Using EEG to discriminate cognitive workload and performance based on neural activation and connectivity*

James R. Williamson[1], Thomas F. Quatieri[1], Christopher J. Smalt[1], Joey Perricone[1], Brian J. Helfer[1], Michael A. Nolan[1], Marianna Eddy[2], Joseph Moran[2]

[1]MIT Lincoln Laboratory, Lexington, Massachusetts, USA
[2]Natick Soldier RD&E Center (NSRDEC), Natick, Massachusetts, USA

Abstract. A major goal of noninvasive brain sensing is to ascertain both the *workload* and the *efficacy* of cognitive processing. Realizing this goal will assist in monitoring cognitive readiness under different levels of cognitive workload and fatigue. Our approach to discriminating a person's cognitive state is predicated on the idea that cognition depends on coordinated neural activations, operating over a range of frequencies, that link functional networks across multiple brain regions. Therefore, our approach focuses on characterizing neural activation and connectivity patterns across the brain within multiple frequency bands. In each band, neural activations are characterized using spatial distributions of power across EEG channels, and neural connectivities are characterized using the eigenspectra of EEG connectivity matrices. The connectivity matrices are constructed using two measures: coherence and covariance. We use an auditory working memory task to vary cognitive workload by altering the number of digits held in memory during the simultaneous retention of a sentence in memory. Cognitive efficacy is assessed based on accuracy in recalling digits from memory. A Gaussian classifier is used to discriminate cognitive load and performance from EEG recorded during each experimental trial, and quantify discrimination accuracy with the area under the receiver operating characteristic curve (AUC) statistic. For cognitive load discrimination, AUC values of 0.59, 0.56, and 0.60 are obtained using power-, coherence-, and covariance-based feature sets, respectively. For cognitive performance discrimination, AUC values of 0.49, 0.62, and 0.63 are obtained for the same feature sets. Therefore, neural activation features (EEG band power) are shown to be relatively effective in discriminating workload but relatively ineffective in discriminating efficacy, compared to the other feature sets. Coherence-based connectivity features produce the opposite result, being relatively ineffective in discriminating load but effective in discriminating efficacy. Covariance-based connectivity features, on the other hand, are relatively effective in both tasks. We advance the hypothesis that robustness is obtained in covariance-based connectivity features due to the fact that they jointly capture information about both neural activations and connectivities.

*This work is sponsored by the Assistant Secretary of Defense for Research & Engineering under Air Force contract #FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

1. Introduction

Cognitive load is defined loosely as the mental demand experienced for a particular task. Efficient and effective methods are needed to monitor cognitive load under cognitively and physically stressful situations. In many scenarios, environmental and occupational stressors can produce cognitive overload, thereby degrading task performance and endangering safety. Examples of mental stressors are repetitive and/or intense cognitive tasks, psychological stress, and lack of sleep. Physical stressors include intense long-duration operations and heavy physical exertion. Both mental and physical stressors can contribute to cognitive load. In operational settings, the objective is often to quickly assess cognitive ability and readiness under cognitively loaded conditions, regardless of their etiology. Effective monitoring therefore requires the ability to simultaneously assess both the *level* of cognitive workload and the *effectiveness* of cognition under the existing load.

One major factor that impacts cognitive load is the amount of working memory required in a task (Lively et al, 1993, Yin et al, 2008). In this study we investigate the ability to discriminate the level of both cognitive workload and cognitive performance based on EEG signals recorded during an auditory working memory task. Cognitive load is quantified by the numbers of digits held in memory, and cognitive performance by the success of digit recall. Our guiding principle is that successful cognition requires coordinated neural activations, over a range of frequencies, in functional networks linking multiple brain regions. Therefore, in order to discriminate cognitive workload and performance from EEG recordings we use feature approaches that extract patterns of neural activation and connectivity in multiple frequency bands.

Each extracted feature vector encodes summary statistics from recorded EEG data during a single experimental trial, using recorded EEG data intervals that range between 1.9 and 5.7 seconds in duration. Patterns of neural activity are represented using EEG log power across the 64 channels in each frequency band. Patterns of neural connectivity are represented based on 64 x 64 element connectivity matrices in each frequency band. To characterize these high dimensional connectivity patterns, an approach is used that is invariant to the channel identities, encoding instead the *structure* of the set of connections, based on rank-ordered eigenvalues of the connectivity matrices (Schindler et al, 2007; Williamson et al, 2011, 2012; Ma and Bliss, 2014; Quatieri et al, 2016; Helfer et al, 2016). The motivation for this approach is that a structural representation may provide a deeper and more generalizable representation of the state of a complex system by encoding its state space dimensionality. Statistical models are used to associate these feature vectors with outcome variables, which are defined at each trial. The two outcome variables are 1) cognitive load level (*low, medium, high*) and 2) cognitive performance (*correct recall, incorrect recall*).

Our paper is organized as follows. In Section 2, the auditory working memory data collection is described, which uses a novel cognitive load protocol that taxes auditory working memory by eliciting recall of sentences and digits under varying levels of cognitive load, with potential effects of load being measurable using EEG, audio, video, and physiological sensing. The EEG preprocessing methods are also described. In Section 3 the feature extraction and machine learning methodologies are described. In Section 4 the effect of cognitive load and performance levels on feature distributions are analyzed and illustrated using brain activation and connectivity maps. Discrimination results for cognitive load and accuracy are also presented. Finally, in Section 5 the implications for real-time EEG monitoring and directions for future work are discussed.

2. Protocol and Materials

2.1 Working memory task

Subjects gave informed consent to our working memory-based protocol approved by the MIT Committee on the Use of Humans as Experimental Subjects (COUHES). EEG signals were collected with a 64-element Neuroscan device, as well as audio and video signals (Quatieri et al, 2015). Following setup and training, each subject engaged in the primary task of verbally recalling sentences with varying levels of cognitive load, as determined by the number of digits being held in working memory (Levitt, 1971; Le et al, 2009; Harnsberger et al, 2008). A single trial of the auditory working memory task comprises: the subject hearing a string of digits, then hearing a sentence, then waiting for a tone eliciting spoken recall of the sentence, followed by another tone eliciting recall of the digits, as shown in Fig. 1. This task is administered with three difficulty levels, involving 108 trials per level. The same set of 108 sentences is used in each difficulty level. The order of trials (sentences and difficulty level) is randomized. The multi-talker PRESTO sentence database is used for sentence stimuli (Park et al, 1958). EEG was recorded from 16 subjects.

In the working memory task an initial calibration test was done to assess each subject's ability. During calibration, the maximum number of digits that a subject can accurately recall was estimated using an adaptive tracking algorithm (Levitt, 1971). This maximum number, *n*, was allocated among the 16 subjects as follows: *n=4* (four subjects), *n=5* (six subjects) *n=6* (four subjects), and *n=7* (two subjects). Three load levels were selected based on *n*. For the first four subjects enrolled in the study, the numbers of digits for the three load levels were selected as: *n*, *n-2*, and max(*1,n-4*). For the final 12 subjects enrolled in the study, in order to explore finer-grained differences in load, the numbers of digits were selected as: *n, n-1*, and *n-2*. The analysis in this paper focuses on these small load differences, with three cognitive load levels, *low*, *medium*, and *high,* defined as *n-2, n-1*, and *n* digits, respectively. Therefore, low- and high-load data are available from all 16 subjects, but medium-load data are available from only the final 12 subjects. Given that a subject's total auditory memory load comprises both the sentence and the digits, the differences in absolute load between the three load classes are relatively small.

A separate subset of this database was carved out for evaluating cognitive performance. Task difficulty was restricted to a fixed load level of 4 digits. This was done to assess cognitive efficacy while avoiding a possible confound with task difficulty. It also simplifies interpretation of digit recall performance. Since digit recall is scored as correct only if all digits are named in their correct order, scoring is inherently more stringent for longer digit lists. In total, the cognitive performance data set consists of 108 trials each from 14 different subjects, of which 1,124 are correct recall trials and 388 are incorrect recall trials.

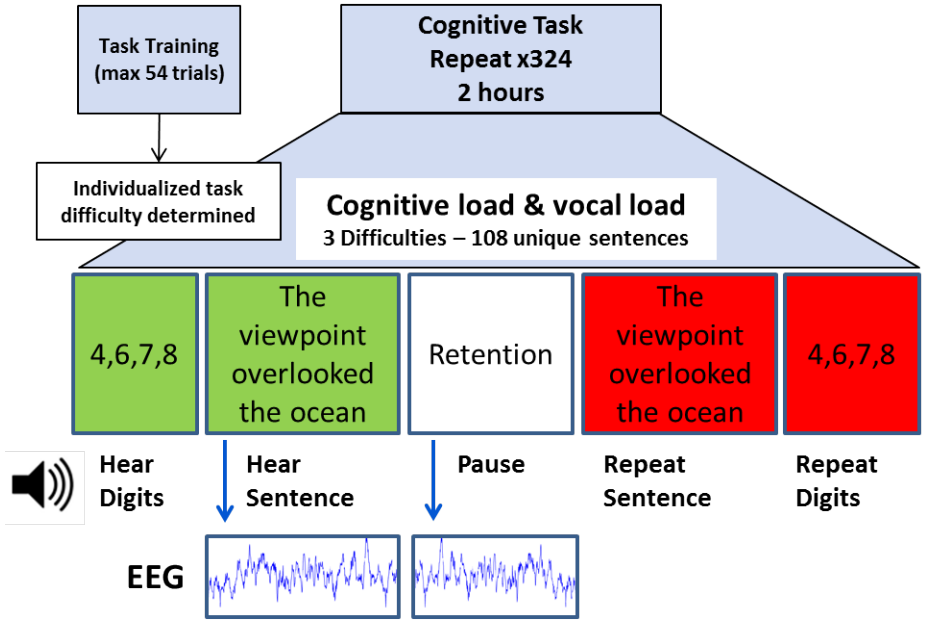


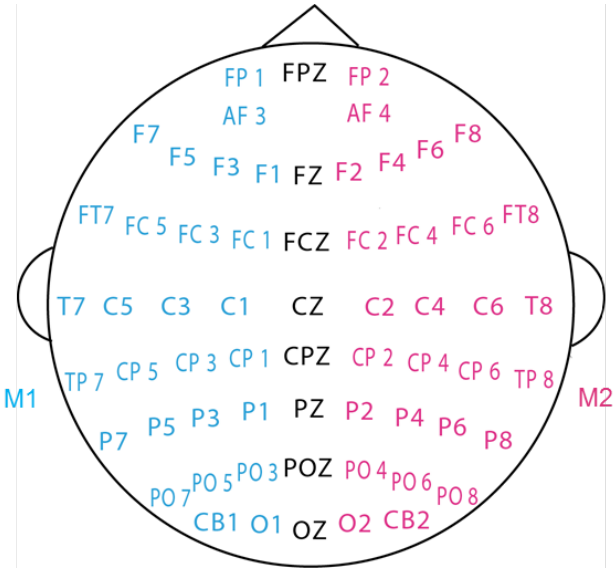
Fig. 1. Auditory working memory protocol.

