Journal of Neural Engineering

PAPER

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

RECEIVED
25 June 2019

REVISED
25 November 2019

ACCEPTED FOR PUBLICATION
10 December 2019

PUBLISHED
5 February 2020

Rakib Al-Fahad[1], Mohammed Yeasin[1,2] and Gavin M Bidelman[2,3,4,5]

[1] Department of Electrical and Computer Engineering, University of Memphis, Memphis, TN, United States of America
[2] Institute for Intelligent Systems, University of Memphis, Memphis, TN, United States of America
[3] School of Communication Sciences and Disorders, University of Memphis, Memphis, TN, United States of America
[4] University of Tennessee Health Sciences Center, Department of Anatomy and Neurobiology, Memphis, TN, United States of America
[5] Author to whom any correspondence should be addressed.

E-mail: gmbdlman@memphis.edu

Keywords: categorical speech perception, machine learning, speech processing, stability selection, functional connectivity

Abstract

Objective. 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. In this study, we introduce a novel approach to address these questions. *Approach*. 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. *Main results*. 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 RTs, (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. *Significance*. Findings show 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.

Introduction

When identifying speech, listeners naturally group sounds into smaller sets of discrete (phonetic) categories through the process of categorical perception (CP) (Liberman *et al* 1967, Pisoni 1973, Harnad and Bureau 1987, Pisoni and Luce 1987). 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 (Prather *et al* 2009). 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' of categorical decisions (Pisoni and Tash 1974, Bidelman *et al* 2013). 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) (Bidelman and Lee 2015c). RTs also slow near perceptual speech boundaries, where listeners shift from hearing one linguistic

class to another (e.g. /u/ versus /a/ vowel) and presumably require more time to access the 'correct' mental speech template (Pisoni and Tash 1974, Liebenthal *et al* 2010, Bidelman *et al* 2013, Reetzke *et al* 2018). 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 (Lively *et al* 1993, Liebenthal *et al* 2010, Bidelman and Walker 2017), auditory plasticity of short- (Liberman *et al* 1967) and long-term (Bidelman *et al* 2014b, Bidelman and Alain 2015b, Bidelman and Lee 2015a) experience, and neuropathologies and language-learning disorders (e.g. Calcus *et al* (2016), Hakvoort *et al* (2016) and Bidelman *et al* (2017, 2014a)). 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 (Pisoni and Luce 1987, Phillips 2001, Bidelman *et al* 2013).

The neuronal elements of the brain organize in complicated structural networks (Cajal 1995). Increasingly, it is appreciated that anatomical substrates constrain the dynamic emergence of coherent physiological activity that can span multiple spatially distinct brain regions (Bressler 1995, Fries 2005, Bullmore and Sporns 2009). 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 (Tononi *et al* 1994, Newman 2003, Bassett and Bullmore 2006, Honey *et al* 2007). 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. Binder *et al* (2004), Guenther Frank *et al* (2004), Myers *et al* (2009), Chang *et al* (2010), Liebenthal *et al* (2010), Bidelman and Lee (2015c), Alho *et al* (2016), Toscano *et al* (2018) and Bidelman and Walker (2019)). Previous studies also suggest that more neurons are preferentially activated by the prototypes of the speech categories compared to those at category boundaries (Guenther and Gjaja 1996). Similarly, improved discriminability at category boundaries could reflect an increased number of neurons encoding sensory cues at these perceptual transitions (Bauer and Der 1996, Guenther *et al* 1999). 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 listeners' perceptual speech identification as RTs are highly sensitive to individual differences in CP (Bidelman *et al* 2014a, 2014b, Bidelman and Alain 2015b, Bidelman and Walker 2017) 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. State-of-the-art studies usually use naive approaches to discover and analyze each edge individually and then compensate for possible errors arising from multiple comparisons (e.g. family-wise error or false discovery rate). These studies mostly yield an unstable set of network edges that are highly sensitive to changes in the hyperparameters within and between datasets (e.g. neural responses from different populations). In this regard, variable selection attempts to identify the most salient subset of variables from a larger set of features mixed with irrelevant variables. This problem is especially challenging when the number of available data samples is smaller compared to the number of possible predictors. Using generic subsampling and high-dimensional selection algorithms, stability selection can yield a stable set of features that distinguish subgroups of the data (e.g. here, listeners with slow versus fast perceptual decisions). It has widely been used in diverse fields of science, including gene selection and neuroimaging. One of the downsides of multivariate approaches are that outcomes often depend on model parameters (e.g. regularization factor). Compared to conventional multivariate approaches, stability selection produces more reliable estimations because of its internal randomization implemented as bootstrap-based subsampling (Meinshausen and Bühlmann 2010, Shah and Samworth 2013). Here, 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) and Price *et al* (2019)). Here, we take an entirely different, comprehensive *data-driven* approach to test whether individual differences in speeded 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 first 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. A second 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 recoded high-density electroencephalograms (EEGs) while listeners rapidly classified speech in a speeded vowel identification task (Bidelman *et al* 2013, Bidelman and Walker 2017). 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 speeded 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, whereas fast and medium responders show less utilization.

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 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. Figures 1(A) and (B) show the distribution of demographic measures (gender and age) of participants.

Speech stimulus continuum and behavioral task

We used a synthetic five-step vowel continuum to investigate the neural correlates of CP (figure 1(C)). 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/ (Bidelman *et al* 2013). 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 1(D)). 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 2(A) shows the BIC scores while tuning parameters. The '∗' indicates the optimal combination of components. The probability of each component

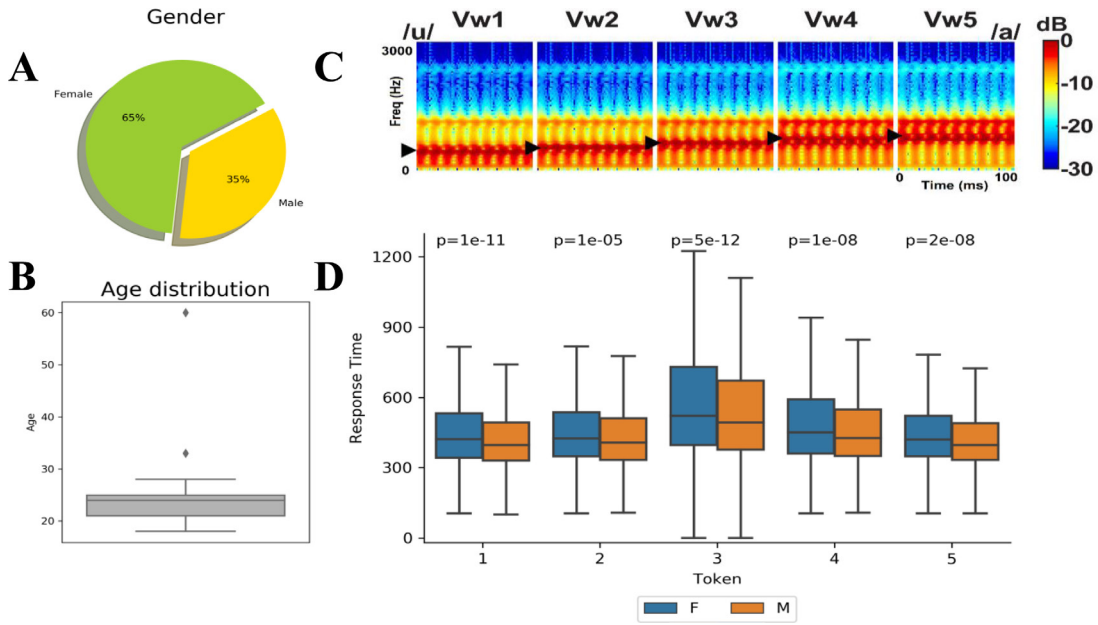

Figure 1. (A) and (B) Demographic gender and age distributions. (C) Acoustic spectrograms of the speech stimuli: The stimulus continuum was created by parametrically changing vowel first formant frequency over five equal steps from 430 to 730 Hz (►), resulting in a perceptual-phonetic continuum from /u/ to /a/. (D) Token-wise RTs for auditory classification. Listeners are slower to label sounds near the categorical boundary (i.e. Token 3). Females (F) have significantly slower RTs than males (M).

