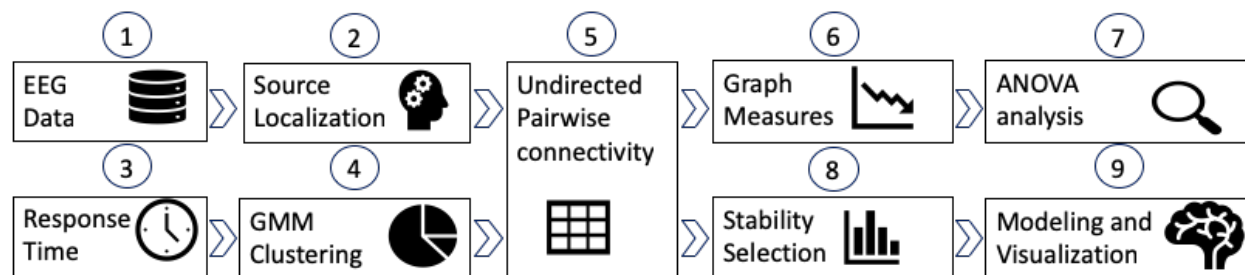404 Figure 5.

405
406 *Figure 5: Schematic diagram of the processing pipeline. The 64-ch EEG data is first preprocessed, and then source localization*
407 *is adapted to convert skull surface data to cortical surface time series data (68 ROIs defined by the Desikan-Killany Atlas*
408 *parcellation). Pairwise correlations were calculated to derive the connectivity matrix for each trial of the speech CP task.*
409 *Behavioral response times (RTs) were clustered with Bayesian non-parametric (GMM) clustering. These clusters were labeled as*
410 *Fast, Medium and Slow RT. ANOVA analysis of Graph measures w adopted to test significance among RT groups. Stability*
411 *selection and machine learning approaches were then used to find significant properties of the brain's functional connectivity*
412 *related to behavioral speeds (RTs) in speech CP.*

413 **RESULTS**
414

415   Figure 1D shows behavioral results in the speech categorization task. Generally speaking,
416 listeners were slower to label sounds near the categorical boundary (token 3), consistent with the
417 higher ambiguity of the mid-continuum stimuli (Bidelman et al., 2013; Liebenthal et al., 2010;
418 Pisoni and Tash, 1974; Reetzke et al., 2018). On average, females also showed slower RTs than
419 males across the continuum (Welch's t-test; p<0.0001). Bayesian nonparametric clustering
420 revealed four distinguishable clusters in the speed (RTs) of listeners' CP (Fast: 120~476 ms,
421 Medium: 478~722 ms, Slow: 724~1430 ms, and Outliers: 1432~2500 ms) (Figure 2C). These
422 clusters were even present at the individual token level.
423   Having established that listeners could be distinguished based on their speed in speech
424 categorization, our next goal was to determine whether network properties of the brain accounted
425 for these behavioral differences. We applied graph theory techniques to construct and analyze the
426 functional brain connectome underlying CP. We considered both individual trial- as well as
427 group-based analyses. For group-based analysis, data were averaged across subjects within each
428 RT cluster. Group means were computed by concatenating group-wise trials and calculating their
429 mean. We then calculated seven global network connectivity features using the BCT toolbox
430 (Rubinov and Sporns, 2010) (see Methods).
431   We used non-parametric ANOVAs (Kruskal-Wallis H-test) to determine if individual
432 trial-based global graph measures varied across RTs (Table 1) This non-parametric test was used
433 given the unequal sample size per group (Lowry, 2014). These analyses revealed that
434 Assortativity and Global Efficiency were modulated depending on behavior speed. Table 2
435 shows a comparison of the graph measures across three RT groups. Global efficiency measures
436 were relatively small, and assortativity had a negative tendency. All other network features were
437 not discriminatory among the RT groups. Therefore, modeling with those features (using SVM
438 with 'RBF' kernel described in method section) showed expectedly poor accuracy (38%).
439
440
441
442
443
444
445

12

446   *Table 1: Significant (bold) global network measures (Kruskal-Wallis H-test tests) (trial-level)*
447

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

448
449   *Table 2: Group comparison of graph measures of functional connectivity between RT groups.*
450

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
Assortativity	-0.0086	-0.0128	-0.0118
Global Efficiency	0.1909	0.1934	0.1944
Transitivity	0.1329	0.1362	0.1354
Maximized Modularity	0.1872	0.1845	0.1875

451
452   Besides analyzing global network properties, we next aimed to identity the most
453   significant properties of functional brain connectivity that were related to behavioral RTs.
454   Functional connectivity for each trial is a high dimensional sparse matrix. Some studies have
455   suggested that properties of functional brain networks are most consistent with the actual brain
456   anatomy when network density is 8–16% (Li et al., 2016; Salvador et al., 2005; Wang et al.,
457   2010). To determine the most behaviorally-relevant arrangement of sparse connectivity, we used
458   stability selection with Randomized Lasso to detect and rank the most important, consistent, and
459   relevant functional connectivity measures that were invariant (stable) over a range of model
460   parameters. Stability selection discarded 88% (total 273) of network edges that were not related
461   to behavioral RTs, but still achieved 92% classification accuracy with AUC=0.9. From Table 3,
462   It was observed that only 7% error tolerance from the optimal value (accuracy from 92% to 85%)
463   allow 80% less edge and 22% less associated nodes. Hence, the selection threshold 0.51 with
464   reasonable performance (ACC=85%, AUC=0.9) were chosen for network visualization as
465   performance declined precipitously above this threshold (Figure 4).
466   Figure 6 shows a visualization of the 54 nodes among 53 ROIs identified via stability
467   selection using BrainNet (Xia et al., 2013). The resulting network revealed a highly dense
468   connectome reflective of listeners' behavioral RTs in speech categorization. Connectivity was
469   particularly strong between the occipital, parietal, and bilateral frontal lobes. As an additional
470   means of data reduction, we further threshold (=0.68) the stability-selected connectome. This
471   resulted in eight highly ranked connectivity edges among 13 nodes across the brain (Figure 7).
472   Even with this sparse network of only 8 edges, model classification was still 57%, meaning this
473   small set of features accuracy predicted RTs. We then ranked the contribution of these stable
474   nodes that described behavior Table 4. We found that three edges (rank: 3, 4, and 6) were in left