2.2 EEG recordings

For analysis of cognitive load and performance, EEG from two data intervals (see Fig. 1) are analyzed: the *sentence* interval when the subjects hears a sentence, and the *retention* interval, when the subject holds both digits and sentence in memory. The sentence duration varies from 1.9 to 5.7 seconds, with a mean of 3.3 seconds. The retention duration is two seconds. Data from the digit-hearing interval are not used because the number of digits is confounded with the load level. Likewise, data intervals during which the subject is speaking are avoided due to motion and muscle artifacts.

EEG data (64 channels) were collected at a 1,000 Hz sample rate using Neuroscan's QuikCap and SynAmps 2 amplifier. Scalp electrodes (62) were arranged in an Extended 10/20 international labeling system (Compumedicsneuroscan). In addition to 62 scalp electrodes, two electrodes were placed at the mastoids, as well as facial electrodes placed above, below, and next to the eyes to capture vertical and horizontal eye movements. Fig. 2 illustrates the layout of EEG channels on a head map and lists, row by row, the correspondence between indices and channels. During recording, data were referenced to an electrode located just posterior to the central, midline electrode (Cz), then re-referenced using an average reference. After re-referencing, data were resampled to 500 Hz and high-passed filtered with a lower edge at 1 Hz using EEGLAB's default FIR filter (EEGLAB). Additionally a notch filter was applied

between 59.75 and 60.25 Hz to remove 60 Hz line-noise. Finally, data were submitted to Infomax independent components analysis using the runica() function in EEGLAB. ICA produces a set of spatially fixed, temporally independent components that are useful for identifying artefactual sources such as blinks. The blink-component was manually identified for each subject by looking at component scalp-maps and time-series and removed from the data by back-projecting non-artefactual sources to channels (EEGLAB).



Indices	Channels
1 – 3	FP1 – FP2
4 – 5	AF3 – AF4
6 – 14	F7 – F8
15 – 23	FT7 – FT8
24 – 32	T7 – T8
33 – 43	M1 – M2
44 – 52	P7 – P8
53 – 59	PO7 – PO8
60 – 64	CB1 – CB2

Fig. 2. Left: International system for 64-channel EEG cap. Right: correspondence between indices and EEG channels, on row-by-row basis.

3. Methods

3.1 Feature extraction

The feature sets used in this paper follow two basic approaches for representing EEG data at each frequency band. The first approach represents neural activation patterns based on log power across the 64 EEG channels. The second approach represents connectivity structure based on the rank-ordered eigenvalues of 64 x 64 EEG connectivity matrices. Connectivity, in turn, is measured in two different ways, coherences and covariance. In Fig. 3, the differences between power features and covariance-based connectivity features are illustrated based on different hypothetical distributions of signals in two EEG channels at the same frequency band. In the EEG channel power approach, the features (indicated by the blue arrow lengths) represent the variance in each data axis (EEG channel). Observe that this representation is invariant to any covariation between the channels. The covariance-based connectivity approach, termed *covariance structure*, is based on the eigenspectrum of the multichannel EEG covariance matrix at each frequency band. In this approach, the eigenvalue features represent the variances, ordered from largest to smallest, of the orthogonal principal axes of the multivariate EEG scatter distribution. These features are indicated by the lengths of the red arrows. Observe that this

representation is invariant to how the principal axes of the distribution project onto the data axes (EEG channels).

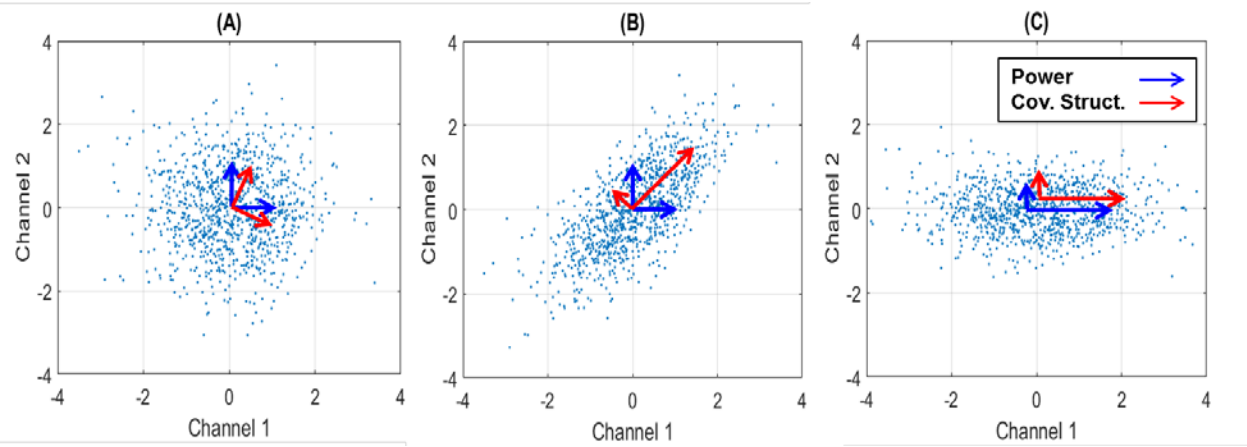


Fig. 3. Notional example showing power (blue) and covariance structure (red) features given three data distributions from two EEG channels. A) EEG channels with unit variance and zero covariance results in identical feature vectors. B) EEG channels with unit variance and large covariance leaves power features unchanged but causes covariance structure features to encode the elongated shape of the distribution. C) If the same principal axes of the distribution align with the channel axes, then power features also encode the elongated shape.

Frequency band power features:

Features based on the power in EEG channels at discrete frequency bands are a standard univariate method for EEG analysis and have been successfully used for estimating cognitive load (Zarjam et al, 2015) and for other applications such as epileptic seizure detection (Shoeb et al, 2004) and seizure prediction (Park et al, 2011). The EEG signals are decomposed into five frequency bands (delta, theta, alpha, beta, gamma), with band ranges of 0-4, 4-8, 8-16, 16-32, and 32-49 Hz, respectively (Zarjam et al, 2015). In each frequency band a 64-dimensional vector is computed, consisting of the logarithm of the power in that band.

Specifically, the spectral density estimate is computed at each frequency using Welch's averaged, modified periodogram method, in which the EEG data is divided into eight sections with 50% overlap, a Hamming window is applied to each section, and eight modified periodograms are computed and averaged. A rectangular approximation is used to integrate the power spectral density over frequencies within each frequency band.

Fig. 4 provides an illustration of the power features obtained from a single subject, showing the features in the beta band from the sentence interval of four different trials. The beta band is shown because it proves to be the most discriminative frequency band for cognitive load. Examples of high and low cognitive load are shown in Fig. 4 (top), where there is more power in left frontal channels in the low load condition. This result is broadly consistent with previous experimental results (Zarjam et al, 2015) and with our experimental results, which are described in Section 4.1: on average, high loads result in

lower power across all channels, particularly in frontal and midline areas. This indicates an ability to discriminate load using these features, as is demonstrated in Section 4.2.

In the high and low load trials illustrated in Fig. 4 (top), the digits were correctly recalled. In Fig. 4 (bottom), the features are shown from two medium load trials with different recall outcomes (correct and incorrect), showing the effect of accuracy variation at the same load. In these two trials, there is little difference in the spatial distribution of power across channels. The net experimental result, described in Section 4.1, is consistent with this, with similar power levels found in the beta band for correct and incorrect digit recall. This indicates an inability to discriminate cognitive performance based on these features, which is what is demonstrated in Section 4.2.

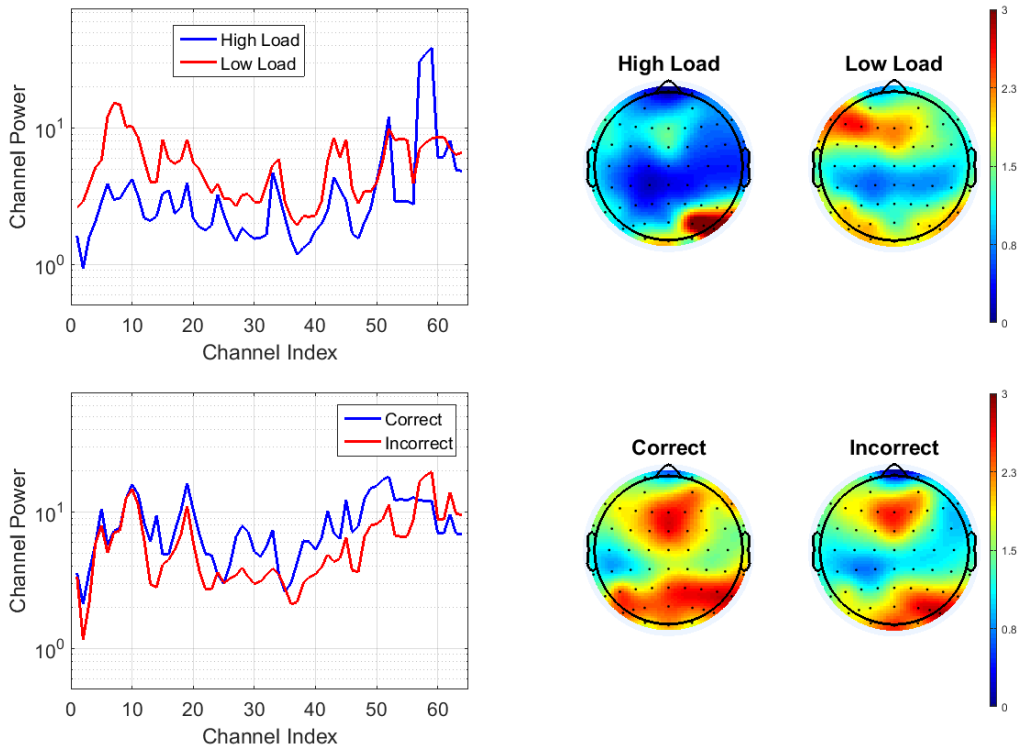


Fig. 4. Beta band power features obtained from sentence interval of single trials for the same subject. Top: power features for high load (5 digits) and low load (3 digits), plotted as function of channel indices (left) and as head map images of log power (right). Bottom: power features from medium load (4 digits) trials with correct and incorrect digit recall.