(see figure 2(B)) 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 2(C) shows token-wise RTs. 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 (2015a) and Bidelman and Walker (2017)). Briefly, neuroelectric activity was recorded from 64 sintered Ag/AgCl electrodes at standard 10–10 locations around the scalp (Oostenveld and Praamstra 2001). 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 $<10\,k\Omega$ during data collection. During acquisition, electrodes were referenced to an additional sensor placed ~1 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) (Picton *et al* 2000). Cleaned EEGs were then filtered (bandpass: 1–100 Hz; notch filter: 60 Hz), epoched (−200 to 800 ms)[7] into single trials, baseline corrected to 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 (Bidelman and Dexter 2015b, Bidelman and Howell 2016), we performed a distributed source analysis to more directly assess the neural generators underlying behavioral decisions related to CP. Source reconstruction was implemented

[7] 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 6). 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*.

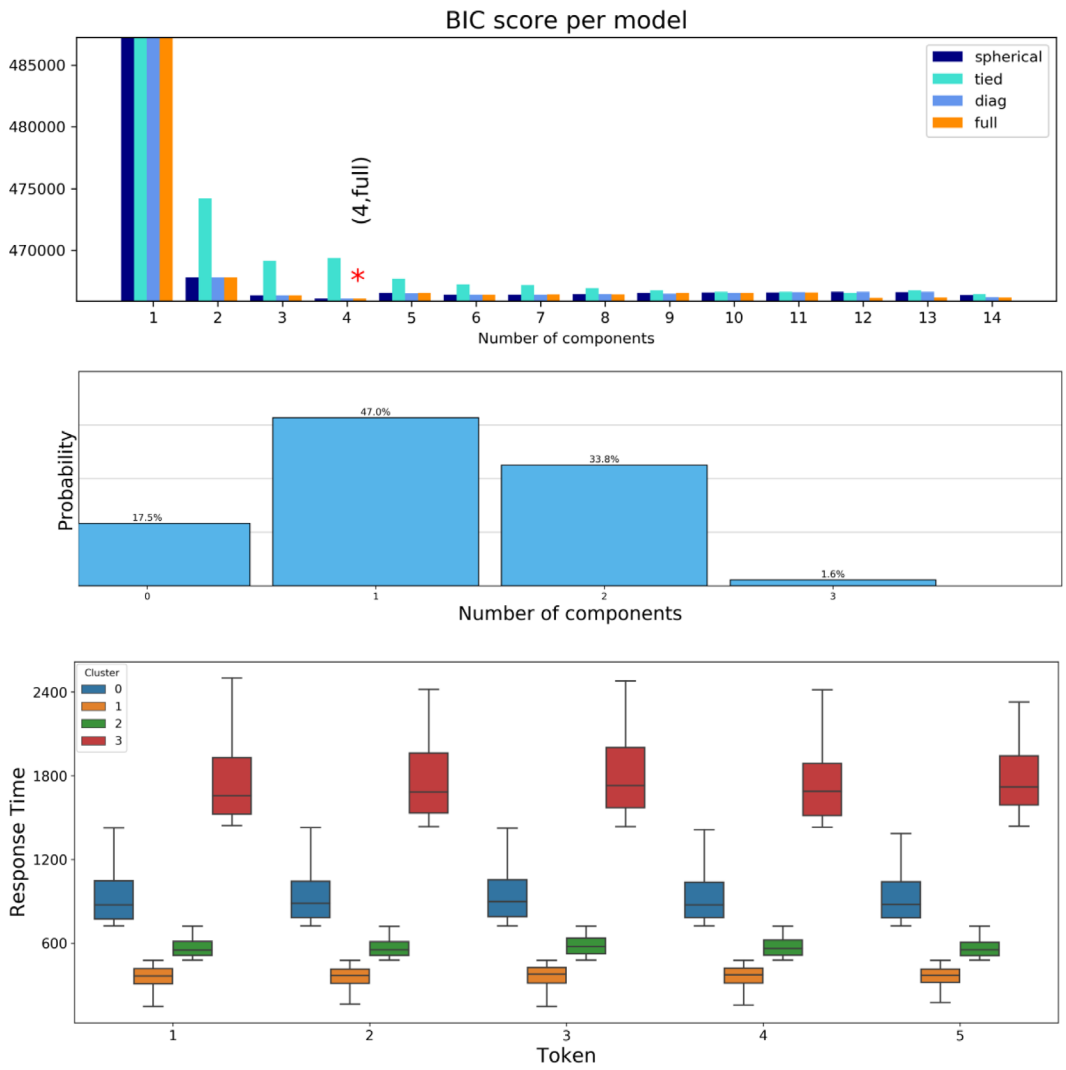

Figure 2. Clustering RT data using GMM and BIC criteria. Model selection concerns both the covariance type and number of components in the model. Brute-force based empirical analysis shows that $n = 4$ components with unique covariance matrix is optimal. The '$*$' marked position of (A) shows the optimal 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.

in the MATLAB package Brainstorm (Tadel *et al* 2011). We used a realistic, boundary element model (BEM) volume conductor (Fuchs *et al* 2002, 1998) standardized to the MNI template brain (Mazziotta *et al* 1995)[8]. The BEM head model was created using the OpenMEEG (Gramfort *et al* 2010) as implemented in Brainstorm on the MNI template brain (Tadel *et al* 2011). A BEM is less prone to spatial errors than other head models (e.g. concentric spherical conductor) (Fuchs *et al* 2002). 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 (Tadel *et al* 2011). The sLORETA provides a smoothness constraint that ensures the estimated current changes little between neighboring neural populations (Picton *et al* 2000, Michel *et al* 2004). This method is better than other inverse solutions because of its smaller average localization error. While higher channel counts improve source localization, for a 64-ch electrode array as used here, best-case estimates of localization error for sLORETA are as low ~1 mm (Song *et al* 2015).

From each single-trial sLORETA map, we extracted the time-course of source activity within 68 regions of

[8] Spatial accuracy of inverse source modeling from EEG can be improved by incoprating MRIs and electrode digitization at the single subject level. Our source reconsruction 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 ~5 mm (Acar and Makeig 2013) 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 (Shirazi and Huang 2019).

