Assessing Cognitive State Adaptations using Predictive Models

Stephen M. Gordon
DCS Corporation
Alexandria, VA 22310
sgordon@dcscorp.com

Vernon J. Lawhern, J. Touryan
US Army DEVCOM ARL
Aberdeen Proving Ground, MD 21005
vernon.j.lawhern.civ@army.mil, jonathan.o.touryan.civ@army.mil

ABSTRACT

Humans constantly adapt to new tasks and circumstances. Such adaptations include changes in both physical behavior and cognitive state. In each case, adaptation requires more than just sequencing components, but balancing costs associated with each component to develop sequences that are consistent with human capabilities while meeting task demands. For instance, it is well known that many cognitive states (e.g., vigilance) cannot be maintained indefinitely and thus optimizing task performance requires adapting a dynamic system that does not allow all possible sequencies. Here we use advanced cognitive modeling and prediction tools, built using strategies from the generative AI community, to analyze and assess cognitive state adaptations during a task in which 100+ participants (individually and in pairs) must search for, identify, and defuse simulated IEDs dispersed in a visually complex environment. The prediction tool first estimates cognitive state directly from physiological data. These estimates are then passed to a generative model that analyzes the sequence and predicts how the state will unfold in the future. The generative tool was trained offline on a corpus of data that does not include the current task or participant pool.

We then analyze data around IED detection and defusal events. The results show that over the course of the experiment a consistent pattern of cognitive state changes emerge across participants. During the first task run, the emergent pattern is not well-predicted by the generative model. In other words, the IED response appears more exogenously driven and not a function of ongoing dynamics. However, by the second run, the model's prediction errors decrease, and the IED response appears to shift from exogenously to endogenously driven. The results highlight how cognitive state adaptations synchronize to task demands to maintain task performance while operating within the bounds of the system's limited capabilities.

ABOUT THE AUTHORS

Stephen M. Gordon (PhD) is a scientist at DCS Corporation. Dr. Gordon received his Ph.D. in Electrical Engineering from Vanderbilt University in 2009. His recent work includes developing machine learning approaches for neurophysiological modeling and investigating human adaptation in complex scenarios. He currently manages a team of scientists and engineers that support fundamental and applied research initiatives at the Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL).

Vernon J. Lawhern (PhD) is a senior scientist at DEVCOM ARL. Dr. Lawhern received his Ph.D. in Statistics from the Florida State University, Tallahassee, FL in 2011. His prior research efforts focused on developing novel machine learning approaches for modeling EEG recordings from human users in complex, dynamics tasks. His current research has focused on developing human-interactive machine learning and reinforcement learning methods for enabling effective human-autonomy integration.

Jonathan Touryan (PhD) is a senior scientist at DEVCOM ARL. Dr. Touryan received his Ph.D. from the University of California, Berkeley in 2004 where he studied the neurophysiology of vision. He has published research in a number of leading neuroscience journals and has experience in both human and animal neuroimaging techniques. His prior research explored the neural correlates of human performance in real-world tasks and the use of multimodal human-sensing technologies to elucidate the endogenous and exogenous sources of variability in visual search. He now manages applied research efforts that use humans-sensing and neurotechnologies to improve Soldier-system performance and human-autonomy integration.

Assessing Cognitive State Adaptations using Predictive Models

Stephen M. Gordon
DCS Corporation
Alexandria, VA 22310
sgordon@dcscorp.com

Vernon J. Lawhern, J. Touryan
US Army DEVCOM ARL
Aberdeen Proving Ground, MD 21005
vernon.j.lawhern.civ@army.mil, jonathan.o.touryan.civ@army.mil

INTRODUCTION

Operating in the real-world requires adapting to constantly changing task demands. This can include learning new behaviors and when to apply those behaviors. The same adaptation can also apply to cognition – i.e., learning when and where to direct cognitive resources to match task demands. While an individual can choose to be in one of several possible cognitive states at any given time, many of these states (e.g., high workload or high vigilance) cannot be maintained indefinitely. Research has shown that there are strict limits to what one can do cognitively (Grier, et al., 2003; Jex, 1988), just as there are limits to what one can do behaviorally. For instance, it is known that individuals cannot maintain high levels of vigilance, or sustained attention, for extended periods of time (Parasuraman & Mouloua, 1987; Smit, et al., 2004). With increased time on task, vigilance levels will decrease, and the probability of an attention-related error will increase. This effect, known as the vigilance decrement (Parasuraman & Davies, 1977), has been studied in multiple paradigms ranging from driving to air traffic control. However, paradoxically humans have repeatedly shown the ability to perform reasonably well at extended-duration, attention-requiring tasks (e.g., long haul truck drivers, air-traffic controllers). To explain this, it is theorized that individuals covertly wax/wane use of their cognitive resources to match task demands (Oken, et al., 2006). In this way, cognitive activity is almost directly analogous to physical behavior that can be deployed for brief durations of time so long as sufficient rest is available in between each deployment.

To understand these dynamics researchers have developed continuous tasks probes, such as the psychomotor vigilance task (PVT) or the gradual-onset continuous performance task (gradCPT) (Drummond, et al., 2005; Fortenbaugh, et al., 2018) and have theorized that certain cognitive states, such as attention, are oscillatory and periodic (Busch & VanRullen, 2010; Gregoriou, et al., 2015). Oscillatory theories of attention have been used to describe our tendency to start in vigilance (high attentive) states, slowly drift into less attentive states, and then oscillate between high and low levels of attentiveness (Esterman, et al., 2013; Kucyi, et al., 2017; Rosenberg, et al., 2015). This implies that at least some cognitive states, and their transitions, are governed by, and should be modeled as, a dynamic system in which resource availability waxes/wanes on its own inherent cycle. Naturally, such a cycle would not be matched to the dynamics of every environment, thus it is an open question: how do these oscillatory dynamics fit with the notion of deploying cognitive resources to task demands? While abrupt changes in task demands could be addressed by rapid reorganization of the cognitive system in a dynamic systems model, such impulse/step responses would necessarily lead to "ripple effects" downstream. Here we define "impulse/step" response as an abrupt change in cognitive state due to a, largely, unpredictable task demand. We define "ripple effect" to be a secondary change in cognitive state that is a response to the impulse/step change and not related to the task. As a simple example, overriding one's circadian rhythm to stay up all night and complete a last-minute tasker at work, would have the ripple effect of induced drowsiness on the drive to work the following morning.

Within the cognitive neurosciences most research focuses on high construct validity paradigms that are often randomized and designed to produce impulse/step responses. Therefore, to understand cognitive state as the outcome of a dynamic system that is deployed to meet task demands arising from an environment with its own unique dynamics requires addressing a few critical research gaps. First, there are few (if any) computational models available that capture (and, thus, accurately predict) sequential cognitive state changes beyond the initial impulse response. Second, it is difficult to observe adaptation of the dynamic system as we are often limited to observing changes in the impulse responses – thus, we do not know the nature of the adaptation, e.g., is it phase shifting of the cognitive system, changing of the underlying oscillatory frequency, or overall suppression of oscillatory activity?