Coherence structure features:

While channel power features can provide useful indications of regional neural activity at different frequency bands, they do not provide direct measures of interactions across brain regions. Features that quantify the structural properties of brain connectivity may therefore provide complementary information by characterizing how communication among brain regions is distributed. The connectivity structure feature approach is explained using two connectivity measures: *coherence* and *covariance*.

For coherence measures, auto-spectral and cross-spectral density estimates are computed at each frequency using Welch's method (described above). The average coherence is then computed, which is the cross-channel power relative to within-channel power over the frequency band,

$$s_{x,y}(j) = \frac{1}{n_f} \sum_{f \in \omega_j} \frac{\left|G_{xy}(f)\right|^2}{G_{xx}(f) G_{yy}(f)}. \quad (1)$$

$G_{xy}(f)$ is the cross-spectral density between the signals x and y at frequency f, and $G_{xx}(f)$, $G_{yy}(f)$ are the auto-spectral densities. The magnitude of the spectral density is denoted by $|G|$.

In order to characterize the entire pattern of multichannel coherences, we construct an $M \times M$ coherence matrix in each frequency band, j, with $M = 64$. Each matrix element contains the pairwise coherence between two channels,

$$S_j = \begin{bmatrix} s_{1,1}(j) & \cdots & s_{1,M}(j) \\ \vdots & \ddots & \vdots \\ s_{M,1}(j) & \cdots & s_{M,M}(j) \end{bmatrix}. \quad (2)$$

The distributional properties of the set of coherences are then quantified using the matrix eigenspectra, which are the eigenvalues ordered from largest to smallest,

$$\left\{\lambda_j^S(1), \ldots, \lambda_j^S(M)\right\} = \mathrm{eig}\left(S_j\right). \quad (3)$$

As shown in Section 4, coherence structure features prove to be relatively less effective in discriminating cognitive load but more effective in discriminating cognitive performance at the same load. Fig. 5 provides an illustration of this finding by plotting the coherence matrices obtained at the gamma band from four different trials of the same subject, using the sentence interval. These are the same trials for which the power features are shown in Fig. 4.

In Fig. 5 (top), matrices from a high-load trial (5 digits) and a low-load trial (3 digits) are shown. In both of these trials the digits were recalled correctly. In Fig. 5 (bottom), matrices from two medium load trials (4 digits) are shown from a digit correct and a digit incorrect recall trial. For both matrix comparisons, the matrix eigenspectra are plotted on the right. In the load comparison, the differences in eigenvalues from the high and low load conditions are small. In our experimental results higher loads are on average associated with large values in the low rank eigenvalues (Section 4.1). However, these features are relatively less effective for discriminating load compared to power and covariance structure features (Section 4.2).

In the accuracy comparison, in Fig. 5 (bottom), there are larger values in the low-rank eigenvalues in the correct recall trial. These differences are consistent with the average experimental results (Section 4.1)

which presage a relatively strong ability for coherence structure features to discriminate digit recall accuracy (see Section 4.2).

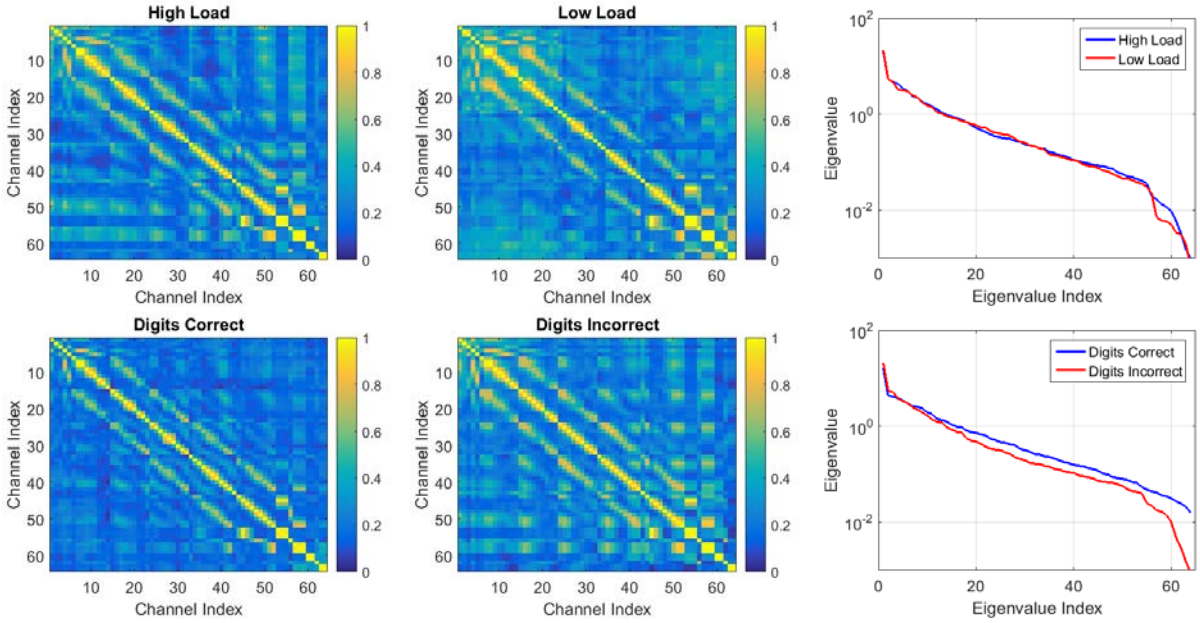


Fig. 5. Gamma band coherence features obtained from sentence interval for same subject. Top: matrices from high load (left) and low load (middle), and resulting eigenspectra (right). Bottom: matrices from medium load for correct recall (left) and incorrect recall (middle), and resulting eigenspectra (right).

Each row (or column) of the matrices depicted in Fig. 5 represents the coherences between a single EEG channel and all 64 EEG channels. The pattern of coherences for one of these rows can be viewed using a head map image. Fig. 6 displays the head map images from row 50 (electrode P4), which is the channel with the largest average root mean square difference between the high load and low matrices and between the correct recall and incorrect recall matrices. For both low load and correct recall, there is a tighter pattern of coherences centered at P4 and weaker coherences at left frontal areas, compared to the patterns obtained with high load and incorrect recall. The tighter spatial distribution of coherences for low versus high load is not consistent with the general pattern shown in Section 4.1, but it is consistent with the general pattern for correct versus incorrect recall, where low rank eigenvalues are larger in the correct recall cases (see Section 4.1). This is because tighter spatial distributions mean, on average, greater independence between channels and thus larger values in low-rank eigenvalues (Schindler et al, 2007).

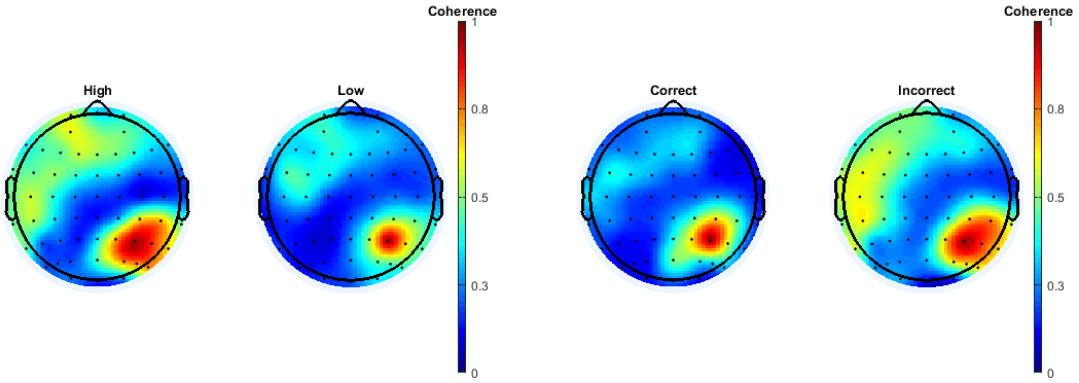


Fig. 6. Head map images of single-trial coherence patterns in the gamma band from the P4 electrode. Left: High load and low load. Right: Correct and incorrect digit recall.

Covariance structure features:

Covariance is an alternative connectivity measure, which is a combined measure of both the correlation between two signals and their power. While we have investigated both correlation and covariance structure features, only covariance structure results are reported here because significantly better results were found with these features for both load and accuracy discrimination. Unlike coherence, covariance can be either positive or negative, retaining sensitivity to the phase relations between signals. Before computing the covariance features, bandpass filtering is applied at each frequency band, *j,* to obtain filtered time-domain signals. Independent z-scoring is then applied for each subject at each frequency band to equalize within-channel variance across all trials. Next, the covariance between each channel pair (*x,y*) at frequency band *j* is computed in the time domain,

$$c_{x,y}(j) = \frac{1}{n}\sum_{i=1}^{n}\left(x_j[i] - \bar{x}_j\right)\left(y_j[i] - \bar{y}_j\right). \tag{4}$$

where *i* is the discrete time index and *n* is the number of data points in the trial.