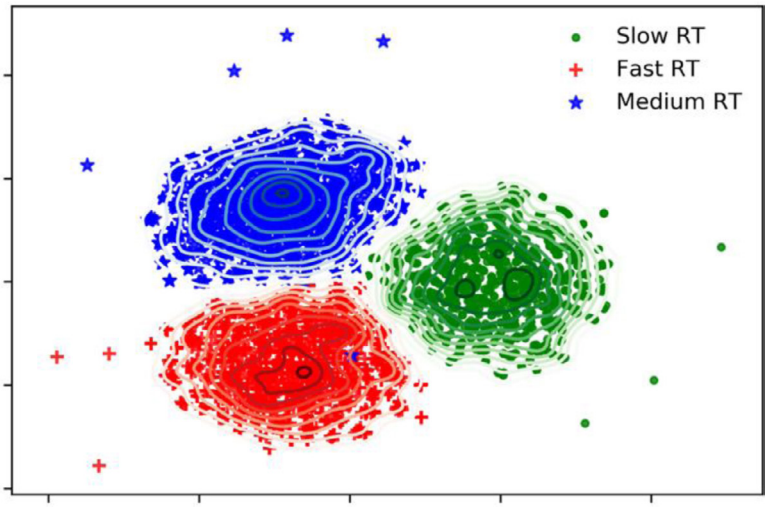

Figure 3. The t-SNE embedded higher dimensional functional connectivity data are represented by a 2-dimensional scatter and KDE plot. The green lines with '·', blue lines with '∗', and red lines with '+' represent data points for slow, medium, and fast RT participants, respectively.

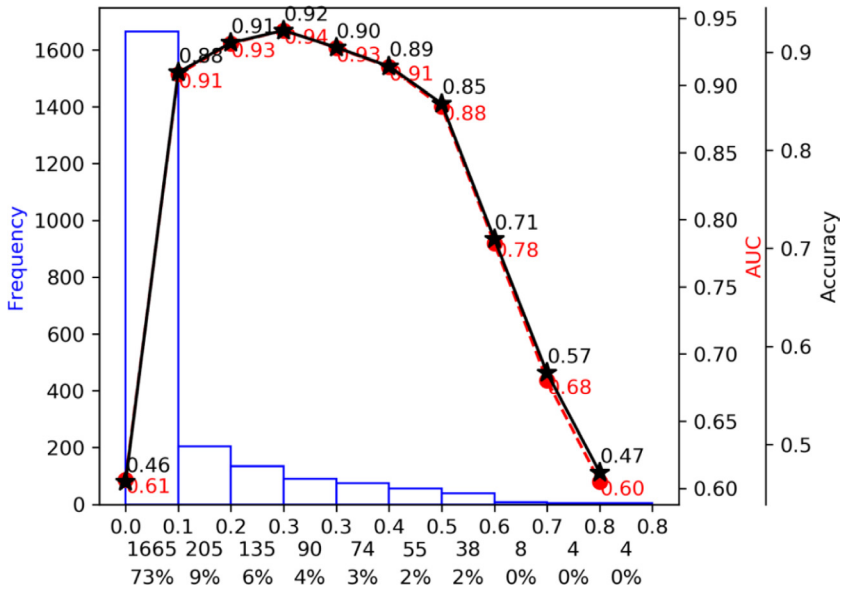

Figure 4. Effect of selection threshold on model performance prediction. The three *x*-labels represent (top) the range of each bin of features score (range: 0–1), (middle) the number of features falling in each bin, and (bottom) the corresponding percentage.

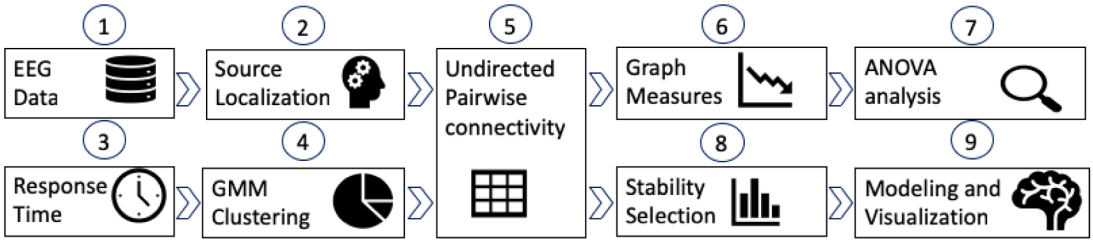

Figure 5. Schematic diagram of the processing pipeline. The 64-ch EEG data is first preprocessed, and then source localization is adapted to convert skull surface data to cortical surface time series data (68 ROIs defined by the Desikan-Killany Atlas parcellation). Pairwise correlations were calculated to derive the connectivity matrix for each trial of the speech CP task. Behavioral 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.

Table 1. 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
Assortativity	**0.0052**
Global Efficiency	**0.0290**
Transitivity	0.8424
Maximized Modularity	0.6617

Table 2. 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
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

interest (ROI) defined by the Desikan-Killany Atlas parcellation (Desikan *et al* 2006) 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 (Bashivan *et al* 2017)), 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 (Brunner *et al* 2016). Still, to ensure the robustness of our connectivity measures, we used bootstrapping to reduce the uncertainty of our connectivity estimates (James *et al* 2013). 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 35 106 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 (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 interactions at different time scales (Zhou *et al* 2009). Popular approaches to quantify functional connectivity are Correlation, Coherence (CH), imaginary part of coherency (iCH), Phase Locked Value (PLV), Phase Slope Index (PSI) (Lachaux *et al* 1999, Nolte *et al* 2004, Stam *et al* 2014). A comprehensive comparison of these methods showed that correlation-based connectivity out-performed the others in classify behavioral RTs (see Appendix for details; figure A1). 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. $3150 * 2278$) for each participant.

Seven global network connectivity features were estimated from each network graph using the BCT toolbox (Rubinov and Sporns 2010): (i) Characteristics path, (ii) Global efficiency, (iii) Average clustering coefficient, (iv) Transitivity, (vi) Small-worldness, (vi) Assortativity coefficient, and (vii) Maximized modularity (see Appendix for mathematical definitions and interpretation of these network features).

Machine learning: identifying behaviorally-relevant aspects of functional connectivity

To first visualize the data, we used the t-distributed stochastic neighbor embedding (t-SNE) (van der Maaten and Hinton 2008), which is a widely used unsupervised learning algorithm to visualize high-dimensional data. t-SNE converts similarities between higher dimensional data point to joint probabilities, providing 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 adopted LDA on our three-class connectivity dataset (i.e. fast, medium, slow responders identified from the behavioral data) and considered 50 dimensions for