13

475   hemisphere, two edges were in right hemisphere (rank: 2, and 5) and three edges were inter-
476   hemispheric (rank: 1, 7 and 8). Notably, these edges included connections between motor
477   (paracentral), visual (lateral occipital/ lingual), linguistic (pars triangularis), auditory (superior
478   temporal gyrus), and parietal areas both within and between hemispheres.
479
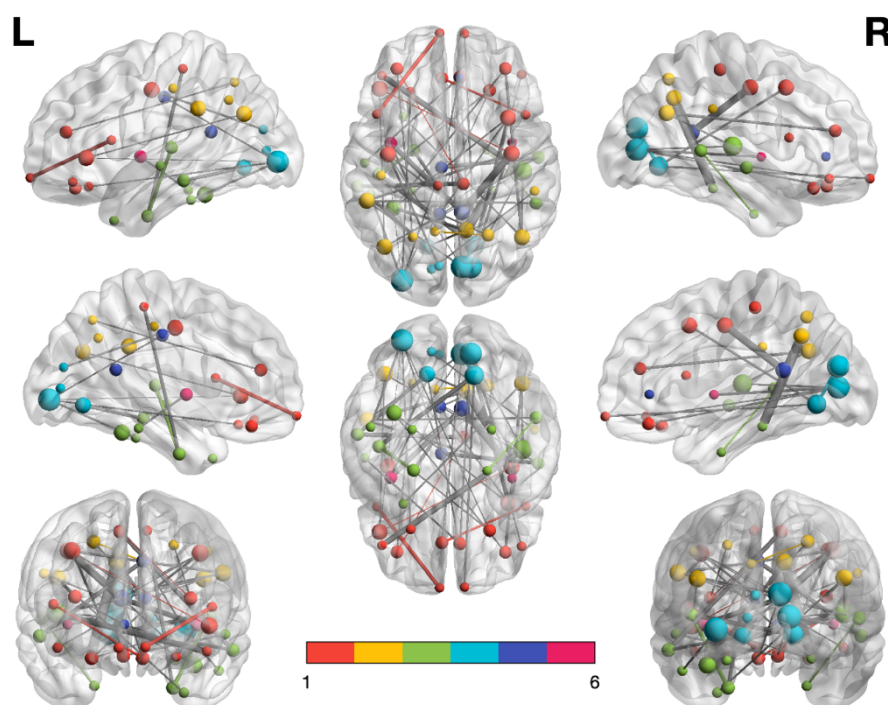
480   *Figure 6: BrainNet visualization (top to bottom: lateral, medial, and dorsal view) of the brain network (54 edges) identified via*
481   *stability selection. Color map 1-6 indicates, 1: Frontal (22 ROI), 2: Parietal (10 ROI), 3: Temporal (18 ROI), 4: Occipital (8*
482   *ROI), 5: Cingulate (8 ROI), 6: Insula (2 ROI) regions. Node size varies with its degree of connectivity. Connectivity among the*
483   *same lobe are colored with similar node color. Edge widths represent the weight of absolute correlation (connectivity strength).*
484
485   *Table 3: Effect of selection threshold of stability selection (Threshold) on model performance. Pairwise correlation between two*
486   *brain regions (functional connectivity edge) were consider as features. Number of unique nodes are brain regions associated*
487   *with selected features. ACC, accuracy; AUC, area under curve.*
488

Threshold	ACC	AUC	Number of Unique Edge (features)	Number of Unique Node
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

14

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 (Bassett and Bullmore, 2006; Honey et al., 2007; Newman, 2003; Tononi et al., 1994). 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 1 and Table 2). 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 (Stam et al., 2014). 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 (Betzel et al., 2018). 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.

Figure 7: 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 6.

15

517   In addition, we found that the functional CP network underlying speeded decisions
518   increased in negative assortativity (i.e., became disassortative) for slower RTs (Table 2). This
519   indicates that brain nodes were more likely to connect with nodes having different degree during
520   slower RTs, implying that important hubs of the CP network communicated with insignificant
521   hubs during states of slower decisions. Based on the interpretation of these graph metrics (see
522   Appendix), we infer that slower, more taxing categorical speech decisions cause excessive use of
523   neural resources and more aberrant information flow within the CP circuitry. Supporting this
524   interperion, we found that Network utilization (Global efficiency) also differentiated RT groups.
525   Higher Global efficiency indicates that the routing of information among nodes with different
526   degree was significantly higher for slow RT trials. In short, we find that slower perceivers tended
527   to utilize functional brain networks excessively (or inappropriately) whereas fast perceivers (with
528   lower global efficiency) utilized the same neural pathways but with more restricted organization.
529   Presumably, these dynamic changes in brain connectivity account for the variations in RTs we
530   find during speech categorization at the behavioral level (Figure 1D).
531   Our data show that global graph measures fail to fully explain the behavioral relevance of
532   important connectivity edges. We observed the functional connectivity matrix underlying speech
533   CP is highly sparse and dynamic. Indeed, only ~12% of all possible edges in the Desikan-Killany
534   Atlas were needed to explain variation in behavioral RTs. In this vein, we adopted stability
535   selection to find edges that were most consistent in distinguishing different network states related
536   to perception. By performing this two stage of randomization iteratively (e.g., 1000 bootstraps),
537   stability selection with randomized lasso assigned high scores to features that were repeatedly
538   selected across randomizations, yielding the most meaningful connections within the CP
539   connectome that describe behavior.
540   Collectively, our results showed that neural classifiers (SVM) coupled with stability
541   selection can correctly classify behavioral RTs related to CP from functional connectivity alone
542   with over 90% accuracy (AUC=0.9). The resulting edges composing the RT-related networks
543   were distributed in both hemispheres and both intra- and inter-hemispheric edges were evident.
544   More interestingly, we found that only 8 edges among 13 ROIs were needed to distinguish RTs
545   well above chance (Figure 7). ROIs composing this sparse but behaviorally-relevant network
546   included (1) Caudalmiddlefrontal R, (2) Inferiorparietal L, (3) Lateraloccipital L, (4) Lingual L,
547   (5) Lingual R, (6) Middletemporal L, (7) Paracentral R, (8) Parahippocampal R, (9)
548   Parstriangularis L, (10) Precuneus R, (11) Rostralmiddlefrontal L, (12) Superiorparietal L, (13)
549   and Superiortemporal R. Previous neuroimaging studies have demonstrated a distributed fronto-
550   temporo-parietal neural network supporting auditory categorization (e.g., (Alho et al., 2016;
551   Bidelman and Lee, 2015; Binder et al., 2004; Chang et al., 2010; Feng et al., 2018; Golestani et
552   al., 2002; Golestani and Zatorre, 2004; Lee et al., 2012; Liebenthal et al., 2010; Luthra et al.,
553   2019; Myers et al., 2009)). Our data corroborate these previous studies by confirming
554   engagement of similar temporal (STG), parietal, motor, and prefrontal regions in CP using an
555   entirely data driven approach (machine learning with stability selection).
556
557
558
559
560
561
562