Cognitive State Modeling

To address these gaps requires building testable models that predict how cognitive states should unfold over time. The recent advent of generative AI (GAI) solutions for problem spaces such as natural language processing (NLP) offers one path for building such a sequential, dynamic model. In this view, well-controlled isolated studies of cognitive function provide the lexicon of distinct states for use by large-scale models that can piece together and predict cognitive state evolution. However, constructing these models requires adapting current foundational approaches. Since we cannot explicitly know cognitive state of a test participant at each time point, we must estimate state using observable variables – such as recorded physiology. Recent work has introduced several large-scale (across task and participant), foundational models for physiological analysis (Cui, et al., 2023; Jiang, et al., 2024; Tegon, et al., 2025; Yang, et al., 2023), but these solutions tend to assess, or predict, raw signal properties with limited generative abilities to forward project patterns over time. In addition, these models use embedding layers that preserve signal features but do not map onto known cognitive states, or constructs. In other words, when compared to similar generative approaches used for NLP, these models attempt to simultaneously solve both the "speech-to-text" problem and the "text generation" problem in the same step.

Lexicon

To adapt GAI approaches to model cognitive states, we must first define a meaningful lexicon. In the current work, we focus on modeling attention-related processes. Attentional systems prioritize relevant information while deprioritizing irrelevant background information (Posner, et al., 1980). To quantify time-varying changes in attentional processing researchers have used multiple paradigms ranging from ecologically-valid designs (Ko, et al., 2017; Lin, et al., 2005), to well-controlled tasks, such as the PVT and gradCPT (Drummond, et al., 2005; Fortenbaugh, et al., 2018). Using variants of the gradCPT researchers in (Esterman, et al., 2013) were able to observe the moment-to-moment fluctuations in task performance and, subsequently, proposed an oscillatory thoery of attention involving the integration of arousal, attentional allocation, and information processing subsystems (Esterman & Rothlein, 2019). Alternative theories have also been proposed that leverage current knowledge of the high frequency neuronal oscillations observed at the cortical surface (Clayton, et al., 2015), such as amplitude and cross-frequency phase coupling of posterior medial frontal Theta (4-7 Hz), lateral prefrontal Alpha (8-13 Hz), Gamma (>30 Hz) band processes, when using scalp electroencephalogram (EEG).

Speech-to-Text

Second, we need a solution for the speech-to-text problem so that we can map high-frequency raw physiological signals onto the lexicon. For the current work we will use EEG as our physiological signal and build upon prior efforts by the Brain Computer Interface (BCI) community to classify attention-related states. Many BCI approaches, including transfer learning and domain adaptation methods require some participant and/or task specific data (Hajinoroozi, et al., 2015; Lotte, 2015; Nguyen, et al., 2017). This inhibits their use in a large-scale, generative approach. Domain generalization approaches, on the other hand, focus on developing machine learning models that can transfer from one domain (i.e., participant or task set) to a new domain (Gordon, et al., 2017) without the need for domain specific knowledge from that new domain. These methods can utilize models with relatively few parameters, rather than the large pretrained transformers, thus these approaches can be tailored to model physiological activity related to specific cognitive constructs using a limited amount of data, such as that contained in a single study. These properties make domain-generalization methods the ideal starting point for importing the lexicon to any subsequent foundational, or generative, model.

Big Data

Third, we need data to train the model. As the result of community-based efforts there is a growing number of large scale opensource data repositories, such as OpenNeuro (Markiewicz, et al., 2021), OpenBCI (https://openbci.com/community/publicly-available-eeg-datasets/), and the Standardized Annotated Neurophysiological Data Repository (SANDR) (Touryan & Lee, 2021), to name a few. However, given that many studies are designed to elicit specific cognitive state changes in a controlled, impulse/step response manner it may not be reasonable to simply combine all available data. Again, drawing inspiration from how foundational models are trained for NLP, we recognize that the training data is largely unscripted with samples often representing complete statements, or dialogues. To obtain similar data describing sequential cognitive state activity, we argue it is best to start with ecologically-valid paradigms that minimize the use of unpredictable events that disrupt cognitive state. It is beyond the scope of the current paper to rigorously evaluate this argument, however, in the methods section we will select ecologically-valid paradigms for our training set.

Approach

We have developed a Generative Cognitive Modeling Tool (GCMT) that uses a decoding layer to estimate cognitive states from recorded scalp electroencephalogram (EEG) and then uses a generative layer to predict how those states will unfold in the future. For the current work we focus on modeling and predicting an attention-related state, defined empirically through a previously conducted laboratory study. The GCMT makes predictions up to 25 secs into the future, which is an order of magnitude greater than alternative approaches for modeling EEG dynamics (Cui, et al., 2023; Jiang, et al., 2024).

Using the GCMT we will investigate the downstream impacts of task-induced state changes, i.e., given a task-induced change, what does the GCMT predict will happen next? Does this prediction come true? We will also measure the errors between predicted state and observed state to assess adaptation over time. In other words, does adaptation involve phase aligning the dynamic system to the task or creating new sequences of states? If the latter, models errors will increase (or stay the same). If the former, model errors will decrease.

The trained GCMT will have no knowledge of the current task in our test set, nor will it have been exposed to data from the test participants. The GCMT can only predict future cognitive states given a finite history of observed states and it *does not learn, or fine-tune, to the test set*. The GCMT is trained offline using publicly available data and then applied, as is, to the test set. Model weights are static. The test set includes a task in which individuals worked alone, or in teams of two, to locate and defuse explosive devices in a simulated factory setting.

METHODS

Generative Cognitive Modeling Tool (GCMT)

The current GCMT is a prototype system initially presented in (Gordon, et al., 2025). It is designed to read EEG data, transform that data into a low-dimensional sequence of probability values for predetermined cognitive states (lexicon) and then make generative assertions on how those states will unfold in the future. We use domain-generalization methods to project the EEG data onto an axis whose values denote the probability that a specific, predefined cognitive state is present. The GCMT is shown graphically in Figure 1 and has been trained on a corpus of 435 individuals using 315 hours of data. This is substantially less than prior foundational approaches that have used up to 25,000 hours of data as the goal with the GCMT is to establish a proof-of-concept for modeling and predicting cognitive activity.