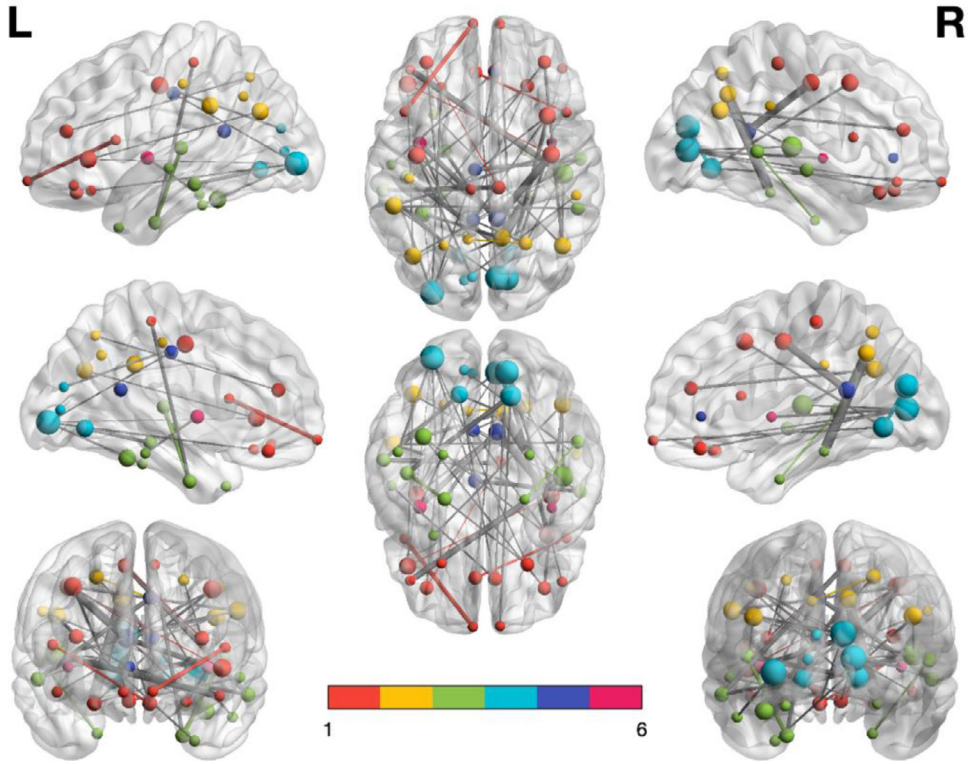

Figure 6. BrainNet visualization (top to bottom: lateral, medial, and dorsal view) of the brain network (54 edges) identified via stability selection. Color map 1–6 indicates, 1: Frontal (22 ROI), 2: Parietal (10 ROI), 3: Temporal (18 ROI), 4: Occipital (8 ROI), 5: Cingulate (8 ROI), 6: Insula (2 ROI) 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).

t-SNE[9]. The hyperparameters of t-SNE were tuned with a grid search approach. Figure 3 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 and is used here to visualize the trend of the data distribution for each different class (data points for fast, medium, and slow RTs). 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 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.

[9] Laurens van der Maaten *et al* (van der Maaten and Hinton 2008) recommend using another dimensionality reduction method (e.g. PCA or LDA) before applying t-SNE. This approach helps to suppress some noise and speed up data processing. Linear Discriminant Analysis (LDA) is a 'supervised' algorithm. It computes the directions that maximize the separation between classes. In short, LDA increases the inter class distance to make visualization more interpretable. Hence, we adopted LDA on our three-class dataset and consider 50 dimensions for t-SNE visualization, as recommended by Laurens van der Maaten *et al* (van der Maaten and Hinton 2008).

Table 3. 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 (features)	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

Feature selection

Feature selection attempts to identify the most salient subset of variables from a larger set of features mixed with irrelevant variables. This problem is especially challenging when the number of available data samples is small compared to the number of possible features. Conventional filter methods identify a consistent set of variables outside of the predictive model based on some filtering criteria, e.g. the variables are individually

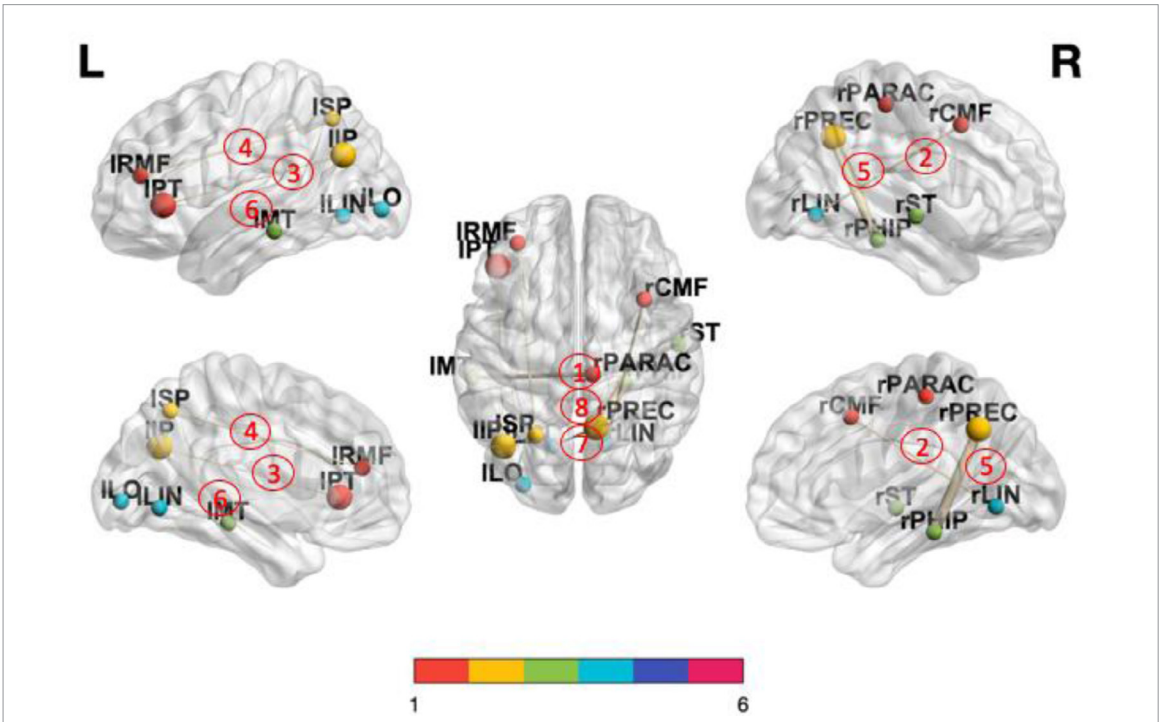

Figure 7. A sparse brain network (8 edges) predicts listeners' speed (RTs) of speech categorization (57% model accuracy). Red numbers are the ranked importance of the edges describing behavior. Otherwise, as in figure 6.

evaluated to check the probable relationship between classes. The sets of variables in this technique are selected based on a threshold of importance. Commonly filter-based methods include correlation, F-test, chi-square test, ANOVA analysis. The highly-correlated or redundant features may be selected, and significant interactions and relationships between variables may not be able to be quantified. However, one of the downsides of the multivariate approaches (e.g. PCA, LDA, Lasso, Elasticnet SVM ranking, Wrapper based methods, GA Wrapper, Forward Backward based methods) is that outcomes often depend on model parameters (e.g. regularization factor). Compared to conventional filter and multivariate approaches, stability selection produces more reliable estimations and yields a stable set of features because of its internal randomization implemented as bootstrap based subsampling. It was reported that even if the necessary conditions needed for consistency of the original Lasso (L1 norm penalized linear models) method are violated, stability selection will be consistent in variable selection (Meinshausen and Bühlmann 2010). The main advantages of this algorithm are (1) it works efficiently with the high-dimensional data, (2) stability selection provides finite sample control with error rates of false discoveries and is a transparent method to choose an amount of regularization for structure estimation; and (3) it is extremely general and has a very wide range of applicability.