16

563 *Table 4: Eight most important edges that govern speeded speech classification. Collectively, these edges achieve a model*
564 *accuracy of 57% in segregating listeners' speeded decisions (RTs) in the perceptual task. Here, a score of 0.85 means that out of*
565 *1000 iterations, the edge was selected by stability selection 850 times.*
566

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

567

568   Notably, we found functional connectivity between right paracentral and left
569 middletemporal gyrus (MTG) was the most important connection describing the speed of
570 behavioral CP (Table 4). MTG has been associated with accessing word meaning while reading
571 (Acheson and Hagoort, 2013) and has been described as an early lexical interface that is heavily
572 involved in sound-to-meaning inference (Hickok and Poeppel, 2007, 2004). Some studies
573 indicate that lesions of the posterior region of the middle temporal gyrus, in the left cerebral
574 hemisphere, may result in certain forms of alexia and agraphia (Sakurai et al., 2008), indicating
575 its role in the language production network (Blank et al., 2002). The strong link between MTG
576 and paracentral gyrus implies a direct pathway between the neural substrates that map sounds to
577 meaning and sensorimotor regions that execute the motor command and therefore govern
578 response speeds (indexed by RTs). The leftward laterality of the MTG node is consistent with the
579 left lateralized nature of language processing in the brain. Still, why left MTG so strongly
580 interfaces with *right* motor areas in our data is unclear, especially given the right-handedness of
581 our participants and expected left (contralateral) motor involvement. Differences in brain
582 connectivity have been observed between sexes (Ingalhalikar et al., 2014) and females may have
583 a more diffuse, bilateral neural system for language processing than males (Shaywitz et al.,
584 1995). Speculatively, the strong communication between left linguistic (MTG) and right motor
585 brain areas we find may reflect the higher preponderance of females in our sample.
586   Relatedly, stability selection identified the second ranked edge between lingual and
587 caudal-middlefrontal gyrus. While the functional role of lingual (occipital) gyrus in speech
588 processing is not apparent prima facie, this region is involved in visual word processing,
589 especially letters (Mechelli et al., 2000). It has also implicated in stimulus naming (Bookheimer
590 et al., 1995; Howard et al., 1992), an operation at the core of our speech categorization (i.e.,
591 sound labelling) task. We also found a third ranked edge predictive of behavioral CP between
592 parstriangularis and inferior parietal cortex. Previous functional neuroimaging and connectivity
593 studies have shown strong engagement of frontal-parietal networks during CP (Feng et al., 2018;
594 Liebenthal et al., 2010; Luthra et al., 2019). Our results corroborate these findings by similarly
595 implicating a strong interface between linguistic (IFG) and parietal (IPL) brain regions in
596 modulating the speed of listeners' categorical decisions. Indeed, decision loads IFG during
597 effortful speech listening (Binder et al., 2004; Bouton et al., 2018; Du et al., 2014) and the IFG-
598 IPL pathway is upregulated when speech material is perceptually confusable (Feng et al., 2018).
599 Therefore, the network organization of brain connectivity observed for slow RTs and importance

600 of IFG-IPL in describing behavior may reflect a similar state of perceptual confusion during
601 rapid categorical speech labeling.
602   One limitation of our study was that our sample contained more females than males (2:1
603 ratio). This is relevant since RTs were significantly different among genders (Figure 1D). Thus, a
604 natural question that emerges from our data is the degree to which our machine learning
605 techniques segregated data based on gender rather than different RTs (i.e., fast vs. slow
606 perceivers), per se. Still, this is probably not the case. Conventional filter-based group analysis
607 can bias classification and feature selection results whereas with our Lasso-based bootstrapped
608 analysis, this becomes less likely (Bach, 2008). Moreover, stability selection with randomized
609 lasso is a similar but more robust approach that produces consistent variable selection with
610 minimal bias. Hence, the impact of our unbalanced sample size on feature selection is probably
611 negligible.
612   Taken together, our novel approach to neuroimaging data demonstrates the derivation of
613 small, yet highly meaningful patterns of brain connectivity that dictate speech behaviors using
614 solely EEG. More broadly, the functional connectivity and machine learning techniques used
615 here could be deployed in future studies to identify the most meaningful changes in
616 spatiotemporal brain activity that are modulated by development, normal learning, or those
617 which decline in neuropathological states.

618 **CONCLUSION**
619
620   We developed an efficient computational framework to investigate whether individual
621 differences in speeded speech categorization could be explained in terms of network-level
622 descriptions of functional brain connectivity. Our EEG data-driven approach reveals that the
623 speed of listeners' ability to categorize and properly label speech sounds is directly related to
624 dynamic variations in their brain connectomics. These findings contribute in several ways to our
625 understanding of how the brain works in categorical perception and provide a basis for further
626 research. In future iterations of the work, we plan to improve our approach by including
627 directional and dynamic connectivity analysis to better delineate the temporal emergence of the
628 phenomena observed here.
629
633
634 **REFERENCE**
635
636 Acheson, D.J., Hagoort, P., 2013. Stimulating the brain's language network: syntactic ambiguity
637   resolution after TMS to the inferior frontal gyrus and middle temporal gyrus. Journal of
638   cognitive neuroscience 25, 1664–1677.
639 Al-Fahad, R., Yeasin, M., Anam, A.I., Elahian, B., 2017. Selection of stable features for
640   modeling 4-D affective space from EEG recording, in: Neural Networks (IJCNN), 2017
641   International Joint Conference On. IEEE, pp. 1202–1209.
642 Alho, J., Green, B.M., May, P.J., Sams, M., Tiitinen, H., Rauschecker, J.P., Jääskeläinen, I.P.,
643   2016. Early-latency categorical speech sound representations in the left inferior frontal
644   gyrus. Neuroimage 129, 214–223.