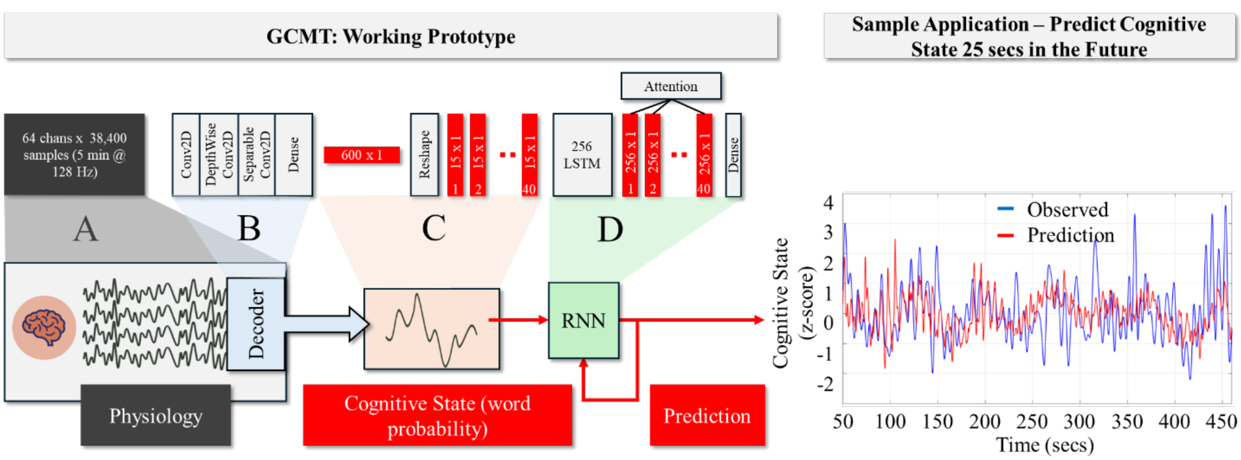

Figure 1: GCMT modeling approach. Raw physiological data (A) is decoded (B) into low-dimensional probability estimates associated with the words in our lexicon (C). The probability estimates are passed to a recurrent network (D) and used to make predictions about how the low-dimensional state will unfold.

Creating the Lexicon

We model a single cognitive state, defined empirically through a previously conducted, controlled research study. The original study is described in (Lin, et al., 2005) and required participants to drive a simulated vehicle along an empty, closed-circuit at night. Experimental manipulations were used to induce drowsiness. During the driving task, a lateral perturbation (i.e., simulated wind) would push the vehicle out of the lane and the time required for the participants to respond and correct the vehicle's trajectory was recorded as a reaction time. From this behavioral recording, periods of drowsy versus alert driving were extracted by identifying the top and bottom 10% of reaction times. Thus, the two words in our lexicon are "drowsy' and "alert". After dividing the data to create the lexicon we preprocessed the data by referencing to the average mastoid signal, bandpass filtering between [0.3, 50] Hz, and downsampling to 128 Hz. We then created cleaned versions of the data using ICA decomposition techniques to remove noisy and artifactual components (Makeig, et al., 1995). We concatenated the cleaned version with the minimal processed version as a form of data augmentation (Lotte, 2015). We refer to this data set as the Lexicon Source (LS) data set. Figure 2 (top) provides an overview of the process. Figure 2 (bottom) highlights the patterns of physiology indicative of changes in this specific state as originally identified in (Lin, et al., 2005) – i.e., increases in Delta, Theta, or Alpha power along frontal midline sections are considered closer to "drowsy" while the reverse is considered closer to "alert". This provides a solution to the "speech-to-text" problem where the complex, multidimensional patterns of physiological activity are the "speech" and the projection of the data onto a low-dimensional lexicon is the "text". Prior work in (Gordon, et al., 2023) established that such cross-experiment decoding can be used to interpret novel data sets from the perspective of previously defined cognitive constructs (or words).

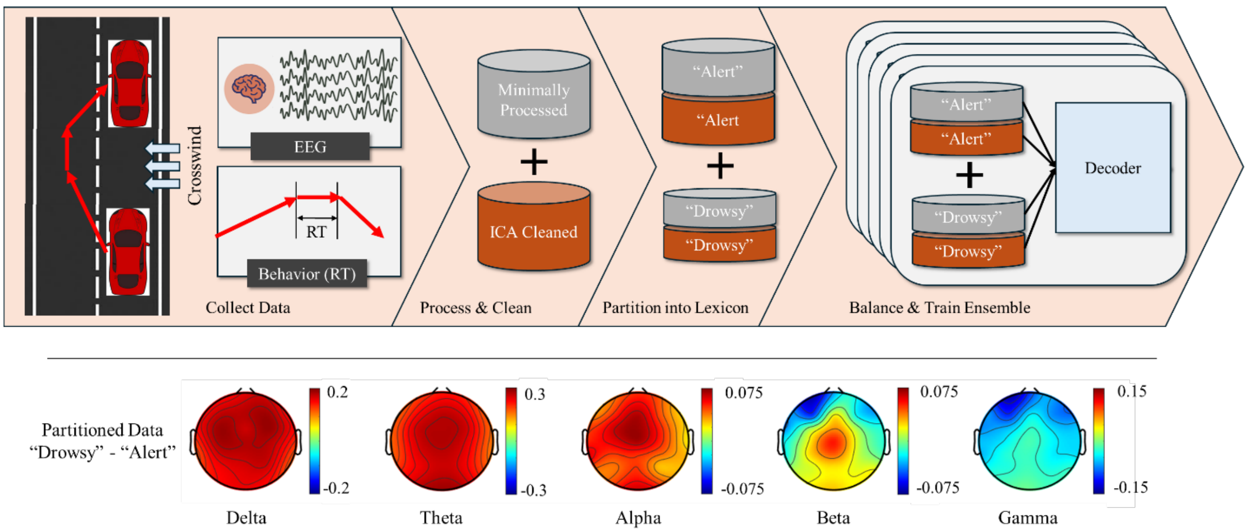

Figure 2. Top – Physiological data is collected and paired with behavior. Once processed the data is partitioned into exemplars matched to the lexicon. The data is concatenated and used to train an ensemble of decoders. Bottom – Physiological features in EEG space distinguishing the words in our lexicon.

Extracting the Lexicon

We used the EEGNet architecture (Lawhern, et al., 2018) to implement the decoder. A full description of how to use EEGNet for decoding cognitive processes is available in (Solon, et al., 2019). Here, we provide a brief overview.

All of the EEG data (training and test) must be preprocessed in a similar fashion. This includes reference signal, bandpass filter settings, and channel montage (or interpolated montage). The training data is then partitioned based on trial or condition. Once labeled, the training samples are balanced via randomized downsampling of the majority class and concatenated across participants. To better utilize all the available training data, we repeat the randomized downsampling 5x to produce an ensemble of 5x decoders (Figure 2 B, right).

To interrogate novel test data (Figure 1 A) the ensemble of trained decoders (Figure 1 B) is stepped across the new data at a sampling rate of 0.5 Hz. A group average is then obtained. This produces a one-dimensional time series signal (Figure 1 C) that can be viewed as a continuous probability estimate that the underlying physiological data 'at that moment' is conceptually more like one (or the other) of the two words in our lexicon, i.e., "drowsy" or "alert".