As with the coherence structure features, an *M* × *M* matrix is constructed in each frequency band (*M = 64*), with each matrix element containing the pairwise covariance between two channels,

$$C_j = \begin{bmatrix} c_{1,1}(j) & \cdots & c_{1,M}(j) \\ \vdots & \ddots & \vdots \\ c_{M,1}(j) & \cdots & c_{M,M}(j) \end{bmatrix}. \tag{5}$$

A covariance structure feature vector is then obtained using the logarithm of the matrix eigenspectra,

$$\log\left(\left\{\lambda_j^C(1),\ldots,\lambda_j^C(M)\right\}\right) = \log\left(\mathrm{eig}\left(C_j\right)\right). \tag{6}$$

Unlike the coherence-based eigenspectra in Equation (3), a log measure of the covariance-based eigenspectra was found to produce better discrimination of cognitive load and performance. The use of a log measure is consistent with previous EEG work involving covariance-based eigenspectra (Williamson et al, 2011, 2012; Ma and Bliss, 2014).

As will be shown in Section 4, covariance structure features are relatively effective in both load and performance discrimination. Fig. 7 illustrates this finding with plots of the covariance matrices obtained at the beta band from the sentence interval of the same four trials that are illustrated in Figs. 3-5. In the high load case (Fig. 7, top) there seems to be greater separation between anterior channels (# 1-20) and posterior channels (# 44-64) than in the low load case. In Fig. 7 (bottom), there is also greater separation between anterior and posterior channels, as well as stronger positive and negative covariances, in the digit correct case than in the digit incorrect case. The matrix differences are quantified in the eigenspectra, where there are larger values in the low rank eigenvalues in both the high load trial and the correct digit recall trial. In Section 4.1 these trends are revealed to be broadly consistent with the average experimental results, in which there are larger values in the low rank eigenvalues for high load and for correct digit recall.

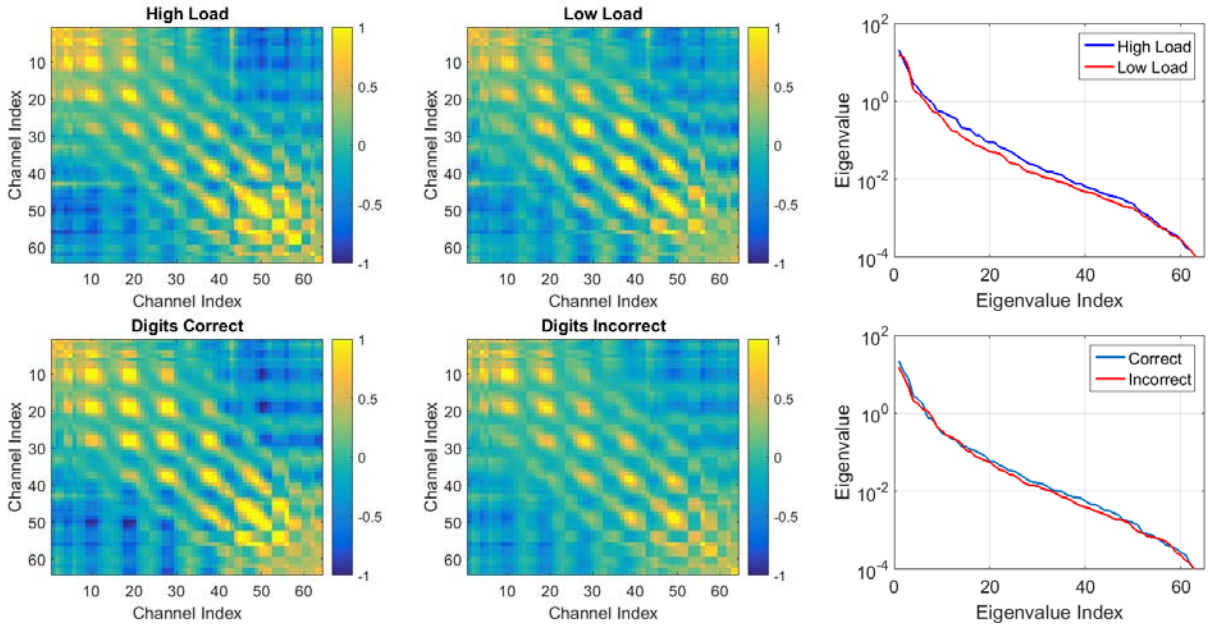


Fig. 7. Beta band covariance features obtained from sentence interval for same subject. Top: matrices from high load (left), low load (middle), and resulting eigenspectra (right). Bottom: matrices from medium load for correct recall (left), incorrect recall (middle), and resulting eigenspectra (right).

Each row (or column) of the matrices depicted in Fig. 7 encodes the covariances between a single EEG channel and all 64 EEG channels. As with the coherence head maps shown in Fig. 6, the covariance patterns from a single row of the matrix can be viewed using a head map image. Fig. 8 displays covariance head map images from the same row as in Fig. 6 (row 50, source electrode P4). In both the high load and correct recall trials, greater positive covariance is seen nearby to P4, as well as greater

negative covariance in the left frontal regions, reflecting stronger functional separation between these regions in the high load and correct recall cases.

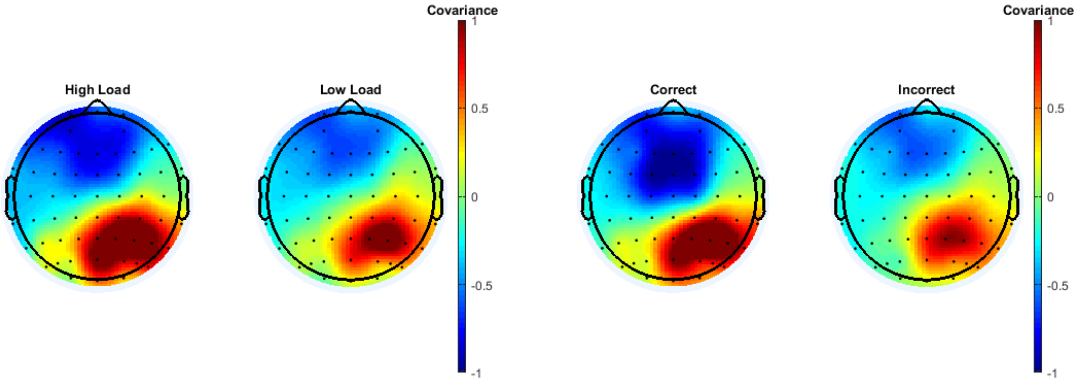


Fig. 8. Head map images of single-trial covariance patterns in the beta band from the P4 electrode. Left: High load and low load. Right: Correct and incorrect digit recall. In contrast to the coherence-based head maps, these head maps show greater similarity between the high load trial (left) and the correct recall trial (right), and greater similarity between the low load trial (left) and the incorrect recall trial (right).

3.2 Machine learning and discrimination

Cross-validation:

The goal of machine learning is to construct statistical models from EEG features for two tasks: 1) detecting the differences in cognitive load levels, defined by the number of digits in working memory, and 2) detecting differences in cognitive performance, defined by recall accuracy for a fixed load of four digits. In each trial, EEG features are estimated from the sentence and retention intervals. For each subject there are 108 trials of each load condition (low, medium, high), with the medium load condition being available in only 12 of the 16 subjects, as described in Section 2.1. To test the ability to generalize to novel data, leave-one-subject-out cross validation is used, so that detecting load or performance differences for a test subject is done using a statistical model trained only on data from the other subjects. For the cognitive load problem, high load trials are assigned to Class 1 and low load trials to Class 2. For the cognitive performance problem, correct digit recall trials are assigned to Class 1 and incorrect recall trials to Class 2.

Feature normalization:

Discriminating load levels requires sensitivity to load-related feature differences *within* the same subject. Therefore, a key processing step is individualized feature normalization, which mitigates inter-subject feature variability. Normalization involves z-scoring each subject's features across all trials (both high and low loads). This processing step implies that the ability to discriminate load conditions in a subject requires baseline EEG features from each subject. Discriminating digit recall accuracy, on the other hand, requires sensitivity to feature differences *across* subjects. Therefore, within-subject feature

normalization is not used because it could partially wash out informative between-subject EEG differences.

Dimensionality reduction:

Prior to construction of statistical models that map the high dimensional power or eigenspectra features to cognitive load and performance outcomes, dimensionality reduction is done to obtain lower dimensional features that are more manageable. Principal component analysis (PCA) is used, which maximally preserves feature variability for any chosen dimensionality of the reduced feature space. A key question is how many principal component dimensions to use for each feature type. In order to avoid overfitting of the statistical models to the 15 feature types under consideration (15 feature types = three feature sets and five frequency bands per feature set; see Table 1), we need to use a consistent rule for selecting the number of PCA dimensions per feature type.

Two simple PCA selection rules were considered: 1) selecting the minimum number of principal components that explains a fixed percentage of variance in each feature type (Helfer et al, 2016), and 2) selecting the same number of principal components for each feature type. Both approaches were explored and option 2 was selected based on finding that a fixed number of principal components (four) produced the best overall performance across feature sets, with improved results using this method on the power and covariance-based features, and about the same results on the coherence-based features. Four principal components explain between 60% and 80% of the variance for power features and more than 80% of the variance for the coherence and covariance features. These percentages are listed in Table 1.

Table 1. Each feature is reduced from 64 to four dimensions using PCA. For each feature type, the percentage of total variance explained by the first four principal components is listed. The percentages are computed as: 100 multiplied by the summed variance of PCA components, and divided by the total variance.

Frequency Band	Percentage of total variance in first four PCA components		
	Power features	Coherence features	Covariance features
Delta	64 %	89 %	88 %
Theta	69 %	87 %	83 %
Alpha	72 %	85 %	88 %
Beta	69 %	80 %	93 %
Gamma	75 %	85 %	94 %

Statistical model:

Discrimination of cognitive load and recall accuracy is done, within leave-subject-out cross-validation, using a Gaussian classifier (GC). For load discrimination there is one Gaussian for Class 1 (high load) and one for Class 2 (low load); for accuracy discrimination there is one Gaussian for Class 1 (accurate recall) and one for Class 2 (inaccurate recall). In load discrimination, equivalent full-data covariance matrices

are used for both Gaussians for improved regularization. In each trial, a GC produces a two-class log-likelihood test statistic. The area under the receiver operating characteristic (ROC) curve, the AUC, is used to characterize the discrimination performance across all subjects over all cross-validation folds. The AUC indicates the likelihood that a randomly selected trial from Class 1 will have higher likelihood than one from Class 2.