645   Bach, F., 2008. Bolasso: model consistent Lasso estimation through the bootstrap.
646   arXiv:0804.1302 [cs, math, stat].
647   Bashivan, P., Yeasin, M., Bidelman, G.M., 2017. Temporal progression in functional
648   connectivity determines individual differences in working memory capacity, in: 2017
649   International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 2943–2949.
650   Bassett, D.S., Bullmore, E.D., 2006. Small-world brain networks. The neuroscientist 12, 512–
651   523.
652   Bauer, H.-U., Der, R., 1996. Controlling the magnification factor of self-organizing feature
653   maps. Neural computation 8, 757–771.
654   Betzel, R.F., Bertolero, M.A., Bassett, D.S., 2018. Non-assortative community structure in
655   resting and task-evoked functional brain networks. bioRxiv 355016.
656   Bidelman, G.M., Alain, C., 2015. Musical training orchestrates coordinated neuroplasticity in
657   auditory brainstem and cortex to counteract age-related declines in categorical vowel
658   perception. Journal of Neuroscience 35, 1240–1249.
659   Bidelman, G.M., Dexter, L., 2015. Bilinguals at the "cocktail party": Dissociable neural activity
660   in auditory–linguistic brain regions reveals neurobiological basis for nonnative listeners'
661   speech-in-noise recognition deficits. Brain and language 143, 32–41.
662   Bidelman, G.M., Howell, M., 2016. Functional changes in inter-and intra-hemispheric cortical
663   processing underlying degraded speech perception. Neuroimage 124, 581–590.
664   Bidelman, G.M., Lee, C.-C., 2015. Effects of language experience and stimulus context on the
665   neural organization and categorical perception of speech. NeuroImage 120, 191–200.
666   https://doi.org/10.1016/j.neuroimage.2015.06.087
667   Bidelman, G.M., Lowther, J.E., Tak, S.H., Alain, C., 2017. Mild Cognitive Impairment Is
668   Characterized by Deficient Brainstem and Cortical Representations of Speech. J.
669   Neurosci. 37, 3610–3620. https://doi.org/10.1523/JNEUROSCI.3700-16.2017
670   Bidelman, G.M., Moreno, S., Alain, C., 2013. Tracing the emergence of categorical speech
671   perception in the human auditory system. NeuroImage 79, 201–212.
672   https://doi.org/10.1016/j.neuroimage.2013.04.093
673   Bidelman, G.M., Villafuerte, J.W., Moreno, S., Alain, C., 2014a. Age-related changes in the
674   subcortical–cortical encoding and categorical perception of speech. Neurobiology of
675   Aging 35, 2526–2540. https://doi.org/10.1016/j.neurobiolaging.2014.05.006
676   Bidelman, G.M., Walker, B.S., 2017. Attentional modulation and domain-specificity underlying
677   the neural organization of auditory categorical perception. European Journal of
678   Neuroscience 45, 690–699. https://doi.org/10.1111/ejn.13526
679   Bidelman, G.M., Weiss, M.W., Moreno, S., Alain, C., 2014b. Coordinated plasticity in brainstem
680   and auditory cortex contributes to enhanced categorical speech perception in musicians.
681   European Journal of Neuroscience 40, 2662–2673. https://doi.org/10.1111/ejn.12627
682   Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D., 2004. Neural correlates of
683   sensory and decision processes in auditory object identification. Nature Neuroscience 7,
684   295. https://doi.org/10.1038/nn1198
685   Blank, S.C., Scott, S.K., Murphy, K., Warburton, E., Wise, R.J., 2002. Speech production:
686   Wernicke, Broca and beyond. Brain 125, 1829–1838.
687   Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of
688   communities in large networks. Journal of statistical mechanics: theory and experiment
689   2008, P10008.