For this work, we used the EEGNet architecture with the settings (8-4-2), as described in (Gordon, et al., 2023). The EEG data was sampled at 128 Hz and we used a decoder input window size of 385 samples (3 secs of data). The EEG montage was set to the 30-channel intersection of the Neuroscan NuAmps system (LS data set) and BioSemi Active II system (all other data sets).

Generative Model

We chose a standard recurrent neural network (RNN) approach to handle the generative task. We made this choice for a few reasons: 1) RNNs tend to use fewer parameters and, thus, require less training data than transformers, 2) RNNs also tend to be more sensitive to the precise ordering of sequences while transformers tend to capture the gist with less focus on order specifics, and finally 3) RNNs tend to perform better at, relatively, short range predictions. The RNN is designed to accept a finite sequence of 600 samples (i.e., 5 mins of data) from the decoder. This 600-sample sequence is reshaped into 40 blocks with 15 samples per block (Figure 1 C). This was done to reduce the effect of vanishing gradients on the RNN. The RNN (Figure 1 D) was implemented using the LSTM layer from Tensorflow (v2.4.1) with input settings: 384 recurrent units, 30% dropout and recurrent dropout, L1 recurrent regularization = 1e-6, L1 kernel regularization = 1e-6, and a tanh activation function. Following the LSTM was an attention layer (Vaswani, 2017), and then a final dense layer with a tanh activation function. The prediction window was set to 25 secs in the future (Figure 1 D).

Training the RNN

We used a total of 9 data sets for RNN training and validation, encompassing 435 participants, and approximately 314.83 hours of recorded EEG data. This is in addition to the LS data set used to train the decoder. All of the data sets (except the LS data set) were collected as part of research programs funded, at least in part, by the Army Research Laboratory. All of the RNN training data sets can be accessed publicly via two locations: 1) SANDR via direct request, and 2) OpenNeuro.

Table 1 highlights relevant information for each of the RNN training data sets. During design of the GCMT we found that data set selection was an important factor in determining the accuracy of the final model. Specifically, data sets with fewer “experimental controls” tended to produce better models. That is, data sets that allowed more naturalistic behavior and had higher levels of ecological validity. We believe this to be the case for at least two reasons: 1) data sets with high construct validity tended to involve repeated trials delivered in short-bursts and thus the context windows could not be very long and 2) there were less unpredictable interruptions to the participants’ cognitive state. However, as previously mentioned it is beyond the scope of the current paper to fully investigate this issue.

It is important to note, though, that for training the RNN we did not use any information regarding the experimental design or manipulations in the training data. We only extracted the time series estimates for the lexicon from the physiological data and then trained the RNN directly on these signals without regard to experimental factors. This decision was made for both pragmatic reasons (i.e., to avoid the difficulty in trying to match experimental conditions across unique studies) and to stay consistent with the GAI assumptions of a common, underlying dynamic model. This enables our current approach to maximally transferable because it is completely agnostic to the structure of the experiment or task.

Like the approach with the decoder ensemble, we trained an ensemble of 5x RNNs. We used data sets 1-5 (Table I) for initial development, validation, and parameter tuning of the general RNN design. During this process we adjusted model parameters to achieve the best fit possible across these 5 data sets. This resulted in the final parameters settings and model design presented in the prior section. We then began training the ensemble of 5 RNNs using data sets 1-9 but holding out one data set at a time from data sets 1-5. This was done to produce more unique models. Data set 9 was used as the final validation set. Each RNN instance was trained for 10,000 epochs using a logcosh error function and a learning rate of 0.001. Final model selection was performed by evaluating the ensemble of RNNs directly on data set 9 and selecting the epoch that produced the best performance after averaging across the ensemble. For the current work, the optimal model occurred around epoch 3,800.

Table 1. Data Sets used to Train GCMT

Data Set	# Participants / Hours of EEG	OpenNeuro ID
1	10 (10.78 hrs)	ds004853
2	14 (16.32 hrs)	ds004842
3	13 (27.61 hrs)	ds004843
4	158 (52.40 hrs)	ds004123

5	109 (116.71 hrs)	ds004122
6	28 (22.37 hrs)	ds004121
7	24 (33.60 hrs)	ds004120
8	13 (16.53 hrs)	ds004855
9	66 (18.51 hrs)	ds004851
SUM	**435 (314.83 hrs)**	

Test Data

Full EEG data was collected from a total of 101 individuals (56F, 42M, 1O, 2None) with average range 18-67 years, average age 35.05, std. 14.1, producing 36 dyad runs. All participants were recruited from third party online platforms in and around greater Los Angeles, California. All participants provided voluntary, informed written consent as required by Title 32, Part 219 of the CFR and Army Regulation 70-25. All human subjects testing was approved by the Institutional Review Board of the United States Army Research Laboratory (protocol number 21-110). To be eligible for the study, participants had to be 18 years or older with normal hearing, normal (or corrected to normal) vision, normal color vision, and fluent in English. Exclusion criteria included history of brain trauma, heart problems, and/or pregnancy.

In the main experiment, all questionnaires and task instructions were presented on a Windows Surface Pro 3 tablet and all tasks were presented using a 24" Dell Ultrasharp U2414H desktop monitor running Microsoft Windows 7. The task was programmed using Unreal Engine 4 (UE4) (https://www.unrealengine.com/en-US).

Data collection for the main experiment occurred in pairs. Each participant was greeted by a different experimenter and escorted to separate lab spaces. Participants were screened in person using an Ishihara color vision test and Snellin Chart (20/40). Those who did not pass in-person screening were excluded. Each participant was then prepared for physiological data collection including EEG recording using a 64 channel BioSemi Active II. Two separate data recording set-ups were used that were synchronized using software the protocol Lab Streaming Layer (LSL) (Kothe, et al., 2024).

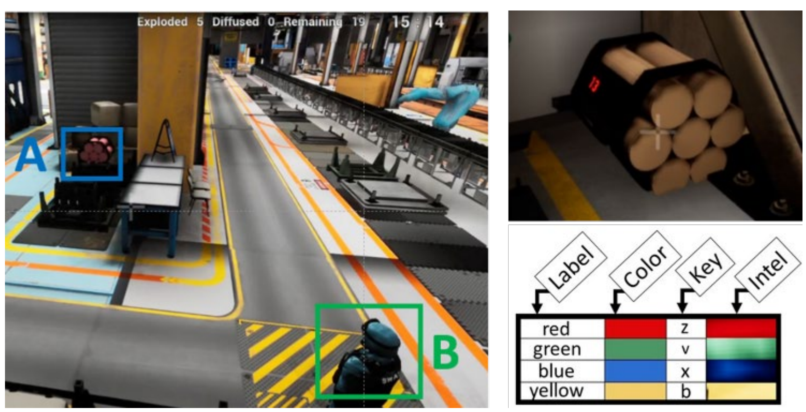

Figure 3. Left: Sample of Player 1's point-of-view during IED Defusal Task. A IED is highlighted with Box A. Player 2's avatar is highlighted with box B. Right (upper): A close-up view of a IED stimulus used for this task. Right (lower): An example look-up code to defuse the IED.