4 Results

4.1 Feature distributions

In Section 3.1 the feature differences from single trials are illustrated for the three feature sets, based on a comparison of individual trials with differences in load and differences in accuracy. Next is illustrated the *average* feature differences, across trials and subjects, due to load and accuracy differences. For the load comparison the average values are plotted from trials with high loads, medium loads, and low loads from the 12 subjects with trials at all three loads. Specifically, to show how big the feature differences are for different loads relative to within-load variability, the expected values of normalized (z-scored) features are plotted for high load (blue), medium load (green) and low load (red). Because of normalization, the vertical distance between plotted points for any two load classes *i* and *j* indicates the Cohen's *d* effect size, $(\mu_i-\mu_j)/\sigma$. Similarly, to show how features differ based on cognitive performance, the average normalized feature values are plotted for correct digit recall (blue) and for incorrect digit recall (red). These plots are obtained from the 4-digit trials of 14 different subjects.

Fig. 9 shows these feature averages from the power features in the sentence interval, with load results in the top row and accuracy results in the bottom row. The load results are obtained in the beta band and the accuracy results in the theta band, which are the most effective bands for discriminating with these features (see Section 4.2). On the left the average results are plotted as a function of channel index. On the right, the results are displayed in two dimensional head maps. There is a monotonic effect of load on channel power, with higher load being associated with less power in all channels, and with the effect most pronounced in left frontal and midline regions. High accuracy is also associated with less power in all channels, but this effect is more evenly spread out across all brain regions. The effects of load and accuracy on the power features in other frequency bands and in the other data interval (retention), which are not shown, are qualitatively similar to the feature patterns shown in Fig. 9, with the following exception: in the two highest frequency bands (beta and gamma), there is little difference in the average power features obtained from correct and incorrect recall trials.

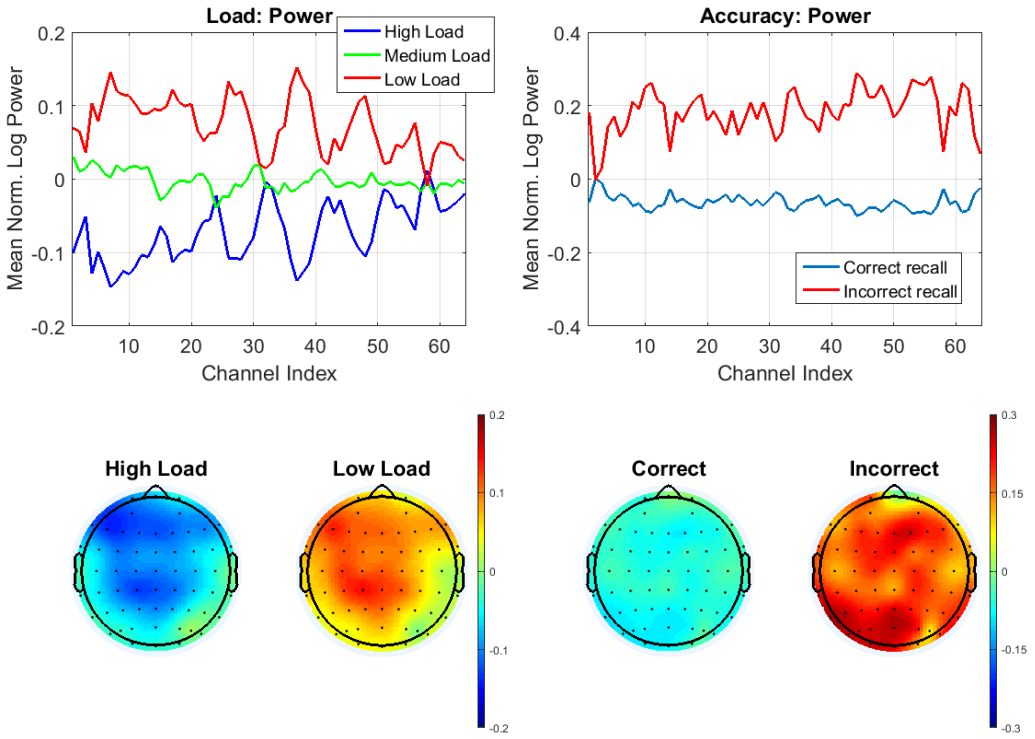


Fig. 9. Top row: Averages are shown for the power features in the beta band due to three load conditions: high (blue), medium (green), and low (red), computed across 12 subjects. Results are shown as a function of channel index (left) and in head maps for the high and low load cases (right). Bottom row: Averages are shown for the power features in the theta band due to correct digit recall (blue) and incorrect recall (red), computed across 14 subjects.

Fig. 10 shows the average results for coherence and covariance structure features on load and accuracy. For coherence structure, high load and high accuracy are both associated with larger values in the lower rank eigenvalues (Fig. 10, left). Due to the normalization in Equation (1), all of the diagonal elements of coherence matrices have unit value, and thus the sum of coherence matrix eigenvalues is constant. Therefore, if the low rank eigenvalues are larger on average in the high load and high accuracy conditions, then high rank eigenvalues must necessarily be smaller for these conditions (Fig. 10, left). Observe that the high load versus low load patterns in Figs 9 and 10 are nearly symmetrical, whereas the correct recall and incorrect recall patterns are not. This is due to the fact that there are the same number of trials in the high and low load classes (12 × 108 = 1,296 trials per class) but far fewer incorrect recall trials than correct recall trials (388 incorrect trials vs 1,124 correct trials).

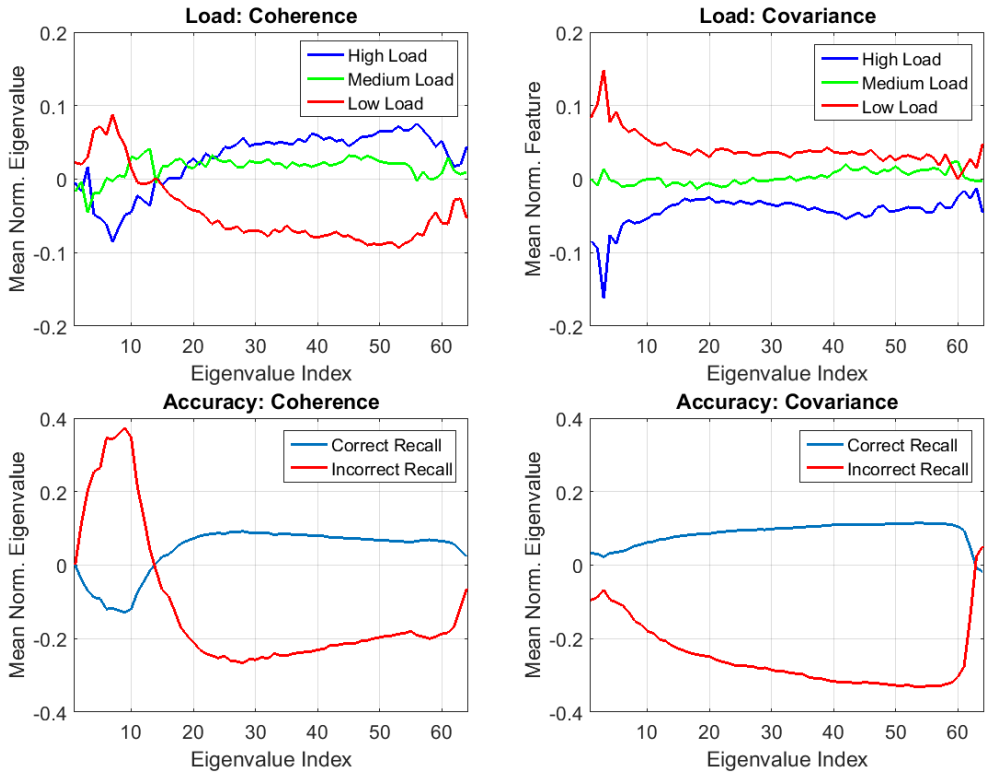


Fig. 10. Top row: Averages for the coherence and covariance structure features in the beta band due to three load conditions: high (blue), medium (green), and low (red), computed across 12 subjects. Bottom row: Averages for the coherence and covariance structure features in the gamma band due to correct digit recall (blue) and incorrect recall (red), computed across 14 subjects.

Because the diagonal elements of covariance matrices vary from trial to trial based on changes in channel signal power, the covariance structure eigenvalues contain information about power as well as connectivity. Therefore, while the relative pattern, across eigenvalue rank, of covariance eigenvalues *within* the same load or accuracy condition is similar to what was found for coherence structure features, the relative pattern at the same rank *across* different load levels or different accuracy levels is quite different. Specifically, high rank eigenvalues are relatively smaller in both the high load and high accuracy conditions, compared to low load and low accuracy. On average the entire high load eigenspectrum is smaller than the low load eigenspectrum, but the entire high accuracy eigenspectrum is larger than the low accuracy eigenspectrum (Fig. 10, right). Therefore, because covariance structure features respond differently to changes in cognitive load versus changes in cognitive accuracy, due to different effects of average signal power, they potentially enable specificity in discriminating the two conditions.

The patterns in Fig. 10 are qualitatively consistent with the patterns for these feature types that were found in the other frequency bands and the other data interval (retention), with the following exception: in the two lowest frequency bands the high rank covariance structure eigenvalues associated with correct digit recall have smaller values than the eigenvalues associated with incorrect recall. Therefore, in these two frequency bands, the correct-recall and incorrect-recall lines cross over, just as

they do for coherence structure in all frequency bands. This result is due to the fact that average power levels tend to be lower in the lower frequency bands for correct digit recall trials.