An attractive feature of Lasso (L1 regularization on least squares) is its computational feasibility for the high-dimensional data with many more variables than samples since the optimization problem of lasso estimator is convex. Furthermore, the Lasso can select variables by shrinking certain estimated coefficients exactly to 0. Hence Lasso was used for stability selection. Applying Randomized Lasso many times and looking for variables that are chosen is a very powerful procedure to select consistent or stable features (Tibshirani 1996, Meinshausen and Bühlmann 2006, Shah and Samworth 2013, Al-Fahad *et al* 2017). Despite its simplicity, it is consistent for variable selection even though the 'neighborhood stability' condition is violated. More about stability section, interpretation, and mathematical definition are explained in the Appendix.

We used Randomized Logistic Regression for stability selection with randomized lasso. It works by subsampling the training data and fitting an L1-penalized Logistic Regression model where the penalty of a random subset of coefficients has been scaled. 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 need to specify the score to find out the best representative set of stable features. Hence, threshold selection is a 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%).

Table 4. 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

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.
4. Selected best models are evaluated on unseen test data. Accuracy (ACC) and area under curve (AUC) were considered for performance measures.

Figure 4 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 was not related to speeded 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 AUC = 0.9. The bell-shaped solid black and red dotted lines of figure 4 shows 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 3 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 ML 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 here for feature selection. Moreover, 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 ML field (Hwang *et al* 2013, Lotte *et al* 2018, Al-Fahad *et al* 2019). A schematic diagram of the method pipeline is shown in figure 5.

Results

Figure 1(D) 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 (Pisoni and Tash 1974, Liebenthal *et al* 2010, Bidelman *et al* 2013, Reetzke *et al* 2018). 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 2(C)). 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 (Rubinov and Sporns 2010) (see Methods).

We used non-parametric ANOVAs (Kruskal-Wallis H-test) to determine if individual trial-based global graph measures varied across RTs (table 1). This non-parametric test was used given the unequal sample size per group (Lowry 2014). These analyses revealed that Assortativity and Global Efficiency were modulated depending on behavior speed. Table 2 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 described in method section) showed expectedly poor accuracy (38%).

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%–16% (Salvador *et al* 2005, Wang *et al* 2010, Li *et al* 2016). 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 3, It was observed that only 7% error tolerance from the optimal value (accuracy from 92% to 85%) allows 80% less edge and 22% less associated nodes. Hence, the selection threshold 0.51 with reasonable performance (ACC = 85%, AUC = 0.9) were chosen for network visualization as performance declined precipitously above this threshold (figure 4).

Figure 6 shows a visualization of the 54 nodes among 53 ROIs identified via stability selection using BrainNet (Xia *et al* 2013). 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 threshold (=0.68) the stability-selected connectome. This resulted in eight highly ranked connectivity edges among 13 nodes across the brain (figures 7 and A2). Even with this sparse network of only eight edges, model classification was still 57%, meaning this small set of features accuracy predicted RTs. We then ranked the contribution of these stable nodes in table 4. 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.

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 (Tononi *et al* 1994, Newman 2003, Bassett and Bullmore 2006, Honey *et al* 2007). 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 (tables 1 and 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.

In addition, we found that the functional CP network underlying speeded decisions increased in negative assortativity (i.e. became disassortative) for slower RTs (table 2). 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 interpterion, we found that Network utilization (Global efficiency) also differentiated RT groups. Higher Global efficiency indicates that the routing of information among nodes with different degree 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 1(D)).

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.

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 (figure 7). 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. Golestani *et al* (2002), Binder *et al* (2004), Golestani and Zatorre (2004), Myers *et al* (2009), Chang *et al* (2010), Liebenthal *et al* (2010), Lee *et al* (2012), Bidelman and Lee (2015c), Alho *et al* (2016), Feng *et al* (2018), Luthra *et al* (2019)). 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 4). MTG has been associated with accessing word meaning while reading (Acheson and Hagoort 2013) and has been described as an early lexical interface that is heavily involved in sound-to-meaning inference (Hickok and Poeppel 2007, 2004). 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 (Sakurai *et al* 2008), indicating its role in the language production network (Blank *et al* 2002). 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 (Ingalhalikar *et al* 2014), and females may have a more diffuse, bilateral neural system for language processing than males (Shaywitz *et al* 1995). 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 (Mechelli *et al* 2000). It has also implicated in stimulus naming (Howard *et al* 1992, Bookheimer *et al* 1995), 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 (Liebenthal *et al* 2010, Feng *et al* 2018, Luthra *et al* 2019). 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 (Binder *et al* 2004, Du *et al* 2014, Bouton *et al* 2018) and the IFG-IPL pathway is upregulated when speech material is perceptually confusable (Feng *et al* 2018). 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 1(D)). 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 versus 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 (Bach 2008). 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 speeded 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 CP 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.

Acknowledgments

This work was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under award number NIH/NIDCD R01DC016267 (GMB).

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) (Haufe *et al* 2013). 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. 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:

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

The equation of Imaginary coherence (Nolte *et al* 2004) is given by:

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

The equation of Phase-Locking Value (Lachaux *et al* 1999) is given by:

$$\mathrm{PLV} = \left|E\left[S_{xy}/\left|S_{xy}\right|\right]\right|.$$

Each line in figure A1 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.

Graph mining

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 (Rubinov and Sporns 2010) 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 (Watts and Strogatz 1998). 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:

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

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 (Honey *et al* 2007). Global efficiency is the average of inverse shortest path length hence inversely

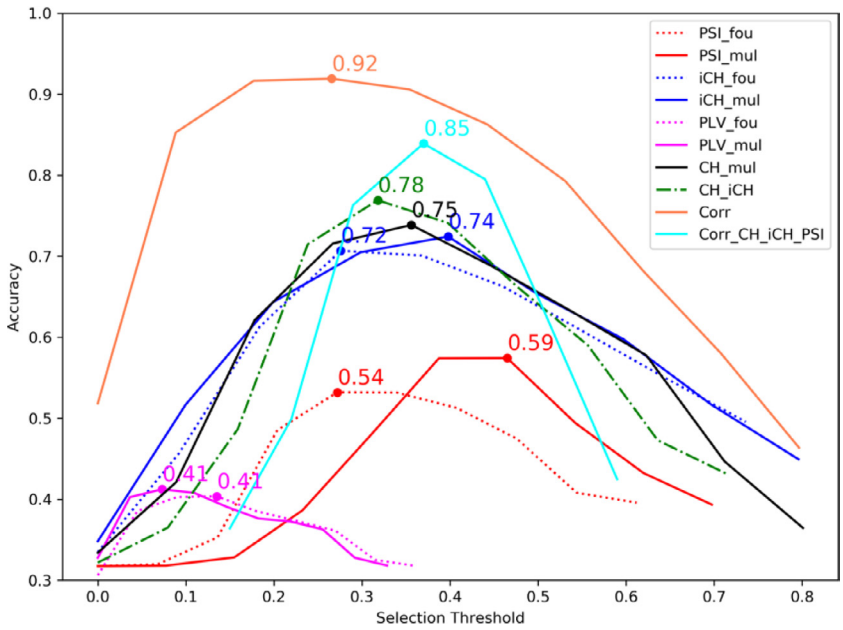

Figure A1. Accuracy curves of stability selection (as in figure 4). Stability selection was applied to Correlation, CH, iCH, PLV PSI based-edge matrix, as well as combinations of CH and iCH, combination of correlation, CH, iCH, and PSI based-edge matrix. Here 'mul' and 'fou' represents multitaper and 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.

related to the average characteristic path length. E is defined as:

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

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 (Watts and Strogatz 1998) and defined as:

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

Here, C_i is the clustering coefficient of ith node. Let k_i is the number of neighborhood nodes, and t_i is the number of triangles created around ith 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:

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

Transitivity

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

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

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 (Newman 2003).