The main task that participants performed was a timed Improvised Explosive Device (IED) task. In this task, participants had to locate IEDs that had been dispersed throughout a factory setting (see Figure 3). Upon locating an IED, the participants had to defuse the device using the appropriate key code. Each IED was color coded with the color corresponding to the specific key that needed to be pressed to disarm the device. Timers located on each IED would begin countdown once an individual was within a specific radius of the device even if the participant did not have line of sight to the device. Timer duration was randomly selected from 15, 20, 25, and 30 secs. To disarm an IED participants had to press and hold the correct key for 5 seconds while standing near the device. Incorrect key presses, or failure to defuse before timer countdown, would result in the IED exploding. There were two versions of the IED task: an individual version and a team version. Each participant first completed a practice version with 4 IEDs, then the individual version (first run), and finally the team version (second run). Each run lasted approximately 15-20 minutes and had 16-24 IEDs randomly distributed in the environment.

Analysis Methods

Impulse Response to IED Events

Throughout the analysis we assess both the observed cognitive state estimate (obtained from the decoder) and the predicted cognitive state value (obtained from the GCMT). As previously noted, the predictions were made using a context window of 5 minutes and prediction window of 25 secs (into the future). Once the entire sequence of observed cognitive states (or predictions) is obtained for an individual for a given run we z-score the data to facilitate better cross-participant comparison and aggregation. We compute grand average responses time-locked to IED events and assess changes in estimated cognitive state around the events.

Model Error Reduction

We assess two types of errors in the model: magnitude and phase errors. To derive these errors we compute the Hilbert transform of the observed (or predicted) z-scored state signal. We then compute the magnitude differences and the phase differences (as absolute values) for each sample and average the errors over specified windows of time.

RESULTS

Comparison of Impulse Response and Ripple Effects

Figure 4 shows the observed and predicted state values for all IED events (i.e., successfully disarmed and exploded) for the first and second runs completed by the participants. The results are time-locked to the end of the IED event (disarm or explosion). Observed state values are shown in blue and predicted state values are shown in red. These are grand averages – averaged first within participants and then across. We assessed state changes in the region of time [-15, 0] secs before the IED was disarmed or exploded. This would have corresponded to the region of time in which the IED had been either detected visually or triggered but the participant(s) were unaware of the IED location. In either case, there would be a visual cue indicating an IED. We refer to this as the impulse response since state changes should be in response to that external environmental cue. From Figure 4 (left) we see that in the first run, there is a significant negative deflection in observed cognitive state in the moments leading up to the disarm/explode event but no change in the predicted state. This makes sense since the participant would not have known an IED was about to be triggered (until it happened) and, thus, the history of their cognitive state dynamics would not enable prediction of such an upcoming random event.

We also assessed state changes in the region [5, 20] secs after the IED exploded or was disarmed. The two regions are highlighted in light and dark grey, respectively. Below each primary figure is the average state value in these regions. In this second region the context window provided to the GCMT would have included data from the first, impulse region (recall: the GCMT makes predictions 25 secs into the future). In the second region, we see a positive deflection in the predicted state from the GCMT that, at first, appears erroneous as there is no associated change in the observed state. It is this predicted change that we suspect is a ripple effect. Performing a one-way ANOVA test using the distribution of average state values from our population we find that the observed changes in the impulse region and the predicted changes in the ripple region are significant (indicated with * in Figure 4 left, bottom).

Figure 4 (right) shows the data from the second run, in which we see that as participants become familiar with the task and environment there is change in both the observed and predicted cognitive state activity. The observed activity produces larger positive deflections after the IED event (i.e., during the ripple region), which better matches the GCMT predictions in this region. What is more interesting, though, is that the predicted state begins to show negative deflections prior to the IED event. Repeating our statistical analysis, we find that the distribution of observed state changes is significant before and after the event and predictions remain significant after the event. The predictions before the event approach significance ($p \sim 0.11$) but do not become significant. The key result here, though, is that the GCMT predicted a shift in cognitive state would occur in the ripple region and that in the later stages of the task, this shift was evident. The GCMT was not only making predictions 25 secs into the future, but also asserting that a state shift should occur, even when that shift was not immediately present.

To further understand this, we divided the data from each run into halves and computed the observed state shift, during the ripple region, for each half. These results are shown in Figure 5 (left). While there is a high degree of variance, we see shift that is close to monotonic. Figure 5 (right) shows the average number of IED events per half. Here we see that probability of finding an IED is skewed towards the first half of each run, which is natural as participants would likely find several IEDs quickly and then spend longer amounts of time searching for the remaining few. The skew in the probability of finding an IED, though, does not correlate or explain the shift in observed state changes.

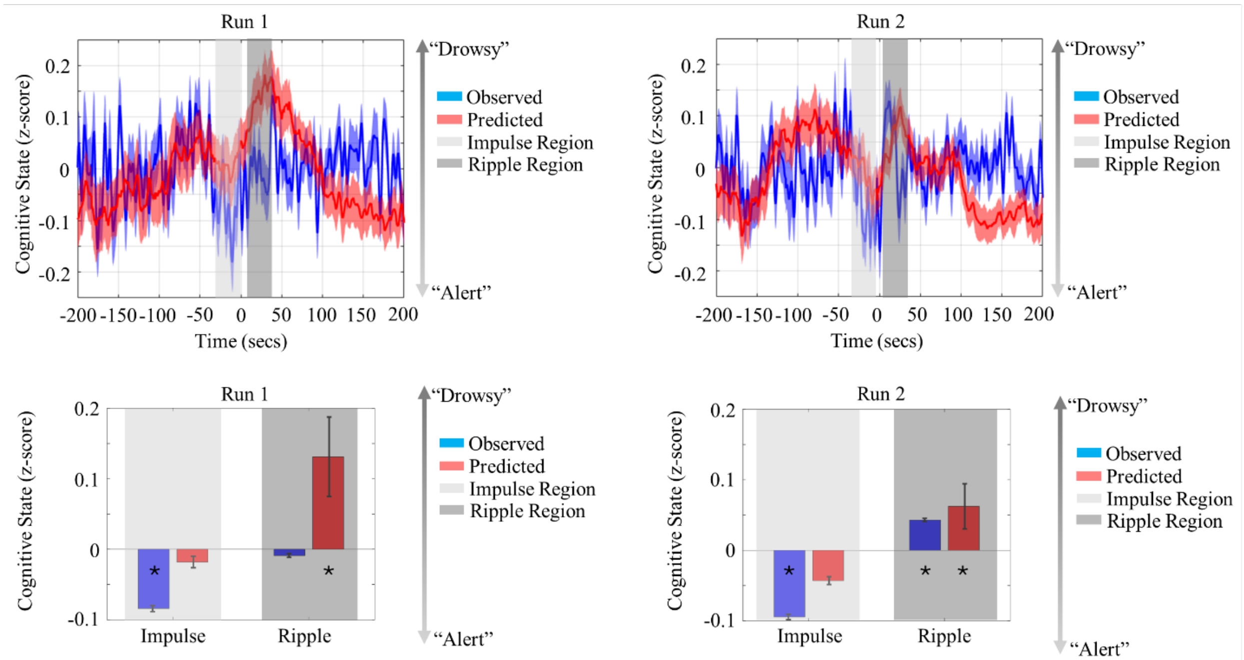

Figure 4. Top – Observed (blue) and predicted (red) state estimates are extracted at the moment of an IED events. Data is presented as grand averages. Bottom – Average state changes are shown during impulse ([-15, 0] secs before event) and ripple ([5, 20] secs after event) regions. Regions with significant changes are indicated with (*).

