COMP 8150 Term Project Report

Project Title:
Generating Sufficient and Less Noisy EEG Data Through Bootstrap Sampling and GAN to Train Deep CNN model for Cognitive Load Classification

Felix Havugimana
Submitted to: Dr. Andrew Neel

November 20, 2022

Abstract

Cognitive load refers to the amount of used working memory resources, which is limited in both capacity and duration. Predicting cognitive load from raw electroencephalogram (EEG) recordings remains a challenge because of the high degree of noise due to technical variations in the recording process and the multi-factorial nature of the mapping between the EEG data and cognitive load. I present deep convolutional neural network (CNN) models for predicting cognitive load and address issues related to noise, data representation, and a small number of samples. I reduce the noise by computing event-related potential (ERP) from raw data. Then I transform time-series signal into a spatial-spectral representation called Topomap, which maintains both spatial (electrode location) and spectral (frequency) information embedded in EEG recordings. I developed bootstrap sampling and deep Generative Adversarial Network (GAN) techniques to generate enough samples to train deep CNN models.

I use two different strategies to predict four different levels of cognitive load. First, I use power spectral densities of three individual frequency bands (Theta, Alpha, Beta) of the encoding period to create a spatial-spectral representation called topomap. Second, I combine all three bands to develop a composite topomap to compare the predictive power of individual and composite representation. I use CNN models to map spectral-spatial information and cognitive load for both individual and composite topomaps. I performed Empirical evaluations to determine the role of different frequency bands in predicting four cognitive load levels. The prediction accuracy of CNN models built using Theta, Alpha, Beta bands, and composite representation are 89%, 89%, 91%, 92%, respectively. The results suggest that the Beta band has the most predictive power and composite representation produces higher accuracy than the individual frequency bands.

1 Introduction

Human Working memory (WM) provides temporary storage and processing of information necessary for performing cognitive tasks such as reasoning, decision-making, and language comprehension [1]. The amount of used WM resources is known as a cognitive load (CL). The WM is limited in both capacity and duration. According to Hick's law, the reaction time to a stimulus increases logarithmically with the number of available choices [2]. Excessive WM usage can lead to cognitive overload, which has adverse effects on performing cognitive tasks [3].

Predicting cognitive load from raw electroencephalogram (EEG) recordings remains a challenge because of the high degree of noise due to technical variations in the recording process and the multi-factorial nature of the mapping between the EEG data and cognitive load. Most of the reported literature uses classical machine learning models such as Support Vector Machines (SVM), K-nearest neighbor (KNN), and random forest to predict cognitive load. Due to the noise in raw EEG data, these methods are limited to using a sub-optimal set of single-frequency power spectral density(PSD) features. Moreover, due to the insufficient experimental EEG data, only a few works have attempted to use deep learning for cognitive load classification.

In this project, I attempted to solve these challenges by proposing a data-driven approach to noise reduction and data augmentation for cognitive load classification. I adopted a commonly used noise reduction technique in EEG known as Event-Related Potential (ERP) to reduce noise from EEG data. ERP signal is obtained by averaging signals from multiple trials that belong to the same task[**luck2014introduction**].

However, due to the small number of participants, computing ERP samples would not generate enough samples for training deep learning models. To solve this problem, I obtained my ERP data through bootstrap sampling by averaging 20 single-trial signals selected randomly with replacement from the original 45trials per event. Nevertheless, sampling with replacement may generate redundant data since some trials can be repeatedly selected in multiple iterations. To reduce the effect of redundancy in the data, I trained a generative adversarial network (GAN) on a subset of bootstrap samples to generate more data. Therefore, half of the data will be generated using GAN. GAN helps me generate new data that are very close but not equal to the original dataset. GAN-generated samples add good variance to the data and eventually improve model generalization.

2 Related Work

Predicting cognitive load from EEG signals using machine learning involves modeling the relationship between features extracted from the signal and different mental load levels. Connectivity, entropy, and power spectral density features are the most reported EEG features used in predicting cognitive load [4, 5, 6, 7]. Among these features, power spectral density (PSD) features are the most admired due to the relationship between different frequency bands and WM activities at different cognitive load states, as reported in various literature [7, 8, 9, 10]. Classical machine learning models such as SVM, K-nearest neighbors (KNN), and Random forest have been extensively applied to PSD features for cognitive load prediction [11, 6, 12, 13, 5].