4.2 Cognitive workload discrimination

Discrimination of high load from low load is done on a single-trial basis. A single trial consists, on average, of a combined total of 5.3 seconds of EEG data while a subject is hearing a single sentence and then holding the sentence and a string of digits in working memory. Discrimination is done using statistical models trained in leave-one-subject-out cross-validation on data comprising the high load (*n* digits) and low load (*n-2* digits) trials. Table 2 summarizes the average ability of the models, across all 16 subjects, to discriminate high from low loads at each frequency band. Separate discriminations are done using the EEG data recorded during the sentence and retention intervals. The power and covariance structure features perform at similar levels, but the coherence structure features perform worse. This difference is probably due to the fact that coherence features do not retain any information about differences in average power levels between high and low loads. The power and covariance structure features perform best in the beta band, with AUC = 0.59 and 0.56 for power features and AUC = 0.58 and 0.56 for covariance structure features on the sentence and retention intervals, respectively.

Table 2. Area under ROC curves from single-trial discrimination (16 subjects) of high versus low loads from sentence and retention intervals for each frequency band feature type (*p<0.05, **p<0.01; one-sided Wilcoxon rank sum test).

Frequency Band	Power Features		Coherence Features		Covariance Features	
	Sent. AUC	Ret. AUC	Sent. AUC	Ret. AUC	Sent. AUC	Ret. AUC
Delta	0.56**	0.54**	0.52*	0.50	0.57**	0.54**
Theta	0.56**	0.54**	0.53**	0.53**	0.55**	0.54**
Alpha	0.57**	0.55**	0.55**	0.54**	0.57**	0.54**
Beta	0.59**	0.56**	0.54**	0.51	0.58**	0.56**
Gamma	0.56**	0.54**	0.54**	0.52*	0.53**	0.52*

Table 3 summarizes discrimination of high versus low loads for each feature set after fusing across frequency bands, which is done by concatenating the PCA feature vectors. Fusion across the sentence and retention intervals is also done, as is fusion across the three feature sets (bottom row). These latter fusions are done by multiplying classifier likelihoods. The best result is obtained by fusing all the feature sets across both data intervals, resulting in AUC = 0.61. Statistical significance in comparisons between different results was evaluated using one-sided t-tests based on AUC standard errors (Hanley and McNeil, 1982). The coherence structure results are significantly worse than the results from the other feature sets compared on the same data intervals ($p < 0.01$). On the other hand, none of the differences among the power, covariance structure, and combined-feature results, are significant when compared on the same data interval ($p > 0.05$).

Table 3. Area under ROC curves on single-trial discrimination (16 subjects) of low versus high loads from sentence and retention intervals, and combined results across both intervals (*p<0.05, **p<0.01; one-

sided Wilcoxon rank sum test). Each result is obtained using concatenated PCA features from all five frequency bands. Combination results are also obtained via likelihood fusion across data intervals and feature sets.

Feature Sets	AUC of Sentence	AUC of Retention	AUC of Sentence & Retention
Power	0.58**	0.56**	0.59**
Coherence	0.54**	0.53**	0.55**
Covariance	0.60**	0.55**	0.60**
Combined-feature	0.60**	0.57**	0.61**

To estimate how well these features would extend to estimation of cognitive load over longer duration tasks, the single-trial load estimates are combined, via likelihood fusion, across multiple randomly selected trials that have the same load. Fig. 11 depicts the average ROC curves obtained in this way from the three feature sets by fusing the sentence and retention results in each trial, and by combining 1, 5, 10, 20, and 30 randomly selected trials. For all three feature sets, performance monotonically improves with additional trials. Average results after 30 trials are AUC=0.90, 0.77 and 0.90, respectively, for the three feature sets.

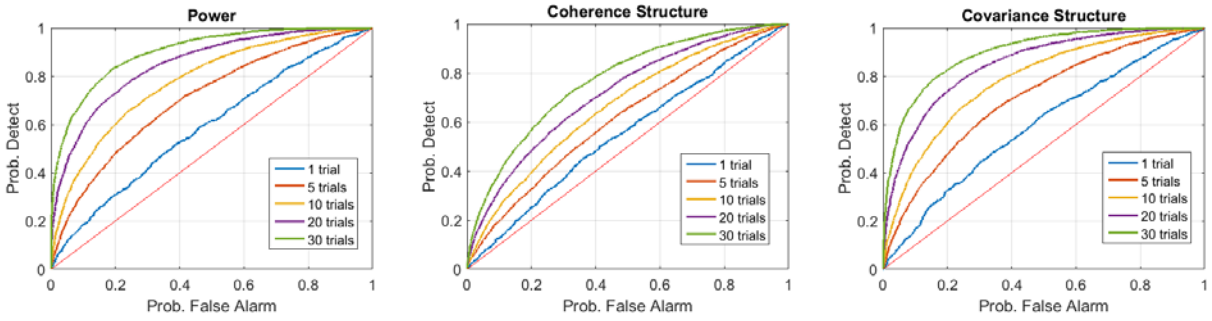


Fig. 11. Average ROC curves obtained by fusing the combined-segment results for each feature set across multiple randomly selected trials. After fusion of 30 trials, average AUC values of 0.90 (left), 0.77 (middle), and 0.90 (right) are obtained for power, coherence, and covariance features.

A further analysis was done for a subset of 12 subjects who also had medium load trials with *n-1* digits, in addition to the high load (*n* digits) and low load (*n-2* digits) trials. In this analysis, classifiers were trained only on high and low loads, and the ability to discriminate finer-grained 1-digit distinctions between high and medium loads and also between medium and low loads was investigated. Fig. 12 plots the results of this analysis, showing AUC values with standard error bars for 2-digit discrimination (high vs. low load) and for 1-digit discrimination (high vs. medium and medium vs. low load). These results demonstrate an ability to generalize to smaller load distinctions. The reduction in accuracy is consistent with the increase in difficulty, with accuracy reduced by about 50% when the digit load difference is cut in half.

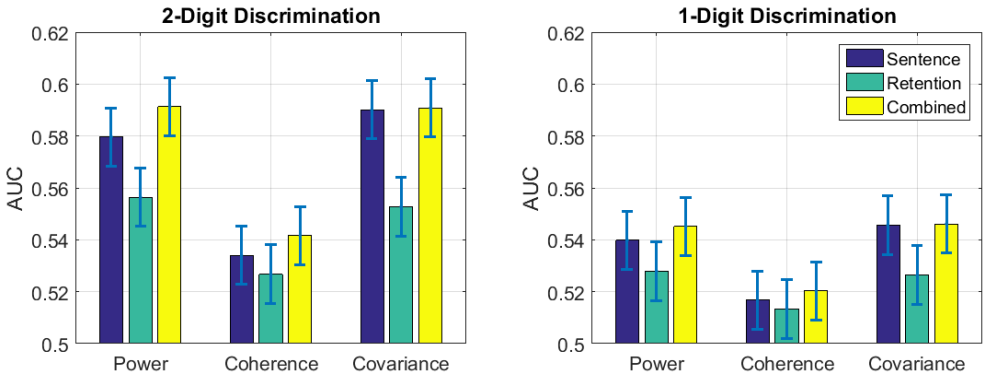


Fig. 12. Single trial cognitive load discrimination results for 12 subjects. Area under ROC curve (AUC) is shown for three feature sets given sentence, retention and combined intervals. Results are shown for discriminating load differences of two digits (left) and one digit (right).

The overall pattern of cognitive load discrimination suggests that EEG power is a strong discriminant of load, with greater power being associated with lower load, particularly in frontal and midline areas. This result is consistent with (Zarjam et al, 2015). Another finding is that connectivity structure is also a discriminant of load, with greater independence among channels associated with higher loads. This finding is based on using the eigenspectrum of connectivity matrices, where larger values in low rank eigenvalues are associated with higher loads. This finding is consistent with the hypothesis that higher loads are associated with more modular representations, which are spatially more tightly distributed, with activity being more highly correlated within the same brain region, and more highly anti-correlated between different regions.

One indication of the robustness of the load discrimination system is that the feature sets and classifiers were developed based solely on their ability to discriminate high loads from low loads. Medium load trials were treated as "out-of-sample" test data. The feature distributions for medium loads were found to be consistently intermediate to the distributions for high and low loads (Figs. 9 and 10). The system was also able to detect that medium loads are higher than low loads and that high loads are higher than medium loads, with an accuracy that is commensurate with the magnitude of the 1-digit load differences.

4.3 Cognitive performance detection

Cognitive performance is defined as accuracy of digit recall. Correct recall means the digits are recalled in the same order that they are heard. In order to avoid effects due to changes in load, discrimination of cognitive performance is evaluated only with trials that have the same load of 4 digits, which is the load level from which the largest number of trials are available. The data set consists of 14 subjects with 1512 trials. Of these, 1124 are correct digit recall trials and 388 are incorrect digit recall trials. Separate discriminations are done using the EEG data recorded during the sentence and retention intervals.

Table 4 summarizes the average ability across the 14 subjects to discriminate cognitive performance at each frequency band for the three feature sets. Separate discrimination is done based on EEG recorded during the sentence interval and the retention interval. The coherence and covariance structure

features perform at similar levels. In contrast with cognitive load discrimination, the power features perform poorly both on the sentence and retention interval. The coherence and covariance structure features perform similarly to each other on the sentence interval, and the covariance structure features perform slightly better on the retention interval.

Table 4. Area under ROC curves on single-trial discrimination (14 subjects) of correct and incorrect digit recall from sentence and retention intervals for each frequency band feature type (*p<0.05, **p<0.01; one-sided Wilcoxon rank sum test).

Frequency Band	Power AUC		Coherence AUC		Covariance AUC	
	Sent.	Ret.	Sent.	Ret.	Sent.	Ret.
Delta	0.45	0.53*	0.55**	0.57**	0.52	0.61**
Theta	0.51	0.49	0.59**	0.53	0.59**	0.51
Alpha	0.45	0.44	0.53*	0.52	0.56**	0.55**
Beta	0.45	0.47	0.60**	0.56**	0.59**	0.61**
Gamma	0.46	0.49	0.61**	0.59**	0.60**	0.62**

Table 5 summarizes incorrect/correct recall discrimination at the feature set level after fusing across frequency bands, which is done by multiplying classifier likelihoods. Fusion across the sentence and retention intervals is also done, as is fusion across the three feature sets (bottom row). The best overall result of AUC = 0.64 is obtained using the covariance features on the retention interval. Statistical significance in comparisons between different results was evaluated using one-sided t-tests based on AUC standard errors (Hanley and McNeil, 1982). The power results are significantly worse than the results from the other feature sets compared on the same data intervals ($p < 0.01$). On the other hand, none of the differences among the coherence structure, covariance structure, and combined-feature results are significant when compared on the same data interval ($p > 0.05$).