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 (Watts and Strogatz 1998). Mathematically S can be expressed as:

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

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 $S > 1$, $C \gg C_{\text{random}}$ and $L \gg L_{\text{random}}$. 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 (Foster *et al* 2010). 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 (Piraveenan *et al* 2012). Mathematically, the assortativity coefficient is the Pearson correlation coefficient of degree between pairs of linked nodes (Newman 2002). 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:

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

Where, σ_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 (Chen *et al* 2014). 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 (Rubinov and Sporns 2010). 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* (Newman 2004) and Blondel *et al* (2008).

Stability selection with randomized Lasso

Randomized Lasso (RL) (Meinshausen and Bühlmann 2010)is a straightforward two-step approach. Instead of applying 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' (Zhao and Yu 2006) is violated. This approach is similar to the concept of bagging (Breiman 1999) and sub-bagging (Büchlmann and Yu 2002) 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 not too high, 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 (Temlyakov 2000). Let W_k be i.i.d. random variable in a range from $(\alpha, 1)$ for $k = 1, \ldots, p$. The estimator of Randomized Lasso can be written as Meinshausen and Bühlmann (2010):

$$\breve{\beta}^{\lambda,W} = \arg\min_{\beta \in \mathbb{R}^p} \|Y - X\beta\|_2^2 + \lambda \sum_{k=1}^{p} \frac{|\beta_k|}{W_k}. \qquad (A.1)$$

Here, Y and X are the class label and feature matrix respectively. Implementation of equation: (A.1) is a straightforward two-stage process:

1. Re-scaling of the feature variables (with scale factor W_k for the kth variable),
2. LARS algorithm is applied to re-scaled variables (Efron *et al* 2004).

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 find out stable feature (Meinshausen and Bühlmann, 2010).

By performing this double randomization several times, RL assigns high scores to features that are repeatedly selected across randomizations. if we run the Lasso for several bootstrapped replications of a given sample, then intersecting the supports of the Lasso bootstrap estimates leads to consistent model selection (Bach 2008, Meinshausen and Bühlmann 2010)

Figure

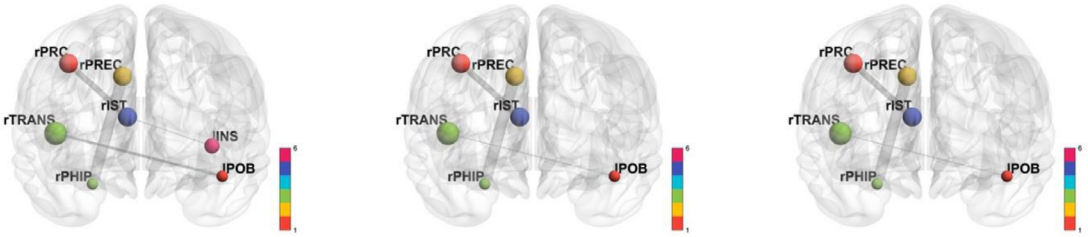

Figure A2. Brain network underlying Slow RT listeners (left), Medium RT listeners (middle), and Fast RT listeners. Shown here are the most highly correlated (absolute correlation $\geqslant 0.5$) network edges. Otherwise as in figures 6 and 7. INS, insula; IST, isthmus of cingulate; TRANS, transverse temporal gyrus (auditory cortex); POB, pars orbitalis; PRC, precentral gyrus (motor cortex); PHIP, parahippocampal gyrus; PREC, precunus; l/r, left/right hemisphere.

ORCID iDs

Rakib Al-Fahad https://orcid.org/0000-0002-0124-7254
Gavin M Bidelman https://orcid.org/0000-0002-1821-3261

References

Acar Z A and Makeig S 2013 Effects of forward model errors on EEG source localization *Brain Topograp.* **26** 378–96
Acheson D J and Hagoort P 2013 Stimulating the brain's language network: syntactic ambiguity resolution after TMS to the inferior frontal gyrus and middle temporal gyrus *J. Cognit. Neurosci.* **25** 1664–77
Al-Fahad R, Yeasin M, Anam A I and Elahian B 2017 Selection of stable features for modeling 4D affective space from EEG recording *2017 Int. Joint Conf. On Neural Networks* (IEEE) pp 1202–9
Al-Fahad R, Yeasin M, Glass J O, Conklin H M, Jacola L M and Reddick W E 2019 Early imaging-based predictive modeling of cognitive performance following therapy for childhood ALL *IEEE Access* **7** 146662–74
Alho J, Green B M, May P J, Sams M, Tiitinen H, Rauschecker J P and Jääskeläinen I P 2016 Early-latency categorical speech sound representations in the left inferior frontal gyrus *NeuroImage* **129** 214–23
Bach F 2008 Bolasso: model consistent Lasso estimation through the bootstrap *Proc. 25th Int. Conf. Mach. Learn.* pp 33–40 (arXiv:0804.1302)
Bashivan P, Yeasin M and Bidelman G M 2017 Temporal progression in functional connectivity determines individual differences in working memory capacity *2017 Int. Joint Conf. on Neural Networks* (IEEE) pp 2943–9
Bassett D S and Bullmore E D 2006 Small-world brain networks *Neuroscientist* **12** 512–23
Bauer H-U and Der R 1996 Controlling the magnification factor of self-organizing feature maps *Neural Computa.* **8** 757–71
Betzel R F, Bertolero M A and Bassett D S 2018 Non-assortative community structure in resting and task-evoked functional brain networks bioRxiv 355016
Bidelman G M and Alain C 2015a Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception *J. Neurosci.* **35** 1240–9
Bidelman G M and Dexter L 2015b 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 Lang.* **143** 32–41
Bidelman G M and Howell M 2016 Functional changes in inter-and intra-hemispheric cortical processing underlying degraded speech perception *NeuroImage* **124** 581–90
Bidelman G M and Lee C-C 2015c Effects of language experience and stimulus context on the neural organization and categorical perception of speech *NeuroImage* **120** 191–200
Bidelman G M and Walker B 2019 Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network *NeuroImage* **201** 116022
Bidelman G M and Walker B S 2017 Attentional modulation and domain-specificity underlying the neural organization of auditory categorical perception *Eur. J. Neurosci.* **45** 690–9
Bidelman G M, Lowther J E, Tak S H and Alain C 2017 Mild cognitive impairment is characterized by deficient brainstem and cortical representations of speech *J. Neurosci.* **37** 3610–20
Bidelman G M, Moreno S and Alain C 2013 Tracing the emergence of categorical speech perception in the human auditory system *NeuroImage* **79** 201–12
Bidelman G M, Villafuerte J W, Moreno S and Alain C 2014a Age-related changes in the subcortical–cortical encoding and categorical perception of speech *Neurobiol. Aging* **35** 2526–40
Bidelman G M, Weiss M W, Moreno S and Alain C 2014b Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians *Eur. J. Neurosci.* **40** 2662–73
Binder J R, Liebenthal E, Possing E T, Medler D A and Ward B D 2004 Neural correlates of sensory and decision processes in auditory object identification *Nat. Neurosci.* **7** 295
Blank S C, Scott S K, Murphy K, Warburton E and Wise R J 2002 Speech production: wernicke, broca and beyond *Brain* **125** 1829–38
Blondel V D, Guillaume J-L, Lambiotte R and Lefebvre E 2008 Fast unfolding of communities in large networks *J. Stat. Mech.: Theory Exp.* **2008** P10008
Bookheimer S Y, Zeffiro T A, Blaxton T, Gaillard W and Theodore W 1995 Regional cerebral blood flow during object naming and word reading *Hum. Brain Mapp.* **3** 93–106
Bouton S, Chambon V, Tyrand R, Guggisberg A G, Seeck M, Karkar S, van de Ville D and Giraud A-L 2018 Focal versus distributed temporal cortex activity for speech sound category assignment *Proc. Natl Acad. Sci.* **115** E1299–308
Breiman L 1999 Using adaptive bagging to debias regressions *Technical Report* 547 Statistics Dept. UCB
Bressler S L 1995 Large-scale cortical networks and cognition *Brain Res. Rev.* **20** 288–304
Brunner C, Billinger M, Seeber M, Mullen T R and Makeig S 2016 Volume conduction influences scalp-based connectivity estimates *Frontiers Comput. Neurosci.* **10** 121
Büchlmann P and Yu B 2002 Analyzing bagging *Ann. Stat.* **30** 927–61
Bullmore E and Sporns O 2009 Complex brain networks: graph theoretical analysis of structural and functional systems *Nat. Rev. Neurosci.* **10** 186