19

690   Bookheimer, S.Y., Zeffiro, T.A., Blaxton, T., Gaillard, W., Theodore, W., 1995. Regional
691   cerebral blood flow during object naming and word reading. Human Brain Mapping 3,
692   93–106. https://doi.org/10.1002/hbm.460030206
693   Bouton, S., Chambon, V., Tyrand, R., Guggisberg, A.G., Seeck, M., Karkar, S., Ville, D. van de,
694   Giraud, A.-L., 2018. Focal versus distributed temporal cortex activity for speech sound
695   category assignment. PNAS 115, E1299–E1308.
696   https://doi.org/10.1073/pnas.1714279115
697   Breiman, L., 1999. Using adaptive bagging to debias regressions. Technical Report 547,
698   Statistics Dept. UCB.
699   Bressler, S.L., 1995. Large-scale cortical networks and cognition. Brain Research Reviews 20,
700   288–304.
701   Brunner, C., Billinger, M., Seeber, M., Mullen, T.R., Makeig, S., 2016. Volume Conduction
702   Influences Scalp-Based Connectivity Estimates. Front. Comput. Neurosci. 10.
703   https://doi.org/10.3389/fncom.2016.00121
704   Büchlmann, P., Yu, B., 2002. Analyzing bagging. Annals of Statistics 927–961.
705   Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural
706   and functional systems. Nature Reviews Neuroscience 10, 186.
707   Cajal, S.R., 1995. Histology of the nervous system of man and vertebrates. History of
708   Neuroscience (Oxford Univ Press, New York).
709   Calcus Axelle, Lorenzi Christian, Collet Gregory, Colin Cécile, Kolinsky Régine, 2016. Is There
710   a Relationship Between Speech Identification in Noise and Categorical Perception in
711   Children With Dyslexia? Journal of Speech, Language, and Hearing Research 59, 835–
712   852. https://doi.org/10.1044/2016_JSLHR-H-15-0076
713   Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T., 2010.
714   Categorical speech representation in human superior temporal gyrus. Nature
715   Neuroscience 13, 1428–1432. https://doi.org/10.1038/nn.2641
716   Chen, M., Kuzmin, K., Szymanski, B.K., 2014. Community detection via maximization of
717   modularity and its variants. IEEE Transactions on Computational Social Systems 1, 46–
718   65.
719   Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L.,
720   Dale, A.M., Maguire, R.P., Hyman, B.T., 2006. An automated labeling system for
721   subdividing the human cerebral cortex on MRI scans into gyral based regions of interest.
722   Neuroimage 31, 968–980.
723   Du, Y., Buchsbaum, B.R., Grady, C.L., Alain, C., 2014. Noise differentially impacts phoneme
724   representations in the auditory and speech motor systems. PNAS 111, 7126–7131.
725   https://doi.org/10.1073/pnas.1318738111
726   Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., others, 2004. Least angle regression. The
727   Annals of statistics 32, 407–499.
728   Feng, G., Gan, Z., Wang, S., Wong, P.C.M., Chandrasekaran, B., 2018. Task-General and
729   Acoustic-Invariant Neural Representation of Speech Categories in the Human Brain.
730   Cereb Cortex 28, 3241–3254. https://doi.org/10.1093/cercor/bhx195
731   Foster, J.G., Foster, D.V., Grassberger, P., Paczuski, M., 2010. Edge direction and the structure
732   of networks. Proceedings of the National Academy of Sciences 107, 10815–10820.
733   Fries, P., 2005. A mechanism for cognitive dynamics: neuronal communication through neuronal
734   coherence. Trends in cognitive sciences 9, 474–480.
735   Friston, K.J., Jezzard, P., Turner, R., 1994. Analysis of functional MRI time-series. Human Brain
736   Mapping 1, 153–171. https://doi.org/10.1002/hbm.460010207

737 Fuchs, M., Drenckhahn, R., Wischmann, H., Wagner, M., 1998. An improved boundary element
738   method for realistic volume-conductor modeling. IEEE Transactions on Biomedical
739   Engineering 45, 980–997. https://doi.org/10.1109/10.704867

740 Fuchs, M., Kastner, J., Wagner, M., Hawes, S., Ebersole, J.S., 2002. A standardized boundary
741   element method volume conductor model. Clinical Neurophysiology 113, 702–712.
742   https://doi.org/10.1016/S1388-2457(02)00030-5

743 Golestani, N., Paus, T., Zatorre, R.J., 2002. Anatomical Correlates of Learning Novel Speech
744   Sounds. Neuron 35, 997–1010. https://doi.org/10.1016/S0896-6273(02)00862-0

745 Golestani, N., Zatorre, R.J., 2004. Learning new sounds of speech: reallocation of neural
746   substrates. NeuroImage 21, 494–506. https://doi.org/10.1016/j.neuroimage.2003.09.071

747 Gramfort, A., Papadopoulo, T., Olivi, E., Clerc, M., 2010. OpenMEEG: opensource software for
748   quasistatic bioelectromagnetics. Biomedical engineering online 9, 45.

749 Guenther, F.H., Gjaja, M.N., 1996. The perceptual magnet effect as an emergent property of
750   neural map formation. The Journal of the Acoustical Society of America 100, 1111–
751   1121.

752 Guenther, F.H., Husain, F.T., Cohen, M.A., Shinn-Cunningham, B.G., 1999. Effects of
753   categorization and discrimination training on auditory perceptual space. The Journal of
754   the Acoustical Society of America 106, 2900–2912.

755 Guenther Frank H., Nieto-Castanon Alfonso, Ghosh Satrajit S., Tourville Jason A., 2004.
756   Representation of Sound Categories in Auditory Cortical Maps. Journal of Speech,
757   Language, and Hearing Research 47, 46–57. https://doi.org/10.1044/1092-
758   4388(2004/005)

759 Hakvoort Britt, de Bree Elise, van der Leij Aryan, Maassen Ben, van Setten Ellie, Maurits
760   Natasha, van Zuijen Titia L., 2016. The Role of Categorical Speech Perception and
761   Phonological Processing in Familial Risk Children With and Without Dyslexia. Journal
762   of Speech, Language, and Hearing Research 59, 1448–1460.
763   https://doi.org/10.1044/2016_JSLHR-L-15-0306

764 Harnad, S., Bureau, A.S.L., 1987. Categorical perception : the groundwork of cognition, 2nd ed.
765   ed. Cambridge ; New York : Cambridge University Press.

766 Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nature Reviews
767   Neuroscience 8, 393–402. https://doi.org/10.1038/nrn2113

768 Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding
769   aspects of the functional anatomy of language. Cognition, Towards a New Functional
770   Anatomy of Language 92, 67–99. https://doi.org/10.1016/j.cognition.2003.10.011

771 Honey, C.J., Kötter, R., Breakspear, M., Sporns, O., 2007. Network structure of cerebral cortex
772   shapes functional connectivity on multiple time scales. Proceedings of the National
773   Academy of Sciences 104, 10240–10245.

774 Howard, D., Patterson, K., Wise, R., Brown, W.D., Friston, K., Weiller, C., Frackowiak, R.,
775   1992. THE CORTICAL LOCALIZATION OF THE LEXICONSPOSITRON
776   EMISSION TOMOGRAPHY EVIDENCE. Brain 115, 1769–1782.
777   https://doi.org/10.1093/brain/115.6.1769

778 Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T.D., Elliott, M.A., Ruparel, K.,
779   Hakonarson, H., Gur, R.E., Gur, R.C., Verma, R., 2014. Sex differences in the structural
780   connectome of the human brain. PNAS 111, 823–828.
781   https://doi.org/10.1073/pnas.1316909110