Overall, these results indicate that features characterizing connectivity structure are more effective than power features for discriminating cognitive performance. Covariance structure features may additionally benefit by the fact that they also represent the underlying EEG power levels. In a separate analysis (not shown here), correlation structure features performed significantly worse than covariance structure features on this task. Correlation structure features are identical to covariance structure features except that they are invariant to changes in channel power due to the fact that correlation coefficients are invariant to signal magnitude.

Table 5. Area under ROC curves on single-trial discrimination (14 subjects) of correct versus incorrect digit recall from sentence and retention intervals (*p<0.05, **p<0.01; one-sided Wilcoxon rank sum test). Each result is obtained using likelihood fusion across results from each feature band. Combination results across data intervals and feature sets are also obtained using likelihood fusion.

Feature Set	Sentence AUC	Retention AUC	AUC of Sentence & Retention
Power	0.46	0.49	0.49
Coherence	0.62**	0.60**	0.62**

Covariance	0.61**	0.64**	0.63**
Combined-feature	0.60**	0.61**	0.61**

An earlier analysis of cognitive performance on this data set is reported in (Helfer et al, 2016), in which power, coherence structure, and graph-based features were used to predict digit and sentence recall accuracy. In that study, EEG data was used from a single data interval combining both the sentence-hearing and the data retention intervals. It was found that coherence structure features outperformed the graph-based features in predicting digit recall accuracy, and that the coherence structure features were effective in all evaluated frequency bands (theta, alpha, beta, gamma), whereas the graph-based features were only effective in the beta and gamma band.

5 Discussion

5.1 Results summary and interpretation

This paper presents two approaches for characterizing scalp EEG measurements, using: 1) spatial patterns of neural activations, and 2) structural analysis of neural connectivity. The spatial pattern approach is based on signal power at the EEG channels. The connectivity structure approach is based on the eigenspectra of connectivity matrices constructed using two different measures of neural connectivity: pairwise channel coherences and pairwise channel covariances. All of the feature approaches are applied at five standard frequency bands, which are known to play important roles in cognitive processing and working memory (Roux et al, 2014; Voytek and Knight, 2015).

These EEG feature techniques are applied to an audio working memory task to discriminate levels of cognitive load and cognitive performance. Cognitive load is defined by the number of digits held in working memory, and cognitive performance is defined by digit recall accuracy. EEG power features are found to be relatively effective for discriminating load but ineffective for discriminating accuracy. Coherence structure features show the opposite pattern, being less effective than the power features at discriminating load but more effective at discriminating performance. Covariance structure features achieve the best of both worlds, being as effective as the power features in discriminating load and as effective as the coherence structure features in discriminating performance.

Fig. 3 illustrates the differing properties of channel power and covariance structure features. The channel power features represent the size of each dimension of the multivariate EEG distribution projected onto the original data axes (i.e., EEG channels), whereas covariance structure features jointly represent both the size and the shape of the EEG distribution, defined on its (rank-ordered) principal axes. Coherence structure features are similar to covariance structure, but with a couple key differences. First, coherence contains no measure of absolute channel power due to the fact that power is factored out in Eq. (1). Second, coherence values represent an average correlation magnitude across all relative time delays of two channels. Correlation (covariance) coefficients, on the other hand,

represent the correlation (covariance) between two signals at a specified relative time lag. In this paper, relative time lags of zero are used.

To directly test the importance of signal power information in the discriminative results obtained by the covariance structure features, we have also evaluated correlation structure features, which are identical to the covariance structure features except that channel power is normalized out in computing correlation coefficients instead of covariance coefficients. We found that the correlation structure features perform similarly to coherence structure features in discriminating cognitive load, and perform significantly worse than both coherence structure and covariance structure features in discriminating accuracy. These findings, which are not shown here due to space limitations, support the hypothesis that a joint encoding of both the size and the shape of multivariate EEG distributions enables discrimination of both cognitive load and accuracy.

Before drawing further conclusions from our experiment and data analysis, we first note with reference to Fig. 3 that if multichannel EEG time domain data within a particular frequency band are conceptualized as a multivariate distribution, in which each EEG channel is a data axis, then: 1) channel power features represent the signal variance along the data axes, and 2) covariance structure eigenvalues represent the (rank ordered) variances along the principle axes of the multichannel distribution. With these points in mind, Table 6 summarizes conclusions from our experiment and data analysis. These conclusions hold, except where noted, for all frequency bands and both data intervals (sentence and retention).

Higher cognitive load and higher cognitive performance are both associated with less overall power. For cognitive load, this effect is particularly strong in the left frontal and midline areas. Therefore, these outcome variables are associated with less scatter in EEG signal distributions, with the effect for cognitive load being seen most strongly in the axes that correspond to left frontal and midline channels. Higher load and performance are also associated with, on average, a lower correlation between channels. Combining these observations, it is apparent that higher load and performance correspond to more tightly concentrated activity patterns, as proposed by (Zarjam et al, 2015) to explain their cognitive load results. This indicates an increase in modularity of neural representations and a concomitant reduction in cross-talk, which enables more reliable storage of a larger number of items in working memory. For cognitive load (but not for cognitive performance), these effects seem to be particularly prominent in left-frontal and mid-line areas.

The above results are therefore consistent with the hypothesis that higher cognitive load and cognitive performance are associated with greater neural suppression (Zanto and Gazzaley, 2009; Zarjam et al, 2015), with the suppression being particularly strong in left frontal and midline areas during high cognitive load. Neural suppression, the attenuation of irrelevant neural representations, is often associated with low level perceptual processing and with selective attention, which is the ability to focus cognitive resources on relevant internal representations. There is converging evidence that selective attention and working memory are highly overlapping constructs (Gazzaley and Nobre, 2012; Wolmensdorf 2015), and that neural suppression, in the form of divisive normalizing inhibition (Williamson, 2001; Grossberg and Williamson, 2001; Reynolds and Heeger 2009; Carandini and Heeger,

2012; Melnick et al, 2012) is a canonical neural mechanism for filtering out irrelevant information, not only in perceptual processing but also in higher level processes such as working memory.

Table 6. Summary of major EEG feature discriminants for the two outcome variables, cognitive load and cognitive performance.

Outcome Variables	EEG Feature Discriminants
Higher Cognitive Load	Less overall power
	Less power in particular channels (especially left frontal, midline)
	Less correlated activity (greater modularity)
Higher Cognitive Performance	Less overall power in first three frequency bands
	Less correlated activity (greater modularity)

5.2 Comparison to previous results

Cognitive Load:

(Zarjam et al, 2015) obtained similar findings to our own on the effects of cognitive load on neural activity patterns. They found, over seven load levels defined by the number of mental operations required in arithmetic tasks, a monotonic reduction in power in frontal areas over seven load levels. Strong effects were found in the delta band, and weaker effects at higher frequencies. In our experiment and analysis, strong effects were found in all frequency bands, with the strongest effect found in beta (Table 2). It is important to note the differences in methodology between the two studies. In (Zarjam et al, 2015), EEG channel selection was done based on correlations between features and outcome variables over the entire data set of 12 subjects. In our analysis, on the other hand, unsupervised dimensionality reduction (PCA) was done in a cross-validation training procedure. Therefore, the high discrimination performance of (Zarjam et al, 2015) could be biased due to the fact EEG channel selection was done using testing data. A second difference in the experimental methodologies is that the load levels in (Zarjam et al, 2015) were presented in ascending order of load, resulting in a possible confound between fatigue/learning effects and load effects.

Finally, the performance results in (Zarjam et al, 2015) were obtained by fusing multiple 5-second data intervals, and therefore are comparable only to our multi-trial fused results shown in Fig. 11, as each of our trials used 5.3 seconds of EEG data on average. Despite these differences, our results replicate the main effects in (Zarjam et al, 2015) study in the delta band, finding monotonically reduced power with greater load, with the effect being stronger in frontal channels. A question for future research is to investigate the relative benefits in detecting load using supervised channel selection, as in the (Zarjam et al, 2015) study, versus using unsupervised feature combinations, as in our study.

Other studies have found that frontal brain areas are strongly implicated in working memory and cognitive load, with critical roles being played by different frequency bands. As summarized by (Roux et al, 2014), alpha has been implicated in frontally controlled top-down neural suppression, theta in the organization of sequential ordering, and gamma in general maintenance of working memory. As summarized by (Womelsdorf and Everling, 2015), there is evidence for attentional networks comprising prefrontal, cingulate, and striatal circuits operating at theta and beta frequencies, that control stimulus selection and employ inhibitory gain control in the service of goal directed behavior. Long-range beta synchronization has been implicated in carrying object and location specific information, with beta coherence being evident in frontal, prefrontal, and parietal cortices. These findings are of interest given that the power and covariance structure features are most discriminative for load in the beta band. Overall, while there is support for specific functionality being most strongly associated with specific frequency bands and cortical regions, the fact remains that demanding working memory tasks produce changes in EEG across many frequency bands and brain regions. In fact, in our analysis there was a remarkable consistency in feature changes across frequency bands in all feature sets, during both the sentence listening interval and the retention interval.

There are some previous findings that show increased activity at particular frequency bands and brain regions due to increases in cognitive load. These findings are in the opposite direction to the results in our study and in (Zarjam et al, 2015). For example, (Jensen and Tesche, 2002) report increases in theta activity in frontal areas as a function of number of items in short term memory, and (Onton et al, 2005) report increases in frontal and left temporal theta during trials where memorization is required. Both of these tasks use the Sternberg paradigm, involving recognition of items in working memory as opposed to recall. Also, these experiments appear to be less fatiguing and mentally taxing than our experimental protocol. Further research is needed to understand the relationship between experimental conditions and increases versus decreases in EEG band power.