A few works have adopted deep learning for working memory analysis and cognitive load classification. Sahal [14] used Stacked Denoising Autoencoder (SDAE) and multilayer perceptron to classify cognitive load from EEG signal. They attempted to reduce the noise by using a denoising autoencoder with only four participants' data, which may not be enough to train deep learning models and achieve a high and reliable classification. Further, in [15], 1D and 2D CNN were applied to vectors and matrices of EEG spectral features and achieved accuracy about 93%, and 91%, respectively. However, they ignored the spatial information embedded in the EEG signal. An attempt to preserve the EEG signal's spatial-spectral structure was made in [12] where authors used Recurrent CNN to classify cognitive load. However, the authors did not provide a framework for EEG noise reduction. In addition, the data augmentation approach used in the paper by randomly adding noise to the images failed to improve classification performance.

This work built on what has been done to provide a more robust approach capable of reducing noise from raw EEG data, increasing the number of samples, and achieving a reliable cognitive load classification performance.

3 Design and Methodology

3.1 Data

In this analysis, I used, with permission, auditory WM data utilized in [16]. Experimental details and behavior results are given in [16]. In brief, a total of 15 participants (female: 8 and male: 7) were engaged in the auditory WM experiment. The continuous EEG signals were recorded with 64 sintered Ag/AgCl electrodes placed around the scalp at standard 10-10 locations (Neuroscan, Quik-cap) with a sampling rate of 500Hz.

During the experiment, participants listened to a series of English characters (SET), each 300ms long with a 700 ms delay in between. After playing SET, there was a 3 seconds delay during which participants were instructed to memorize the characters. After the delay, a test character (TEST) was played, and participants were asked to press one of the two buttons to indicate whether the TEST was among SET or not. Per request, we received pre-processed Audio task EEG data from 11 participants.

Among the three stages of the WM experiment (encoding, maintenance, recall), we focus on the encoding stage, which covers the SET characters' presentation time. In each of the 45 experimental trials, the number of SET characters is 2, 4, 6, and 8. The SET size reflects the level of cognitive load corresponding to the mental effort put into encoding SET in the memory. Throughout this report, I label SET sizes 2,4,6, and 8 as cognitive load levels 1, 2, 3, and 4, respectively. Therefore, this work's classification task is to predict these four levels of cognitive load using spatial-spectral features extracted from their corresponding EEG recordings.

3.2 Event Related Potential and Bootstrap Sampling

Excessive noise in EEG recordings and insufficient samples are the common roadblocks that hinder the adoption of deep learning techniques for EEG data classification tasks. The common approach to reducing noise from single-trial data is by extracting Event-Related Potential(ERP). ERP is computed by averaging a participant's single-trial EEG signals, which removes random brain activities and produces low noise signal [17]. However, due to the small number of participants, computing ERP using all trials would not generate enough samples for training deep CNN models. I obtained our ERP data through bootstrap sampling by averaging 20 single-trial signals selected randomly with replacement from the original 45 trials per task to solve this problem. We repeated this process 1000 times to generate 1000 samples for each participant per one cognitive load level.

3.3 Spatial-spectral Feature Extraction

Various frequency bands of EEG signals have been linked to the brain's WM processes. Standard five EEG frequencies are Delta (0.5 to 3Hz), Theta (3 to 7.5Hz), Alpha (7.5 to 12.5Hz), Beta (12.5 to 30Hz), and Gamma (> 30Hz). Frequency bands that are reported to be mostly correlated with the increase or decrease of cognitive load are Theta, Alpha, and Beta [8, 9, 10]. The traditional representation of EEG spectral features uses feature vector aggregating mean PSD values from all 64 electrodes. However, this representation ignores the spatial structure of the EEG signal. In this work, I use a spatial-spectral representation of EEG data by transforming the signal into 2D images (topomap), preserving both the spatial and spectral structure of the data.

I applied Continuous Fast Fourier Transform (FFT) on time series EEG and calculated average power spectral density (PSD) for theta, alpha, and beta bands to obtain spectral features. To generate topomaps, I projected 64 average PSD values to the brain scalp topographical map and interpolated values between EEG channels. The above process is summarized in Fig.1.

My objective is to use spatial-spectral features and learning by the convolution neural network in predicting four cognitive load levels.