Cajal S R 1995 Histology of the nervous system of man and vertebrates *History of Neuroscience* (New York: Oxford University Press)
Calcus A, Lorenzi C, Collet G, Colin C and Kolinsky R 2016 Is there a relationship between speech identification in noise and categorical perception in children with dyslexia? *J. Speech, Lang. Hear. Res.* **59** 835–52
Chang E F, Rieger J W, Johnson K, Berger M S, Barbaro N M and Knight R T 2010 Categorical speech representation in human superior temporal gyrus *Nat. Neurosci.* **13** 1428–32
Chen M, Kuzmin K and Szymanski B K 2014 Community detection via maximization of modularity and its variants *IEEE Trans. Comput. Soc. Syst.* **1** 46–65
Desikan R S, Ségonne F, Fischl B, Quinn B T, Dickerson B C, Blacker D, Buckner R L, Dale A M, Maguire R P and Hyman B T 2006 An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest *NeuroImage* **31** 968–80
Du Y, Buchsbaum B R, Grady C L and Alain C 2014 Noise differentially impacts phoneme representations in the auditory and speech motor systems *Proc. Natl Acad. Sci.* **111** 7126–31
Efron B *et al* 2004 Least angle regression *Ann. Stat.* **32** 407–99
Feng G, Gan Z, Wang S, Wong P C M and Chandrasekaran B 2018 Task-general and acoustic-invariant neural representation of speech categories in the human brain *Cereb Cortex* **28** 3241–54
Foster J G, Foster D V, Grassberger P and Paczuski M 2010 Edge direction and the structure of networks *Proc. Natl Acad. Sci.* **107** 10815–20
Fries P 2005 A mechanism for cognitive dynamics: neuronal communication through neuronal coherence *Trends Cognit. Sci.* **9** 474–80
Friston K J, Jezzard P and Turner R 1994 Analysis of functional MRI time-series *Hum. Brain Mapp.* **1** 153–71
Fuchs M, Drenckhahn R, Wischmann H and Wagner M 1998 An improved boundary element method for realistic volume-conductor modeling *IEEE Trans. Biomed. Eng.* **45** 980–97
Fuchs M, Kastner J, Wagner M, Hawes S and Ebersole J S 2002 A standardized boundary element method volume conductor model *Clin. Neurophysiol.* **113** 702–12
Golestani N and Zatorre R J 2004 Learning new sounds of speech: reallocation of neural substrates *NeuroImage* **21** 494–506
Golestani N, Paus T and Zatorre R J 2002 Anatomical correlates of learning novel speech sounds *Neuron* **35** 997–1010
Gramfort A, Papadopoulo T, Olivi E and Clerc M 2010 OpenMEEG: opensource software for quasistatic bioelectromagnetics *Biomed. Eng. Online* **9** 45
Guenther F H and Gjaja M N 1996 The perceptual magnet effect as an emergent property of neural map formation *J. Acoust. Soc. Am.* **100** 1111–21
Guenther F H, Husain F T, Cohen M A and Shinn-Cunningham B G 1999 Effects of categorization and discrimination training on auditory perceptual space *J. Acoust. Soc. Am.* **106** 2900–12
Guenther Frank H, Nieto-Castanon A, Ghosh Satrajit S and Tourville Jason A 2004 Representation of sound categories in auditory cortical maps *J. Speech Lang. Hear. Res.* **47** 46–57
Hakvoort B, de Bree E, van der Leij A, Maassen B, van Setten E, Maurits N and van Zuijen T L 2016 The role of categorical speech perception and phonological processing in familial risk children with and without dyslexia *J. Speech Lang. Hear. Res.* **59** 1448–60
Harnad S and Bureau A S L 1987 *Categorical Perception: the Groundwork of Cognition* 2nd edn (Cambridge: Cambridge University Press)
Haufe S, Nikulin V V, Müller K-R and Nolte G 2013 A critical assessment of connectivity measures for EEG data: a simulation study *NeuroImage* **64** 120–33
Hickok G and Poeppel D 2004 Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language *Cognition* **92** 67–99
Hickok G and Poeppel D 2007 The cortical organization of speech processing *Nat. Rev. Neurosci.* **8** 393–402
Honey C J, Kötter R, Breakspear M and Sporns O 2007 Network structure of cerebral cortex shapes functional connectivity on multiple time scales *Proc. Natl Acad. Sci.* **104** 10240–5
Howard D, Patterson K, Wise R, Brown W D, Friston K, Weiller C and Frackowiak R 1992 The cortical localization of the lexiconspositron emission tomography evidence *Brain* **115** 1769–82
Hwang H-J, Kim S, Choi S and Im C-H 2013 EEG-based brain-computer interfaces: a thorough literature survey *Int. J. Hum.-Comput. Interact.* **29** 814–26
Ingalhalikar M, Smith A, Parker D, Satterthwaite T D, Elliott M A, Ruparel K, Hakonarson H, Gur R E, Gur R C and Verma R 2014 Sex differences in the structural connectome of the human brain *Proc. Natl Acad. Sci.* **111** 823–8
James G, Witten D, Hastie T and Tibshirani R 2013 *An Introduction to Statistical Learning* (Berlin: Springer)
Lachaux J-P, Rodriguez E, Martinerie J and Varela F J 1999 Measuring phase synchrony in brain signals *Hum. Brain Mapp.* **8** 194–208
Lee Y-S, Turkeltaub P, Granger R and Raizada R D S 2012 Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis *J. Neurosci.* **32** 3942–8
Li W, Wang M, Li Y, Huang Y and Chen X 2016 A novel brain network construction method for exploring age-related functional reorganization *Comput. Intell. Neurosc.* **2016** 5
Liberman A M, Cooper F S, Shankweiler D P and Studdert-Kennedy M 1967 Perception of the speech code *Psychol. Rev.* **74** 431–61
Liebenthal E, Desai R, Ellingson M M, Ramachandran B, Desai A and Binder J R 2010 Specialization along the left superior temporal sulcus for auditory categorization *Cereb Cortex* **20** 2958–70
Lively S E, Logan J S and Pisoni D B 1993 Training Japanese listeners to identify English /r/ and /l/. II: the role of phonetic environment and talker variability in learning new perceptual categories *J. Acoust. Soc. Am.* **94** 1242–55
Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A and Yger F 2018 A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update *J. Neural Eng.* **15** 031005
Lowry R 2014 *Concepts and Applications of Inferential Statistics* (http://vassarstats.net/textbook/)
Luthra S, Guediche S, Blumstein S E and Myers E B 2019 Neural substrates of subphonemic variation and lexical competition in spoken word recognition *Lang. Cogn. Neurosci.* **34** 151–69
Mazziotta J C, Toga A W, Evans A, Fox P and Lancaster J 1995 A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM) *NeuroImage* **2** 89–101
Mechelli A, Humphreys G W, Mayall K, Olson A and Price C J 2000 Differential effects of word length and visual contrast in the fusiform and lingual gyri during *Proc. R. Soc.* B **267** 1909–13
Meinshausen N and Bühlmann P 2006 High-dimensional graphs and variable selection with the lasso *Ann. Stat.* **xx** 1436–62
Meinshausen N and Bühlmann P 2010 Stability selection *J. R. Stat. Soc.* B **72** 417–73
Michel C M, Murray M M, Lantz G, Gonzalez S, Spinelli L and Grave de Peralta R 2004 EEG source imaging *Clin. Neurophysiol.* **115** 2195–222
Myers E B, Blumstein S E, Walsh E and Eliassen J 2009 Inferior frontal regions underlie the perception of phonetic category invariance *Psychol. Sci.* **20** 895–903
Newman M E 2002 Assortative mixing in networks *Phys. Rev. Lett.* **89** 208701
Newman M E 2003 The structure and function of complex networks *SIAM Rev.* **45** 167–256
Newman M E 2004 Fast algorithm for detecting community structure in networks *Phys. Rev.* E **69** 066133
Nolte G, Bai O, Wheaton L, Mari Z, Vorbach S and Hallett M 2004 Identifying true brain interaction from EEG data using the imaginary part of coherency *Clin. Neurophysiol.* **115** 2292–307
Oldfield R C 1971 The assessment and analysis of handedness: the edinburgh inventory *Neuropsychologia* **9** 97–113