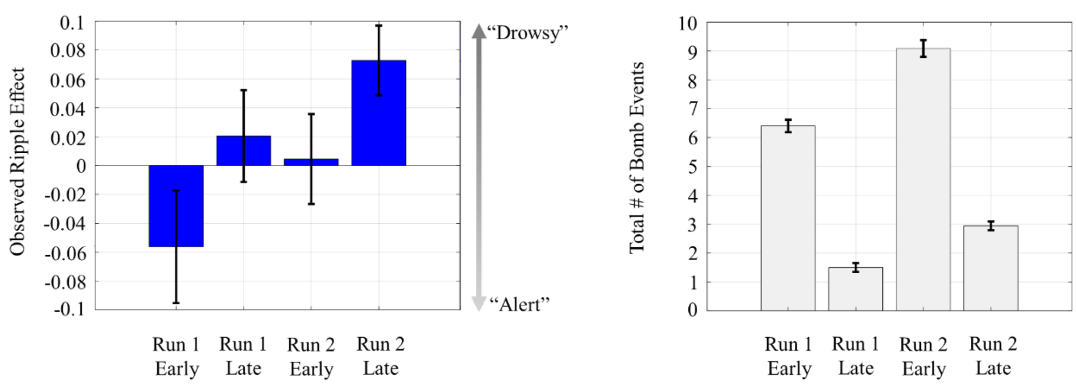

Figure 5. Left – Change in observed state response during the ripple region. There is a general increase in state values that with increasing experience match the predictions from the GCMT. Right – Histogram of bombs counts. In both runs, participants find more bombs in the early portion of the run.

Model Errors and Adaptation

The results in Figures 4 (right) and 5 (left) strongly suggest adaptation is occurring and the adaptation is to fit a common model to the task demands. Therefore, we then assessed magnitude and phase errors between the observed and GCMT predicted state with increasing experience in the task. These results are shown in Figure 6. Overall, we see a net reduction in phase errors over time with significant changes between the runs and an overall significant downward trend over time using a linear regression test. Again, these changes are not correlated with the number of IED events. We do not see decreases in the magnitude errors. The reduction in phase errors between the observed and predicted state indicates that as participants gain experience with the task their cognitive state dynamics moves towards that of a common model (i.e., the GCMT).

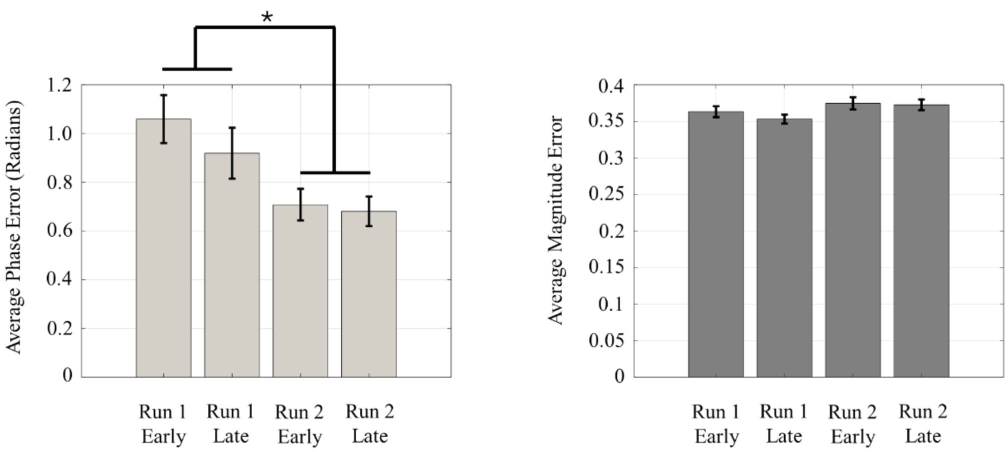

Figure 6. Left – phase errors between observed and predicted state values with increasing experience with the task. There is a significant reduction in phase errors as time on task increases. Right – magnitude errors between observed and predicted state values. There is no significant relationship between time on task and change in magnitude errors.

Figure 7 shows the observed and predicted state changes (previously presented in Figure 4) but now only for those IEDs that were not successfully disarmed, and only for the second run. We downselected the data to this subset because in the second run, participants were working with a partner and this led to an increase in the number of successfully disarmed IEDs. We also hypothesized that these IEDs were the most likely to produce true, unpredictable impulse responses, since participants would not have visually located these IEDs (i.e., since the IEDs ultimately were not disarmed). From Figure 7 (left), we see that there is a very strong correlation between the observed and predicted state values. Further investigation showed that the distribution of correlation values was significantly greater than chance, which is predicted by the results in Figure 6. In Figure 7, though, it is the directional pattern that matters. Prior to the IED event the shift is towards "Alert". After the event, the shift is towards "Drowsy" indicating that not only are participants' state dynamics moving to that of a common model, but also optimizing to the task, where having attentional resources ready to deploy prior to an event is functionally better than having them ready after the event.

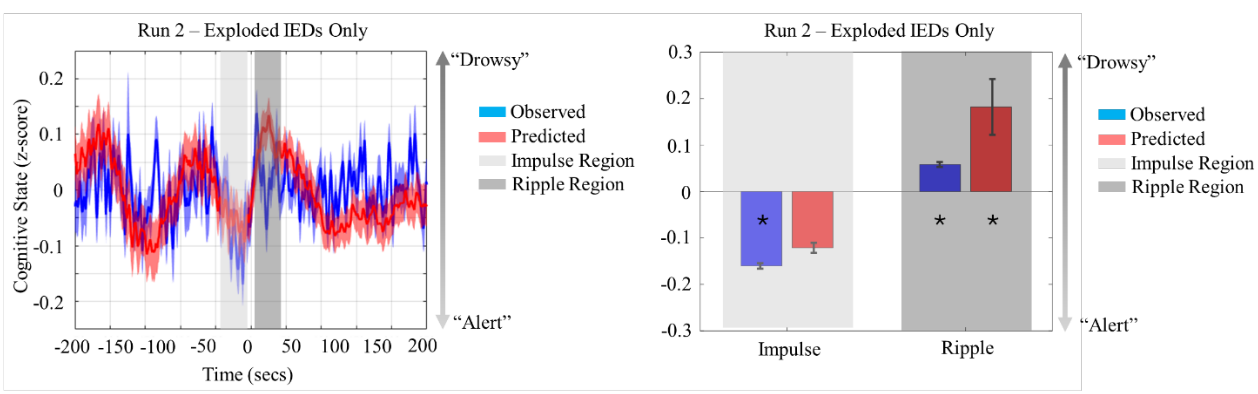

Figure 7. Left - observed and predicted state changes from run 2 and assessing only exploded IEDs. Right – average state changes during the impulse and ripple regions. Significant changes are marked with (*).