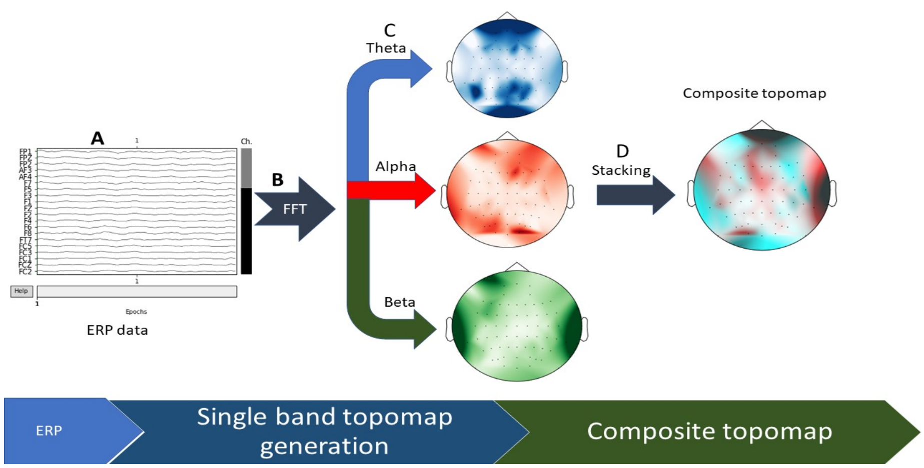


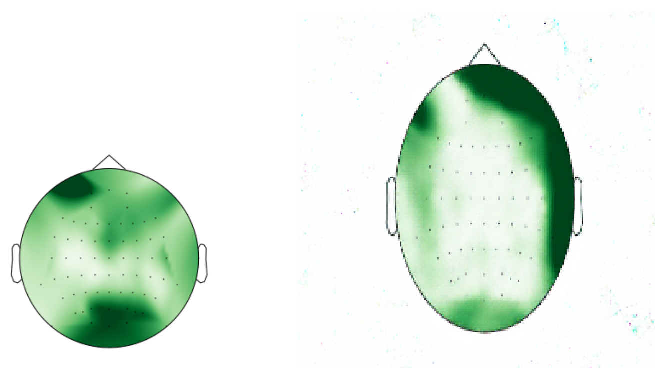
Figure 1: Overview of EEG signal to spatial-spectral representation: A) I extract ERP signal from raw EEG data. B) I apply FFT to ERP signal to get average PSD density from three frequency bands. C) I project PSD values over scalp surface to obtain spectral topography maps.D) Single-frequency topomaps are stacked horizontally to form a composite topomap

3.4 Data Augmentation with GAN

The goal of using GAN is to solve the problem of overfitting the CL classification due to possible duplicate images due to a larger number of bootstrap iterations. Therefore, the current data generation process will involve both bootstrap sampling followed by GAN. So, half of the required dataset was generated using the bootstrap sampling method and the rest through GAN.

GAN refers to the " Generative Adversarial Network." It is used to generate fake images as close as possible to the real images. GAN consists of two competing convolution neural networks, Discriminator and Generator. Discriminator tries to discriminate fake from real images. The generator network generates fake images as close as possible to the real images to fool the discriminator. Therefore, by using the feedback from the discriminator, the generator updates its weights to produce more images as close as possible to the actual images.

I trained GAN on topomap images from each frequency band. After training I used the generator network to generate 22000 images which will be combined with another 22000 images generated through bootstrap sampling to create a 44000 images data set. Therefore, up to this point we have 44000 images for alpha, theta, beta, and composite band representations which will be used to train CNN model for classify four levels of cognitive load. Fig.2 shows an example of real and GAN-generated images.