In addition to finding effects of load on the spatial patterns of EEG band power, load effects were also found related to changes in connectivity structure at multiple frequency bands, with connectivity measured using coherence and covariance. Coherence structure features, which essentially encode the average shape of multichannel EEG distributions across all relative time lags, are less discriminative at all frequency bands (Table 3). Covariance structure features, which encode both the shape and size of multichannel EEG distributions at a relative time lag of zero, are more discriminative of load, achieving accuracy levels similar to the channel power features (Table 3). This is an interesting result, as the covariance structure features are completely untethered from the identities of the EEG channels. Therefore, the features may provide complementary information to the EEG channel power features, and could provide indications of cognitive load that are invariant to cognitive tasks that stress different functional networks and brain regions.

Cognitive Performance:

There is much previous work using graph analytic approaches to relate neural connectivity properties to outcome measures such as cognitive aptitude, working memory performance, and neurological state. Results have been somewhat inconsistent, with higher cognitive aptitude being correlated with reduced

"small-world" graph properties, specifically lower clustering coefficients and higher path lengths (Micheloyannis et al, 2006), and in other studies higher IQ being correlated with shorter path lengths (Li et al, 2009; van den Heuvel et al, 2009), and longer path lengths in Alzheimer's disease (Stam et al, 2007). One possible source of inconsistency in these results is that path length is an unstable measure with respect to changes in graph connectedness. This problem can be avoided using measures of network efficiency based on reciprocal path lengths (Achard and Bullmore, 2007). From this study lower network efficiency, which essentially translates to longer average path lengths, is positively correlated with subject age and with pharmacological blockade of dopamine neurotransmission.

The above results compare long term performance outcomes with measures of average connectivity. In our study the focus is on trying to relate dynamically changing connectivity patterns with the changes in stimulus conditions or behavioral outcomes on a trial by trial basis. In previous work, increases in task demands on short time scales have been found to produce higher clustering, higher modularity, and a smaller proportion of long-distance connections in graph measures (Bullmore and Sporns, 2012). In previous results obtained by our group on our cognitive performance data set, several graph-based metrics derived from coherence-based connectivity matrices were used to predict recall performance (Helfer et al, 2016). Most graph metrics were found to perform poorly in discriminating cognitive performance, although two different measures performed moderately well in the two highest frequency bands.

The connectivity structure approach used here and in (Helfer et al, 2016), based on a variety of connectivity measures, has been used for EEG analysis in a variety of other application domains. Connectivity structure derived from EEG correlation matrices has been used in characterizing seizure dynamics from EEG (Schindler et al, 2007), in detecting seizures (Pan et al, 2012), and in predicting seizures from EEG (Williamson et al, 2011, 2012; Ma and Bliss, 2014). In the seizure detection and prediction approaches, the correlation and covariance matrices were expanded in dimensionality using time delay embedding at multiple delay scales.

Because the connectivity structure and graph-analytic approaches are both based on identical connectivity matrices, we strongly recommend in future work that the effectiveness of the two approaches should be compared. Connectivity structure features could serve as a useful baseline to determine what, if any, value is added by extracting higher order structural information using graph analytic approaches, above and beyond the second order statistical information embodied in matrix eigenspectra.

Audio and video sensor modalities:

Finally, it is important to note that audio- and video-based features from our auditory working memory data set have also been analyzed, which reflect qualities of speech and facial expression during the spoken sentences (Quatieri et al, 2015; Quatieri et al, 2016). Coherence, covariance, and correlation structure feature approaches based on these modalities have been analyzed, and the same phenomena have been observed as in EEG, to wit: higher cognitive load is associated with larger values in the middle- or low-rank eigenvalues. Thus, using the same experimental manipulation, similar changes in

dynamics have been found in three different signal modalities: EEG, audio, and video. This suggests that connectivity structure features provide a useful basis not only for probing changes in the dynamical complexity of brain networks, but also in their ramifications in complex motor behaviors such as speech.

6 References

Achard, S., & Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. PLoS Comput Biol, *3*(2), e17.

Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. Nature Reviews Neuroscience, 13(5), 336-349.

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. Nature Reviews Neuroscience, 13(1), 51-62.

Compumedicsneuroscan http://compumedicsneuroscan.com/scan-acquire-configuration-files

Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004; 134: 9–21.
EEGLAB http://sccn.ucsd.edu/eeglab

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. Trends in cognitive sciences, 16(2), 129-135.

Grossberg, S., & Williamson, J. R. (2001). A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual grouping and learning. Cerebral cortex, 11(1), 37-58.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143(1), 29-36.

Harnsberger, J. D., Wright, R., & Pisoni, D. B. (2008). A new method for eliciting three speaking styles in the laboratory. Speech communication, 50(4), 323-336.

Helfer, B. S., Williamson, J. R., Miller, B. A, Perricone, J., Quatieri, T. F. (2016). Assessing functional neural connectivity as an indicator of cognitive performance. Neural Information Processing Systems 2016 (in press).

Jensen, O., & Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. European journal of Neuroscience, 15(8), 1395-1399.

Le, P. N., Ambikairajah, E., Choi, E. H., & Epps, J. (2009, December). A non-uniform subband approach to speech-based cognitive load classification. In Information, Communications and Signal Processing, 2009. ICICS 2009. 7th International Conference *on* (pp. 1-5). IEEE.

Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. The Journal of the Acoustical society of America, 49(2B), 467-477.

Li, Y., Liu, Y., Li, J., Qin, W., Li, K., Yu, C., & Jiang, T. (2009). Brain anatomical network and intelligence. PLoS Comput Biol, 5(5), e1000395.

Lively, S. E., Pisoni, D. B., Van Summers, W., & Bernacki, R. H. (1993). Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. The Journal of the Acoustical Society of America, 93(5), 2962-2973.

Melnick, M. D., Harrison, B. R., Park, S., Bennetto, L., & Tadin, D. (2013). A strong interactive link between sensory discriminations and intelligence. Current Biology, 23(11), 1013-1017.

Micheloyannis, S., Pachou, E., Stam, C. J., Vourkas, M., Erimaki, S., & Tsirka, V. (2006). Using graph theoretical analysis of multi channel EEG to evaluate the neural efficiency hypothesis. Neuroscience letters, 402(3), 273-277.

Onton, J., Delorme, A., & Makeig, S. (2005). Frontal midline EEG dynamics during working memory. Neuroimage, 27(2), 341-356.

Pan, Y., Guan, C., Ang, K. K., Phua, K. S., Yang, H., Huang, D., & Lim, S. H. (2012, June). Seizure detection based on spatiotemporal correlation and frequency regularity of scalp EEG. In Neural Networks (IJCNN), The 2012 International Joint Conference on (pp. 1-7). IEEE.

Park, H., Felty, R., Lormore, K., & Pisoni, D. B. (2010). PRESTO: Perceptually robust English sentence test: Open-set—Design, philosophy, and preliminary findings. The Journal of the Acoustical Society of America, 127(3), 1958-1958.

Park, Y., Luo, L., Parhi, K. K., & Netoff, T. (2011). Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. Epilepsia, 52(10), 1761-1770.

Quatieri, T. F., Williamson, J. R., Smalt, C. J., Patel, T., Perricone, J., Mehta, D. D., ... & Palmer, J. (2015). Vocal biomarkers to discriminate cognitive load in a working memory task. In Sixteenth Annual Conference of the International Speech Communication Association.

Quatieri, T. F., Williamson, J. R., Smalt, C. J., Perricone, J., Patel, T., ... & Moran, J. (2016). Multi-modal biomarkers to discriminate cognitive state. In press.

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. Neuron, 61(2), 168-185.

Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: alpha–gamma versus theta–gamma codes for distinct WM information?. Trends in cognitive sciences, 18(1), 16-25.

Schindler, K., Leung, H., Elger, C. E., & Lehnertz, K. (2007). Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG. Brain, 130(1), 65-77.

Shoeb, A., Edwards, H., Connolly, J., Bourgeois, B., Treves, S. T., & Guttag, J. (2004). Patient-specific seizure onset detection. Epilepsy & Behavior, 5(4), 483-498.

Stam, C. J., Jones, B. F., Nolte, G., Breakspear, M., & Scheltens, P. (2007). Small-world networks and functional connectivity in Alzheimer's disease. Cerebral cortex, 17(1), 92-99.

van den Heuvel, M. P., Stam, C. J., Kahn, R. S., & Pol, H. E. H. (2009). Efficiency of functional brain networks and intellectual performance. The Journal of Neuroscience, 29(23), 7619-7624.

Voytek, B., & Knight, R. T. (2015). Dynamic network communication as a unifying neural basis for cognition, development, aging, and disease. Biological psychiatry, 77(12), 1089-1097.

Williamson, J. R. (2001). Self-organization of topographic mixture networks using attentional feedback. Neural Computation, 13(3), 563-593.

Williamson, J. R., Bliss, D. W., & Browne, D. W. (2011, May). Epileptic seizure prediction using the spatiotemporal correlation structure of intracranial EEG. In Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on (pp. 665-668). IEEE.

Williamson, J. R., Bliss, D. W., Browne, D. W., & Narayanan, J. T. (2012). Seizure prediction using EEG spatiotemporal correlation structure. Epilepsy & Behavior, 25(2), 230-238.

Womelsdorf, T., & Everling, S. (2015). Long-range attention networks: Circuit motifs underlying endogenously controlled stimulus selection. Trends in neurosciences, 38(11), 682-700.

Yin, B., Chen, F., Ruiz, N., & Ambikairajah, E. (2008, March). Speech-based cognitive load monitoring system. In Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on (pp. 2041-2044). IEEE.

Zanto, T. P., & Gazzaley, A. (2009). Neural suppression of irrelevant information underlies optimal working memory performance. The Journal of Neuroscience, 29(10), 3059-3066.

Zarjam, P., Epps, J., & Lovell, N. H. (2015). Beyond Subjective Self-Rating: EEG Signal Classification of Cognitive Workload. Autonomous Mental Development, IEEE Transactions on, 7(4), 301-310.