DISCUSSION

Most research in the cognitive sciences focuses on well-controlled paradigms with high construct validity. Such studies tend to utilize experimental manipulations to perturb cognitive state. This, in turn, produces impulse/step responses. While prior work has shown that individuals can reorient cognitive state to meet such task demands it is also known that certain states (e.g., attention) are oscillatory and periodic. Basic dynamic systems theory suggests, then, that impulse/step responses (i.e., rapid reorientation) passed into an oscillatory dynamic system would yield downstream ripple effects, which could lead to suboptimal (or disastrous) outcomes when these ripples are not matched to task demands. Therefore, we hypothesized that adaptation must involve both picking the right cognitive state for the right circumstance and fitting cognitive state dynamics to task demands to minimize the potentially disastrous outcomes of ripple effects. Yet, it was an open question how individuals would perform this adaptation.

We found that in our task, there was an initial deflection in cognitive state preceding IED events that was not predicted by the GCMT. This was the original impulse response. In the first task run this impulse response produced no observable ripple effect, however, there was a ripple effect predicted by the GCMT indicating that as far as cognitive state dynamics were concerned a ripple effect was expected. We suspect that participants were suppressing this ripple effect during the initial exposure to the task, which would be consistent with prior neuroscientific theories of activity suppression during periods of novel, or high, task demands (Klimesch, et al., 2007; Schroeder & Lakatos, 2009). As participants' familiarity with the task and environment increased, the ripple effect predicted by the GCMT began to manifest in the observed signal as well.

At the same time, the phase errors between the observed and predicted state decreased (Figure 6) with significant correlations occurring even before an IED event (Figure 7). While this result may seem confusing (i.e., how can the GCMT model with no information about the task predict that an event is about to occur), we believe the explanation is straightforward and that participants allowed their cognitive system to run and matched their behavior to the state (i.e., ebbing and flowing their search strategies with state). It also is important to note that our experimental design did not enforce truly randomized IED onsets. While IED locations and distribution were unknown to the participants, the devices were static and did not "pop-up" or appear out of thin air. The devices were separated enough such that no two devices had overlapping radii to trigger explosions. In other words, like most ecologically-valid environments there was structure (Schroeder & Lakatos, 2009). This leads to the most significant result of the current work, which is that in an ecologically-valid environment participants were phase-shifting their cognitive state dynamics, and potentially behavior, to the structure of the environment and the task demands. We were able to observe and measure that adaptation using a generative, foundational model for predicting cognitive state dynamics; however, it was beyond the scope of the current work to investigate the downstream impacts on behavior. Prior work in (Gordon, et al., 2025) using the GCMT has shown that predicted cognitive state does influence behavior.

CONCLUSION

We constructed the GCMT to address specific research gaps related to how the community can measure and understand cognitive state dynamics. Once trained on the data in Table I, the weights in the GCMT remained static, thus the adaptation observed in the participants for the current task was towards a group model. Naturally, it is an open question how far such a common, group model can go and when do individual differences appear. It is also clear (from visual inspection of the data and results in Figures 4 and 7) that the GCMT is tending towards lower frequency state changes – well below the 0.3 Hz high pass setting on the EEG data.

If verified through further research, we believe the GCMT, or similar approaches, could be used as a more accurate model of human performance by moving away from a static representation of state (e.g., high/low workload) and towards a dynamic concept in which current states predictably lead to new states. This could improve training by measuring how well operators match resources to task needs by measuring synchrony between predicted model outputs and task demands or predicted model outputs and observed state changes. In addition, the GCMT could be used to assess whether observed operator performance levels are achieved by exploiting internal dynamics or suppressing them, which could improve the value of skill-based assessments by understanding who picks up on the structure of the task. In our current results, we saw that task performance during the first run was largely driven by suppressing state dynamics while in the second run state dynamics were not suppressed but rather synchronized to task needs. From a dynamic systems theory, observing large discrepancies between predicted and observed cognitive state would be a sign that individuals are not anticipating the world around them and not properly matching resources to task demands. Of course, in some cases it is not possible to anticipate all possible changes in the world and, in such cases, one would expect to see some deviations between predicted and observed cognitive state – and these deviations would lead to subsequent ripples later in time. Another key implication of the GCMT is that these ripples are just the oscillations of the underlying state and, essentially, an unavoidable by-product of the (human) system design. Thus, another application area would be to improve human-system interaction through intelligent task allocation. In prior work using the GCMT we found that participants were more likely to engage autonomous controls (e.g., speed and lane control) for a vehicle if their current cognitive state did not match their GCMT predicted cognitive state. In that work, participants leveraged the vehicle's autonomy during the "ripple" period.

In both that prior work and the current work the GCMT can be viewed as verifying that, at least some, cognitive states oscillate, or wax/wane, on their own cycle. Given the wealth of prior work highlighting the impact of such oscillations on attention and task performance we believe the ultimate application of the GCMT would be to facilitate the mitigation of such necessary oscillations through improved system design. For example, during the everyday task of driving a car down a busy street the GCMT could anticipate when an individual's cognitive state is about to drift into a suboptimal region. This anticipation would lead to a critical design choice: should the system alert the driver and risk a ripple effect or engage additional self-driving features of the vehicle for a short amount of time? While the former option may seem intuitive the results presented in this current work suggest the latter option may, at times, be more effective – since waxing/waning through attentional states is a fundamental feature of the system and interrupting such processes can produce larger subsequent ripples.

Finally, we believe it is possible to extend the GCMT to model additional states and processes and include additional physiological signals. Additional states could include various definitions of task load, alternative definitions of attention (or alertness), working memory functions, or even non-cognitive processes such as emotional states. Each of these states have well-documented physiological markers. To expand to additional physiological signals, and perhaps even moving away from EEG, would require developing the necessary "speech-to-text" decoders to map new physiological signals onto a lexicon that can be ingested by the GCMT. Again, there are many documented analyses that link physiological signals to cognitive constructs. Thus, dynamic cognitive state models, such as the GCMT, are just the next step in understanding the human biological system so that scientists and human factors engineers can better design technologies that account for capabilities that constantly wax/wane on multiple timescales.