(a) Real topomap (b) GAN generated topomap

Figure 2: Example of real topomap (a) (b) GAN generated topomap

3.5 Classification with Convolution Neural Networks

I used CNN to classify four cognitive load levels using spatial-spectral images corresponding to the average PSD distribution over brain scalp surface. CNN has gained popularity in recent years due to its ability to extract and learn low-level and high-level features from an image signal. Before CNN became prevalent, classical machine learning models such as SVM, random forest were used for image classification. However, images to be used with these models

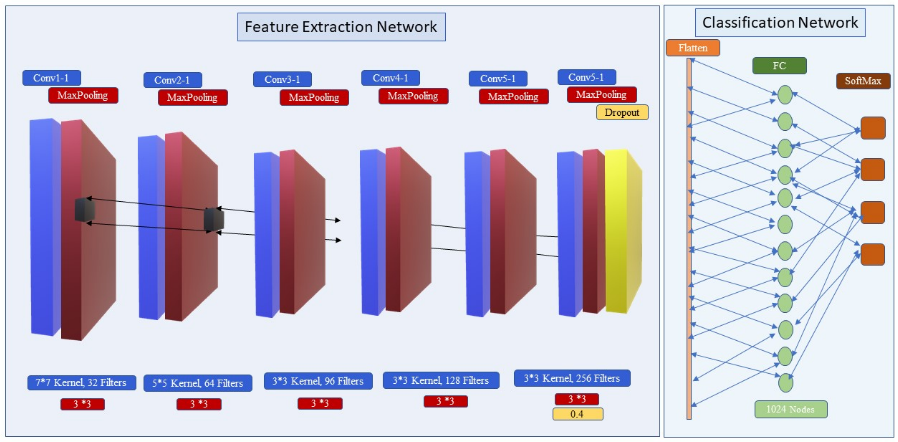


Figure 3: CNN architecture. Feature extraction network consists of six Convolution and MaxPooling sets followed by a dropout layer , and classification network consists of Flatten layer, a single Dense layer with 1024 nodes and a SoftMax layer with 4 nodes

need to be flattened into vector space resulting in a dataset with a very high dimension that increases with the image's size. Further, an image carries useful spatial information that can be lost when converted into a vector. CNN is the best method to address these challenges. As its name suggests, a 2D Convolution Neural network model performs a 2-D convolution on an image extracting and learning features using its spatial filters. The CNN filters can extract and learn low-level features such as edges and high-level features such as eye patterns in a human face image. A CNN model consists of a series of successive convolution and pooling layers. After convolution layers, the learned features are flattened to form a dense layer fed to the neural network layers for image classification.

The architecture of my CNN models is shown in Fig.3. The model's architecture consists of six Convolution layers as a feature extractor. Each Convolution layer is followed by the batch normalization, Relu, and MaxPooling layer. The kernel size of all Convolution layers is 3×3 except for the first layer (7×7) and the second (5×5). The number of filters in the Convolution layer started at 32, doubled the number of filters for each successive layer, and stopped at 256 for the last two Convolution layers. For the classification network, I used a single fully connected layer with 1024 nodes and a SoftMax dense layer to learn the decision surface to separate different levels of topomap images. The Softmax layer has an output shape of four, representing the four cognitive load classes.

I applied an L2-norm weight regularization penalty of 0.01 to the last three Convolution layers to account for over-fitting. Also, I used a single dropout layer with a dropout ratio of 0.4 before the Soft-Max layer. I trained the CNN model to minimize a categorical cross-entropy loss function using the Adam optimizer and trained using a decreasing learning rate criterion starting from 0.0001 with a decreasing factor of 0.2 for a batch size of 32 images. I randomly split the available images using a 70% - 30% split ratio for training and validation to train the CNN architecture.

4 Results

In the previous sections, I described my approach for noise reduction, data augmentation, the transformation of the signal into spatial-spectral representations, and modeling of a CNN architecture capable of learning these representations. The performance metrics used in evaluating the models are accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (ROC AUC) score. The results show how the above metrics values change from one frequency band to another, reflecting their respective predictive power.

First, I trained the deep CNN model on topomap images from three frequency bands (Theta, Alpha, Beta). Second, I trained the model on composite topomap images resulting from the fusion of individual band topomaps.I trained the models for 100 epochs with a batch size of 32% images.

Fig.4 shows cognitive load prediction performance in terms of accuracy, precision, recall, and F1-score. From the figure, I can deduce that the Beta band carries more cognitive load predictive power than Theta and Alpha since it outperformed the other two bands in predicting four cognitive levels with an accuracy of 91%. Further, I did not find a significant difference between alpha and theta bands' predictive power as the performance of CNN models trained on their corresponding topomaps is approximately the same with the accuracy of = 89% and = 89% respectively. Though we achieved a good classification performance from individual frequency bands, combining the individual bands improved the performance significantly up to the accuracy of 92%.