782 James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning.
783   Springer.

784 Lee, Y.-S., Turkeltaub, P., Granger, R., Raizada, R.D.S., 2012. Categorical Speech Processing in
785   Broca's Area: An fMRI Study Using Multivariate Pattern-Based Analysis. J. Neurosci.
786   32, 3942–3948. https://doi.org/10.1523/JNEUROSCI.3814-11.2012
787 Li, W., Wang, M., Li, Y., Huang, Y., Chen, X., 2016. A novel brain network construction
788   method for exploring age-related functional reorganization. Computational Intelligence
789   and Neuroscience 2016, 5.
790 Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of
791   the speech code. Psychol Rev 74, 431–461.
792 Liebenthal, E., Desai, R., Ellingson, M.M., Ramachandran, B., Desai, A., Binder, J.R., 2010.
793   Specialization along the Left Superior Temporal Sulcus for Auditory Categorization.
794   Cereb Cortex 20, 2958–2970. https://doi.org/10.1093/cercor/bhq045
795 Lively, S.E., Logan, J.S., Pisoni, D.B., 1993. Training Japanese listeners to identify English /r/
796   and /l/. II: The role of phonetic environment and talker variability in learning new
797   perceptual categories. The Journal of the Acoustical Society of America 94, 1242–1255.
798   https://doi.org/10.1121/1.408177
799 Lowry, R., 2014. Concepts and applications of inferential statistics.
800 Luthra, S., Guediche, S., Blumstein, S.E., Myers, E.B., 2019. Neural substrates of subphonemic
801   variation and lexical competition in spoken word recognition. Language, Cognition and
802   Neuroscience 34, 151–169. https://doi.org/10.1080/23273798.2018.1531140
803 Maaten, L. van der, Hinton, G., 2008. Visualizing data using t-SNE. Journal of Machine
804   Learning Research 9, 2579–2605.
805 Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J., 1995. A probabilistic atlas of the
806   human brain: theory and rationale for its development. The International Consortium for
807   Brain Mapping (ICBM). Neuroimage 2, 89–101.
808 Mechelli, A., Humphreys, G.W., Mayall, K., Olson, A., Price, C.J., 2000. Differential effects of
809   word length and visual contrast in the fusiform and lingual gyri during. Proceedings of
810   the Royal Society of London. Series B: Biological Sciences 267, 1909–1913.
811 Meinshausen, N., Bühlmann, P., 2010. Stability selection. Journal of the Royal Statistical
812   Society: Series B (Statistical Methodology) 72, 417–473.
813 Meinshausen, N., Bühlmann, P., 2006. High-dimensional graphs and variable selection with the
814   lasso. The annals of statistics 1436–1462.
815 Myers, E.B., Blumstein, S.E., Walsh, E., Eliassen, J., 2009. Inferior Frontal Regions Underlie the
816   Perception of Phonetic Category Invariance. Psychol Sci 20, 895–903.
817   https://doi.org/10.1111/j.1467-9280.2009.02380.x
818 Newman, M.E., 2004. Fast algorithm for detecting community structure in networks. Physical
819   review E 69, 066133.
820 Newman, M.E., 2003. The structure and function of complex networks. SIAM review 45, 167–
821   256.
822 Newman, M.E., 2002. Assortative mixing in networks. Physical review letters 89, 208701.
823 Oldfield, R.C., 1971. The assessment and analysis of handedness: The Edinburgh inventory.
824   Neuropsychologia 9, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4
825 Oostenveld, R., Praamstra, P., 2001. The five percent electrode system for high-resolution EEG
826   and ERP measurements. Clinical Neurophysiology 112, 713–719.
827   https://doi.org/10.1016/S1388-2457(00)00527-7
828 Phillips, C., 2001. Levels of representation in the electrophysiology of speech perception.
829   Cognitive Science 25, 711–731. https://doi.org/10.1207/s15516709cog2505_5

830 Picton, T.W., van Roon, P., Armilio, M.L., Berg, P., Ille, N., Scherg, M., 2000. The correction of
831   ocular artifacts: a topographic perspective. Clinical Neurophysiology 111, 53–65.
832   https://doi.org/10.1016/S1388-2457(99)00227-8
833 Piraveenan, M., Prokopenko, M., Zomaya, A., 2012. Assortative mixing in directed biological
834   networks. IEEE/ACM Transactions on Computational Biology and Bioinformatics
835   (TCBB) 9, 66–78.
836 Pisoni, D.B., 1973. Auditory and phonetic memory codes in the discrimination of consonants and
837   vowels. Perception & Psychophysics 13, 253–260. https://doi.org/10.3758/BF03214136
838 Pisoni, D.B., Luce, P.A., 1987. Acoustic-phonetic representations in word recognition.
839   Cognition, Special Issue Spoken Word Recognition 25, 21–52.
840   https://doi.org/10.1016/0010-0277(87)90003-5
841 Pisoni, D.B., Tash, J., 1974. Reaction times to comparisons within and across phonetic
842   categories. Perception & Psychophysics 15, 285–290.
843   https://doi.org/10.3758/BF03213946
844 Prather, J.F., Nowicki, S., Anderson, R.C., Peters, S., Mooney, R., 2009. Neural correlates of
845   categorical perception in learned vocal communication. Nature Neuroscience 12, 221–
846   228. https://doi.org/10.1038/nn.2246
847 Reetzke, R., Xie, Z., Llanos, F., Chandrasekaran, B., 2018. Tracing the Trajectory of Sensory
848   Plasticity across Different Stages of Speech Learning in Adulthood. Current Biology 28,
849   1419-1427.e4. https://doi.org/10.1016/j.cub.2018.03.026
850 Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and
851   interpretations. Neuroimage 52, 1059–1069.
852 Sakurai, Y., Mimura, I., Mannen, T., 2008. Agraphia for kanji resulting from a left posterior
853   middle temporal gyrus lesion. Behavioural neurology 19, 93–106.
854 Salvador, R., Suckling, J., Coleman, M.R., Pickard, J.D., Menon, D., Bullmore, E.D., 2005.
855   Neurophysiological architecture of functional magnetic resonance images of human
856   brain. Cerebral cortex 15, 1332–1342.
857 Shah, R.D., Samworth, R.J., 2013. Variable selection with error control: Another look at
858   Stability Selection. Journal of the Royal Statistical Society: Series B (Statistical
859   Methodology) 75, 55–80. https://doi.org/10.1111/j.1467-9868.2011.01034.x
860 Shaywitz, B.A., Shaywltz, S.E., Pugh, K.R., Constable, R.T., Skudlarski, P., Fulbright, R.K.,
861   Bronen, R.A., Fletcher, J.M., Shankweiler, D.P., Katz, L., Gore, J.C., 1995. Sex
862   differences in the functional organization of the brain for language. Nature 373, 607.
863   https://doi.org/10.1038/373607a0
864 Stam, C.J., Tewarie, P., Van Dellen, E., van Straaten, E.C.W., Hillebrand, A., Van Mieghem, P.,
865   2014. The trees and the forest: Characterization of complex brain networks with
866   minimum spanning trees. International Journal of Psychophysiology 92, 129–138.
867   https://doi.org/10.1016/j.ijpsycho.2014.04.001
868 Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M., 2011. Brainstorm: a user-friendly
869   application for MEG/EEG analysis. Computational intelligence and neuroscience 2011,
870   8.
871 Temlyakov, V.N., 2000. Weak greedy algorithms [*] This research was supported by National
872   Science Foundation Grant DMS 9970326 and by ONR Grant N00014-96-1-1003.
873   Advances in Computational Mathematics 12, 213–227.
874 Tibshirani, R., 1996. Regression Shrinkage and Selection via the Lasso. Journal of the Royal
875   Statistical Society. Series B (Methodological) 58, 267–288.