Oostenveld R and Praamstra P 2001 The five percent electrode system for high-resolution EEG and ERP measurements *Clin. Neurophysiol.* **112** 713–9
Phillips C 2001 Levels of representation in the electrophysiology of speech perception *Cogn. Sci.* **25** 711–31
Picton T W, van Roon P, Armilio M L, Berg P, Ille N and Scherg M 2000 The correction of ocular artifacts: a topographic perspective *Clin. Neurophysiol.* **111** 53–65
Piraveenan M, Prokopenko M and Zomaya A 2012 Assortative mixing in directed biological networks *IEEE/ACM Trans. Comput. Biol. Bioinformat.* **9** 66–78
Pisoni D B 1973 Auditory and phonetic memory codes in the discrimination of consonants and vowels *Percept. Psychophys.* **13** 253–60
Pisoni D B and Luce P A 1987 Acoustic-phonetic representations in word recognition *Cognition* **25** 21–52
Pisoni D B and Tash J 1974 Reaction times to comparisons within and across phonetic categories *Percept. Psychophys.* **15** 285–90
Prather J F, Nowicki S, Anderson R C, Peters S and Mooney R 2009 Neural correlates of categorical perception in learned vocal communication *Nat. Neurosci.* **12** 221–8
Price C N, Alain C and Bidelman G M 2019 Auditory-frontal channeling in α and β bands is altered by age-related hearing loss and relates to speech perception in noise *Neuroscience* **423** 18–28
Reetzke R, Xie Z, Llanos F and Chandrasekaran B 2018 Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood *Curr. Biol.* **28** 1419–27.e4
Rubinov M and Sporns O 2010 Complex network measures of brain connectivity: uses and interpretations *NeuroImage* **52** 1059–69
Sakurai Y, Mimura I and Mannen T 2008 Agraphia for kanji resulting from a left posterior middle temporal gyrus lesion *Behav. Neurol.* **19** 93–106
Salvador R, Suckling J, Coleman M R, Pickard J D, Menon D and Bullmore E D 2005 Neurophysiological architecture of functional magnetic resonance images of human brain *Cereb. Cortex* **15** 1332–42
Shah R D and Samworth R J 2013 Variable selection with error control: another look at stability selection *J. R. Stat. Soc.* B **75** 55–80
Shaywitz B A *et al* 1995 Sex differences in the functional organization of the brain for language *Nature* **373** 607
Shirazi S Y and Huang H J 2019 More reliable EEG electrode digitizing methods can reduce source estimation uncertainty, but current methods already accurately identify brodmann areas bioRxiv 557074
Song J, Davey C, Poulsen C, Luu P, Turovets S, Anderson E, Li K and Tucker D 2015 EEG source localization: sensor density and head surface coverage *J. Neurosci. Methods* **256** 9–21
Stam C J, Tewarie P, Van Dellen E, van Straaten E C W, Hillebrand A and Van Mieghem P 2014 The trees and the forest: characterization of complex brain networks with minimum spanning trees *Int. J. Psychophysiol.* **92** 129–38
Tadel F, Baillet S, Mosher J C, Pantazis D and Leahy R M 2011 Brainstorm: a user-friendly application for MEG/EEG analysis *Comput. Intell. Neurosci.* **2011** 8
Temlyakov V N 2000 Weak greedy algorithms [*] This research was supported by National Science Foundation Grant DMS 9970326 and by ONR Grant N00014-96-1-1003 *Adv. Comput. Math.* **12** 213–27
Tibshirani R 1996 Regression shrinkage and selection via the Lasso *J. R. Stat. Soc.* B **58** 267–88
Tononi G, Sporns O and Edelman G M 1994 A measure for brain complexity: relating functional segregation and integration in the nervous system *Proc. Natl Acad. Sci.* **91** 5033–7
Toscano J C, Anderson N D, Fabiani M, Gratton G and Garnsey S M 2018 The time-course of cortical responses to speech revealed by fast optical imaging *Brain Lang.* **184** 32–42
van der Maaten L and Hinton G 2008 Visualizing data using t-SNE *J. Mach. Learn. Res.* **9** 2579–605
Wang L, Li Y, Metzak P, He Y and Woodward T S 2010 Age-related changes in topological patterns of large-scale brain functional networks during memory encoding and recognition *NeuroImage* **50** 862–72
Watts D J and Strogatz S H 1998 Collective dynamics of 'small-world' networks *Nature* **393** 440–42
Xia M, Wang J and He Y 2013 BrainNet Viewer: a network visualization tool for human brain connectomics *PLoS One* **8** e68910
Zhao P and Yu B 2006 On model selection consistency of Lasso *J. Mach. Learn. Res.* **7** 2541–63
Zhou D, Thompson W K and Siegle G 2009 MATLAB toolbox for functional connectivity *NeuroImage* **47** 1590–607