In Fig.5, I show the confusion matrix of my CNN models for composite representation. Looking at the confusion matrices, we can see that my models have more misclassifications at the upper intermediate cognitive load level (CL-3), which could result from high signal variations during the transition from low to high cognitive loads. On the other hand, the CNN models easily discriminated other loads: load-1(SET size = 2), load-2(SET size = 6), and load-4(SET size = 8), which shows the steadiness of EEG signals at low and high memory loads.

Overall, my results show that without tedious hand-crafted feature selection, and while preserving the spatial-spectral structure of EEG data, the CNN model can learn the mapping between EEG signal and different levels of cognitive load with high accuracy. Moreover, we can reach higher performance beyond individual frequency bands by training the CNN model on the composite band representation.

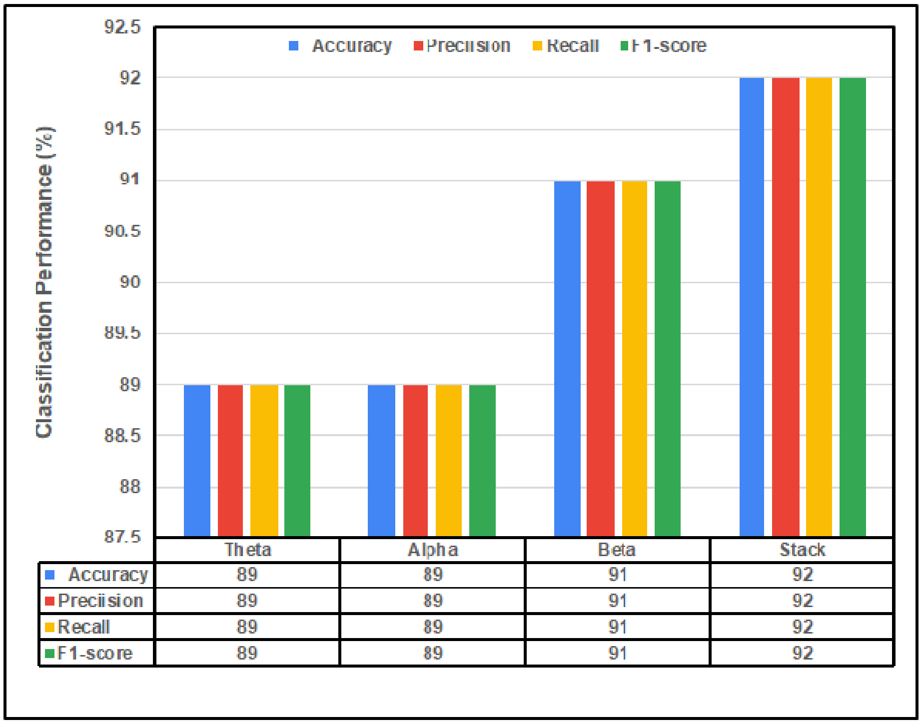


Figure 4: Cognitive load classification results with CNN model: Accuracy, precision, recall, and F-1 scores were used to evaluate the model model on spatial-spectral topomap from theta, alpha, beta bands, and composite topomap in predicting four levels of cognitive load

5 Conclusions

This paper presented a data-driven approach to learn the spatial-spectral representation of EEG signals recorded from participants performing an auditory working memory experiment. CNN models were used to learn and map the EEG representations to four levels of CL corresponding to the complexity introduced into WM tasks. Data augmentation using eigenspace-based bootstrap sampling allowed me to find bias-variance trade-offs in building CNN models. Also, the transformation of ERP data into spectral topomap images allowed us to preserve the spatial-spectral structure of EEG signals. Further, I used deep CNN architecture to alleviate the need for tedious feature extraction and selection commonly used in classical machine learning.