876 Tononi, G., Sporns, O., Edelman, G.M., 1994. A measure for brain complexity: relating
877   functional segregation and integration in the nervous system. Proceedings of the National
878   Academy of Sciences 91, 5033–5037.
879 Toscano, J.C., Anderson, N.D., Fabiani, M., Gratton, G., Garnsey, S.M., 2018. The time-course
880   of cortical responses to speech revealed by fast optical imaging. Brain and Language 184,
881   32–42. https://doi.org/10.1016/j.bandl.2018.06.006
882 Wang, L., Li, Y., Metzak, P., He, Y., Woodward, T.S., 2010. Age-related changes in topological
883   patterns of large-scale brain functional networks during memory encoding and
884   recognition. Neuroimage 50, 862–872.
885 Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world'networks. nature 393,
886   440–442.
887 Xia, M., Wang, J., He, Y., 2013. BrainNet Viewer: a network visualization tool for human brain
888   connectomics. PloS one 8, e68910.
889 Zhao, P., Yu, B., 2006. On model selection consistency of Lasso. Journal of Machine learning
890   research 7, 2541–2563.
891 Zhou, D., Thompson, W.K., Siegle, G., 2009. MATLAB toolbox for functional connectivity.
892   NeuroImage 47, 1590–1607. https://doi.org/10.1016/j.neuroimage.2009.05.089
893

894 **APPENDIX**

895 **GRAPH MINING**

896
897 Mathematical definitions and interpretation of network features are given below:

898
899 **Characteristics path**
900   A fundamental property of brain networks is functional integration, which indicates how
901 integrated a network is and, thus, how easily information flows (Rubinov and Sporns, 2010)
902 among nodes. A widely-used approach to estimate properties of functional integration between
903 nodes is based on the concept of characteristic path length. The characteristic path length is
904 defined as the average shortest path length in the network (Watts and Strogatz, 1998). Hence,
905 small characteristic path values imply dense connectivity and stronger potential for integration
906 among nodes. Let, L_i is the average distance between node i and all other nodes of a network,
907 Average Characteristic path is defined as:

$$908 \quad L = \frac{1}{n}\sum_{i \in N} L_i = \frac{1}{n}\sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{ij}}{n-1}$$

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

911
912 **Global efficiency**
913   Global efficiency (E) is used to find, how cost-efficient a particular network construction
914 and how fault tolerant the network is. Hence, high global efficiency, implying the excellent use
915 of resources. In brain connectivity analysis, structural and effective networks are similarly
916 organized and share high global efficiency. On the other hand, functional networks have weaker
917 connections and consequently share lower global efficiency (Honey et al., 2007). Global
918 efficiency is the average of inverse shortest path length hence inversely related to the average
919 characteristic path length. E is defined as:

920
$$E = \frac{1}{n}\sum_{i \in N} E_i = \frac{1}{n}\sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{i,j}^{-1}}{n-1} .$$

921

922 **Average clustering coefficient**

923 The average clustering coefficient for the network reflects, how close its neighbors are to

924 being a clique or complete graph. The average clustering coefficient of a node is defined as the

925 fraction of triangles around a node (Watts and Strogatz, 1998) and defined as:

926
$$C = \frac{1}{n}\sum_{i \in N} C_i .$$

927 Here, C_i is the clustering coefficient of ith node. Let k_i is the number of neighborhood node, and t_i

928 is the number of triangles created around i^{th} node. If a node has k neighbors, there are

929 $k(k-1)/2$ edges could exist among the nodes within the neighborhood. Hence, C can be

930 defined as:

931
$$C = \frac{1}{n}\sum_{i \in N} \frac{2t_i}{k_i(k_i-1)} .$$

932 **Transitivity**

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

934
$$T = \frac{\sum_{i \in N} 2t_i}{\sum_{i \in N} k_i(k_i-1)} .$$

935 The value of average clustering coefficient can be influenced by nodes with a low degree. But

936 transitivity is normalized collectively and consequently hence, does not have such problem

937 (Newman, 2003).

938

939 **Small-worldness**

940 Small-world network (S) is formally defined as networks that are significantly densely

941 clustered and have larger characteristic path length than random networks (Watts and Strogatz,

942 1998). Mathematically S can be expressed as:

943
$$S = \frac{C/C_{random}}{L/L_{random}} .$$

944 Where C and C_{rand} are the clustering coefficients, and L and L_{rand} are the characteristic path

945 lengths of the test network and an equivalent random network with the same degree on average

946 respectively. For a small world network S > 1, C >> C_{random} and L >> L_{random}. Such network tends

947 to contain more densely connected cliques/near-cliques/sub-networks than random network.