ACKNOWLEDGEMENTS

This project was sponsored by the US Army Research Laboratory under contract W911QX-21-D-0004. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the US Government.

REFERENCES

Busch, N. A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences*, *107*(37), 16048-16053.

Clayton, M. S., Yeung, N., & Kadosh, R. C. (2015). The roles of cortical oscillations in sustained attention. *Trends in cognitive sciences*, *19*(4), 188-195.

Cui, W., Jeong, W., Thölke, P., Medani, T., Jerbi, K., Joshi, A. A., & Leahy, R. M. (2023). Neuro-gpt: Developing a foundation model for eeg. *arXiv preprint arXiv:2311.03764*, *107*.

Drummond, S. P., Bischoff-Grethe, A., Dinges, D. F., Ayalon, L., Mednick, S. C., & Meloy, M. J. (2005). The neural basis of the psychomotor vigilance task. *Sleep*, *28*(9), 1059-1068.

Esterman, M., Noonan, S. K., Rosenberg, M., & DeGutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral cortex*, *23*(11), 2712-2723.

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current opinion in psychology*, *29*, 174-180.

Fortenbaugh, F. C., Rothlein, D., McGlinchey, R., DeGutis, J., & Esterman, M. (2018). Tracking behavioral and neural fluctuations during sustained attention: A robust replication and extension. *NeuroImage*, *171*, 148-164.

Gordon, S. M., Jaswa, M., Solon, A. J., & Lawhern, V. J. (2017, March). Real world BCI: cross-domain learning and practical applications. In *Proceedings of the 2017 ACM Workshop on An Application-oriented Approach to BCI out of the laboratory* (pp. 25-28).

Gordon, S. M., Lawhern, V. J., & Touryan, J. (2025), Human performance modeling using generative AI principles, in *Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium* (GVSETS), NDIA, Novi, MI, Aug. 12-14, 2025

Gordon, S. M., McDaniel, J. R., King, K. W., Lawhern, V. J., & Touryan, J. (2023). Decoding neural activity to assess individual latent state in ecologically valid contexts. *Journal of Neural Engineering*, *20*(4), 046033.

Gregoriou, G. G., Paneri, S., & Sapountzis, P. (2015). Oscillatory synchrony as a mechanism of attentional processing. *Brain research*, *1626*, 165-182.

Grier, R. A., Warm, J. S., Dember, W. N., Matthews, G., Galinsky, T. L., Szalma, J. L., & Parasuraman, R. (2003). The vigilance decrement reflects limitations in effortful attention, not mindlessness. *Human factors*, *45*(3), 349-359.

Hajinoroozi, M., Mao, Z., & Huang, Y. (2015, December). Prediction of driver's drowsy and alert states from EEG signals with deep learning. In *2015 IEEE 6th international workshop on computational advances in multi-sensor adaptive processing (CAMSAP)* (pp. 493-496). IEEE

Jex, H. R. (1988). Measuring mental workload: Problems, progress, and promises. In *Advances in psychology* (Vol. 52, pp. 5-39). North-Holland.

Jiang, W. B., Wang, Y., Lu, B. L., & Li, D. (2024). NeuroLM: A Universal Multi-task Foundation Model for Bridging the Gap between Language and EEG Signals. *arXiv preprint arXiv:2409.00101*.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition–timing hypothesis. *Brain research reviews*, *53*(1), 63-88.

Ko, L. W., Komarov, O., Hairston, W. D., Jung, T. P., & Lin, C. T. (2017). Sustained attention in real classroom settings: An EEG study. *Frontiers in human neuroscience*, *11*, 388.

Kothe, C., Shirazi, S. Y., Stenner, T., Medine, D., Boulay, C., Grivich, M. I., ... & Makeig, S. (2024). The lab streaming layer for synchronized multimodal recording. *BioRxiv*, 2024-02.

Kucyi, A., Hove, M. J., Esterman, M., Hutchison, R. M., & Valera, E. M. (2017). Dynamic brain network correlates of spontaneous fluctuations in attention. *Cerebral cortex*, *27*(3), 1831-1840.

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Jrn. of neural engineering*, *15*(5), 056013.

Lin, C. T., Wu, R. C., Liang, S. F., Chao, W. H., Chen, Y. J., & Jung, T. P. (2005). EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Transactions on Circuits and Systems I: Regular Papers*, *52*(12), 2726-2738.

Lotte, F. (2015). Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces. *Proceedings of the IEEE*, *103*(6), 871-890.

Makeig, S., Bell, A., Jung, T. P., & Sejnowski, T. J. (1995). Independent component analysis of electroencephalographic data. *Advances in neural information processing systems*, *8*.

Markiewicz, C. J., Gorgolewski, K. J., Feingold, F., Blair, R., Halchenko, Y. O., Miller, E., ... & Poldrack, R. (2021). The OpenNeuro resource for sharing of neuroscience data. *Elife*, *10*, e71774

Nguyen, T., Ahn, S., Jang, H., Jun, S. C., & Kim, J. G. (2017). Utilization of a combined EEG/NIRS system to predict driver drowsiness. *Scientific reports*, *7*(1), 43933.

Oken, B. S., Salinsky, M. C., & Elsas, S. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clinical neurophysiology*, *117*(9), 1885-1901.

Parasuraman, R., & Davies, D. R. (1977). A taxonomic analysis of vigilance performance. In *Vigilance: Theory, operational performance, and physiological correlates* (pp. 559-574). Boston, MA: Springer US.

Parasuraman, R., & Mouloua, M. (1987). Interaction of signal discriminability and task type in vigilance decrement. *Perception & Psychophysics*, *41*, 17-22.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of experimental psychology: General*, *109*(2), 160.

Rosenberg, M. D., Finn, E. S., Constable, R. T., & Chun, M. M. (2015). Predicting moment-to-moment attentional state. *Neuroimage*, *114*, 249-256.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in neurosciences*, *32*(1), 9-18.

Smit, A. S., Eling, P. A., & Coenen, A. M. (2004). Mental effort causes vigilance decrease due to resource depletion. *Acta psychologica*, *115*(1), 35-42.

Solon, A. J., Lawhern, V. J., Touryan, J., McDaniel, J. R., Ries, A. J., & Gordon, S. M. (2019). Decoding P300 variability using convolutional neural networks. *Frontiers in human neuroscience*, *13*, 201.

Tegon, A., Ingolfsson, T. M., Wang, X., Benini, L., & Li, Y. (2025). FEMBA: Efficient and Scalable EEG Analysis with a Bidirectional Mamba Foundation Model. *arXiv preprint arXiv:2502.06438*.

Touryan, J. and Lee, T., (2020) Cognition and Neuroergonomics Collaborative Technology Alliance Final Report.

Yang, C., Westover, M., & Sun, J. (2023). Biot: Biosignal transformer for cross-data learning in the wild. *Advances in Neural Information Processing Systems*, *36*, 78240-78260.