I conducted empirical analyses to compare the predictive power of individual frequency bands (theta, alpha, beta) and composite representation in classifying cognitive load. The results show that the Beta band(12.5 Hz - 30 Hz) has more predictive power than Theta (3 - 7 Hz) and Alpha (8 - 12 Hz) in classifying

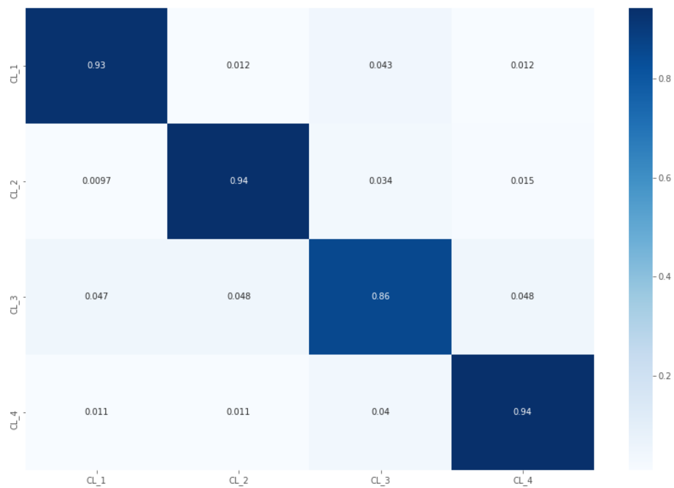


Figure 5: Confusion of CNN model trained on spectral topomap from composite topomap

cognitive load with accuracy $> 91\%$. The classification accuracy is $> 92\%$ for combinations of 3-frequency bands. Overall, the results show that the combination of data transformation, data augmentation, and use of CNN models are highly accurate ($> 92\%$) in predicting CL and robust against typical noise present in the EEG recordings.

References

[1] Alan Baddeley. "Working memory". In: *Science* 255.5044 (1992), pp. 556–559.

[2] William E Hick. "On the rate of gain of information". In: *Quarterly Journal of experimental psychology* 4.1 (1952), pp. 11–26.

[3] Fred Paas et al. "Cognitive load measurement as a means to advance cognitive load theory". In: *Educational psychologist* 38.1 (2003), pp. 63–71.

[4] Nir Friedman et al. "EEG-based Prediction of Cognitive Load in Intelligence Tests". In: *Frontiers in human neuroscience* 13 (2019), p. 191.

[5] Pega Zarjam et al. “Estimating cognitive workload using wavelet entropy-based features during an arithmetic task”. In: *Computers in biology and medicine* 43.12 (2013), pp. 2186–2195.

[6] Leonard J Trejo et al. “EEG-based estimation and classification of mental fatigue”. In: *Psychology* 6.05 (2015), p. 572.

[7] Naveen Kumar and Jyoti Kumar. “Measurement of cognitive load in HCI systems using EEG power spectrum: an experimental study”. In: *Procedia Computer Science* 84 (2016), pp. 70–78.

[8] Zhongxiang Dai et al. “EEG cortical connectivity analysis of working memory reveals topological reorganization in theta and alpha bands”. In: *Frontiers in human neuroscience* 11 (2017), p. 237.

[9] C Richard Clark et al. “Spontaneous alpha peak frequency predicts working memory performance across the age span”. In: *International Journal of Psychophysiology* 53.1 (2004), pp. 1–9.

[10] Adriano BL Tort et al. “Theta–gamma coupling increases during the learning of item–context associations”. In: *Proceedings of the National Academy of Sciences* 106.49 (2009), pp. 20942–20947.

[11] Jason K Johannesen et al. “Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults”. In: *Neuropsychiatric electrophysiology* 2.1 (2016), p. 3.

[12] Pouya Bashivan et al. “Learning representations from EEG with deep recurrent-convolutional neural networks”. In: *arXiv preprint arXiv:1511.06448* (2015).

[13] Joseph K Nuamah, Younho Seong, and Sun Yi. “Electroencephalography (EEG) classification of cognitive tasks based on task engagement index”. In: *2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*. IEEE. 2017, pp. 1–6.

[14] Anushri Saha et al. “Classification of eeg signals for cognitive load estimation using deep learning architectures”. In: *International Conference on Intelligent Human Computer Interaction*. Springer. 2018, pp. 59–68.

[15] Yuetian Liu and Qingshan Liu. “Convolutional neural networks with large-margin softmax loss function for cognitive load recognition”. In: *2017 36th Chinese control conference (CCC)*. IEEE. 2017, pp. 4045–4049.

[16] Pouya Bashivan. *Commonality and singularity in working memory network predicting performance and individual differences*. The University of Memphis, 2016.

[17] Michael GH Coles and Michael D Rugg. *Event-related brain potentials: An introduction*. Oxford University Press, 1995.