948 Those sub-networks are interconnected by one or more edge.

949

950 **Assortativity coefficient**

951 Despite the importance local and community structure, it is essential to study global

952 diversity in networks. Hence the tendency to connect nodes with similar numbers of edges. This

953 tendency, called assortativity, described crucial dynamic and structural properties of real-world

954 networks, such as epidemic spreading or error tolerance (Foster et al., 2010). A positive

955 assortativity coefficient indicates that nodes tend to link to other nodes with the same or similar

956 degree, on the other hand negative values indicate relationships between nodes of different

957 degree. Biological networks typically show negative assortativity coefficient as high degree

958 nodes tend to attach to low degree nodes (Piraveenan et al., 2012). Mathematically, the

959 assortativity coefficient is the Pearson correlation coefficient of degree between pairs of linked

960 nodes (Newman, 2002). Consider an undirected graph of N vertices and M edges with degree

961 distribution p_j. That is p_j is the probability that a randomly chosen node on the graph will have

962 degree k and q_k is the distribution of the remaining degree. This q_k captures the number of edges

25

963 leaving the node, other than the one that connects the pair. The assortativity coefficient (r) is
964 defined as:

965
$$r = \frac{\sum_{jk} jk(e_{jk} - p_j q_k)}{\sigma_q^2} .$$

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

968

969 **Modularity Index**
970   Modularity refers to the ability of subdivision the network into non-overlapping groups of
971 nodes (known as modules or community) in a way that maximizes the number of within-group
972 edges. Networks with high modularity have dense connections between the nodes within the
973 modules but sparse connections between nodes in different modules. Hence, modularity
974 quantifies the community strength of a test network by comparing the fraction of edges within
975 the community with respect to random network (Chen et al., 2014). It is widely used to discover
976 anatomical modules correspond to groups of specialized functional area which is previously
977 determined by physiological recordings. Usually, anatomical, effective and functional modules
978 in brain connectivity show extensive overlap (Rubinov and Sporns, 2010). Modularity index of a
979 given network is the fraction of the edges that fall within the given groups minus the expected
980 fraction if edges were distributed at random. Finding optimal modular structure is an
981 optimization problem. Any optimization approach generally sacrifices some degree of accuracy
982 for computational speed. Widely used algorithm to find optimal modular structure are proposed
983 by Newman et al. (Newman, 2004), and Blondel et al. (Blondel et al., 2008).

984

985 **Stability selection with Randomized Lasso**
986   Randomized Lasso (RL) (Meinshausen and Bühlmann, 2010) is a straightforward two
987 step approach. Instead of applying specific algorithm to the whole data set to determine the
988 selected set of variables based on the weight of coefficient, RL applied randomized lasso several
989 times to random subsamples of the data of size n/2 (n = number of samples) and chose those
990 variables that are selected consistently across subsamples. By performing this double
991 randomization several times, the method assigns high scores to features that are repeatedly
992 selected across randomizations. In short, features selected more often are considered good
993 features even though the "irrepresentable condition" (Zhao and Yu, 2006) is violated. This
994 approach is similar to the concept of bagging (Breiman, 1999) and sub-bagging (Büchlmann and
995 Yu, 2002) algorithm.
996   We know, Lasso has sparse solutions. For higher dimensional data, many estimated
997 coefficients of variables become zero. Removing the variables can be used to reduce the
998 dimensionality of the data. There are some limitations of Lasso-based feature selection are:

999   1. Lasso has a tendency to select an individual variable out of a group of highly correlated
1000   features.
1001   2. When the correlation between features is not too high, the performance of Lasso is
1002   restrictive.
1003 Lasso penalizes the absolute value of coefficients $|\beta|_k$ of every component with a penalty term
1004 proportional to the regularization parameter $\lambda \in \mathbb{R}$. On the other hand, Randomized Lasso
1005 penalizes using randomly chosen values in a range $[\lambda, \lambda/\alpha]$ where, $\alpha \in (0,1)$ is the weakness
1006 parameter. The concept of weakness parameter is closely related to weak greedy algorithms

26

1007   (Temlyakov, 2000). Let W_k be an i.i.d. random variable in a range from $(\alpha, 1)$ for k = 1, …., p.
1008   The estimator of Randomized Lasso can be written as (Meinshausen and Bühlmann, 2010):
1009
1010

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

1011
1012   Here, Y and X is the class label and feature matrix respectively. Implementation of equation:
1013   1 is a straightforward two-stage process:

1014   1. Re-scaling of the feature variables (with scale factor W_k for the k-th variable),
1015   2. LARS algorithm is applied on re-scaled variables (Efron et al., 2004).

1016
1017   In this approach, the reweighting is simply chosen at random. It is not sensible to expect
1018   improvement from randomization with one random perturbation. However, applying
1019   Randomized Lasso with many iterations (e.g. 1000 times) and looking for variables that are
1020   chosen frequently is a useful tool to find out stable feature (Meinshausen and Bühlmann, 2010).
1021
1022   By performing this double randomization several times, RL assigns high scores to
1023   features that are repeatedly selected across randomizations. if we run the Lasso for several
1024   bootstrapped replications of a given sample, then intersecting the supports of the Lasso bootstrap
1025   estimates leads to consistent model selection (Bach, 2008; Meinshausen and Bühlmann, 2010)
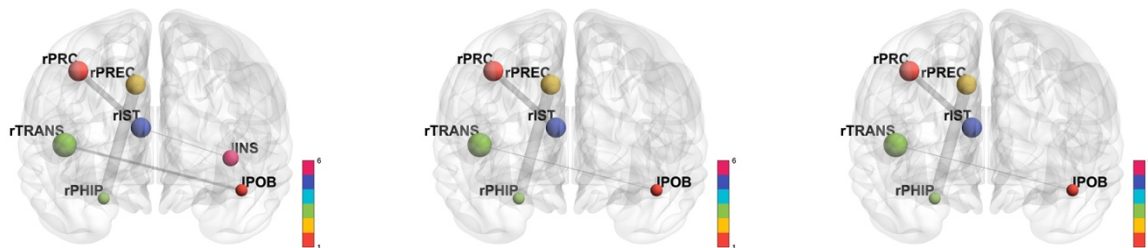1026
1027   **FIGURE**
1028

Figure 8: Brain network underlying Slow RT listeners (left), Medium RT listeners (middle), and Fast RT listeners. Shown here are the most highly correlated (absolute correlation ≥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.

1029